

WOFEX 2014

proceedings of the 12th annual workshop

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

ISBN 978-80-248-3458-0

WOFEX 2014 wofex

<http://wofex.vsb.cz/2014/>

12th annual workshop,
Ostrava, 9th September 2014
Proceedings of papers

Organized by

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



WOFEX 2014

© Michal Krátký, Jiří Dvorský, Pavel Moravec, editors

ISBN 978-80-248-3458-0

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: 711
Impression: 200
Edition: 1st
First published: 2014

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 2014 (PhD workshop of Faculty of Electrical Engineering and Computer Science) was held on September 9th September 2014 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 112 papers of faculty PhD students and 4 papers of external authors. The proceedings of WOFEX 2014 are also available at WOFEX Web site <http://wofex.vsb.cz/2014/>. 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 2014

Michal Krátký
Program Committee Chair
WOFEX 2014

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

9th September 2014

<http://wofex.vsb.cz/2014/>

Table of Contents

Study Programme P2649 Electrical Engineering

Electrical Power Engineering (3907V001)

MV Substation Protection Design and Setting Coordination	1
<i>Eduard Beránek</i>	
Current Load of an Electrification Mast	6
<i>Jiří Cigánek</i>	
Operation of a Local Distribution System	12
<i>Jiří Drholec</i>	
A Train Simulation	18
<i>Maroš Ďurica</i>	
Measurement of disturbing glare caused by interior luminaires with LED modules	24
<i>Barbara Helštýnová, Karel Sokanský, Tomáš Novák, and Zbyněk Carbol</i>	
Determination of Coefficients for Matrix of Importace	30
<i>Vít Houdek</i>	
Research and measurement of active and reactive power supplied by the biogas station into the distribution network	36
<i>Jiří Janša</i>	
Concept of charging stations for electric cars	42
<i>Oliver Marcincin</i>	
Methods of Small Hydropower Plants Connections in Water Supply System	48
<i>Tomáš Mozdřeň</i>	
Statistical Evaluation of Voltage and Current Asymmetry in AC Low Voltage Part of Photovoltaic Plant	54
<i>Michal Ney</i>	
Precising of failure rate valuation	60
<i>Martin Slivka</i>	
Power Analysis of Photovoltaic Power Plant	66
<i>Martin Smoček and Zdeněk Hradílek</i>	
Simulation of Active Energy Unit with respecting Smart-Grid requirements	72
<i>Jindřich Stuchlý</i>	

Applications of using the thermal energy from cogeneration unit for accumulate into the technology	78
<i>Michal Špaček</i>	
Solver platform – optimization tool for distribution grid operation.....	84
<i>Marián Uher</i>	
Hydrogen as a storage of electrical energy	90
<i>Jan Vaculík</i>	
Adaptive relay protection system for Off-Grid systems	96
<i>Jakub Vramba</i>	

Electrical Machines, Apparatus and Drives (2642V004)

Identification of System Parameters using Genetic Algorithms	102
<i>Jiri Hajovsky</i>	
The Contribution of 3-Phase Induction Motor Parameters Identification Using the Equivalent Circuit	108
<i>Jan Otýpka</i>	
Design of Luenberger Observer for Sensorless Control of Induction Motor	114
<i>Ondřej Petrtýl</i>	
Induction Generator Optimized Operation for Small-Scale Wind Power in Program LabVIEW	120
<i>Stanislav Rusnok, Pavel Sobota, and Michal Kelnar</i>	
Identification and Optimization of Parameters of Synchronous Machines with PM in axis D and Q	126
<i>Ladislav Slíva</i>	
Control of Induction Motor Using Artificial Neural Networks	132
<i>Tomáš Verner</i>	

Electronics (2612V015)

The converter for coupling to photovoltaics with accumulation	138
<i>Katerina Baresova</i>	
The Properties of Boost Converter with SiC MOSFET.....	144
<i>Martin Borski and Jiří Špička</i>	
Actuators for Intake Air Regulation of Modern Turbo Diesel Engines	150
<i>Tomas Mrovec</i>	
Full – Wave Rectifier Using Functional Structure Based on Current Conveyors.....	156
<i>Radek Šulák</i>	

Technical Cybernetics (2612V045)

Design of a controller for the motor speed	162
<i>Tomáš Dočekal and Kristýna Friedrischková</i>	
Software diagnosis and drivers for photovoltaic power station	168
<i>Petr Drábek</i>	
Management of energy systems combined with renewable or alternative energy resources	174
<i>Kristýna Friedrischková, Bohumil Horák, and Tomáš Dočekal</i>	
Autonomous mode generator unit start choke control	181
<i>Kristýna Friedrischková, Tiantian Yu, Boris Vukojević, Jakub Rawicki, and Shen Junwen</i>	
Image Signal Processing and Face Detection	187
<i>Jiri Haska</i>	
Advanced Method for Selective Extraction of Lutein and Zeaxanthin from Plant Material	193
<i>Barbora Hrvolová and Jiří Kalina</i>	
Application of fuel cell technologies	198
<i>Jiří Kazárik and Bohumil Horák</i>	
Universal embedded platform for low-cost industrial image processing analysis	203
<i>Michal Kelnar, Ondřej Horáček, and Miloslav Holba</i>	
Interactive Rehabilitation System for Balance Exercises	209
<i>Jana Krohova and Maros Smondrk</i>	
WiFi Localization of Exploration Vehicle Based on Multilateration	215
<i>Ales Kurecka</i>	
The Bioelectrical Impedance Spectroscopy and Electrical Impedance Tomography Measurement Device	221
<i>Marketa Michalíková</i>	
Methods verification of image recognition in metallurgy industry	227
<i>Martin Mikolajek</i>	
Usage of Genetics Algorithms in Task of System Identification	233
<i>Jana Nowaková and Miroslav Pokorný</i>	
User identification by biometric methods	239
<i>Pavlina Nudzikova</i>	
Remote control of electric car system	245
<i>Vít Otevřel and Bohumil Horák</i>	

Pulse transit time: promising method for continuous non-invasive blood pressure monitoring	251
<i>Lukas Peter</i>	
The color analysis of transfusion bag	257
<i>Petra Rajmanová</i>	
Measurements for remotely controlled power generator	263
<i>Jakub Rawicki, Vít Otevřel, and Bohumil Horák</i>	
Remote Control of Power Line Socket	269
<i>Junwen Shen, Vít Otevřel, and Bohumil Horák</i>	
Distance Determination of Objects of Cluster of n-dimensional Structure .	275
<i>Roman Slabý, Jana Nowaková, and Martin Hasal</i>	
Exercise-based Games for Balance Rehabilitation	281
<i>Maros Smondrk and Jana Krohova</i>	
Fast sequence anomaly detection	287
<i>Andrey V. Sukhanov, Sergey M. Kovalev, and Jan Otýpka</i>	
Advanced method for control of magnetization of AMR sensing element . .	294
<i>David Vala</i>	
Artificial Neural Network Based Derivation of Vectorcardiographic Leads	300
<i>Michal Vozda, Martin Cerny, and Jindrich Cernohorsky</i>	
Remote-controlled generator	305
<i>Boris Vukojević, Jiří Kazárik, and Bohumil Horák</i>	

Study Programme P1807 Computer Science, Communication Technologies and Applied Mathematics

Computer Science (1801V001)

Dynamic Network Analysis of DBLP	311
<i>Alisa Babskova</i>	
Processing of Range Query in R-tree Using CPU and GPU	316
<i>Pavel Bednář</i>	
Stains detection on nanoparts images	322
<i>Petr Berek, Jan Platos, and Vaclav Snasel</i>	
The stock market's solution synthesized by Analytical programming	328
<i>Marian Bielik and Tomas Vantuch</i>	
Flexible Neural Tree in Pseudo-Periodic Time-Series Forecasting	334
<i>Tomas Burianek</i>	

Human Activity Recognition: Classifier Performance Evaluation on Multiple Datasets	340
<i>Pavel Dohnálek</i>	
Distance and Energy-based Descriptors for Object Detection	346
<i>Radovan Fusek and Eduard Sojka</i>	
Eye and Iris Detection in Images with Active Near-infrared Illumination	352
<i>Michael Holuša and Radovan Fusek</i>	
LIDAR Function Transformation: A New Algorithm for Registration of LIDAR Data via SOMA Optimization	358
<i>Branislav Holý</i>	
Fast Parallel Artificial Bee Colony on CUDA	364
<i>Jan Janoušek</i>	
Finding Nearest Neighbors in Graph Layout Algorithm	370
<i>Tomáš Ježowicz</i>	
Formal Software Process Framework	376
<i>Michael A. Košinár</i>	
Cost-based Optimizations of XML Queries	383
<i>Petr Lukáš</i>	
Data Structures for Indexing Triple Table	389
<i>Roman Meca</i>	
Chaos Level Measurement in Logistic Map Used as the Chaotic Numbers Generator in Differential Evolution	395
<i>Lenka Skanderova and Ivan Zelinka</i>	
Familiarization with Distributed Defence System for Public Networks	401
<i>Zbyněk Složil and Daniel Strěbný</i>	
Approximation of molecular energies using neural networks	407
<i>Martin Stachoň and Aleš Vítek</i>	
An Application of Process Mining to Analysing the User Process on the Web Page	411
<i>Jakub Stolfa and Michal Holis</i>	
Implementation of Distributed Defence System for Public Networks	417
<i>Daniel Strěbný and Zbyněk Složil</i>	
Simulation of Agile Software Process with the aid of Artificial Neural Network	422
<i>Radoslav Štrba</i>	
Kaira: Analysis and Verification	428
<i>Martin Šurkovský and Ondřej Meca</i>	

Automatic Detection and Segmentation of the Carotid Artery in Ultrasound Images	434
<i>Jan Tomeček and Radim Farana</i>	
Visualisation of volumetric data on the GPU	441
<i>Vojtěch Uher</i>	
Dependencies Extraction by Grammatical Evolution approach	448
<i>Tomas Vantuch</i>	
Context Transformations	454
<i>Michal Vašínek</i>	
Comparison of Arrhythmia Classification Methods in ECG Series	460
<i>Lukáš Zaorálek</i>	

Computational and Applied Mathematics (1103V036)

Mapping of a quantum system of particles onto a classical system of cyclic polymers	466
<i>Rajko Čosić</i>	
SPN as a tool for risk modeling in process industries	472
<i>Ondřej Grunt</i>	
Optimization Process for Searching Thermal Conductivity of Mill Scales	478
<i>Martin Hasal</i>	
Optimization methods for inverse problems with energy norm	484
<i>Rostislav Hrtus, Jaroslav Haslinger, and Radim Blaheta</i>	
LiDAR data and eigenvalues	490
<i>Milan Jaroš</i>	
Parallel Matrix-vector Multiplication, its Memory Requirements and Optimizations	496
<i>Michal Kravčenko</i>	
Parallel k -Means Clustering in Computational Astrophysics	502
<i>Lukáš Malý</i>	
Parallel solution of the time dependent wave equation	508
<i>Michal Merta</i>	
Semi-Smooth Newton Method Combined with TFETI for Contact Problems with Coulomb Friction	514
<i>Kristina Motyčková</i>	
Active-set algorithm with spectral steps for semicoercive quadratic problems in particle dynamics	520
<i>Lukáš Pospíšil</i>	

Traffic Flow Modeling Based on Sparse Data	526
<i>Lukas Rapant</i>	
Use of the Simplex Algorithm for Solving the Multiple Traveling Salesman Problem	532
<i>Tomáš Režnar</i>	
Finding an optimal path for collecting items in the store	538
<i>Adam Silber</i>	
Boundary Element Method for Mathematical Homogenization of Composite Materials	544
<i>Matyáš Theuer</i>	
Graph Partitioning using Spectral Methods	550
<i>Radek Tomis</i>	
Elastic image registration based on domain decomposition with mesh adaptation	556
<i>Alena Vasatova and Ales Ronovsky</i>	
Shape Optimization for Free Surface Problems Using Subdivision Techniques	562
<i>Jan Zapletal</i>	
Communication Technologies (2601V018)	
The Influence of Gamma Irradiation on the Optical Coupler	568
<i>Lukas Bednarek, Petr Winkelhofer, and Martin Papes</i>	
Detection of the Network Anomalies by Processing of NetFlow with Machine Learning Systems	574
<i>Miroslav Bures, Pavel Nevud, and Jaroslav Zdrlek</i>	
Optical Vibration Sensor based on Michelson Interferometer Arrangement with Polarization-maintaining Fibers	584
<i>Jakub Cubik</i>	
Speech Quality Measurement of GSM infrastructure built on USRP N210 and OpenBTS project	590
<i>Marcel Fajkus and Martin Mikulec</i>	
Possibility of Computer Simulation of Modulation Formats for Real Environment Conditions of Free Space Optics	596
<i>Lukas Hajek, Ales Vanderka, Tomas Hlavinka, and Jan Latal</i>	
Basic Evaluation and Testing of V2I Communication System	604
<i>Stanislav Hejduk</i>	
Mach-Zehnder interferometer for movement monitoring	610
<i>Stanislav Kepák</i>	

The Temperature Measurement of the Extermination of Dry Rot Fungus Using Microwave Heating	614
<i>Andrej Liner</i>	
SOM Classifier for Speech Stress Detection	620
<i>Pavol Partila, Jaromir Tovarek, and Miroslav Voznak</i>	
OFDM and QAM Modulation Under Different Weather Conditions	626
<i>Frantisek Perecar, Jakub Jaros, Andrej Liner, and Vladimir Vasinek</i>	
Measuring Optimal Length of the Amplifying Fiber in Different Working Conditions of the Amplifier	632
<i>Radek Poboril</i>	
Monitoring of Speech Quality in full-mesh Networks	638
<i>Jan Rozhon</i>	
Classification of VoIP Honey-pot Data	644
<i>Jakub Safarik</i>	
Radio Jamming	650
<i>Martin Tomis</i>	
Precision of Classification Methods for Berlin Database of Emotional Speech	654
<i>Jaromir Tovarek, Pavol Partila, and Miroslav Voznak</i>	
Thermal inducted change of beam profile in bend fibers	660
<i>Petr Zavadný</i>	
Conical Termination of Optical Fibers for Sensor Applications	666
<i>Ondřej Zbořil, Tomáš Kajnar, and David Hrubý</i>	

External Authors

An analysis of the geometry of a power line span on the basis of high-resolution images	671
<i>Rafał Gasz</i>	
Diagnostics of technical condition of boiler tubes	677
<i>Miroslaw Lasar</i>	
Homogenization of Electrical Machines Stator Windings	683
<i>Piotr Mynarek</i>	
StimSys: functional MRI Stimulation System	687
<i>Jan Rydlo, Pavel Dvořák, and Jaroslav Tintěra</i>	

Author Index

MV Substation Protection Design and Setting Coordination

Eduard Beránek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
eduard.beranek.st@vsb.cz

Abstract. The main purpose is to find the optimum relay settings with respect to overload, short circuit in order to disconnect only the faulted part of the network and to minimize the damage caused by fault currents.

It will be shown the practical example of designing and setting of the electric protections.

Keywords: protection, over current protection, short-circuit, voltage, current, short-circuit current, short-circuit ratio, substation

1 Introduction

The protections are essential parts of the power system. Protection devices contribute to protecting people against electrical hazards, avoid damage to equipment, limit thermal, dielectric and mechanical stress on equipment, maintain stability and service continuity in the power system.

2 Calculation of short circuit ratio of substation

For designing and setting of electric protections we have to know short circuit ratio at the substation. We need to know a minimum two-phase short circuit current $I_{k2\min}''$, which is necessary for design and setting of an electric protections and we need to know a maximum three-phase short circuit current $I_{k3\max}''$ for checking the temperature rise of the devices and for setting a maximum time delay.

2.1 Description of substation

The substation in industrial company are supplied with two overhead lines 607 and 608 via two transformers 110/10,5 kV TH1 and TH2 from substation Lískovec R 110 kV to substations 41.03 RX1 and 41.03 RX2. These substations contain two synchronous generators TG1 and TG2 and also two synchronous motors MA2 and MA. New substations 22.12 E RX1 and 40.12 RX1 (which is connected asynchronous motor MA21) are supplied from substation 41.03 RX1. New substation 22.12 E RX2

and older substations 40.12 RX2 and 20.06 RX1 (which are connected further asynchronous motors) are supplied from substation 41.03 RX2.

The different substations can be interconnected to each other, but they are not connected during normal operation.

2.2 Scheme of connection

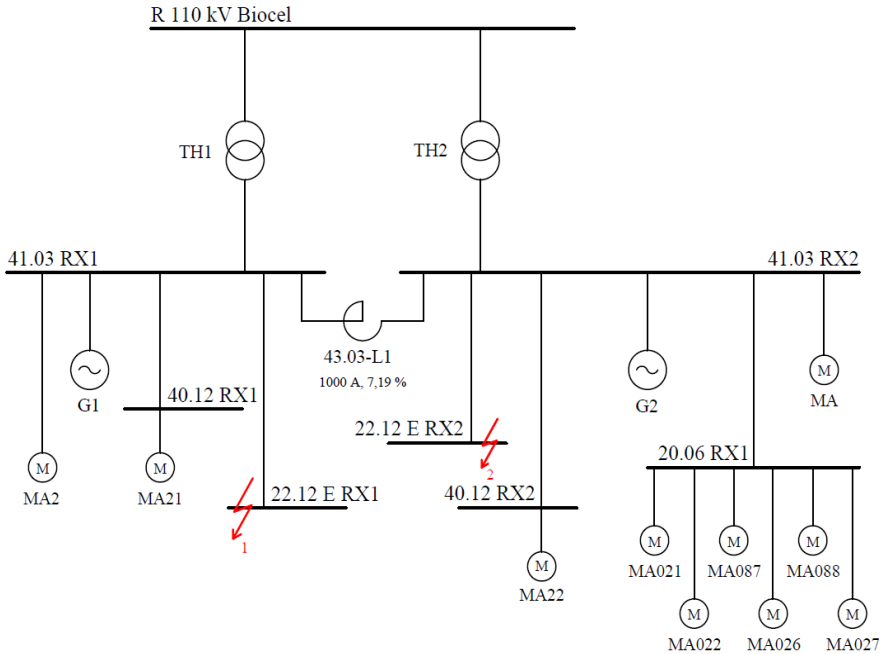


Fig. 1. Scheme of connection in industrial company

For calculation of maximum short circuit ratio, operation at maximum power is being considered. The calculations are performed according to the standard IEC 60909-0. In table below, the different short circuits from each power operation are presented.

Table 1. Summary of calculation of short circuits

Power operation	Place of short-circuit	S''_{k3}	I''_{k3}	I''_{k2}
		[MVA]	[kA]	[kA]
maximum	-	454,9	26,26	-
minimum	1	126,26	7,29	6,31
	2	82,62	4,77	4,13
through	TM1	-	-	1,377
through	TM2	-	-	1,377
through	TS1	-	-	0,462

3 Design and setting of protections

3.1 Used protections

Design and setting of protections is made for two new substations 22.12 E RX1 and 22.12 E RX2.

The substations are equipped with the ABB 615 series units (REF 615, REU 615).

The REF 615 is a specialized feeder IED perfectly aligned for the protection, control, measurement and supervision of utility substations and industrial power systems. REF 615 is a member of ABB's Relion family and a part of its 615 series IEDs are characterized by their compactness and withdrawable design.

The REF 615 provides main protection for overhead lines, cable feeders, and busbar systems of distribution substations. It can be applied for protection and control of grounded and ungrounded distribution systems.



Fig. 2. Protection REF 615

3.2 Protections coordination

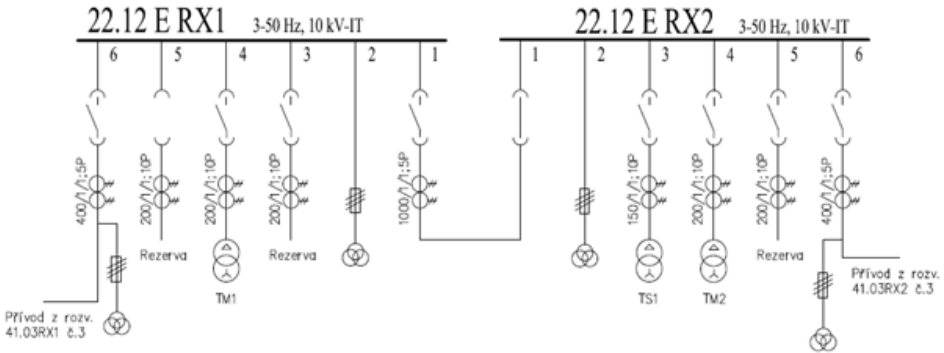


Fig. 3. Single-line diagram of substation 22.12 E RX1 and 22.12 E RX2

The units are used with following protective functions:

Over current protection function PHLPTOC, $3I>$ with normal inverse characteristic and short circuit protection function PHHPTOC, $3I>>$ with definite time.

With the suggested relay settings presented in the setting table, selectivity will be obtained.

Example of calculation:

Protective equipment

- Transfer feeder Al $2 \times 3 \times 240 \text{ mm}^2$, $I_{\text{PER}} = 660 \text{ A}$
- CT: 400/1/1 A
- Electric protection: REF 615

Over current protection function PHLPTOC, $3I>$

$$I_{\text{start}} \geq I_{\text{pmax}} \cdot \frac{k_b}{k_p \cdot p_i} = 660 \cdot \frac{1,1}{0,96 \cdot 400} = 1,89 \text{ A} \quad (1)$$

where

k_b = safety coefficient (between 1,1 up to 1,35)

k_p = reset ratio (typical 0,96)

p_i = rated transformation ratio of current transformers

Short circuit protection function PHHPTOC, $3I>>$

$$I_{\text{start}} < \frac{I''_{k2\text{min}}}{k_c \cdot p_i} = \frac{4,13 \cdot 10^3}{1,5 \cdot 400} = 6,88 \text{ A} \quad (2)$$

where

k_c = sensitivity coefficient (typical 1,5)

When setting the protection we must take into account the setting master protection and current transformer, so that in the final set starting values are lower for both degrees

Table 2. Protection setting

Substation	Field number	Protection function	Starting value	Time factor k	t [ms]	Real current value I [A]
			$I_r \times I_n$			
22.12 E RX1	6	Over current PHLPTOC, 3I>	1,1	0,4	-	440
		Short circuit PHHPTOC, 3I>>	6	-	400	2400

4 Conclusion

The main power components studied are sufficient protected against over current and short circuit currents. Selectivity will be obtained.

Acknowledgements

This work was supported with the financial support of the European Union in the framework of the Partnership in Energy, project No.CZ.1.07/2.4.00/31.0080.

The results of the thesis are published in:

- Energy Partner, Volume II, No. 5 / 2014

References

1. BERÁNEK, E.: Diplomová práce - Návrh a nastavení ochran pro rozvodnu vn průmyslového podniku, 2013.
2. Popis a technická data výrobku, IED pro chránění a ovládání vývodu REF615, REU615, ABB, 18. 2. 2011.
3. ČSN EN 60909-0: Zkratové proudy v trojfázových střídavých soustavách, Květen 2002.
4. GRYM, R.: Chránění II, 1. vyd. Havířov: IRIS, 2004. 305 s. ISBN 80-903540-0-9

Current Load of an Electrification Mast

Jiří Cigánek

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jirka.ciganek@seznam.cz

Abstract. This paper describes the modelling of magnetic fields in and around a concrete base of a railway mast. It also describes the current load on the concrete base of the railway mast during a lightning arrester breakdown. A specific base is chosen for the simulation of a 3D model. The Comsol Multiphysics program is used for the purposes of simulation. The program uses the finite element method for solving partial differential equations. (Current Load of the Rail Mast Pad During Surge Absorber Breakdown).

Keywords: Electrification mast, concrete base, electric potential, current density.

1 Introduction

An electrification system includes both power and signal sources and transmitters which are connected together by the trolley structure, the trackbed and the conductive ground. Tracks are grouped into parallel power conductors and signal and communication conductors, and the entire electrification system behaves like a spatial, linear distribution network.

The ground therefore forms a significant part of the electric traction system. Often, it is exactly this element of the electric traction system which is of a particular interest (e.g. in the case of assessing the risk of corrosion on underground components). Stray currents – the term which designates reverse leakage current of the electric traction – are, however, not the only currents flowing in the ground. Ground current fields have many causes. They can be artificially made or occur naturally, and they can more or less influence energy transfer on an electric traction. Artificial current fields occur primarily in the surroundings of electric facilities which are powered from sources and which are insufficiently insulated from the ground, or which possibly use the ground as a return conductor.

The current density vector J is a basic field quantity describing current fields. It is a differential quantity and it therefore determines the amount of current passing through an elementary area in every point of a solved space. J depends on resistivity ρ , more precisely on conductivity σ according to the differential form of Ohm's law.

$$\mathbf{J} = \sigma \mathbf{E} \quad (\text{Am}^{-2}) \quad (1)$$

The current density vector can be determined in every point where it is possible to determine the intensity of an electric field and where conductivity is known. In an isotropic environment, its direction is identical to the direction of the intensity vector of an electric field.

Determining the current density vector uses a more complicated method in an anisotropic environment where conductivity appears in the form of a tensor. In most cases a tensor has only diagonal components, i.e. different conductivities are in the direction of different coordinates, but the intensity of the electric field in the direction of one coordinate does not affect the current density in the direction of a different coordinate, for example:

$$\mathbf{J}_x = \sigma_{xx} \mathbf{E}_x, \quad \mathbf{J}_y = \sigma_{yy} \mathbf{E}_y, \quad \mathbf{J}_z = \sigma_{zz} \mathbf{E}_z, \quad (\text{Am}^{-2}) \quad (2)$$

Then the differential form of Ohm's law is as follows

$$\mathbf{J} = \sigma \mathbf{E} = \sigma (-\text{grad } \varphi) \quad (\text{Am}^{-2}) \quad (3)$$

$$\text{where} \quad \sigma = \begin{bmatrix} \sigma_{ii} & 0 & 0 \\ 0 & \sigma_{jj} & 0 \\ 0 & 0 & \sigma_{kk} \end{bmatrix} \quad (\text{Sm}^{-1}) \quad (4)$$

Electromagnetic problems associated with solving current fields in a railway surface and a substructure generally work with a total current, more precisely current density, which is connected to intensity, or rather potential. Forced quantities are potentials or the primary (forced) intensity of the electric field \mathbf{E}_v . The primary intensity determines the distribution of charges and currents and it is used for specifying this distribution. For the purposes of the analysis it can be accepted that the primary intensity determines the secondary distribution of charges and currents, and thus the secondary intensity \mathbf{E}_{sec} , i.e.:

$$\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t \quad (\text{Vm}^{-1}) \quad (5)$$

$$\nabla \mathbf{D} = \rho_c \quad (\text{Cm}^{-2}) \quad (6)$$

$$\nabla \times \mathbf{H} = \mathbf{J} \quad (\text{Am}^{-2}) \quad (7)$$

$$\nabla \mathbf{B} = 0 \quad (\text{Vsm}^{-2}) \quad (8)$$

where the symbol ρ_c does not designate resistivity but electric charge density. Similarly to the intensity \mathbf{E} , the current density \mathbf{J} has also two components, i.e. the forced current density component \mathbf{J}_v and the density of eddy currents $\sigma \mathbf{E}$. Material relationships must be added to these equations on the interface

$$\mathbf{D} = \varepsilon \mathbf{E} \quad (\text{Cm}^{-2}), \quad \mathbf{B} = \mu \mathbf{H} \quad (\text{Vsm}^{-2}), \quad \mathbf{J} = \sigma \mathbf{E} \quad (\text{Am}^{-2}) \quad (9)$$

The dynamic solution of Maxwell's equations includes also the influence of eddy currents and it is appropriate to use here a combined vector potential

$$\nabla \times \mathbf{A} = \mathbf{B} \quad (\text{Vsm}^{-2}) \quad (10)$$

$$\nabla \times \mathbf{H} = \mathbf{J}_v + \sigma \mathbf{E} = \mathbf{J}_v - \sigma \partial \mathbf{A} / \partial t \quad (\text{Am}^{-2}) \quad (11)$$

therefore
$$\nabla \times (\nabla \times \mathbf{A}) = \mu(\mathbf{J}_v - \sigma \partial \mathbf{A} / \partial t) \quad (\text{Vs m}^{-1}) \tag{12}$$

A cissoid function is used in problems which involve time changes of the field.

$$e^{jx} = \cos x + j \sin x = \text{cis } x \tag{13}$$

Therefore
$$E(t) = E e^{j\omega t} = E(\cos \omega t + j \sin \omega t) \quad (\text{Vm}^{-1}) \tag{14}$$

$$E'(t) = j\omega E e^{j\omega t} \quad (\text{Vm}^{-1}) \tag{15}$$

It is often suitable to perform a solution of steady current fields for the scalar potential of the electric field. The advantage of its use for large ground fields lies in the substitution of only one scalar variable for the vector field \mathbf{E} , which is given by 3 components. The boundary problem will be therefore described by one scalar variable φ , and after its solution all the components of the intensity of the electric field and subsequently also the components of the current density are obtained. Potential has a significant importance in connection with transferring a spatial model to a perimeter model described by the node voltage method.

Laplace's equation in the Cartesian coordinate system applies to the scalar electric potential in homogeneous environments without sources:

$$\nabla^2 \varphi = \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} = 0 \quad (\text{V}) \tag{16}$$

In modelling 3D current steady fields of the railway traction, the research used the duality of the stationary current field with the electrostatic or magnetostatic field.

2 Description of the Examined Mast

Modelling was performed in the 3kV DC electrification system. A lattice mast made of profile steel was used. A steel board with holes for bolts is welded to the end of the mast; the board is fastened to the footing base (main support) by means of a metal reinforcement which is set in concrete in the mast footing. The ground plan of the concrete footing is shown in Fig. 1.

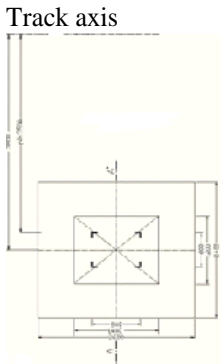


Fig. 1: Ground plan of a mast footing

The A-A cross-section of the mast footing is shown in Fig. 2.

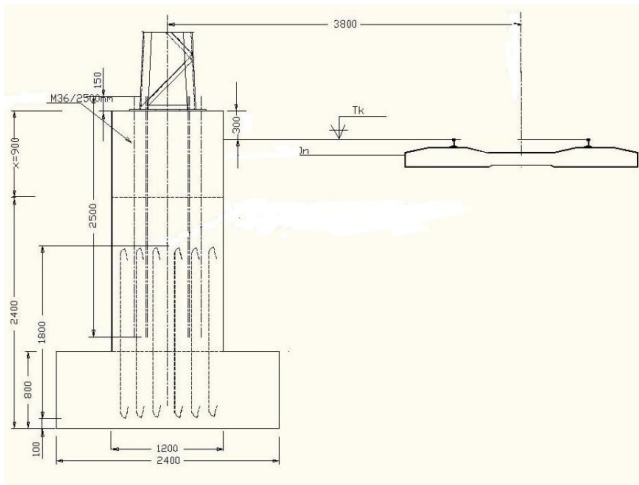


Fig. 2: Cross-section of the mast footing

3 Modelling in COMSOL

The mast footing was modelled in COMSOL Multiphysic in AC/DC Module physics – The Electric Currents interface in the model “DC electric current flow in conductive media” was used. The interface solves standard equations for electric potential. The 3D model counted in approx. 720,000 nodes. The 3D model of the solved area is shown in Fig. 3 in the most illustrative way.

Conductivities of individual solved sub-areas:

soil	$\sigma_s = 0.01 \text{ Sm}^{-1}$
concrete	$\sigma_c = 1.5 \text{ Sm}^{-1}$
steel	$\sigma_e = 1.12 \cdot 10^7 \text{ Sm}^{-1}$

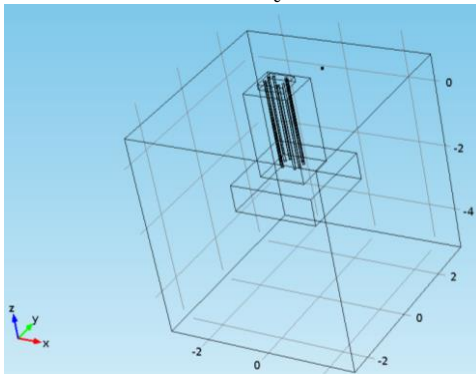


Fig. 3: Model of the solved area

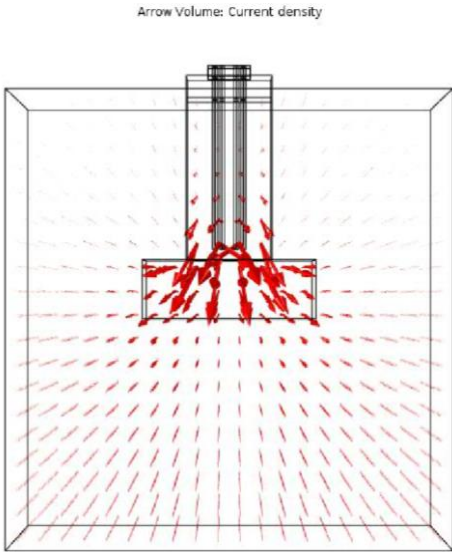


Fig. 4: Current density vectors in the footing

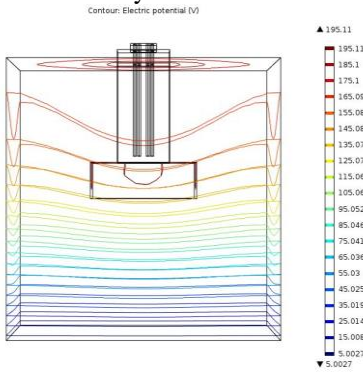


Fig. 5: Distribution of electric potential

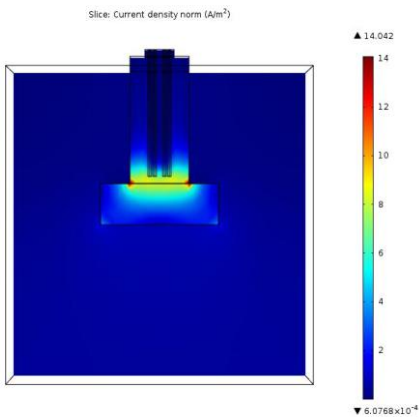


Fig. 6: Distribution of current density

4 Conclusion

According to railway system operators there are frequent cases of damage to mast footings, in particular when a surge arrester is short-circuited for a long time. In such a case the rail voltage gets directly to the mast and therefore also to the bolts. Several calculations for different input parameters - e.g. the voltage changed in the range from 50 to 200 V, the conductivity of the used materials was also changed, etc. – were performed. Using modelling it was found that footings which have sharp edges are really stressed by strong current at the points of these edges – see Fig. 5 in the circle. However, it is the points of contact between the steel bolts and the concrete which are especially stressed.

5 Bibliography

- [1] .Paleček, “Vybrané kapitoly z problematiky elektrotechniky v dopravním inženýrství,” (Selected Chapters from Electrical Engineering Problems in Transportation Engineering) university textbook, VŠB-TU Ostrava 1996
- [2] L. Ivánek, “Modely a přenosové parametry pro šíření zpětných proudů v elektrické trakci,” (Models and Transfer Parameters for Propagation of Reverse Currents in Electric Traction) post-doctoral thesis, Ostrava 1998
- [3] ČSN EN 50122-2 “Dražní zařízení-Pevná trakční zařízení – Část2: Ochranná opatření proti účinkům bludných proudů, způsobených DC trakčními proudovými soustavami,” (Railway applications – Fixed installations - Part 2: Protective provisions against the effects of stray currents caused by DC traction systems) as amended
- [4] Promotion material from the company Humusoft
- [5] V. Kolář, J. Paleček, T. Mlčák, “Vliv provozu vodní elektrárny Lipno na kvalitu elektrické energie v souvislosti s trakčním odběrem,” (The Influence of the Lipno Hydro Power Plant Operation on the Quality of Electric Energy in Connection with Traction Consumption) Electric Power Engineering 2012, Brno, 2012

Operation of a Local Distribution System

Jiří Drholec

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jiri.drholec.st@vsb.cz

Abstract. This paper deals with the agenda associated with management of a power control centre within a power supply plant together with operation of a local distribution system within the premises of such power supply plant. The main part aims at the analysis of power supply reliability concerning a selected section of the system at 6kV distribution plant. The background data used for calculation are sourced from the database comprising events recorded in the control centre logs.

Keywords: reliability, local distribution system, plant, control centre, distribution plant, faultless operation probability

1 Introduction

This paper deals with agenda associated with management of a power control centre in a company concerned with industrial power supply [1]. The company represents a large complex specialised to cater for the needs of metallurgical industry. Its scope of activities includes provision of services within the fields of water economy, combined heat and power production, gas engineering, power engineering and production of technical gases. This paper further deals with operation of the local distribution system (LDS) situated within the premises of the company focused on production and processing of pig iron as well as the metallurgical and machine engineering production. The final part of this paper deals with reliability of power supply to the selected manufacturing plant (6kV distribution plant) in the year 2010.

2 Production and Distribution of Power within Plant Premises

The source is represented by an integrated system for production of steam in steam boilers. These boilers burn dust coal, coke or blast-furnace gas. These gases develop as by-products during production of pig iron and coke, their quantity is proportional to the production volumes handled by these processes. The parameters of such generated steam are: 520 °C at 9.6 MPa.

Power is generated using back pressure and bleeder steam turbines processing superheated steam. Turbines are driven by synchronous generators producing electric power at the voltage level of 6 kV. The power is then run through two cable ducts to reach the two main distribution plants and further to the LDS.

2.1 External Links to the Distribution System

The transmission system (TS) area linked to the industrial premises can be supplied via system transformers EHV / 110 kV from up to four distribution plants. When creating the link to 110 kV system, the power control centre of the distribution system operator complies with the rule involving supply from two TS centres. The LDS of this company is linked to the superior distribution system via two 110 kV distribution plants. Power lines are fed from transformers rated for 110 / 22 kV and situated within the industrial premises.

2.2 Internal Links for Grid Operation

The operation of combined heat and power station with the high voltage (HV) distribution grid is controlled by the shift supervisor on duty at the combined heat and power station control centre. The electric power generation system comprises the compact building of combined heat and power station. The electric power lines are distributed all over the premises (forming the LDS) to ensure supply of power to each production line and match its particular quality and quantity requirements.

The actual power output depends on the production economy, the existing needs of individual enterprises with respect to power supply, it will also depend on the requirements for supply of blast air, compressed air, heat etc.

3 Control and Monitoring of LDS Distribution Plants

The control centre enables operation of the selected LDS distribution plants as well as the very end distribution plants with activation and deactivation of power features. This purpose is served by the system for remote control and monitoring of EHV and HV distribution plants.

Individual distribution plants are fitted with automatic units and data concentrators. Those are linked to the communication server via the technological grid to ensure processing of the operation data, this server is provided with a proper data backup facility. The server allows gathering all the communication with clients at distribution plants, it communicates with client applications located at the control centre, provides information from distribution plants feeding it into the system for automatic secondary control of voltage, transfers information into the server for summary measurements and the control centre board, it saves the values and status details into the databases and sets the exact time.

The control centre is formed by two control workstations. These enable lucid overview of status at each distribution plant or cell, operation of power features, imaging

and saving of failure logs and status changes for particular power features. The system keeps track of technological events in a timeline manner and records all the events accordingly. Images of particular distribution plants show diagrams of cell locations, their fittings, linkage and measured values. These images allow remote control over cells, acknowledgement of statuses and failures.

3.1 Operation of the Main Distribution Plant of the Combined Heat and Power (CHP) Unit

The main distribution plant of the combined heat and power station is split in two sections. It features a cellular-type open design with four lines intervening with five separate areas (bus, switch, reactor, reactor centre and cabling areas). The distribution plant is formed by three systems, split by means of longitudinal bus disconnectors. The main distribution plant is provided with 6kV cable lines connected to the terminal and nodal distribution plants in the industrial premises.

The operation of LDS is managed by the shift supervisor in control centre. He performs the analysis of reliable and safe LDS operation. He also needs to deal with effects developed due to unexpected operation issues within the transmission and distribution systems.

Most handling operations in the main distribution plant are performed from the control centre on remote basis. Emergency control features include the option to operate the systems on site, using controls that are situated right on cell control cabinets. The remote control does not involve earthing blades, arresters and longitudinal disconnectors. Any handling of these are subject to orders from the shift.

Switches are activated by means of three phasing fields. The main phasing field is fitted with both the manual phasing as well as the phasing device Synchronotakt.

3.2 Monitoring of Distribution Plants with CCTV

Performing operations at distribution plants requires visual inspections of the system status after every completed operation. However, this task requires an operator present at the plant. The problem is how to remove a failure that requires performance of several operations at multiple distribution plants. The supply restoration lead-time is affected by both the operations performed as well as the operator's response and arrival time (15 – 20 min). That causes results in greater production losses due to black-outs.

One of the options to eradicate the response time of operator that needs to attend the distribution plant in person would be to install a CCTV system on site. The system should provide the following options:

- recording and playback of image captured by CCTV cameras and accessibility of such data via a web browser,
- working independently of the operation system,
- focusing CCTV cameras on particular cells as needed,
- switching the lighting at the distribution plant on/off.

4 Reliability of a Selected Distribution Plant

The reliability of LDS refers to its ability in ensuring uninterrupted and quality supply of electric power to end users while keeping the required parameters, especially the frequency, output and voltages within intervals and periods defined by particular technical conditions. The entire LDS is required to provide highly reliable service with a very low failure rate to avoid any failures in supply of electric power to end users, supported by the method of operation and backup employed.

Potential quantification of reliability level [2]:

- failure rate λ (year⁻¹)
- mean time to repair τ (h)
- faultless operation probability R (-)

The agenda of reliability can be considered in two aspects. The first one deals with acquisition of input data, the second one is the reliability calculation itself.

Completion of reliability calculations relevant to the selected distribution plant rated for 6 kV in the year 2010 required the analysis of operation, maintenance and failures of the distribution plant monitored together with all the supply lines feeding the distribution plant. The input data has been sourced from operation logs kept at the control centre.

Fig. 1 shows the diagram of potential supply to the industrial enterprise using the distribution plant R3 6 kV B. The aim is to determine reliability of supply provided to this enterprise.

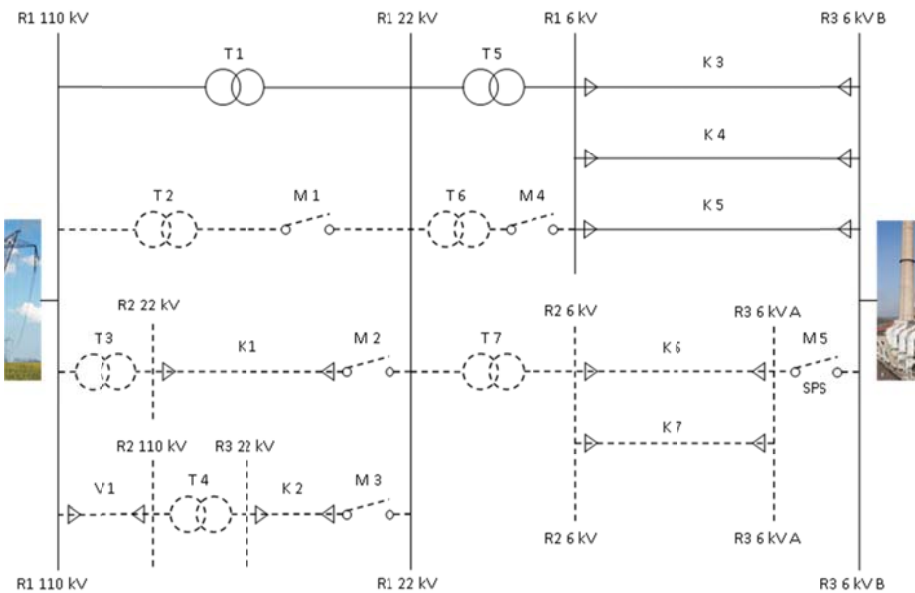


Fig. 1. Reliability diagram for power supply to industrial enterprise using the R3 6 kV B distribution plant

The reliability level has been determined using the modified method of reliability diagrams [3] that assumes the following:

- compilation of a reliability diagram,
- assigning individual elements with the relevant reliability values,
- progressive simplification of the reliability diagram.

The uninterrupted supply of electric power to VHV and HV distribution plants in industrial grids is a redundant prerequisite for faultless production process. Any power blackouts must be removed and the power supply to the affected area restored as soon as possible. That is why parallel connection of elements is used as the cold backup. As far as this connection is concerned, the calculation will also need to include the handling time τ_M .

The analysis helped to determine reliability indicators for power supply lines to the LDS from the superior grid as well as reliability indicators for individual elements [3]. The year 2010 did not see any suspension of bus operations at any 6 kV, 22 kV or 110 kV distribution plants for repair or maintenance purposes. That is why reliability of these buses was considered absolute for calculation processing.

The calculation itself has been completed in several steps. The first aspect to be realised represents the supply options for the particular distribution plant. The next steps involves compilation of the reliability diagram and its further simplification.

The results obtained through calculations have been plotted in Table 1 showing the overall reliability of power supply to the distribution plant R3 6 kV B, including several backup options. The table is split into three vertical parts. The first part defines probability of faultless operation of power supply to the distribution plant R1 22 kV and its backup by three different means, the second one show the probability of faultless operation of power supply to the distribution plant R3 6 kV B with one backup option only. The third part shows the resultant probability of faultless operation, involving the entire supply line to the superior system – the 110 kV distribution plant in Vratimov - up to the distribution plant R3 6 kV B.

Table 1. Overall results of faultless operation probability

Faultless operation probability of						
R1 110 kV → R1 22 kV with the power supply transformer T 1 and cold backup:				R1 22 kV → R3 6 kV B with supply branch T 5, T 6, K 4, K 5, K 6 and cold backup:		Total
-----	T 2	T 3, K 1	V 1, T 4, K 2	-----	T 7, K 6, K 7	
0.98506849				0.99987806		0.98494837
					0.99993567	0.98500513
	0.99982116			0.99987806		0.99969924
					0.99993567	0.99975684
		0.99998002		0.99987806		0.99985808
					0.99993567	0.99991569
			0.99997985	0.99987806		0.99985791
					0.99993567	0.99991552

The resultant values represent the actual probability of faultless operation achieved during the year 2010. However, their magnitude is affected by many different factors. The essential ones include correct evaluation of any failure event by the shift supervisor and implementation of the relevant operative remedy. Inaccurate information has a negative impact on proper evaluation of these events. Example: a loss of certain section of the communication cabling to thieves results in vague, distorted or even missing signals. Another factor to be considered is the method employed to control the distribution plant that requires the remedy. That means either remote or on-site controls. The result is also affected by arrival times of operators when attending the distribution plant, as well as their ability to work in a swift, prudent and especially safe manner to complete the operations required.

5 Conclusion

Gathering and subsequent analysis of a large amount of materials referring to operation of the LDS in 2010 allowed compiling the database to document the overall behaviour of the system with respect to its operation, maintenance and failures, including time stamps for each occurrence. That enables use of an unexacting mathematical tool to enumerate reliability indicators for the selected section of this system. The distribution plant selected for the reliability analysis calculation was rated for 6 kV. The supply of power to each distribution plant with any voltage rating is subject to the approved so-called "modular connections" kept under regular conditions of faultless operation. If needed for maintenance or repair of any power feature, the shift supervisor at control centre decides to switch over to the alternate power supply to the particular distribution plant or the area affected. That means each feature within the LDS is provided with a backup, this is an essential requirement for industrial grids.

ACKNOWLEDGEMENTS

This work was supported by the Grant of SGS VŠB - Technical University of Ostrava (No. SP2014/187).

The results of the thesis are published in:

- conference EPE 2013, Kouty nad Desnou
- journal paper *Przegląd elektrotechniczny*, vol. 89, 11/2013

References

1. Hradílek, Z.: *Elektroenergetika distribučních a průmyslových sítí, VŠB - TU Ostrava*, 2008, 208 p.
2. Brown, R. E.: *Electric power distribution reliability*. Marcel Dekker, New York, 2002, 365 p.
3. Todinov, M. T.: *Reliability and Risk Models: Setting Reliability Requirements*. Wiley, England, 2005, 322 p.
4. Tůma, J., Rusek, S., Martínek, Z., Chemišinec, I., Goňo, R.: *Spolehlivost v elektroenergetice*. ČVUT Praha, 2006. 291 p.

A Train Simulation

Maroš Ďurica

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
d.maro@seznam.cz

Abstract. Demandingness of the energy calculation by means of the rail vehicle momentary output time integration method lies in a need for detailed input data as a tachogram - a time-speed relation and a corresponding passed distance which can be obtained by a calculation or from a train entrepreneur. A locomotive load value, comprising wagons and resistance components of a track, determines a required tractive power value on a wheel circumference. Each traction power value is assigned to a traction motor current value in traction characteristics, which can be further used for a locomotive output calculation. To make the calculation more accurate a model has been created in the EMTP – ATP software, its advantage is to simulate a dynamic behavior of a model which is affected by more influences at the same time. The simulation time can be shortened in an exact ratio to the real time and thus an evaluation and implementation of changes of the model parameters can be performed more quickly.

Keywords: rail transportation, mathematical model, vehicle driving, motors, impedance.

1 Introduction

A run of a train with a 182 type locomotive has been simulated in the Přerov - Česká Třebová track segment (track no.270 according to the Czech Railways nomenclature), to be specific, between the stations with traction substations Grygov, Červenka, as shown in Fig. 3, where a placement of the traction substations corresponding with the actual condition can be seen [4]. By means of the EMTP – ATP programme the progression of voltage and current in traction motors during various periods of a run has been simulated. The output values have been further used for a calculation of the train electrical energy consumption by means of the locomotive momentary output time integration method.

2 Train Data

The 182 type locomotive output control is a resistance one, traction motors are direct-current six-pole series electric motors sequenced in three ways:

- Series (6 motors in a series, $U = 500 \text{ V}$),
- 1. Series-parallel (2 legs of 3 series-connected motors, $U = 1000 \text{ V}$),
- 2. Series-parallel (3 legs of 2 series-connected motors, $U = 1500 \text{ V}$).

The locomotive has been loaded with eleven Faccs type freight wagons of 657 t total weight, as shown in Figure 1 [2, 6].



Fig. 1. 182 type locomotive and Faccs type freight wagon

3 Data of Traction Mains and Traction Substations

The two-side power supply of the traction mains is realized by a Cu trolley wire 150 mm^2 and a Cu suspension cable 120 mm^2 together with an AlFe line feeder 240 mm^2 with $0.06 \text{ } \Omega/\text{km}$ specific resistance. The basis of the traction substation is a three-winding traction transformer $23/ 2 \times 2.5 \text{ kV}$ with a winding in Yyd, as it is shown in Figure 2. In the same Figure there is also a corresponding model of a rectifying block in the EMTP – ATP comprising a BCTRAN transformer with corresponding values of a traction oil transformer of 5.3 MVA output which is made by Power – Energo company [3].

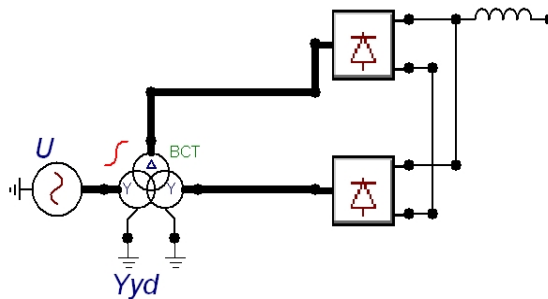


Fig. 2. The traction transformer and twelve-pulse rectifier model in the Emtp – Atp

To save a space in the scheme in Figure 3, the block with the transformer and the rectifier is depicted as a square with a designation GRP. Each of the substations consists of four blocks connected in parallel, connected to the trolley wire through a

choke of 4 mH value, in the simulation it helps to smooth a progression of the rectified voltage and restrains the rate of the short-circuit current rise.

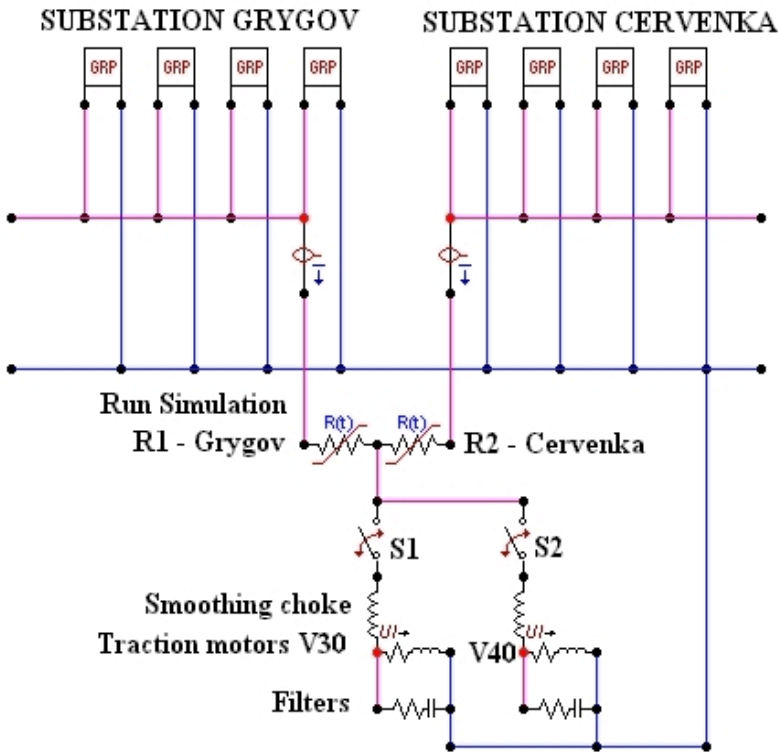


Fig. 3. 182 type locomotive traction motors in the series-parallel connection and a circuit diagram for the train run simulation in the Grygov – Červenka segment in the EMTP – ATP

4 The Simulation of the 182 Type Locomotive Run in the Grygov – Červenka Track Segment

The train run according to the time-speed tachogram begins with starting up to 30 km/h speed in the Grygov station, goes-on at a constant speed for a minute and accelerates again up to 40 km/h speed which is maintained till the arrival to the Červenka station where the train begins to brake to stop. According to time the passed distance and the corresponding train tractive power has been calculated according to the particular phases of the run and the reduced gradient value, as shown in Figure 4.

The train run real time 2640 s is too long for the simulation in the EMTP – ATP programme, therefore it has been speeded-up a thousand times:

$$\frac{\text{Real time}}{1000} = \text{Simulation time (s,-, s)}. \tag{1}$$

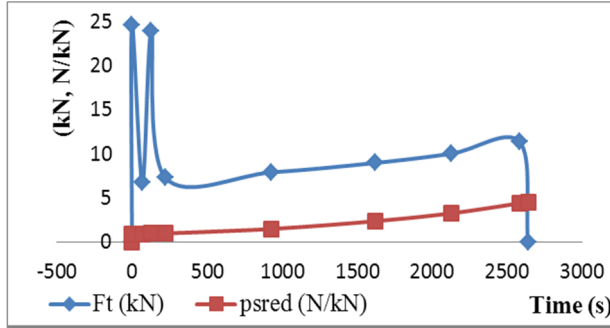


Fig. 4. Required tractive power on a wheel circumference and during the constant speed run, where it is increased depending on the track reduced gradient

For the 27.86 km distance between the stations the line resistance value is 1.67172Ω . The train run simulation according to the progression of voltage and current V30 to V40 between the Grygov – Červenka stations is solved by two time-dependent resistors with a VA characteristic. The R1 resistor Grygov resistance value is changing depending on the distance in the segment (the given distance in km $\times 0.06 \Omega$), while the R2 Červenka resistance value is decreasing, so that the sum of the resistances in each time step is equal to the line resistance 1.67172Ω between the stations. Determination of the motor resistance values for the simulation has been performed according to the locomotive total current, whereas the current of one traction motor needed to be known [5]. Motor groups sequencing for 30 km/h speed is series-parallel with pre-sequencing a resistance step no.25, where one motor current is 225 A (corresponding tractive power is 135 kN) and the locomotive total current is determined according to the formula below:

$$I_{30} = \frac{6 \cdot 225 A}{3} = 450 A . \quad (2)$$

A calculation of the resistance for the simulation:

$$R_{30} = \frac{3000 V}{450 A} = 6.66 \Omega . \quad (3)$$

Motor groups sequencing for 40 km/h speed is series-parallel with a shunt section no.2, where one motor current is 252 A (corresponding tractive power is 150 kN) and the locomotive total current is determined according to the formula below:

$$I_{40} = \frac{6 \cdot 252 A}{3} = 504 A . \quad (4)$$

A calculation of the resistance for the simulation:

$$R_{40} = \frac{3000 V}{504 A} = 5,95 \Omega . \quad (5)$$

5 Calculation of Electrical Energy Consumption of a Freight Train and the Grygov and Červenka Substations

A calculation of electrical energy consumption for the run and the Červenka and Grygov substations has been performed for the run at 40 km/h speed by means of the locomotive momentary output time integration method [1]. A period of a calculation step was used: 0.0001, to which a simulation time step 0.05 s corresponds according to the output data file View LIS file 0.05 s = 50 s real time step = 0.01388 hour.

The electrical energy consumption has been calculated for 40 km/h even speed, for this run phase the traction motor output has been calculated and integration for the simulation time step $t = 0.01388$ h has been performed.

$$f = Power = \frac{V_{40} \cdot I_{40}}{1000} \text{ (kW)} . \quad (6)$$

For E182 energy consumption calculation the integration has been performed using the Mathcad software, the result is a matrix of values of electrical energy consumption for the given time step. The result of the sum of all elements of the matrix is the total electrical energy consumption:

$$\sum E182 = \int f(t) \text{ (kWh)} = 914,5 \text{ kWh} . \quad (7)$$

In the railway transport the specific energy consumption (W h / t.km) is determined which is the total consumption rate (W h) related to t km (ton-kilometre). At energy consumption 914.5 kW, the specific energy consumption is (designated as W_{msee}):

$$W_{msee} = \frac{T_{40} \cdot \sum E182 \cdot 1000}{M \cdot L_{40}} = 29 \text{ Wh/tkm} . \quad (8)$$

where T_{40} is the run time at 40 km/h even speed – 0.6389 h,

M is a weight of the locomotive and wagons – 777 tons,

L_{40} is a length of the passed distance at 40 km/h even speed – 26.07 km.

It is necessary to add the specific energy consumption (designated as W_{pp}) of auxiliary drives as compressors, ventilators, control circuits; in literature, for 182 type locomotive the value $P_{pp} = 52.2$ kW is mentioned as the accessory circuits input.

$$W_{pp} = \frac{T_{40} \cdot P_{pp} \cdot 1000}{M \cdot L_{40}} = 1,6 \text{ Wh/tkm} . \quad (9)$$

The specific energy consumption for the locomotive collector at 40 km/h even speed **in summer** is $W_{msee} + W_{pp} = 30.48$ W h / t km.

6 Conclusion

In the paper a problem of the specific energy consumption calculation has been solved for a selected train – 182 type locomotive with Faccs freight wagons when running at 40 km/h even speed in the Grygov – Červenka track segment. A corresponding simulation in the EMTP – ATP programme has been performed, the resulting voltage and current progression values have been used for the specific energy consumption calculation by input integration for the given time. The specific energy consumption result for summer is 30.48 W h/t km, which corresponds to a Pn through freight train consumption at a track reduced gradient up to 4.5 ‰, as shown in the nomogram on Figure 6.

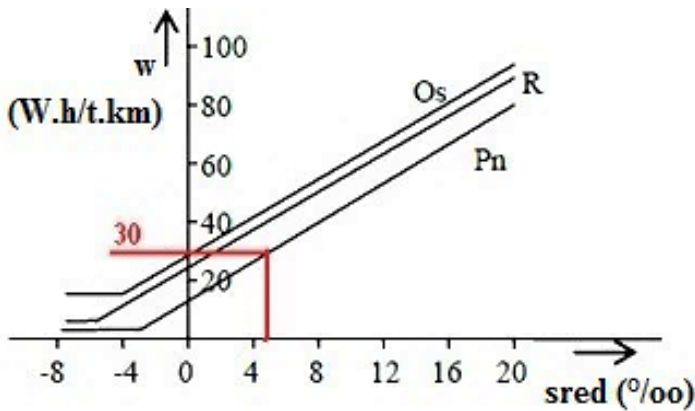


Fig. 6. Nomogram for determination of orientational specific energy consumption for typical trains

References

1. J.Paleček, *Electroenergetics in Transportation, study materials for lectures*, (Book style), Ostrava, VŠB – Technical University of Ostrava, Winter Semester 2010.
2. Encyclopedia of locomotives. *Locomotive 182* [online]. c2004. [retrieved.2013-11-18]. URL :<<http://www.atlaslokomotiv.net/loko-182/>>.
3. Power Energo. *Traction transformer* [online]. c2011. [retrieved.2013-10-14]. URL:<<http://www.power-energo.cz/p114-soubory-ke-stazeni-specialni-transformatory.html/>>.
4. Railway tracks history. *Tracks data* [online]. c2011. [retrieved.2013-09-05]. URL :<<http://historie-trati.wz.cz/>>.
5. A mathematical model for a train run. *Simulation of a train run* [online]. c2008. [retrieved.2014-05-01]. URL :<<http://diplom.utc.sk/wan/1881.pdf/>>.
6. Company RYKO PLUS. *Wagon type Faccs 38 m³* [online]. c2010. [retrieved. 2013-10-01]. URL:<<http://www.rykoplus.cz/doc/typovy-list-faccs-38-m3.pdf/>>.

Measurement of disturbing glare caused by interior luminaires with LED modules

Barbara Helštýnová, Karel Sokanský, Tomáš Novák, and Zbyněk Carbol

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{barbara.helstynova, karel.sokansky, tomas.novak1, zbynek.carbol}@vsb.cz

Abstract. The aim of this paper was to compare the results measured using the UGR method of computer programs and practical implementation. The reason for this comparison is newly deployed LED luminaires. For LED luminaires may happen some applications to the user's subjective complaints of glare, although calculations for project documentation meet the conditions for glare reduction. The basic problem is a very small area from which the LEDs emit. The optical active area has significantly higher luminances than conventional light sources. The energetic advantage of lighting systems using LEDs will lead to more widespread usage, provided the lighting. It is therefore necessary that the issue involving high luminances in the direction of view of the user to solve as soon as possible.

Keywords: disturbing glare, UGR method, luminance analyzer

1 Introduction

This article describes the outcomes of measuring the luminance using the luminance analyzer and follow determination UGR using measured luminance in the new plan office equipped with recessed LED lighting. Office was selected for comparison with LED lighting with maximum attention on the glare reduction. Used luminaires use LED modules as private sources of radiation, which is itself a significantly lower luminance than discrete LED. Lamps are further limited by the reflectors, which significantly prevent the emission of light flow into critical viewing directions. Lighting system with these lights should be no discrepancies between the calculated values and the measured values of the UGR.

2 Parameters installed luminaires

Lamps are down lights with LED light source and reflector without cover glass. Parameters input power 11W/1100lm (S1) and 21W/2000lm (S2). This kind of lumi-

naires was used because of good distribution with minimal distribution to problematic angles from 45°-85°.

3 Requirements of the glare

Norm EN 12464-1:2012, Light and lighting - Lighting of work places - Part 1 - Inside workspaces, specifies the minimum shielding angles for specified highlights of light sources as follows.

Standard specifies the average limit highlights luminaires can mirror lamp shades in flat panel displays at an angle of 65 ° or more as follows:

Table 1. Average luminance limits of luminaires, which can be reflected in flat screens

Screen high state luminance	High luminance screen $L > 200 \text{ cd}\cdot\text{m}^{-2}$
Case A (positive polarity and normal requirements concerning colour and details of the shown information, as used in office, education, etc.)	$\leq 3000 \text{ cd}\cdot\text{m}^{-2}$

3.1 Comparison shielding luminaires at critical angles – 65°

The calculation of the average normal luminance light source provided cosine radiation light source:

- Diameter LED light source : $d = 67 \text{ mm}$
- Luminous flux : $\Phi = 2000 \text{ lm}$
- The luminance of the light source:

$$L = \frac{l}{S} = \frac{\frac{\phi}{\pi}}{\frac{\pi \cdot d^2}{4}} = \frac{4 \cdot \phi}{\pi^2 \cdot d^2} = \frac{4 \cdot 2000}{\pi^2 \cdot 0,067^2} = 181 \text{ kcd} \cdot \text{m}^{-2} \tag{1}$$

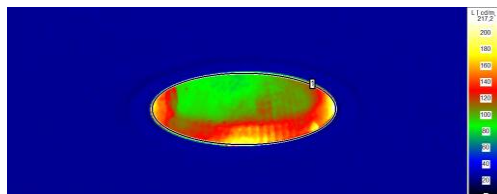


Fig. 1. Luminance map luminaires 2000 lm at an angle of 65 °

- The measured luminance of the lamp according to the luminance maps at an angle of 65 °- $L_{65} = 116.1 \text{ cd.m}^{-2}$.
- Luminaire appropriate for the aperture of the light source: shielding angle is greater than 25 °.
- Luminaire appropriate for the use in workplaces with display, $L_{65} \leq 3000 \text{ cd.m}^{-2}$.

4 UGR method according to CIE 117 – 1995 Interference lighting in indoor lifting

Discomfort glare is defined by the CIE publication 17 4th 1987, "International Lighting Technical Dictionary" as: glare, which gives an unpleasant feeling, but not necessarily worsen visual performance.

Three methods have been developed glare rating:

- The formula for uniform evaluation of glare (UGR).
- Derived table method enabling easy comparison of different lighting options.
- Method of limit lines derived luminance assist engineers with the development of lamps and providing designers approximate information about the appropriateness of luminaries.

In this case a uniform formula glare rating (UGR):

$$UGR = 8 \log \left(\frac{25}{L_b} \cdot \sum \frac{L^2 \cdot \omega}{p^2} \right) \quad (2)$$

- UGR is unified glare rating [-]
- L_b is the background luminance [cd. m^{-2}]
- L luminance of each light in the direction of the eye of the beholder [cd. m^{-2}]
- ω solid angle under which the observer sees the shining luminaries of each [sr]
- p official position by Guth for each luminary (displacement in direction [-])

4.1 Derived table method

This method involves calculating the uncorrected glare factor intended for publication, together with the photometric parameters luminaires, a majority of them for each type of luminaires supplied by the manufacturer.

5 Descriptions of measurement

Light sources were burned-in (last at least 100 hours burn time). Measurements were performed to stabilize luminous flux of lamps. The ceiling height is 3 meters and luminaires too. Lighting system isn't affected by daylight or street lighting. The meas-

ured area was installed recessed luminaires type S1 and S2, their distribution was irregular. In selected rooms with higher expected value UGR photo was taken, which, after evaluation has become a luminance map area. Captured photos were evaluated in software LMK LabSoft. The program was calculated for each luminaire following data: average luminance, solid angle, positions the observer. Height measurement plane UGR is 1,2 m, so the normalized height of the sitting person.

Used instruments:

- The luminance analyzer LMK Mobile Advanced - range 0,3 ÷ 50 000 cd m⁻², kcd.m⁻².
- Luxmeter PRC Krochmann RadioLux 111, S / N: 100612, photometric head light (λ) correction S / N: 121027-1.

5.1 Calculation of reflectivity

The reflectivity factor of the ceiling and walls are the same because of the same painting, the floor was laid carpet. To calculate the reflectivity factor was measured illuminance using a luxmeter and highlights using LMK analyzer. Reflectance factor is calculated as:

$$\rho = \frac{\pi \cdot L}{E} \cdot 100 \tag{3}$$

- ρ is reflectivity factor [%].
- L is the luminance of the wall / ceiling [cd.m⁻²].
- E is the value of illuminance wall / ceiling [lx].

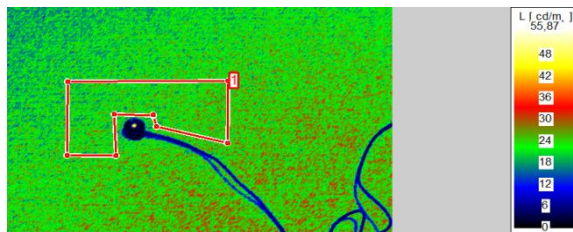


Fig. 2. Example of luminance map of the floor

Table 2. The measured values and the calculated factor of reflectivity value

	Average luminance L _{avg} [cd m ⁻²]	Illuminance E [lx]	Factor of reflectivity ρ [%]
Wall /ceiling	42,3	160,1	83
Floor	20,6	270,2	24

Calculated reflectivities are 83/83/24, which corresponds closely to the first column in Table UGR (Table 2). Expected value for UGR light S2 is in the range of 12.5 to within + 5.3 to - 8.1, as expected spacing fixtures about 2m.

6 Evaluation of measurement

Calculation view 1 of the adaptation luminance of vertical illuminance:

$$L_b = \frac{E_s}{\pi} = \frac{293,5}{\pi} = 93,42 \text{ cd} \cdot \text{m}^{-2} \tag{4}$$

To the next formula are implements dates from Tab. 4.

Table 3. Auxiliary table calculation, view 1

Lamp	Horizontal angles of view [°]	Vertical angles of view [°]	Factor of position p	Average luminance L_{avg} [cd/m ²]	Solid angle Ω_{lamp} [sr]
1	5,55	77,85	1,53	2081	0,00509818
2	40,5	54,6	4,5	2370	0,00136359
3	3,3	51	5	2322	0,00224662
4	63	43,8	8,9	401	0,0002897
5	52,8	29,4	14	420	0,00029382
6	38,25	34,5	10,6	326	0,00059645
7	2,7	31,5	14,7	295	0,0006154

Calculation of the UGR:

$$UGR = 8 \log \left[\frac{0,25}{L_b} \cdot \sum \frac{L^2 \cdot \omega}{p^2} \right] = 11,51 \tag{5}$$

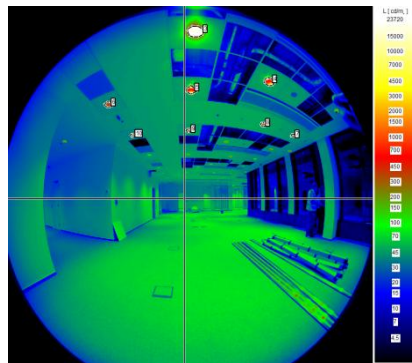


Fig. 3. View 1 from a height of 1.2 m (occupant)

There are information about luminaires, which are visible observes angle of view (Fig. 8).

7 The evaluation

The illumination area was made a set of measurements, calculations and comparisons, whose purpose was to assess the suitability of the lighting system with LED light sources in space agencies in terms of luminance ratios and disturbing glare. Index disturbing glare UGR determined based on the measured reflectance surfaces and standardized UGR table lamp designed for S1 set at 10.6 (+5.3 , -8.1) and for luminaire S2 12.7 (+5.3 , -8 , 1) , this values are from Table II, the spacing between of about 2 m , depending on the position of the observer. From the measured luminance map taken in measured room was calculated index glare UGR , which came to 11.51(S1) and 14.69(S2) , which validates the table method . Use luminaries are suitable for use in offices and workplaces with the display and the whole lighting system implemented appropriate for the index glare UGR according to standard EN 12464-1 Light and lighting - Lighting of work places - Part 1: Indoor work places, 2012 for office space (offices) , pursuit of writing , typing , reading, data processing and other activities with the desired value $UGR < 19$. The aging lighting system will gradually decrease the brightness of light sources and surface reflectance. Therefore, highlights and calculated indices are highest glare and gradually will gradually decrease.

It was the LED luminaires with LED modules with a large area, the results obtained by measuring a comparable table method.

The measured area was without equipment. It should be noted that the factor glare UGR addresses only direct discomfort glare caused by lights.

1. ŠKODA, J.; BAXANT, P.; KRBAL, M.; SUMEC, S.; PAVELKA, T. Photometry of LED sources. *Przeglad Elektrotechniczny*. 2013. 89(6/2013). p. 341 - 344. ISSN\~0033-2097.
2. ŠKODA, J.; BAXANT, P. Nekonvenční metody měření fotometrie svítidel. In *Proceedings of the 9th International Scientific Conference Electric Power Engineering 2008*. První. Brno, BUT, FEEC, Department of Electrical Power Engineering. 2008. p. 419 - 423. ISBN 978-80-214-3650-3.
3. CEN/TC 169 N 835. Light and lighting — Lighting of work places — Part 1: Indoor work places. 2010. 57 p.
4. ČSN EN 12464-1. Světlo a osvětlení – Osvětlení pracovních prostorů – Část 1: Vnitřní pracovní prostory. Praha: Český normalizační institut, 2003. 22 p.
5. TNI 36 0450. *Rušivé oslnění při osvětlení vnitřních prostorů*. Praha: Český normalizační institut, 2004. 36 p.

Determination of Coefficients for Matrix of Importance

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 paper describes restoration methodology for 110 kV power lines. This method is based on reliability centered maintenance (RCM). The distributed energy and effect of line failure are used to determine the importance of line. For determine the technical condition is use database of faults. The results and analysis of first and second version are used to modify the last version of methodology.

Keywords: Power line, RCM, reliability, technical condition, importance of line, restoration, matrix

1 Introduction

To maintain distribution networks, most distributors follow the schedule of preventive maintenance which stipulates procedures and times of preventive maintenance. The modern maintenance is based on the principle of reliability centered maintenance (RCM) which is directly dependent on the technical condition and importance of distribution system equipment. By employing RCM principles, maintenance can be economized and the restoration of distribution network can be properly planned. Our method aims to establish the sequencing of pieces of distribution network equipment for restoration. The new version of methodology is modify by results and analysis of older versions of methodology. This new version is more accurate determination of input parameters.

2 Restoration method for 110 kV power lines

Restoration method for 110 kV lines is based on the principles of RCM [1]. To determine the technical condition, the 110 kV line failure database is used. The line importance is determined with respect to the power transmitted by the line and also the effect of line failure. The outcome of the method is a quantity we call ‘restoration priority’ PO . Restoration priority is a relative value within the range of 0 and 100 %. Higher PO value means increasing urgency of restoration.

$$PO = (100 - TS) \cdot (1 - k_{TSDV}) + DV \cdot k_{TSDV} \quad (1)$$

Restoration priority is calculated in the relation (1), where TS is the technical condition of the line, and DV is the importance of the line. Coefficient k_{TSDV} divides the influence of TS and DV on final restoration priority value. If the value is more than 0.5, line importance is preferred and vice versa.

Not only accurate assessment of technical condition, but also connection of particular maintenance area with the appropriate line is important. The technical condition is relevant to the maintenance area. However, line importance is relevant to the whole line. Most lines are divided in two and more maintenance sections.

2.1 Importance of line

The line importance calculation is based on the yearly distributed energy and the effect of a failure for the whole line. Distributed energy is based on daily readings of average, minimum and maximum values, taken every hour in the whole of the distribution line. Total calculation is automatized with help of macros in Excel and is run with all the items including voltage, current etc. for the whole year and 365 files. Before reading the distributed energy, correctness of number of days on both ends of lines is verified. An approximate value of distributed energy is sufficient to assess importance, therefore all 110 kV lines are assorted in categories, similarly with points of technical condition (Table 1). Values E_x determine the upper value of an interval of the given category.

Table 1. Categorization of distributed energy

Distributed energy	E_1	E_2	E_3	E_4
Category	1	2	3	4

The second input parameter for line importance is the effect of failure. This parameter defines the consequence of a line failure, or in other words, it describes a back-up of a line. In the analysis we determined 7 possible effects of a failure:

- 1 – “No effect” – no manipulation necessary,
- 2 – Back-up by 110 kV line of approx. the same length,
- 3 – Back-up by 110 kV line of approx. the same length with limited power,
- 4 – Back-up by a longer 110 kV line,
- 5 – Back-up by a longer 110 kV line - limited power,
- 6 – Back-up by MV line,
- 7 – Back-up by MV line – limited power.

To assess the line importance is simpler than to assess the technical condition as no calculation is necessary. Both input parameters are categorized, therefore when the effect of failure is e.g. type 3 and distributed energy type 3, then the final line importance must equal 30%. The matrix values are determined on the basis of the input parameters analysis and theoretical analysis.

Table 2. Importance of line - Matrix

MATRIX Importance of line		Distributed energy			
		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 %

The values prove that the line backups that can carry only reduced power and those that can be backed up only by MV lines have higher values of line importance. The higher is distributed energy, the higher is final value of line importance.

2.2 Technical condition

Technical condition is assessed on the basis of a failure database (in xlsx format) provided by the information system of a distribution system operator. One failure can be described in as much as 216 entries. As most of these entries are irrelevant, the database is reduced to the total of 30 important entries.

A failure priority determines the urgency of maintenance. Up to now, we differentiate from 4 failure priorities:

- 1 – URGENT,
- 3 – MONTH,
- 4 – YEAR,
- 5 – CONTROL at ŘPÚ.

A failure priority is used to evaluate the technical condition of maintenance sections; therefore there are both – a verbal and a number code priority indication.

The quantity entry helps to make the technical condition assessment more accurate. It does not indicate a total number of faults in one maintenance section, but a number of faults of the same type found at one inspection (e.g. faulty insulators, missing warning plates, etc.). The key issue of quantity is that it can refer to pieces or meters, depending on the failed object. When the failed element is a line or a cable, then the value is in meters. Other 110 kV line elements are always in pieces (support, bracket). Therefore, quantity is categorized in three types (Table 3).

Table 3. Categories of quantity

Quantity			
Meters		Pieces	
up to 25 m	1	up to 10 pcs	1
from 26 to 50 m	2	from 11 to 30 pcs	2
over 51 m	3	over 31 pcs	3

To assess the technical condition, the number of failures in every maintenance section in relevant priority and category must be defined via contingency table.

The calculation of technical condition has two steps. Firstly, points of technical condition (BTS) are calculated for the relevant section by an equation (2), where VP_x is weight value for priority x. Variable p_{M1x} means the number of failures with x priority in category 1 of quantity; p_{M2x} a p_{M3x} are numbers for higher category of quantity. Coefficients k_2 and k_3 increase the final value of *BTS* with respect to the quantity category.

$$BTS = \sum VP_x \cdot \sum (p_{M1x} + k_2 \cdot p_{M2x} + k_3 \cdot p_{M3x}) \quad (2)$$

Secondly, the technical condition of a maintenance section is assessed according to results in Table 4. The BTS_x values determine the upper limit of points of technical importance span interval which is assigned one per cent value of technical condition.

Table 4. BTS limits

BTS limits				
BTS_1	BTS_2	BTS_3	BTS_4	>BTS_4
Technical condition TS				
95 %	80 %	70 %	60 %	40 %

3 Modification of methodology

The first results and analysis of methodology showed less accuracy of the methodology. Two major modifications are made. In first modification we modify the determination of technical condition – more categories of BTS limits, prioritizing of TS etc. The second modification concerned on the importance of line. Firstly we increase the distributed energy categories from 4 to 6. Secondly we determine new values in matrix.

3.1 New importance matrix

The values in the matrix of importance were based on theoretical assumptions [2], analyses, and theoretical calculations. These calculations drew upon reliability scheme methods, and updated reliability parameters. There was one disadvantage to

this calculation, namely that transmission capability could not be included. However, this drawback was eliminated by software SPOLEH [3].

The software SPOLEH was used to calculate probability of failure-free operation. Unlike in previous calculations, only one backup line was modeled in two lengths - 100 km, and 150 km. The current reliability parameters were applied [3], e.g. $\lambda = 0.286 \text{ year}^{-1}$ failure rate for 110 kV lines and $\tau = 2.198$ hours failure mean time.

There were six different types of modeling of failure effects, one for each of six distributed energy categories. Transformers HV/MV was modeled for MV backup lines. Transmission capability was simulated to secure $R = 0.999$ failure-free operation at the full load. Reduced transmission capability was modeled on $R = 0.9975$. Then in the ratio of distributed energy was reduced probability. Thus probability for all variants of failure effect – distributed energy was calculated. With help of these probabilities the values for the matrix of importance were acquired. (Tab. VIII). The importance values were rounded off to integer per cent. The best importance was set on 0 % and the worst importance on 60 %.

Table 5. New importance matrix

MATRIX Importance of line		Distributed energy					
		1	2	3	4	5	6
Effect of failure	1	0%	2%	3%	5%	6%	8%
	2	3%	5%	8%	11%	14%	16%
	3	9%	14%	18%	23%	28%	36%
	4	5%	7%	10%	13%	16%	20%
	5	15%	21%	27%	36%	42%	48%
	6	12%	15%	18%	22%	26%	34%
	7	19%	26%	32%	41%	51%	60%

4 Conclusion

The paper outlines the issues of 110 kV lines restoration. The input data of the methodology are failure database, distributed energy and effect of failure. The critical issue is determination of two parameters – technical condition and importance of line. Application of method should determine the value priority of restoration.

Verification of method was carried out on the data file of one distribution area. Distributed energy was determined on the basis of a yearly measurement in this area. Assessment of 110 kV line backup was determined from the network scheme. The technical condition was assessed on the basis of 44-month faults database.

The first version proved lower validity and method was modified in two following versions. The method for technical condition assessment was corrected in the second version and method of assessment of line importance was modified in the third ver-

sion using software SPOLEH. This modification is described in this paper. New version of matrix increase the accuracy of the methodology.

ACKNOWLEDGEMENTS

This work was supported by the SGS grant from VŠB - Technical University of Ostrava (No. SP2014/187) and by the project ENET (No. CZ.1.05/2.1.00/03.0069).

The results of the thesis are published in:

- konference CIRED 2013, Tábor, ISBN 978-80-905014-2-3
- článek Przegląd elektrotechniczny 11/2013, ISSN 0033-2097
- konference ELNET 2013, Ostrava, ISBN 978-80-248-3254-8
- konference EPE 2014, Brno, ISBN 978-1-4799-3806-3

5 References

1. Z. Martínek, V. Královacová: The Solution for Repairable Units. Proceedings of the 11th international Scientific Conference Electric Power Engineering 2010, University of Technology Brno.
2. V. Houdek, S. Rusek, R. Goño: Backup Alternatives for 110 kV Lines. Advances in Electrical and Electronic Engineering. 2013, vol.11 no.3. ISSN 1366-1376
3. Manuál programu SPOLEH

Research and measurement of active and reactive power supplied by the biogas station into the distribution network

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 article deals with the evaluation of measurement at a specific biogas station. The introduction described the biogas plant which was held at the measurement, which is discussed in more detail next section. There is described a measuring device that was used, followed by analysis of the measured data output - both active and mainly reactive, which has a very interesting course. The last section is devoted to higher-order harmonic current and voltage. The measured values are evaluated according to quality standard voltage to ČSN EN 50160. This work is part of a longer-term research that addresses the impact of the operation of a biogas station to the distribution network.

1 Introduction

Biogas station (BGS) in which the measurement was carried out within the territory of the Moravia-Silesia Region. BGS is situated within the grounds of an agricultural enterprise engaged in pig and processes mainly maize silage and pig manure, to a lesser extent waste from the processing of sugar beets, beet chips. The main reason was the source of manure for use liquids for wet fermentation technology.[1] One other reason this location was the possibility of using waste heat for heating the adjacent stables, office buildings, and in the summer in a newly built grain dryer.

Biogas station has an installed electrical output in the amount of 1,090 kW and 1,080 kW thermal output. Conversion of biogas into electricity makes four cogeneration units. Three identical compression ignition units have an output of 250 kWe and one spark ignition unit has an output 340 kWe. Power is brought out into the overhead line 22kV distribution network via transformer substations 0.4 / 22kV equipped with one transformer with a maximum power 1250 kVA with a nominal current of 1804 A. [2]

2 Measurement[3]

The measurement was realized at the end of January to March this year, particularly since 9 January 2014 (15:50) until March 7, 2014 (9:00) - a total of 57 days. Meas-

urements were realized of an automatic measuring device - network analyzer ENA 330 from ELCOM, a.s.. Phase voltage was measured by analyzer directly on the bus-bars in switchboard RH1 and currents in each phase using the clamp current transducers using already installed instrument current transformers 1500/5A.

Within the measurement of electrical quantities, the analyzer measured the voltages, currents, angles between voltages and currents and other variables mathematically calculated. It was also an evaluation of power quality according to ČSN EN 50160. An assessment was mainly the voltage, total harmonic distortion, long-term flicker severity, voltage decreases and unbalance.

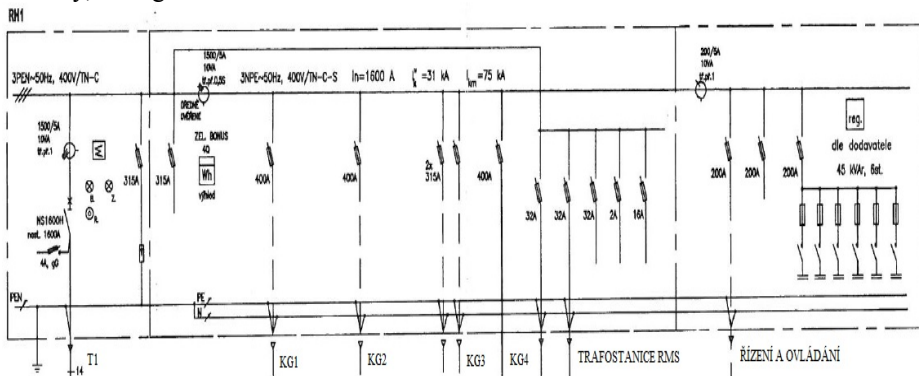


Fig. 1 Single-pole scheme of the main switchboard a biogas station

Analyzer ENA 330 meets the requirements for measuring instruments and measurement procedure given standards ČSN EN 50160, ČSN EN 61000-4-7, ČSN EN 61000-4-15 a ČSN EN 61000-4-30.

2.1 Outputs

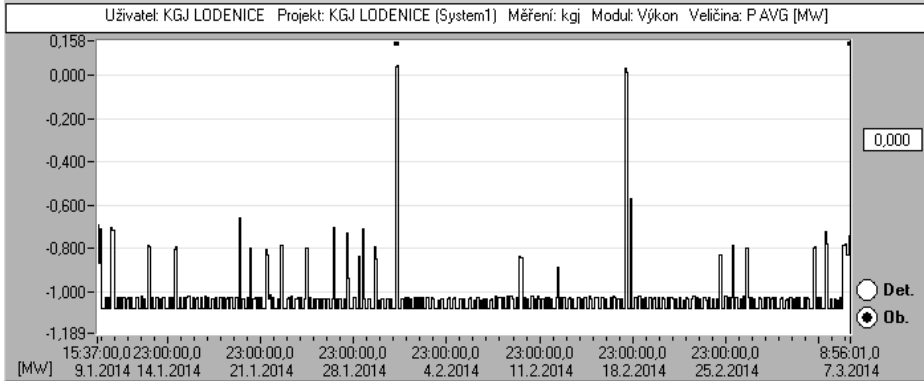


Fig. 2 The course of the active power supply to the grid

In Figure 2 we can see the overall progress of the active power supplied by all cogeneration units to the network for the entire measurement period. In the reporting period, there were only three failures of the grid, while on February 1, 2014, there were two failures in a total duration of 30 minutes. The third failure occurred on February 18, 2014 and lasted nearly one hour (54min).

The first two failures were caused by the failure of 22 kV distribution network, due to adverse weather conditions. The last failure was caused by a technical failure on two cogeneration units, which caused the need for a total shutdown of BGS.

In the period from 01.17. to 31.1., there are several recessions power output, which has been re-caused shutdown of a cogeneration unit in which it occurred is still the same problem, namely the overheating of one cylinder internal combustion engine. This problem is a long time failed to correctly repair and individual servicing help only for a few days. A major service maintenance eliminate this problem altogether in the end.

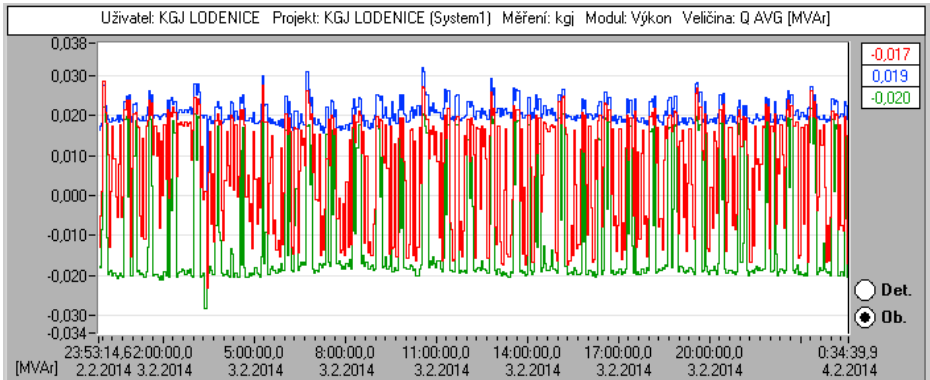


Fig. 3 Detail course reactive power in the individual phases

Figure 3 shows the course of reactive power in each phase at the time of the entire measurement. It is seen that the reactive power of each phase mostly keeps close to some specific value, and sometimes there will be some volatility. Reactive power of the second phase (blue) is around the 20 kvar inductive against the third phase (green), in which the predominant reactive power values around - 20 kvar capacity. The first phase (red) fluctuates between these two values.

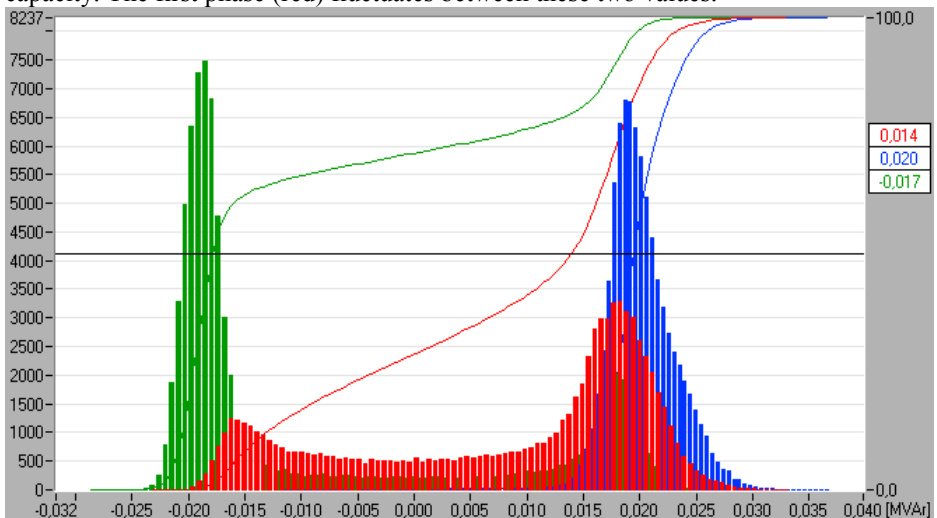


Fig. 4 The histogram of reactive power in the individual phases

Fig. 4 shows histograms of reactive power for each phase. In the first phase shows two smaller peaks, the first at around -16 kVAr, ie capacitive, and another significant peak at around 18 kVAr, ie inductive character. In the second phase the values are distributed mainly around 19 kVAr (inductive). In the third phase values are narrowly distributed around a value of -19 kVAr (capacitive). This disproportion in reactive power within each phase could indicate improper compensation setting.[4]

2.2 Higher-order harmonics

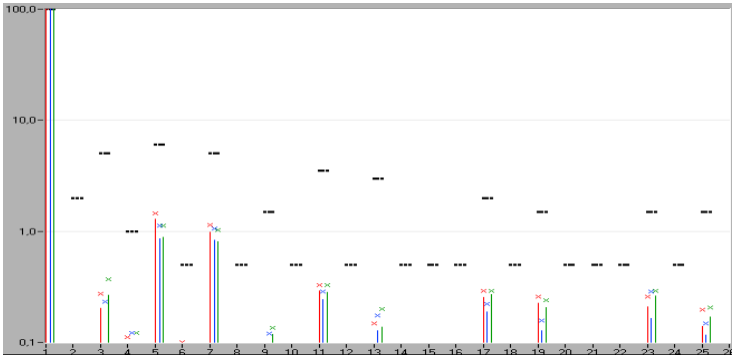


Fig. 5 The spectrum of higher order harmonic voltage marked permissible limits according to ČSN EN 50160

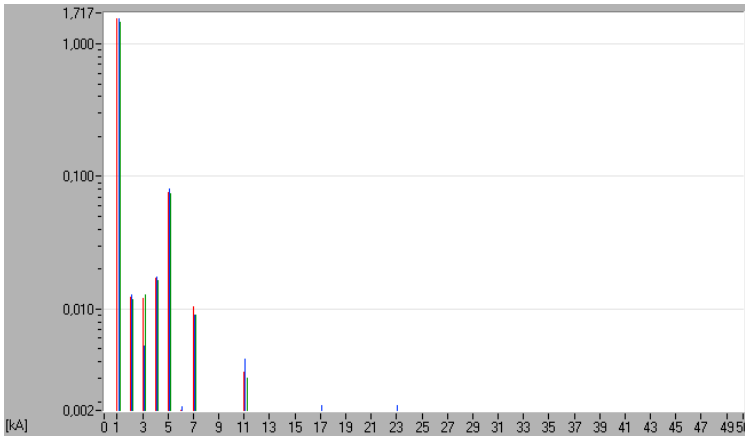


Fig. 6 The spectrum of higher order harmonic current

Fig. 5 shows the spectrum of higher order harmonic voltage marked permissible limits according to ČSN EN 50160. No limit has been exceeded, most them approaching the fifth harmonic, which reaches almost 22% of the permitted limit of $\leq 6\%$, ie 1.3% of the value of the fundamental harmonic and seventh harmonic reaching almost 20% of the permitted limit of $\leq 5\%$, ie 0, 99%. This is a logarithmic scale, so it appears that amounts to more than half of the permissible values.

Fig. 6 shows the harmonic spectrum of higher order flow from the first (baseline) to the fiftieth harmonic. In the entire spectrum of the reflected visible only lower order harmonics, ie the second, third, fourth, fifth, seventh and eleventh.

Since the spectrum of higher order harmonic voltages appear higher orders than in the spectrum of harmonic currents, which are mainly caused by synchronous generators of cogeneration units, then these higher orders voltage probably introduced from the parent network of 22 kV, which is connected to at least one photovoltaic plant.

3 Conclusion

In this article, we focused on the analysis of course reactive power, which has a very interesting course. Is currently underway analysis of the possibilities of this uneven distribution sources in the different phases. Most used motors for pumps, mixers and income-equipment are three-phases, therefore, would not cause an unbalanced load. The most probable appear possibility of poor settings of the compensating device. This variant is discussed with the supplier of the technology biogas plant. Another part of the article deals with the analysis of higher-order harmonics, which biogas station operates at 22 kV distribution network. All observed higher order harmonics fulfill the requirements of ČSN EN 50160, the most significant are the 5th and 7th harmonics which reaches slightly over half the permissible limits. This article is part of a long-term research investigating the effect of the action of a biogas station to the distribution network of 22 kV.

References

1. J. Jansa, Z. Hradilek, "Biogas Station as Storage Unit in Electric Power Grid", Proceedings of the 14th International Scientific Conference Electric Power Engineering 2013, Dlouhé Stráně, 2013, ISBN 978-80-248-2988-3
2. J. Jansa, Z. Hradilek, "Impact of Biogas Plant on Distribution Grid", Proceedings of the International Conference on Environment and Electrical Engineering 2014, Krakow, 2014
3. P. Moldrik "Měření na bioplynové stanici 2014" – Zpráva z měření
4. Biogas from Waste and Renewable Resources: An Introduction, Second Edition, Prof. Dr.-Ing. Dieter Deublein, Dipl.-Ing. Angelika Steinhauser, Publikováno online: 27 OCTOBER 2010, Online ISBN 9783527632794, available from:
<http://onlinelibrary.wiley.com/book/10.1002/9783527632794>

Concept of charging stations for electric cars

Oliver Marcincin

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
oliver.marcincin.st@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 includes an executive summary existing knowledge about the structure and service of electricity distribution network, analysis of electric car industry and its potential impact on the electrical network and also conceptual solution charging stations electric cars with minimal effects on the power network.

Keywords : active charging station, accumulation of electrical energy, electric network, electric car, charging station.

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 to the low density of charging stations, most of which are also operated an amateur way. In case of expansion of electric vehicles will also 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 energy produced by the sun with the possibilities of accumulation and its subsequent delivery to the uniform charging infrastructure.

2 Connecting of charging stations for electric cars to the distribution network

If the number of electric vehicles operators will increase, will be necessary to build sufficiently dense network of charging stations, to ensure not only the range of electric cars in its area of operation, but also allows occasional crossing at medium distances up to about 300 km [2]. This requires the deployment of charging stations along major roads within a maximum of 50-60 km.

2.1 Estimation of the electrical energy amount needed for charging vehicles

The amount of electric power consumption can be estimated on the based expected performance structure of the vehicle and their fail safe distances. The energy needed for an average range 120 km of electric vehicles in three distinctive categories gives us Table 1. The three basic categories of electric vehicles, which are characterized by different engine power were chosen for analysis [3]. Category "A" - a small urban vehicle with traction engine power about 10 kW, category "B" - standard vehicle lower middle class with performance of traction engine about 30 kW and category "C" - a higher class with performance of traction engine about 80 kW. Range is 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 charging current for the stored energy and the expected nominal voltage of 350 V battery, 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 [5], 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:

$$\begin{aligned}
 W_{ak(celk)} &= N_{(A)} \cdot W_{ak(A)} + N_{(B)} \cdot W_{ak(B)} + N_{(C)} \cdot W_{ak(C)} = \\
 &= 17500 \cdot 13300 + 6250 \cdot 30000 + 1250 \cdot 41100 = \\
 &= 471,6 \text{ (MWh)}
 \end{aligned}
 \tag{1}$$

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.

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.

A possible structure of such network shows Fig. 3.

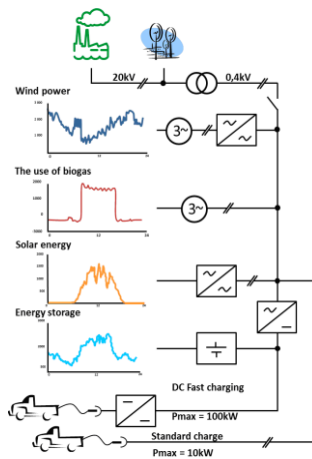


Fig. 1. The integration of active charging station into the structure with the accumulation of renewable sources [1]

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 of charging modes by 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. The problems with increased of power draw in the 3. and 5. harmonic are caused by this simple method of charging with 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 time 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.

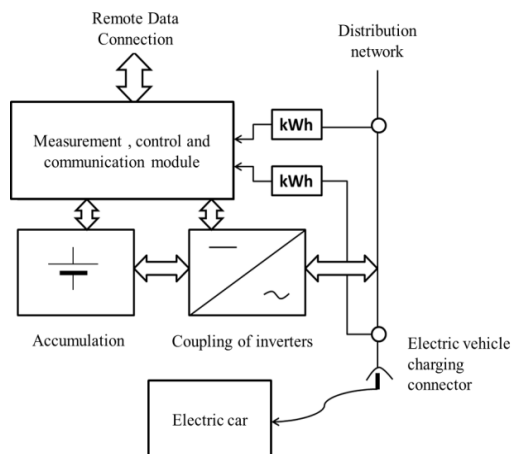


Fig. 2. The structure of the charging station for electric vehicles

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 charg-

ing 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 cover the needs of the charging electric power at a time when the network failure by drawing power from the auxiliary battery of active charging station. This solution also allows in appropriate control of 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.

4 Daily load balance of active charging station

The impact on daily load diagram by small charging stations with one to three charging sockets is minimal; therefore, it doesn't make sense to deal with it. Battery of active charging station has its justification only with regard to the possibility of charging efficient vehicles and doesn't overload the power network.

During the operation of electric vehicles in their vast majority can be expect their operation during daylight hours, allowing basic charge at night, for example, electric energy at reduced overnight rates directly from a garage by the private socket. We assume that the vehicles in larger cities will be charged from public charging stations overnight. In focus to daily load diagram makes sense to think primarily about impact active charging stations with more sockets and with greater energy stored in the battery.

The basic requirement is balancing of electric energy consumption from the network on the mean value power of daily load diagram. If we use to the accumulation all possible energy, it will ensure that the network will be loaded at night to medium power, resulting in more efficient use of the network. However, it doesn't suppress the extent of peak demand because they are made from other appliances. The important thing is that if, during the day electric vehicles need additional charge, they can be charged in a network of charging stations from the energy stored in the battery charging stations, and thus will not contribute to further increase the demand of daily peak.

For an overall proposal power cables is appropriate to determine the values of equivalent powers for morning and evening accumulation, which correspond to a constant (equable) flow of energy throughout accumulation.

If the active charging station is interconnected with a photovoltaic power station, the peak of supply energy from solar panels is equalized by the battery of station. This

contributes to more efficiency of the charging station, but also to suppress the negative impacts of photovoltaic on network stability.

With respect to the power network, at night is irrelevant if it's charged the electric vehicle or the battery of charging station. The main priority is charging an electric vehicle if it is connected, in the remaining time can be charged battery of charging station. If is sufficiently proposed electric connection charging station, they may be charged together. In the case that the charging station is a part of Smart Grids or Micro Grids, should be available remote control by dispatching center or control system of network.

5 Conclusion

The purpose is to create such an active station, which is in equilibrium with the opportunities and needs of daily load diagram at the site. Therefore, the whole calculation is based on the needs 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, such as in terms of its design and in terms of possible sources, is not adequate 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 a suitable geographical distribution of active charging stations in both locations such as location with high density in cities and also in areas that are likely to be implemented in infrastructure between the larger sites, creates new possibilities for managing the operation of distribution network by active charging stations.

This work was supported by the SGS grant from VŠB – technical University of Ostrava (No. SP2014/187).

References

1. 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.
2. Chlebišová, E., Kyzeková, J., Svobodová, H.: Marketing Study of the Electric Vehicles' Diffusion. In 11th International Scientific Conference on Electric Power Engineering 2010, Brno, Czech republic, pp. 807-811, 2010, ISBN 978-80-214-4094-4.
3. Barešová, H.: Vliv nabíjecích stanic elektromobilů na elektrickou síť. DP VŠB-TU Ostrava, Ostrava, 2011.
4. <http://www.hybrid.cz/clanky/rychlounabijeci-stanice-pro-elektromobily>
5. <http://www.mvcr.cz/clanek/centralni-registr-vozidel-865510.aspx?>

Methods of Small Hydropower Plants Connections in Water Supply System

Tomáš Mozdřeně

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
tomas.mozdren.st1@vsb.cz

Abstract. This paper deals with alternatives of connecting small hydropower plants (SHP) with water supply system. The requirements for plants with power output from 30 kW to 100 kW and more are outlined, the SHPs passportization in the city of Ostrava is described. The law amendments referring to the re-distribution of state subsidy for renewable power energy sources are described.

Keyword: water supply, small hydro power plants, connection to distribution networks.

1 Introduction

The priority for water supply companies is drinking water supply to a consumer. Some of these companies are able to use this water driven under pressure in water mains for power generation. Increase in building such small hydropower plants (SHPs) in water supply system has been supported by state-subsidized purchase price for power from RES. Installing a SHP in water supply system does not require long shut-downs; however, it is imperative that any pollution of drinking water is prevented. Operating a SHP brings substantial benefits and optimizes the quality of processed water by aeration during its passage through turbine.

2 Small Hydropower Plants

The principle of a SHP is in building a weir or a dam on the water flow. This concentrates the flow and raises the water head. Water is driven through the screen into the engine room where the hydropower is transformed to mechanical power. The shaft brings mechanical power to the generator which generates electrical power. [4]

The principal parts of a SHP:

- Barriers on water flows
- Penstock
- Screen
- Technical equipment
- Tailrace

SHPs are designed as: [5]

- Run-of-river
- Storage

Depending on water head SHPS are:

- Low-pressure – up to 20m
- High-pressure – over 100m
- Medium-pressure – up to 100m

Depending on the design SHPs are:

- Weir
- Dam
- Derivative

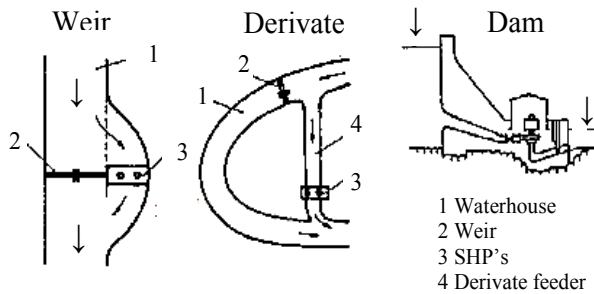


Figure 1. SHP general arrangement [5]

Mostly Banki, Pelton, Francis or pump turbines are employed for SHPs. The amount of usable energy depends on the head (vertical change in elevation between the upstream and the tail stream water level) and flow (volume of water passing through turbines).

2.1 Alternatives of connection SHP's to water supply system

The simplest connection of a machine set is an asynchronous generator and a properly adjusted pump working as a turbine. In Figure 2. there is a more complicated connection with more machine sets working in parallel. This connection is more suitable when flow parameters are fluctuating. The accurate pressure regulation is controlled by servo-valves [6].

2.2 SHP on a water purification plant penstock

A SHP Vyšní Lhota, which was put in operation in 2008, can serve as an example. Specifically, this plant is installed on the penstock duct in the grift chamber of a water pu-

rification plant. The run-of-river plant has three pump machine sets working as turbines. The machine set comprises three asynchronous generators Siemens ETANORM 150-400 s P = 90 kW, ETANORM 200-400 s P = 110 kW and ETANORM 300-400 s P = 132 kW. Only one turbine is activated at a time and their control is fully automated, the inlet and outlet in turbines is controlled via hydro drive and electro-flaps. [6]

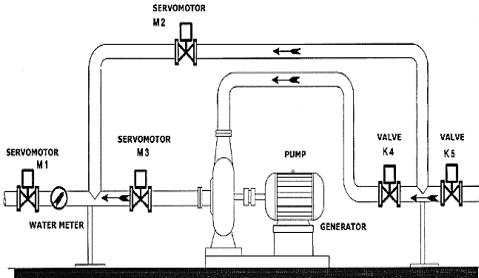


Figure 2. SHP machine set [6]



Figure 3. SHP Vyšní Lhoty [6]

2.3 SHP on the tower penstock

2.3.1 Water tower pressure penstock

A SHP built in Frýdek-Místek in 1997 is used as an example. The plants is run-of-river and has one turbine and one asynchronous generator type MEZ Frenštát F250MK06 at 37 kW is on Figure 4. The turbine inlet is controlled with a flow valve via servo drive. The operation is fully automated. [6]

2.3.2 Water tower pipes feeders

A SHP Hlavatce is located on their premises on a pressureless water tower pipe feeder. The machine set comprises an adjusted pump Sigma Hranice DE 400-LN-S. The flow rate is 500 l/s 8 hours a day. The pressure of 2.2 bar in the system is equal to the head of 2 m [3].



Figure 4. SHP Frýdek-Místek [6]



Figure 5. SHP Hlavatce [3]

2.4 SHP on dam

Concerning the location of SHPs, these can be installed either under the dam barrier or directly in it. They are pumped-storage plants that work in peak or semi-peak hours of daily load diagram. Their operators are not water supply companies as a rule, but the companies managing the dam, or the electricity companies with a river basin authority contract guaranteeing the operation of a SHP within their area.

2.5 SHP on water mains

This type of a SHP is installed on the water grid mains directed to a subscriber. It can be located before the water tower (SHP No. 3) or a terrestrial water reservoir (SHP No. 4.), but also after them (SHP No. 5 and SHP No. 6) e.g. in a break pressure tank. Water towers are built in a level country and terrestrial water reservoirs on elevated ground. Their objectives are to secure water storage and necessary pressure in water mains and also to level the differences between the inflow from the water source and outflow to subscribers. Water is driven into a reservoir either by gravity (water source is higher than water reservoir) or by a force main (water source is lower than reservoir and water is driven in with a pump).

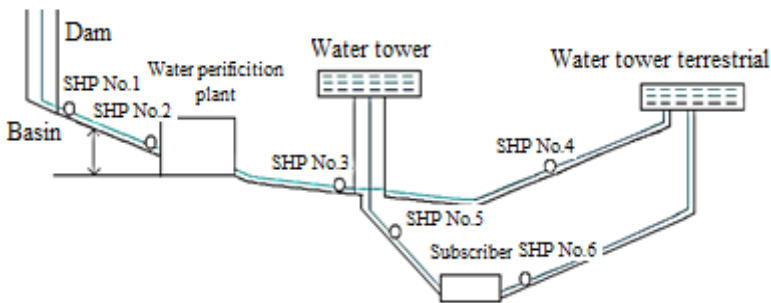


Figure 6. Simplified scheme of SHP connection onto the water supply system

3 Basic equipment of renewable energy sources of plants

The electrical machines controlled from the dispatching are installed primarily due to limits in their production that were determined in Act 458/2000 Coll. In plants with power output from 30 kW to 100 kW a Ripple Control (RC) receiver will be installed for the above given reasons. In plants with power output of 100 kW and more, a RC receiver and a Remote Terminal Unit (RTU) will be installed [1].

3.1 Equipment of plants with 30 kW to 100 kW power output

To disconnect the source from the distribution network, the dispatching will use a RC receiver controlled relay. In areas where RC signal cannot control disconnection, a RTU will be deployed. The RC receiver must be in operable mode even after the disconnection

from parallel operation - therefore it must be powered from the distribution grid. The RC receiver for power output control cannot substitute the RC receiver for switching consumption tariffs.

3.2 Equipment of plants with power output to 100 kW and more

The sources must be able to give rapid and accurate response to command sent from the dispatching. The transmission of readings and signaling from the Remote Terminal Units must be provided for the distribution grid dispatcher. A dispatcher can specify a continuous voltage control and reactive power for an electrical power source. Remote control provides a standardized communication protocol via RTU communication interface.

4 Requirements for control devices

The RES plant should be built near a source compatible with the terminal device in the dispatching center. In ČEZ Distribuce PLC, Moravia region almost all RES deploy RTU by Elvac a.s.

The RTU is a modular unit designed for data collection and control in distribution grids. It is easy to install and can be used for existing switchboards. The system itself consists of modular cards which fit into corresponding units. Except for signaling of status DI and control with RC, measurement of alternation quantities, other events are monitored as well. These are - exceeding of temperature limits, power outages, faults in reading electrical quantities of voltage and current, and moreover, active power, apparent power, and reactive power quantities are calculated. Readings are taken every second and transmitted to the dispatching control system.

The dispatching control system is connected via data concentrators in dispatching centers with monitoring control systems and in electricity plants including RES. For data communication between dispatching control system and monitoring control systems, the protocols such as MDXL, HioCom2 are used, as well as other standardized protocols such as IEC 60870-5-101 and IEC 60870-5-104. The dispatching control system and data concentrators communicate with monitoring and control systems in Moravia region via the aforementioned IEC 60870-5-104 protocol [2].



Figure 7. Remote Terminal Unit by Elvac a.s. [2]

There are various alternatives to the control systems interconnection. Either the deployment of protocols as a fixed point, or by using GSM (GPRS), in which case the RTU is to be equipped with plug-in modules enabling communication via appropriate protocols.

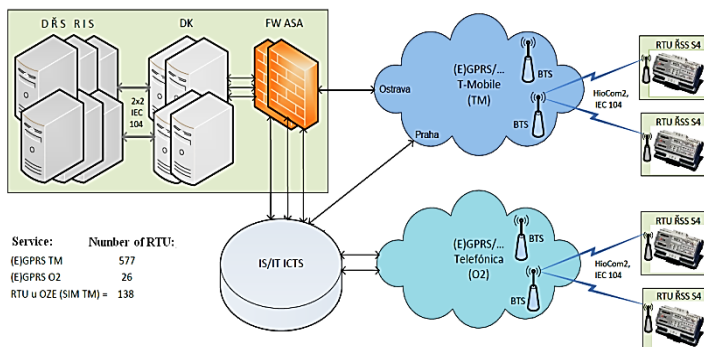


Figure 8. Simplified scheme of communication between Remote Technical Unit and Dispatching Control System [1].

Acknowledgements

This work was supported by the Grant of SGS VŠB - Technical University of Ostrava (No. SP2014/187) and by the project ENET (No. CZ.1.05/2.1.00/03.0069).

The results of the thesis are published in:

- Conference EEEIC 2014, Krakow, Poland, ISBN 978-1-4799-4661-7

References

- [1] ČEZ, a.s.: Requirements for equipment for regulation and control of renewable energy sources connected to distribution system.
- [2] Elvac, a.s.: User's guide RTU7M
- [3] Atlas of the facilities using renewable energy sources [online], [cit. 2014-7-1], Available at: <http://calla.ecn.cz/atlas/list.php?type=1>
- [4] P. Mastný, "Zdroje elektrické energie – vodní energie" [online], [cit.2014-7-1], Available at: <http://www.ueen.feec.vutbr.cz/~mastny/vyuka/mmze/skripta/voda.pdf>
- [5] J. Bouška, P. Knížek, J. Kašpar, "Sborník technických řešení MVE, Svaz podnikatelů pro využití energetických zdrojů," Praha 2000.
- [6] M. Goňo, M. Kyncl, R. Goňo, "Renewable energy sources - new possibilities in using small hydropower stations," Conf. EPE 2012, Brno University of Technology, 2012, vol. 1, pp. 657–659.

Statistical Evaluation of Voltage and Current Asymmetry in AC Low Voltage Part of Photovoltaic Plant

Michal Ney

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`michal.ney.st@vsb.cz`

Abstract. This paper deals with statistical evaluation of voltage and current asymmetry in AC low voltage part of photovoltaic plant. Input data contains voltage and current values measured in a real photovoltaic plant situated in Moravian-Silesian region. Data are first used for graphic representation of measured values during monitored day and evaluation of voltage fluctuation according to ČSN EN 50160. Then follows calculation of voltage and current asymmetry coefficients and their comparison with requirement of the above mentioned standard. Then is performed a statistical evaluation of all measured and calculated values using standard statistical methods, so determination of elementary statistical characteristics, representation of "box and whiskers" plots, analysis of data normality and scattering and finally regression analysis of voltage and current asymmetry coefficients dependence.

1 Introduction

Current asymmetry, so unequal load in each phase occurs especially in low voltage grids due to connection of single-phase loads. Significant single-phase loads are also in medium voltage grids (induction furnaces) and high voltage grids (traction transformers). This unequal load in each phase creates unequal voltage losses in lines and then different voltages at the end of line, so voltage asymmetry. Voltage asymmetry in low voltage grids is easy to reduce by appropriate distribution of single-phase loads to all phases, in medium and high voltage grids are necessary technological devices as symmetrisation and compensation units, Scott transformers etc. Specific requirements for allowed voltage asymmetry coefficient values are given by standard ČSN EN 50160 "Voltage characteristics of electricity supplied by public electricity networks".

2 Input data

Input data come from department's database and contains values of phase voltages (U_1 , U_2 , U_3) and currents (I_1 , I_2 , I_3) in AC part of photovoltaic plant with nominal voltage $3 \times 230/400$ V. Measurement was performed 1.7.2010, data were recorded in minute intervals during all measurement time, so from 0:00 to 23:59. Total range of input data is 1440 lines. Following graph shows voltage shapes during monitored day.

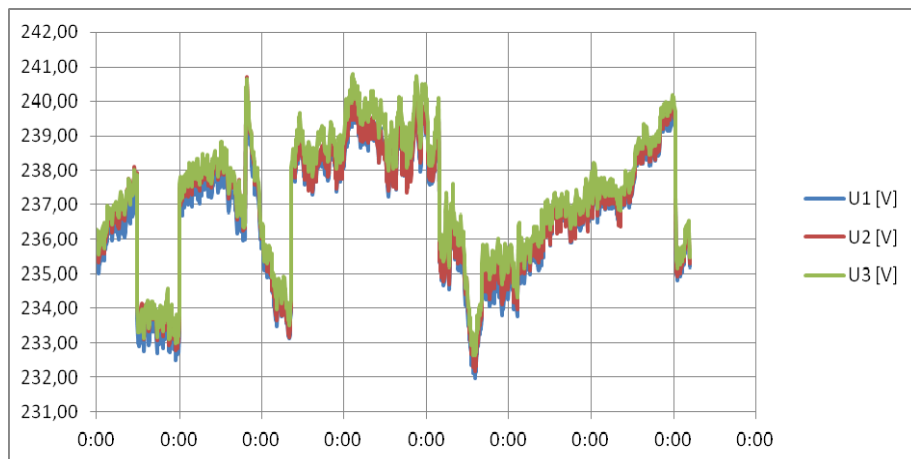


Fig. 1. Voltage shapes during monitored day

As graph shows, all voltage values ranging between 232 V and 241 V, according to allowed fluctuation $\pm 10\% U_n$. Next graph shows current shapes during monitored day.

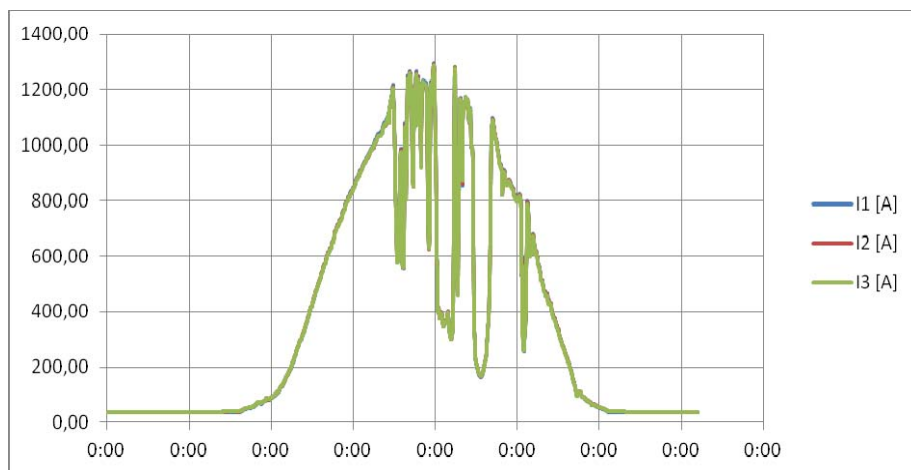


Fig. 2. Current shapes during monitored day

3 Voltage and current asymmetry coefficients calculation

Voltage asymmetry coefficient can be calculated using following equation:

$$\rho_U = \frac{U_{AB}^2 + U_{BC}^2 + U_{CA}^2 - \sqrt{3 \cdot (U_{AB}^2 + U_{BC}^2 + U_{CA}^2)^2 - 6 \cdot (U_{AB}^4 + U_{BC}^4 + U_{CA}^4)}}{U_{AB}^2 + U_{BC}^2 + U_{CA}^2 + \sqrt{3 \cdot (U_{AB}^2 + U_{BC}^2 + U_{CA}^2)^2 - 6 \cdot (U_{AB}^4 + U_{BC}^4 + U_{CA}^4)}} \cdot 100 \quad [\%]$$

where U_{AB} , U_{BC} , U_{CA} are voltages U_1 , U_2 , U_3 . Calculation is performed in source data file in Excel, calculated values of voltage asymmetry coefficient are graphical represented next.

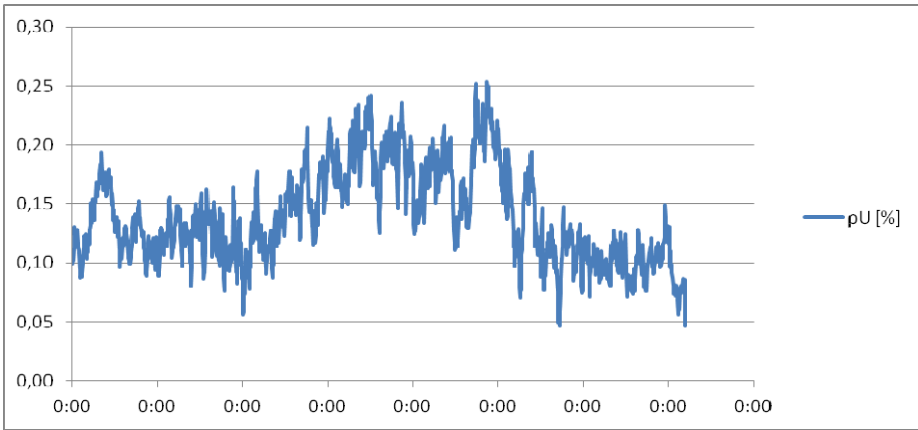


Fig. 3. Voltage asymmetry coefficient shape during monitored day

Calculated values ranging between 0,05 % and 0,25 % according to allowed maximum voltage asymmetry coefficient value of 2 %. Current asymmetry coefficient can be calculated using the same equation as for voltage asymmetry with substitution of current values I_1 , I_2 , I_3 . Calculated values are also graphical represented.

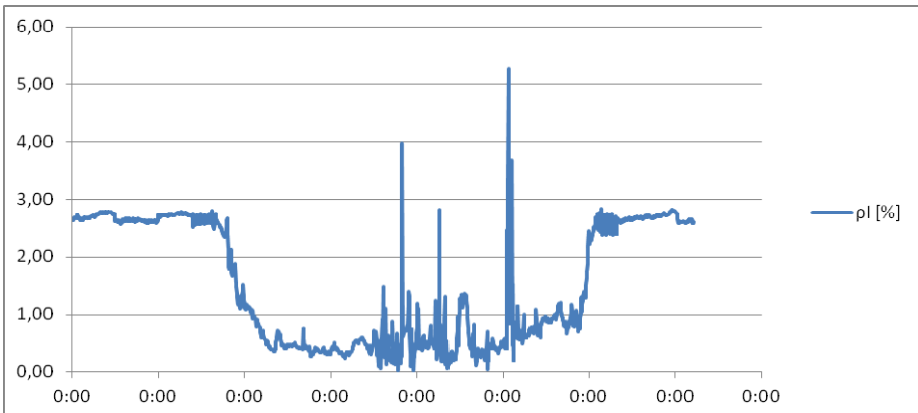


Fig. 4. Current asymmetry coefficient shape during monitored day

4 Statistical characteristics and analysis

From measured values of voltages and currents and calculated values of voltage and current asymmetry coefficients were determined statistical characteristics using statistical functions of Excel.

Statistical characteristics	U1	U2	U3	asymmetry
Mean value	236,68	236,86	237,22	0,14
Mean value error	0,0276	0,0285	0,0006	0,0000
Median	236,97	237,17	237,41	0,13
Modus	235,38	236,77	236,40	0,11
Determinative deviation	1,8513	1,8107	1,8713	0,0405
Sample scattering	3,4298	3,2807	3,5041	0,0016
Curtosis	-0,6981	-0,7118	-0,6585	-0,5455
Skewness	-0,4207	-0,4141	-0,3717	0,4472
Max. and min. variance	8,34	8,55	8,18	0,20
Minimum	232,00	232,16	232,63	0,05
Maximum	240,35	240,71	240,80	0,25
Count	1440	1440	1440	1440

Statistical characteristics	I1	I2	I3	asymmetry
Mean value	362,97	362,96	361,92	1,50
Mean value error	6,0041	5,9950	0,0159	0,0000
Median	120,79	122,01	121,83	1,09
Modus	35,96	37,48	37,62	2,63
Determinative deviation	397,0133	394,5868	393,9859	1,0446
Sample scattering	157729,0845	155806,9323	155332,7399	1,0920
Curtosis	-0,7598	-0,7671	-0,7526	-1,6916
Skewness	0,8605	0,8565	0,8634	0,1697
Max. and min. variance	1260,15	1255,06	1251,67	5,25
Minimum	35,32	36,71	36,80	0,03
Maximum	1295,46	1291,77	1288,47	5,28
Count	1440	1440	1440	1440

As an example of box and whiskers plots serve box and whiskers plots of voltage U1, current I1 and voltage and current asymmetry coefficients. Beginning of the line segment shows minimum, end of this line maximum. Left side of the rectangle area shows lower quartil, right side of this area shows upper quartil and vertical line inside shows median.

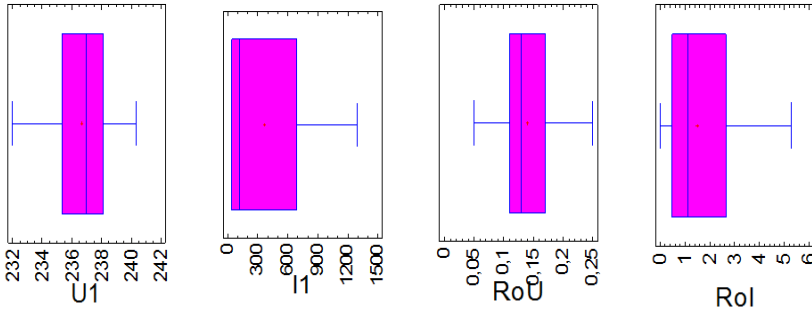


Fig. 5. Box and whiskers plots of voltage U1, current I1 and voltage and current asymmetry coefficients

Data normality and scattering analysis was performed in Statgraphics with result that data distribution isn't responsible to normal distribution and values aren't statistical similar.

Regression analysis of voltage and current asymmetry coefficients dependence was also performed in Statgraphics with this final relation between analyzed quantities.

$$\text{RoU} = 0,172694 - 0,0220729 * \text{RoI}$$

5 Conclusion

Measured voltage values ranging between 232 V and 241 V according to allowed fluctuation $\pm 10\% U_n$. Calculated values of voltage asymmetry coefficient ranging between 0,05 % and 0,25 % according to allowed maximum voltage asymmetry coefficient value of 2 %. All measured and calculated values are in accordance with requirements of standard ČSN EN 50160. But application of statistical methods, specially data normality and scattering analysis shown that measured values aren't statistical similar.

References

- [1] Department of Electric Power Engineering, VŠB-TU Ostrava, Department's database
- [2] M. Ney, "Statistical evaluation of voltage and current asymmetry in photovoltaic plant Starojická Lhota" semestral work of Statistical methods in engineering practise, 6/2014
- [3] R. Briš, M. Litschmannová, "Statistics I." subject syllabus 2004

Precising of failure rate valuation

Martin Slivka

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
s1i162@vsb.cz

Abstract. The paper deals with the computation of distribution network component reliability parameters. Current values of the component reliability parameters in distribution network are used for the reliability computation and also for reliability centered maintenance system. The reliability confidence parameters are calculated for accuracy.

Keywords : Failure database; confidence interval; exponential distribution

1 Interval estimate of failure rate

An extensive database of reliability parameters is developed by collecting of data from various electro-energetic companies. The same companies can use it retrospectively for maintenance scheduling or other reasons. Individual records on failures are not usable at once – they must be calculated. One of most required quantities is the failure rate. A value simply calculated from certain number of failures is a point estimate. The value calculated such a way has low validity as it does not draw upon the full extent of the database and therefore may cause unintentional misinterpretation of the results [1].

An appropriate means for the expression of indefiniteness of the results is an interval estimate. Based on the suitable model, an interval can be calculated. A given value will be within this range with certain reliability predefined by a user, typically 90%. Indefiniteness calculation can be shown on the following example. A piece of equipment monitored for 100 days breaks down ten times. A point estimate of the failure rate is therefore $\lambda \sim 10 / 100 = 0.1$, however by an interval estimate we obtain this interval (0.05413; 0.1696), which is substantial indefiniteness. If the same equipment is monitored 10000 days with 1000 failures, the point estimate will correspond with the previous case $\lambda \sim 1000 / 10000 = 0.1$. However, the interval estimate (0.0949; 0.1054) is much more accurate because the monitored period is longer and the amount of collected data is larger.

2 Exponential model

To determine the confidence interval, the Poisson's or the exponential probability distribution can be applied. The Poisson's probability distribution is discrete and can

express the failure rate directly. Continuous exponential distribution expresses the period of time between accidental events and can be used to model the time up to failure. Both these distribution types are sometimes called distributions without memory because they only depend on the unit of time. Both results have obviously reciprocal relation irrespective of the distribution used. However, the applied model should be respected in calculations and taken into consideration in results. Exponential distribution has one parameter δ and provided that $\delta > 0$, the probability density holds [2]:

$$f(x) = \delta \cdot e^{-\delta \cdot x} \tag{1}$$

Distribution function is in the shape:

$$F(x) = 1 - e^{-\delta \cdot x} \tag{2}$$

for $x > 0, \delta > 0$

The mean value holds:

$$E(X) = \frac{1}{\delta} \tag{3}$$

And the distribution is:

$$D(X) = \frac{1}{\delta^2} . \tag{4}$$

To determine the confidence interval, this statistics can be used $\frac{2n\bar{X}}{\delta}$, with distribution χ^2 with the number of degrees of freedom $\nu = 2n$. For reciprocal reliability interval holds:

$$1 - \alpha = P\left(\chi_{\alpha/2}^2 < \frac{2n\bar{x}}{\delta} < \chi_{1-\alpha/2}^2\right) = P\left(\frac{2n\bar{x}}{\chi_{1-\alpha/2}^2} < \delta < \frac{2n\bar{x}}{\chi_{\alpha/2}^2}\right) \tag{5}$$

3 Current condition of database

Since year 2000 the data on faults and failures in electrical power supply have been collected at the Department of Electrical Power Engineering, VŠB - TU Ostrava. The current database comprises partial databases of different operators of distribution networks (distribution areas). The structures of individual databases differ and therefore for each a different way of conversion and evaluation must be applied. The data collection is still in the process, but for the purpose of evaluation of component reliability in this paper only data acquired until the end of year 2013 were processed.

Currently, the collective database is fed by six Czech distribution areas and one Slovak company.

Distributors deliver their data in 'xls' files twice a year. Today the database contains more than 400 thousand records (from 2000 to 2013) on voltage levels 110 kV, MV and partially LV.

The reliability computation of the whole system is executed on the basis of components reliability that is included in the system. That is the reason why the reliability is computed in two phases. The first phase represents the retrieving of component reliability parameters and the second phase is the reliability computation itself. Other phases may include the evaluation of computed results and an improvement of the supply quality.

In the case of empirical reliability, we need data on operations and outages of components occurring in the reliability diagram. In the case of power system components, data must be available for outages of breakers, disconnectors, transformers, lines, etc. for a set type and voltage level. Moreover, there is another type of data necessary for the reliability computation. We need to have knowledge of the power network itself. For example, we must know the number of pieces of equipment for a set type, the total length of a line type, voltage level and so on. It is possible to compute basic reliability parameters of particular components from this data [3].

4 Latest results

The results of component reliability published in this section are acquired with the methods described in the 'Regulations for Distribution System Operation' [4]. Component reliability is evaluated by the following structure:

1. Equipment reliability

- Equipment type - overhead line, underground line...,
- Network voltage, equipment voltage

2. Component reliability

- Damaged equipment - specific device - conductor, switch, pole, fuse...
- Network voltage, equipment voltage
- Equipment Type - further specification - wooden pole, steely pole...
- Producer
- Production year - age of the component

The calculations take into consideration events which take from 3 minutes to 1 month. The graphic representation of all distribution regions reliability indices from the above-mentioned data for the 22 kV cable is given in Fig. 1. From the significant differences in particular years it is possible to observe the contribution of our analyses. The divergence of reliability indices is eliminated during long-term observation.

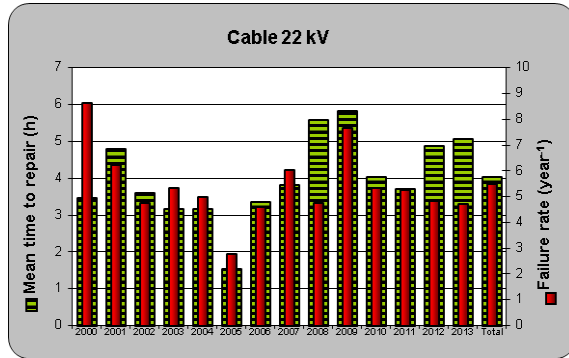


Fig. 1. The value tendency of reliability indices of the 22 kV cable.

Fig. 2 compares failure rates in different areas. This can be useful for the Energy Regulatory Office when looking for evidence for estimation of expenses for the renovation of lines in different distribution areas.

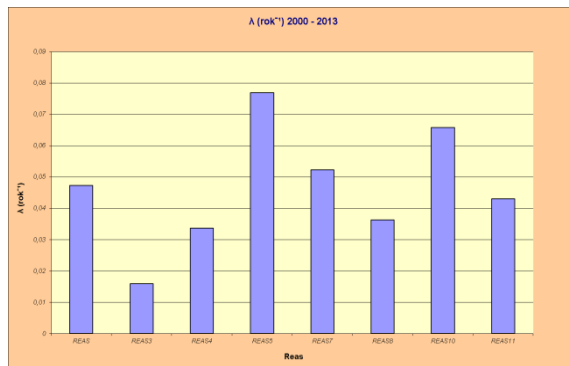


Fig. 2. Comparison of distribution regions.

Component reliability in the above given form and in tables is evaluated for these components:

- Overhead line 22 kV
- Overhead line 110 kV
- Cable 10 kV
- Cable 22 kV
- Circuit breaker 22 kV
- Circuit breaker 110 kV
- Transformer MV/LV
- Transformer 110 kV
- Section disconnector 22 kV
- Instrument transformer 22 kV
- Instrument transformer 110 kV

Further, more detailed analyses can be made with help of the internet application which cooperates with the unified database.

5 COMPARISON OF RELIABILITY DATA

Calculations drawing upon data from the failure database provide information on current technical condition of electrical power system components.

Currently used, but outdated method ČEZ 22/80 determines guide values of failure rate for individual network components. Since this method was developed, however, technological level of components was enhanced and changes in organization took place. This caused changes in record keeping of certain data. The latest comparison of data according to the given method with the data from the database is in Table 1.

Table 1. Comparison of results with old methodology

Equipment		ČEZ 22/80	2000 - 2013
22 kV cable	λ (year ⁻¹)	14.5	4.724
	τ (h)	215	5.063
22 kV overhead line	λ (year ⁻¹)	14	2.856
	τ (h)	3	4.108
110 kV overhead line	λ (year ⁻¹)	5.2	0.274
	τ (h)	3.5	1.804
MV/LV transformer	λ (year ⁻¹)	0.03	0.055
	τ (h)	2500	0.808
110 kV/MV transformer	λ (year ⁻¹)	0.04	0.056
	τ (h)	1300	0.276
22 kV circuit breaker	λ (year ⁻¹)	0.015	0.012
	τ (h)	30	24.707
110 kV circuit breaker	λ (year ⁻¹)	0.01	0.022
	τ (h)	100	23.418

The collection and procurement of materials for the database is subject to the local traditions and directions. Every distribution area sends data in different forms due to different procedures in record keeping. A unified database eliminates the influence of local customs and allows better comparison on the national level. In terms of quality, a well-maintained technical condition has substantial effect on reliability parameters. It is interesting to compare these reliability parameters with those of other countries, especially outside the EU, for example the US [5] (see Table 2.). It should be noted that certain data published in literature [5] are related to the units used in the particular country. For means of comparison, these data must be recalculated. The 'usual' values are given hereby.

Table 2. Comparison CR and USA

Equipment		CZ	USA
22 kV cable	$\lambda(\text{year}^{-1})$	4,724	4,3
	τ (h)	5,063	10
22 kV overhead line	$\lambda(\text{year}^{-1})$	2,856	6,2
	τ (h)	4,108	4
110 kV overhead line	$\lambda(\text{year}^{-1})$	0,274	-
	τ (h)	1,804	-
MV/LV transformer*	$\lambda(\text{year}^{-1})$	0,055	0,04
	τ (h)	0,808	40
110 kV/MV transformer*	$\lambda(\text{year}^{-1})$	0,056	0,03
	τ (h)	0,276	70
22 kV circuit breaker	$\lambda(\text{year}^{-1})$	0,012	0,01
	τ (h)	24,707	12
110 kV circuit breaker	$\lambda(\text{year}^{-1})$	0,022	-
	τ (h)	23,418	-

* In the studied literature the transformers were distributed into categories up to 25 MVA and over 25 MVA. The transformers up to 25 MVA are compared with MV/LV transformers, those over 25 MVA with 110 kV/MV.

6 Conclusion

Employment of confidence intervals for the reliability calculation is beneficial not only concerning the database span. The development of the confidence interval function in online application will be helpful for the user because the extensiveness of the database makes more accurate results possible with selected confidence. In the future the calculated confidence intervals will be more statistically relevant with growing database.

ACKNOWLEDGEMENTS

This work was supported by the Grant of SGS VŠB - Technical University of Ostrava (No. SP2014/187) and by the project ENET (No. CZ.1.05/2.1.00/03.0069).

The results of the thesis are published in:

- conference EEEIC2014
- journal paper *Przegląd elektrotechniczny*, vol. 1, 11/2013

References

1. Z. Martínek, V. Královacová, "The Solution for Repairable Units," Conf. EPE 2010, Brno University of Technology, 2010, pp. 593-597.
2. L. Průcha, "4. Intervalové odhady parametrů rozdělení," [online], [cit. 7.3.2014], Available at: <http://math.feld.cvut.cz/prucha/mstp/4pu.pdf>.
3. M. Slivka, R. Gono, S. Rusek, "Failure duration in distribution networks," *Przegląd elektrotechniczny*, vol. 89, No. 11/2013, pp. 1-3.
4. PPDS - Regulations for Distribution System Operation in the Czech Republic.
5. R. E. Brown, *Electric Power Distribution Reliability*, Dekker 2002, p. 365.

Power Analysis of Photovoltaic Power Plant

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 sources of electrical energy that are considerably dependent on weather conditions, especially on direct sunshine. This is known as the so-called stochastic effect of production. This paper focuses on the description of stochastic production generated by photovoltaic power plants with respect to impact on the distribution power grid. The year-round operation of the plant is examined in each month. The Evaluation is based on real-time synchronous data measuring both on-site photovoltaic power plant operation and on-site relevant electric power line output from the substation, where the plant is connected. Statistical methods have been utilized to compile a methodology to define extreme conditions of power production. This survey assesses the extent of power production in individual months during whole year.

Keywords: Photovoltaic Power Plant, Stochastic Effect, Extreme Conditions, Statistical Methods, Calculation Methodology, Ancillary services.

1 Introduction

This paper discusses the impact of photovoltaic power plant (PVP) with regard to Ancillary Services. In electrical grid, one rule has to be met. It is that the immediately productions of power have to equal the immediately consumptions across electrical grid. Briefly, what is produced has to be consumed at the same time. The main problem occurs in situation when the photovoltaic power plants produce their power energy, depended on meteorological conditions, but there are no consumers in electrical grid or conversely, when the electrical energy is needed in the grid but the sun is not shine and PVP does not work. To solve this problem, other sources have to be in distribution grid in order to balance the stochastic power produced from PVP. These sources have to keep power energy in backup and have to be prepare to compensate power flowing from PVP. To keep these backups is very uneconomical. This study interests in possibility of reduce these power reserves.

2 Analyzed Area

The assessment is based on annual monitoring of not only operation of PVP, yet also the relevant outgoing feeder from the substation. The network topology has a loop system however it is operated in radial system. At the time when the measurement was acquired, the PVP was only electric power source with significant output on the given power line. The grid diagram is shown in Fig. 1.

The first measurement was performed at the outgoing feeder from the substation 110/22 kV using the side of lower voltage level of 22 kV. The second measurement was done at the connection point of the photovoltaic plant to the grid. The data has been obtained at the voltage level of 0.4 kV. The actual measurements are highlighted by red points in Fig. 1. The label (ss), in Fig. 1, means disconnected section switchers on the feeders.

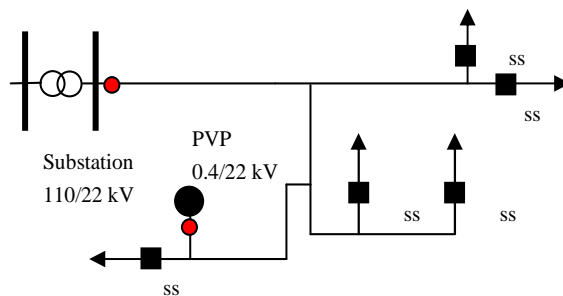


Fig. 1. Analyzed network topology

where: PVP - photovoltaic power plant,
 ss - section switch (disconnected),
 110/22 kV - substation plant 110/22 kV.

3 Transmission Grid Code

The Transmission Grid Code, part II, as amended by revision 14 duly approved by the Energy Regulatory Office on 1.1.2014 in Czech Republic, defines the matters relevant to use of Grid Code with respect to Ancillary Services (AS). The document includes special methodology for determination of summary control backups of AS for the electric grid in whole Czech Republic.

The basis for determination of summary control backups of Ancillary Services is the so called deviation $OD_{OZE(t)}$. This is a parameter to establish the increment of balancing deviation caused by production from newly installed renewable energy resources. This increment is defined by the sum of increase generated by wind power plant $OD_{VE(t)}$ and photovoltaic power plant $OD_{FVE(t)}$. The deviation $OD_{OZE(t)}$ is then used to determine volumes of Ancillary Services and is started by (1).

$$OD_{OZE(t)} = OD_{VIE(t)} + OD_{FVE(t)} \quad [1] \quad (1)$$

The purpose of this survey is asses if and how much the PVP production has influence to calculation of Ancillary Services with respect to each month of the year.

4 Calculation Model

The calculation methodology has been designed for determine extreme conditions of photovoltaic power plant in cooperation with daily load diagram. The process of calculation methodology is described by the block diagram in Fig. 2. The diagram is divided into three horizontal sections highlighted by blue background.

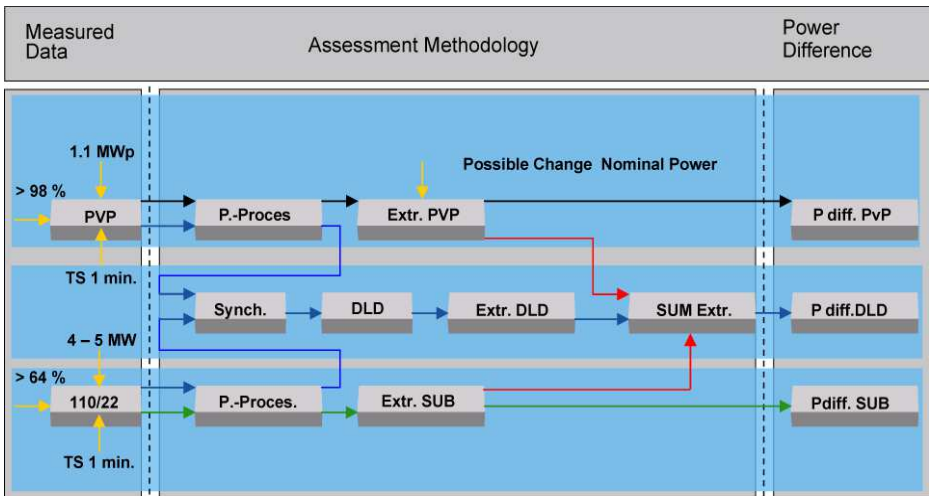


Fig. 2. Block diagram of calculation methodology

The first section comprises blocks linked by black arrows. These describe the steps for assessment of power differences on the PVP itself only. This indicator does not include the impact of daily load diagram.

The last horizontal section contains blocks linked with green arrows. The data measured at substation includes the impact of PVP operation too. Determination of difference of active power in this manner is correct, although dedicated only. It cannot be applied to other model situation in a broader sense.

The second section proves itself as an optimal solution, it is linked to blocks using blue arrows. We need to realize that extreme conditions of power at the substation occur due to extreme conditions on both the PVP as well as the daily load diagram and there is considerable variation between curves extreme conditions of PVP and daily load diagram. This is reason why it is necessary to define extreme conditions separately both PVP (block Extr. PVP in Fig. 2) and daily load diagram (block Extr. SUB). Only then, it is possible to sum extreme conditions PVP as well as daily

load diagram (block SUM Extr.). The understanding these issues is quite challenging. Literature [2] describes it in more detail.

5 Power Difference

The results from calculation methodology, above mentioned, are essential for defining of maximum extreme conditions that may occur at the outgoing feeder from the substation. The exact evaluation procedure is given in the source [3].

The magnitude of difference between maximum and minimum extreme conditions at the substation over time represents the maximum expected differential change to active power. This bar chart of hourly time comparisons for April is shown in Fig. 3.

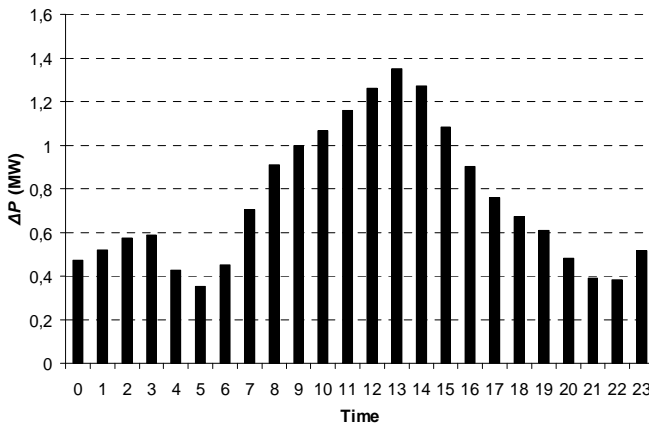


Fig. 3. Hourly histogram with expected change to active power

Next Steps of this study will focus on the application of this methodology to all months. Subsequently, the resulting differences will be respected in the optimization of Ancillary Services for each month. The expected differences in each month are slightly indicated in chapter 6.

6 Annual Extreme Conditions of PVP

On the basis of suggested methodology [3], the maximum production extreme curve and minimum production extreme curve is determined for each month of photovoltaic power plant production.

Figure 4 shows three values for each month. It is maximum value from maximum production extreme curve, maximum value from minimum production extreme curve and differences between these mentioned values. The dimensions of units are in kilowatts.

One can notice that the amplitude of both MAX and MIN is rising with increasing intensity of solar radiation. As expected, it should be from December till June, i.e. the

periods of winter and summer solstice. Of course, conversely, the amplitudes should be decreasing from July till December, i.e. between the summer and winter solstice. However, this expectation is not confirmed. The maximum occurs in May rather than during the summer solstice, as might be expected. This case therefore leads to establishment of optimal ratio between the intensity of solar radiation and average temperature during the particular month (previous months indicate lack of solar radiation and the month of June brings higher ambient temperature with negative impact on panel efficiency).

Values of black columns, represented power differences between MAX and MIN, have the same increasing and decreasing character during the year as well as MAX and MIN.

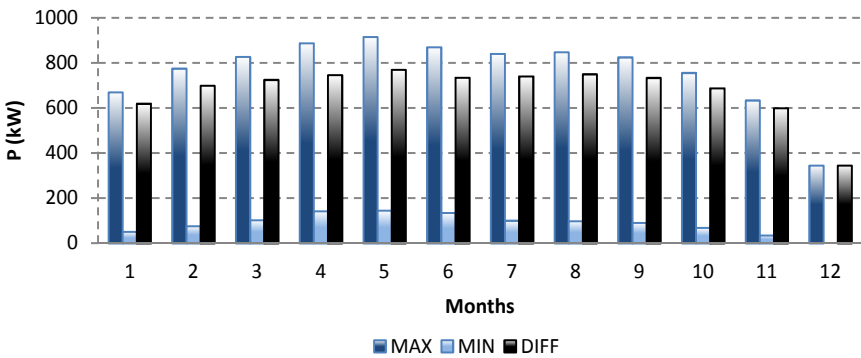


Fig. 4. Histogram for maximum and minimum production of PVP

where: MAX - power represented the maximum of PVP production for each month,
 MIN - power represented the minimum of PVP production for each month,
 DIFF - power represented difference between MAX and MIN.

Figure 5 displays same histogram as Fig. 4, however, there are the values in percentage ratio. It gives better relationships and better observation of individual months.

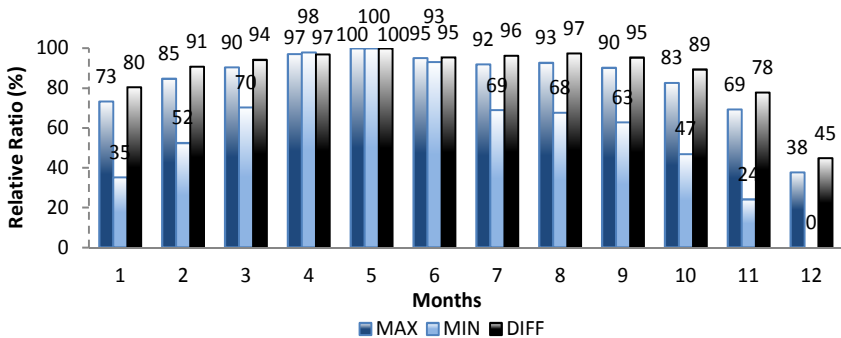


Fig. 5. Histogram for maximum and minimum production of PVP, percentage ratio

7 Conclusion

Briefly, it is essential to note the following. It is well known that the photovoltaic power plants produce more energy in the summer time and while less in winter time periods, but the power difference varies only relatively slightly (except December). This is the reason why for PVPs are placed high demands on Ancillary Services almost throughout the year. The difference has negative impact on immediate power balance between production and consumption within the distribution system and it is highly uneconomical. Next Steps of this study will focus on the application of this methodology on a full database and evaluation the relationship of each month at the level of hourly interval, as shown Fig. 3. These results will exploit to optimize power reserves of Ancillary Services.

Acknowledgements

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2014/187) and by the project ENET (Research and Development for Innovations Operational Programmes No. CZ.1.05/2.1.00/03.0069).

Reference

1. ČEPS, Transmission Grid Code, Review 14 – Part II., Ancillary Services, available from: <https://www.ceps.cz>, 15.1.2014.
2. SMOČEK, M. – HRADÍLEK, Z. Impact of Photovoltaic Power Plants Operation on Differential Change of Distribution Grid Power, In Proceedings of 2014 15th International Scientific Conference on Electric Power Engineering. 1. 2014, VUT Brno, ISBN: 978-1-4799-3806-3.
3. SMOČEK, M. – HRADÍLEK, Z. “Photovoltaic Power Plants, Extreme Change of Power Difference”. The 7th International Scientific Symposium on Electrical Power Engineering, TU - Košice, 2013, ISBN: 978-80-553-1462-4.

Simulation of Active Energy Unit with respecting Smart-Grid requirements

Jindrich Stuchly

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jindrich.stuchly.st@vsb.cz

Abstract. One of the possible way how to increase the efficiency of the renewable energy sources is supplying the objects, where the energy has been generated. In 2011 has been developed the power electric platform in campus of VSB – Technical university of Ostrava. This platform is operated at family house power level and energy consumption is covered from renewable sources of electric energy and energy storage device in the form of battery bank. For this system has been created the monitoring and remote control system in Labview™ in 2012, which is able to be controlled outside of the university, even over the World. And finally in 2013 has been developed and implemented sophisticated automated dispatcher system - Active Demand Side Management (ADSM). The presented paper introduces develop and simulation of ADSM in the Off-Grid systems with the set of a specific requirements, while the requirements are the same like in the Smart-Grids.

Keywords: Active Energy Unit, Active Demand Side Management, Smart-Grid, Simulation

1 Introduction

Recently the creation of the Smart-Grid has been posed as one of the greatest challenges of this century. Smart-Grid generally refers to a class of a technology using to bring utility electricity delivery systems, using computer-based remote control and automation. [1] More specifically, the Smart-Grid can be regarded as an electric system that uses information, two-way, cyber-secure communication technologies, and artificial intelligence in an integrated element across electricity generation, transmission, substations, distribution and consumption to achieve a system that is clean, safe, secure, reliable, durable, efficient, and sustainable. In case of Smart-Grid system, that is district distribution grid with possibility of autonomous running in island-mode (Off-Grid) and running in parallel with external distribution system (On-Grid). [2]

A Smart-Grid is an evolved grid system that manages electricity demand in a sustainable, reliable and economic manner, built on advanced infrastructure and tuned to facilitate the integration of all involved. Smart-Grids possess demand response capacity to

help balance electrical consumption with supply, as well as the potential to integrate new technologies to enable energy storage devices and the large-scale use of electric vehicles. [3]

1.1 Smart-Grid requirements

In most cases, Smart-Grid systems are designed for parallel running with the external distribution system, it means in On-Grid mode. The co-generation unit is used as a basic source of electric and heat energy. Photovoltaic and wind power plants are used as an addition for electric energy generating according to meteorological and geomorphologic condition in specific place of Smart-Grid.

The Smart-Grids are mostly operated in two voltage levels, so in medium and low voltage level with possibility of supply from two sides and with the possibility of bi-directional power flow. There is a specific set Smart-Grid requirements:

- Autonomous operation without dependence on energy from external power grids,
- Operation with well-balanced production and consumption of electric energy,
- Possibility of energy storage,
- Predominant use of RES,
- Capability to serve non-traditional loads,
- New type of grid protection allowing bi-directional power flows, and
- Active Demand Side Management.

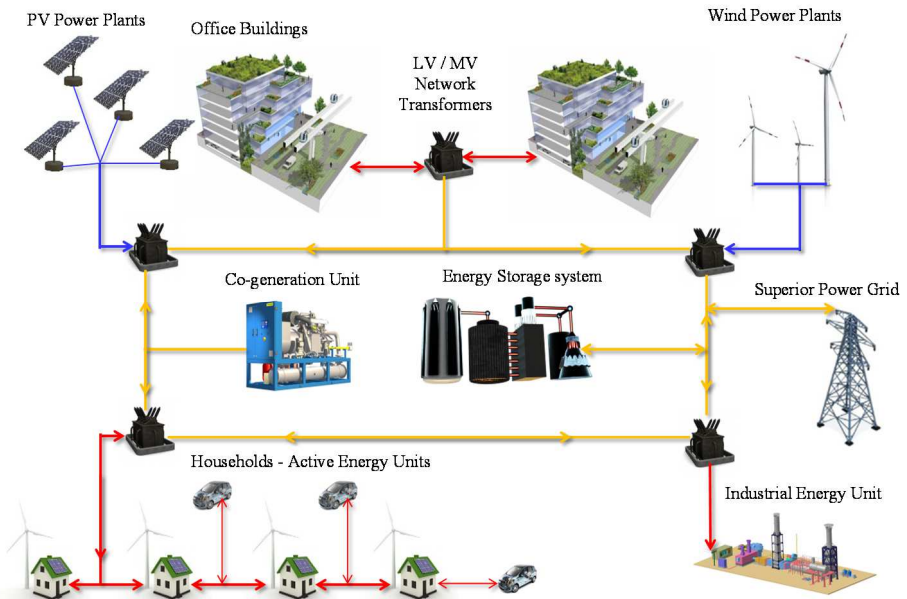


Fig. 1. Smart-Grid block scheme

However, in reality, only the physical platforms of individual components of the energy systems have been created, without sophisticated systems for their coordination and control. Therefore, mostly the pilot systems do not deserve the attribute “Smart”. [4]

2 Active Energy Unit descriptions

To create the physical platform of Active Energy Unit (AEU) was necessary to design and implement a hybrid power sources to power closed system. A power closed system of consumption is currently made up of several home appliances, office equipment and programmable load. This creates a physical model of AEU, where can be simulated various operating conditions of wind and photovoltaic power plants and various demands of electric power without connecting appliances to the standard AC distribution grid. For block scheme of Active Energy Unit see Fig.2. [5, 6, 7, 8]

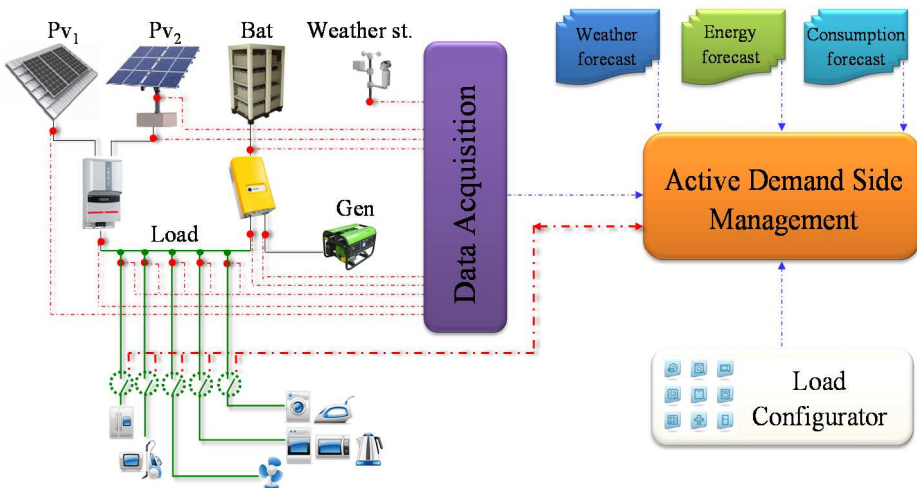


Fig. 2. Block scheme of Active Energy Unit

2.1 Active Demand Side Management

In short, we may describe the function of the Active Demand Side Management in a model situation, where the system will control the power flows in the energy unit with expected changeable weather. Individual sequences ADSM is following:

1. The weather prediction tool determines the 3days weather forecast
2. Expected scenario of electricity consumption will be generated using ties between weather, season and consumer behavior
3. The energy prediction tool determines the value of available energy from the source part and determines, if this amount of energy is enough to cover expected consumption scenarios.
4. Active demand side management system accepts all three information input (weather forecast, consumption forecast, energy forecast) and spread out the energy during this period using the energy storage device. The surplus of energy will be stored and the energy deficit will be covered from energy storage device.
5. The priority of individual home appliances is respecting in switching process based on the supervision and their nature.
6. In the event of lack of energy in the system in selected consumption scenarios has to the Active demand side management system suggest different consumption respecting the load priorities – e. g. delayed start of selected appliances – kettle, washing machine, dishwasher, TV, etc.)

3 Results of Simulation

The Active Demand Side Management model as well as the whole platform of the Smart-Grid was evaluated using practical experiments. A couple months' simulations with the complete usage of the appliances has been used as well as the energy production of the photovoltaic power plants and battery bank as energy storage device. Actually has been used a power profile of regular home appliance and office equipment. Main goal of the platform of the ADSM is to maintain the system without need of energy from external grid, so in the Off-Grid mode.

Therefore, I will evaluate the properties of the designed model and platform which will prove the ability of the algorithm to improve the comfort of the living as well as to reduce the power consumptions or to decrease the price of the platform with cheaper/smaller batteries or smaller power plants.

Table 1. Results of June

June	Without ADSM	With ADSM
First period	544	0
Period Count	210	0
Missing Energy	49570	0
Unconsumed energy	111228	73080

The system without ADSM is not able to perform all appliance cycles and the first problems is in the period 544 (one period is 15min timestamp) and totally 210 periods contain appliance cycles which were not ran, see Table 1. Totally, more than 49 kW·h of energy is missing. In the case of using ADSM, there is no energy missing and even over 73 kW·h is unused.

Table 2. Results of December

December	Without ADSM	With ADSM
First period	76	76
Period Count	1610	1672
Missing Energy	342109	339841
Unconsumed energy	0	0

The results for December are depicted in Table 2. The situation is completely different, because the solar radiance is very low and the photovoltaic power plant has minimum power output even in the midday. The renewable sources of electric energy are necessary to supplement with some additional resources, such Electric power Generator. This generator could use as fuels some ecologic and environment - friendly fuel such as LGP, NG or CNG.

Although the wind power plant is classified as a renewable source of electric power, it is in the context of reliability of power supplies for use in Off-Grid systems less suitable than the LPG generator, especially in terms of investment costs and the need to obtain the necessary permits to build up the wind power plant.

4 Conclusion

As has been mentioned in the introduction chapter of this paper, electric energy and related issues in the areas of energy independence and reliability, together with the growing penetration of renewable energy sources are an important topic in recent years and undoubtedly challenges for researchers and engineers in the near future.

The controlling of the energy flows inside the energy objects using sophisticated methods - artificial intelligence, such as neural network, fuzzy logic give us an appropriate opportunity to solve the lack of energy during the energy peaks or shift the loads in time, increase the loads or even fill up the valley, when the power production is over the consumption. The goal of this paper is to present the simulation results for the selected object. These results are described in the chapters above, and I can simply say, that the ADSM significantly reduces the amount of unused electric energy in the AEU and is able to preserve all appliance cycles and use most energy generated by the power plants without energy wasting in summer time. It means further increasing of the efficiency of renewable energy sources reliability and efficiency of operation and last but not least to make shorter the financial and energy payback time period.

The main benefit of this Active Management System is to improve the comfort of the living as well as to reduce the power consumptions or to decrease the price of the platform with cheaper/smaller batteries or smaller power plants. The future works will be focused on the improvement of the ADSM with more intelligent scheduling of the appliances which cannot be satisfied by the produced energy, especially for winter season. Moreover, the system will be used for more detailed analysis of the appliances usage and for optimization of power sources and storage devices for Off-Grid systems, e.g. Smart houses.

Acknowledgement

This paper 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 Programme '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 projects reg. no. SP2014/110 and SP2014/49, and project LE13011 Creation of a PROGRES 3 Consortium Office to Support Cross-Border Cooperation.

References

1. P. Krömer, A. Abraham, V. Snášel, J. Platoš, H. Izakian, "Differential evolution for scheduling independent tasks on heterogeneous distributed environments", in *Advances in Intelligent Web Mastering - 2*, ser. *Advances in Soft Computing*, vol. 67. Springer, 2010, pp. 127-134
2. HADJISAÏD, Nouredine a Jean-Claude SABONNAIÈRE. *SmartGrids*. London: ISTE, 2012. ISBN 978-1-84821-261-9.
3. G.T. Heydt, "The next generation of power distribution systems", *IEEE Trans. Smart-Grid*, vol. 1, no. 3, Dec. 2010.
4. S. Misak, L. Prokop, *Off-Grid Power Systems. An International Conference on Environment and Electrical Engineering 2010*. ISBN 978-1-4244-5370-2, pp. 23-26, 2010
5. STUCHLY, Jindrich and Jakub VRAMBA: *Off-Grid Operation of Self-Sufficient Smart House Conception: Workshop na podporu sítí*, Ostrava: 2013, pp. 8-11. ISBN 978-80-248-2986-9
6. VRAMBA Jakub, STUCHLY Jindrich and KOSMAK Jakub. Family house off-grid conception. 7th International Scientific Symposium on Electrical Power Engineering, ELEKTROENERGETIKA 2013; Stara Lesna; Slovakia; 18 September 2013 through 20 September 2013; Code 105843. ELEKTROENERGETIKA 2013. Available from: <http://147.232.26.36/viewabstract.php?id=676>
7. STUCHLY, Jindrich, Stanislav MISAK, Jakub VRAMBA, Marian UHER a Petr KUBALIK. Control of autonomous active distribution grid - introduction. *2014 14th International Conference on Environment and Electrical Engineering*. IEEE, 2014, pp. 328-333. DOI: 10.1109/EEEIC.2014.6835888. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6835888>
8. MISAK, Stanislav, Jakub VRAMBA, Jindrich STUCHLY, Peter BRACINIK a Marek HOGER. The protection and the control of a small Off-Grid system. *Proceedings of the 2014 15th International Scientific Conference on Electric Power Engineering (EPE)*. IEEE, 2014, pp. 155-160. DOI: 10.1109/EPE.2014.6839461. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6839461>

Applications of using the thermal energy from cogeneration unit for accumulate into the technology

Michal Špaček

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michal.spacek.st2@vsb.cz

Abstract. In recent times we are trying to dispose with energy more efficiently. When is coming with a nature-friendly technologies and energy savings. One possible solution is replacing gas boilers with a cogeneration unit (CHP) instead of gas boilers. Natural gas or other gases rich in methane such as biogas are combusted in the CHP. The gas is burned in a combustion engine, which is connected by the shaft to the electric generator and the electric generator produces electricity. Cooling engine and flue gases are obtained by thermal energy.

Keywords: cogeneration unit, accumulation.

1 Introduction

One of the variants of saving energy is applied in the company NC-Line. This solution was created by the company Menergo a.s. A gas boiler and gas burners were replaced by the cogeneration unit with the accumulation tank. Thermal and electrical energy is produced by using a cogeneration unit that burns natural gas. The CHP is connected to the pipe leading out of hot water (85 ° C), which is transported through the module to the technology and an accumulation tank. Part of the heat (1/3) is thus supplied directly to technology and the other part of the heat (2/3) is supplied to the accumulation tank. Scheme of CHP technology and the accumulation tank is displayed in the picture number 1. Warm water of main heating circuit of the cogeneration unit operates with a thermal gradient of 85/65°C. For measuring the amount of heat energy is connected calorimeter. The calorimeter is marked with a symbol Q with unit GJ in the picture number 1. The accumulation tank of 60 cubic meters which is thermally insulated was used for accumulation of thermal energy. [1]

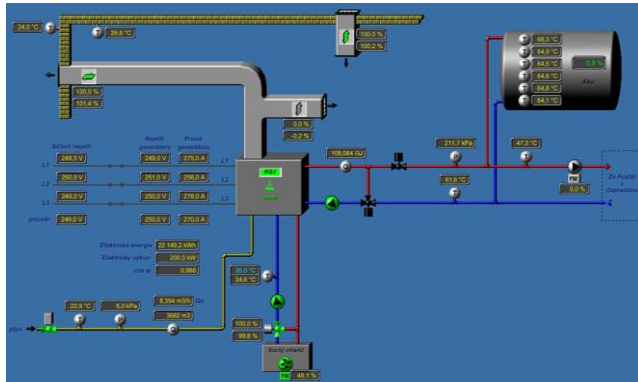


Fig. 1. Visualization of technology cogeneration unit with accumulation tank and other as-
sistive devices

2 Cogeneration unit

The used cogeneration unit was Vitobloc 200 EM-199/263 from company Viessmann. Energy balance is a graphical representation of the flow of energy module BTP. The energy balance shows the conversion of primary energy (natural gas, 100%) to usable electricity and thermal energy. Losses that occur based on the conversion of this energy are also shown. Electrical usable energy arise in the process of combustion in the Otto gas engine, when the rotary motion is converted into a synchronous generator to the current. Thermal usable energy arise also in the process of combustion in the Otto gas engine. Otto gas engine is operated as a combustion engine with turbocharger and two-stage cooling of the mixture with the ratio of air $\Lambda \approx 1.6$. The heat gain from the flue gases, from the manifold, from the engine block and from engine lubricating oil and is used for heating e.g. heating water. [2]

Table 1. Parameters CHP - Vitobloc 200 EM-199/263 [3]

Max. electrical power:	199 kW _e
Max. thermal power:	263 kW _t
Fuel consumption:	53 Nm ³ /hour
Min. overall efficiency:	89,6 %
Temperature gradient of heating circuit CHP :	65/85 C
Temperature gradient of cooling the fuel mixture:	35/38 C

Table 2. Energy balance of module BTP - Vitobloc 200 EM-199/263 [3]

Cooling the mixture	external 50 C
The use of energy – heating value	100 %
Mechanical energy	39,0 %
Thermal energy	61,0 %
Electrical usable power	37,0 %
Thermal usable power	48,9 %
Losses	10,4 %

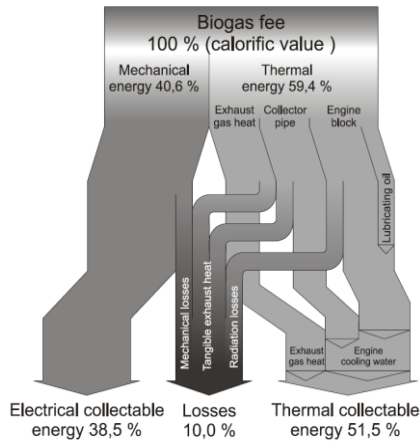


Fig. 2. Energy balance of module BTP [3]

Module of block-type thermal power station (module BTP) is a complete unit ready for connection. It includes air-cooled asynchronous generator to produce three-phase current 400 V, 50 Hz and hot water with a temperature levels of inflow and return flow of 85/65 ° C at full load and a standard temperature difference of 20 K. Each module BTE can be operated thermally and electrically independently in the range of the electrical load between 50-100% (equivalent to 60 - 100% heat output).

In this deployment is run in heat mode at 100. The unit is 199 kW of power and therefore it belongs to the category of plants with power above 100 kW connected to the distribution network (DN). In the case of threats and reliable operation of the power system is necessary for supervisory control temporarily restricts or shut down the active power supply of electricity. The source is adequately (quickly and accurately) able to respond to commands from the control center PDS to reduce active power step-in mode 0, 50, 75 and 100% of installed capacity [3], [4], [5].



Fig. 3. Cogeneration Viessmann Vitobloc 200 EM-199/263

3 Measurement of the cogeneration unit

In the block-type thermal power station there is installed device VitoBloc Gateway, which provides information through the bus RS 485. The bus is connected to the control computer, where the values are recorded. Electrical values are recorded at one-minute intervals.

The obtained data are dated July 2, 2014, when the command to switch on the CHP was at 4:00 to switch off at 12:00.

In the process of electric power (Fig. 4) there can be observed considerably frequent oscillations in the first half of the running time. In the second half of the period of the operation there has not been recorded such oscillations that arise at every start CHP and last so long before it settles machine. Before the end of CHP from 11:40 it is possible to observe a significant instability of performance that is caused by an increase of water temperature in the main heating circuit. Temperature of return water increased due to charge the accumulation tank. While a sudden change in inlet temperature had to reset the internal parameters of the CHP which results in instability of the produced electrical power.

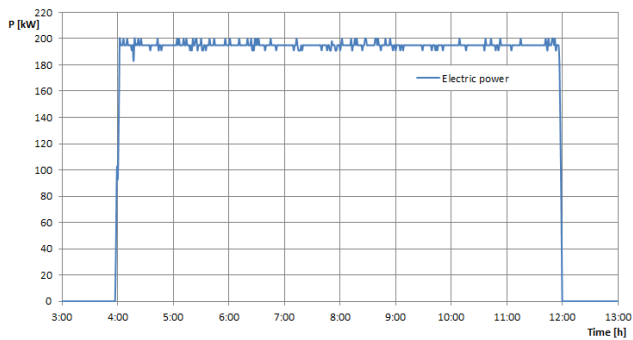


Fig. 4. Process of electric power during the operation of the CHP

In the Fig 5 there are shown progressions of electric current from the generator at each phases. Where it is possible to observe current asymmetry, especially in phase L2.

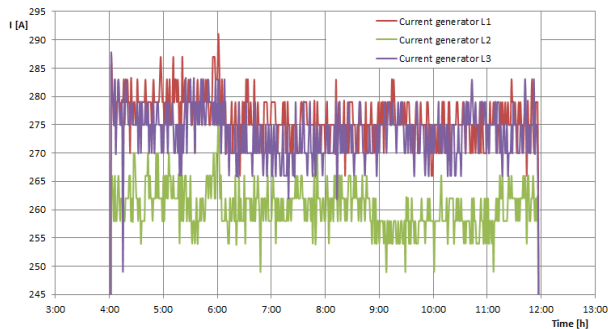


Fig. 5. Process of electric current during the operation of the CHP

From the measurement of the voltage of CHP it is possible to observe the voltage fluctuations that are caused by the nature of the network load. Lasers for metal are connected to the network, which are characterized with instability consumptions.

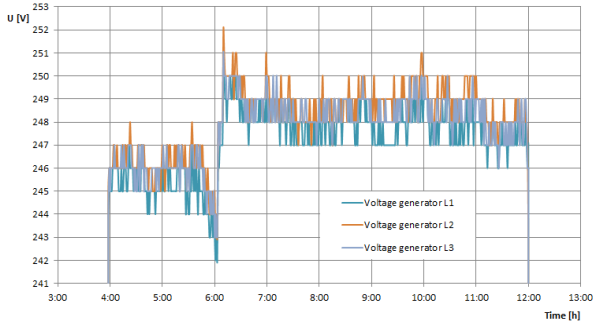


Fig. 6. Process of electric voltage during operation of the CHP

4 Accumulation tank

The tank is circular cross-section of 3.40 meters, a total length of 7,812 meters and volume of 60 cubic meters. Progression of charging accumulation tank is on the Fig 7 and discharging in the Fig 8. Temperature sensors with which the progressions were recorded, are distributed over the entire height of the accumulation tank. The temperature sensors T2, T3, T4 and T5 are positioned in descending order between the sensors T1 and T6.

From Fig 7 you can see that at 9:00 the CHP has been activated, when the temperature began to rise at the highest point of the curve of accumulation tank. Depending on the time began to increase the temperature of each sensor toward the lower end of the curve of accumulation tank.

On the other hand on the Fig 8 there is recorded the progress of discharging the accumulation tank. First, the temperature begins to decrease at the bottom of the tank, which records the temperature sensor T6. Gradual heat extraction from the tank begins to change (decrease) of temperature towards the bottom of the tank to the upper temperature sensor T1. In the period between 12 to 14 hours it is possible to observe that the heat consumption was lower than in the evening. The decrease occurred with a longer interval.

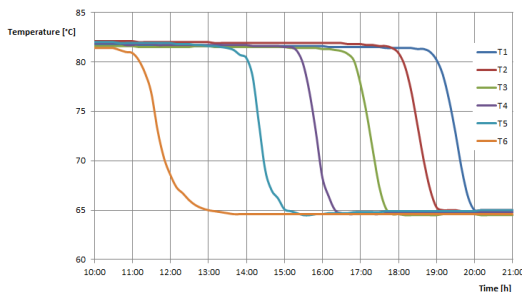


Fig. 7. Charging the accumulation tank

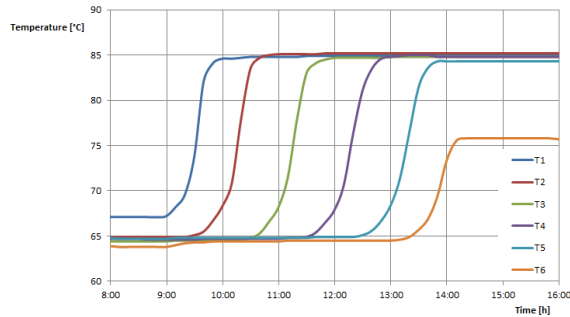


Fig. 8. Discharging the accumulation tank

5 Conclusion

The connection of CHP with accumulation tank has important applications especially in terms of controlling the electric power in the distribution network. As described at the beginning, CHP is also a source of thermal energy, which is necessary to accumulate somewhere or thwart. In this case is particularly relevant accumulation of heat energy to achieve the most economical energy management. In the application of this technology it is possible to set the time mode operation of CHP. Electric power of CHP Viessmann Vitobloc 200 EM-199/263 can be regulated in the range of 50% to 100% of nominal power. This technology CHP is operated in a heat mode, which corresponds to 100% of the nominal electric power. Sample of running CHP from the perspective of electrical power is in Fig 6.

Acknowledgment

The presented research was supervised by prof. Ing. Zdeněk Hradílek, DrSc. from Department of Electrical power engineering, VSB - TU Ostrava, Czech republic.

References

1. Dvorský, E.: "The combined production of electricity and thermal energy," BEN technical literature, Praha 2005, ISBN 80-7300-118-7
2. Krbek, J., Ochrana, L., Polesný, B.: "Industrial Power," PC DIR Real Brno, 1996
3. Dokumentation BHKW Vitobloc 200, 2012-02 / V.1, Viessmann Group, ESS Energie Systems and Service GmbH
4. Krbek, J., Ochrana, L., Polesný, B.: "Heating and cogeneration," PC-DIR Real, s r.o., Brno, 1999, ISBN 80-214-1347-6.
5. Krbek, J., Polesný, B. "Small cogeneration units in municipal and industrial energy," PC-DIR Real, s r.o., Brno, 1999, ISBN 80-85895-23-4.

Solver platform – optimization tool for distribution grid operation

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. The article describes the development of a software tool to calculate the optimal topology of the distribution network under different climatic conditions in response to the revitalization of existing distribution networks SmartGrids. The article includes individual parts of a principle function and genesis and development of the software as well as a demonstration of its practical use. The advantages of its use for energy engineers there are also described.

Keywords: solver platform, renewable sources, feedback effect, distribution power grid, optimal topology

1 Introduction

Integration of distributed generation of electricity from renewable sources into the existing distribution system is a recent trend in the production and transmission of electricity. This fact entails difficulties in the form of feedback effects of renewable sources to the applied grid. Among the feedback effects of renewable energy sources can include voltage change in connecting point, flicker effect, current surges, frequency change, harmonic distortion of voltage and current, bi-directional power flow in line.

It is the most suitable to quantify described feedback effects with regard to limits determined by appropriate standards before connecting the renewable power plant source.

There are many simulation programs and computational method, in which it is possible to simulate power systems operation. However, a problem arises in request to calculate the network steady at a chosen time step continuously, without complicated and time low-effective settings, and also an image of behavior of renewable. Therefore so-called “Solver platform” has been developed to Department of Electric Power of VSB-TUO.

2 Solver platform

It is the software that enables the calculation of the ratios in the analyzed power grid for 24 hours based on input information from the predictors of climatic conditions and consumption of electric power in every consumption point (distribution transformers). The output is then the following value:

- Voltage condition on LV level
- Voltage condition on MV level
- Voltage condition on HV level
- Line current
- Active power
- Reactive power
- Power factor

Of course there are a number of alarm messages of exceeding the limit values set by the user in each hour calculation:

- Overvoltage
- Under voltage
- Overcurrent of lines and transformers
- Uneconomical operation (transformers)

The entire Solver platform consists of two building elements. The most important one is the mathematical model of selected part of (real) power system, designed in the ATPDraw. This program was chosen for its unique setting options for all network parameters using the input text template in a format *.txt, without the use of visual frontend program. Also, this program is useful for its choice of saving the calculated results in the selected sampling frequency to file format *.txt once again .

Another building element is a graphical visualization superstructure and "decision-core" of Solver platform that is created in using LabVIEW software. This element not only allows import of input information from predictors, but also monitor of developments in the analyzed system using the user interface.

The basic principle of the Solver platform function is based on the interaction between the basic core program ATPDraw computing and visualization software LabVIEW. The file with an extension *.atp (*.txt), which contains a description of the mathematical model of the analyzed power grid is saved and then loaded from pre-defined folder. In LabVIEW, this file is imported and converted into text format. Variables selected by the user are overwritten by the "find and replace" using graphical menu in the user interface. The text template is saved and re-imported by computational core software ATPDraw after starting the calculation. The computing core also generates a Result list in the form *.lis (*.txt) file. This entire process is done for each hour of interest.).

2.1 Possibilities of users interface

The user himself can choose the topology of the analyzed power grid either manually using the switches (14 pcs. at MV level and 2 pcs. at HV level) in the block diagram of the system created in the user interface, or chooses the choice of one of the preset options of network topology. It is possible to set limit values of voltage and current of each type of elements in the power grid, i.e. different cable types, each type transformers, voltage limits LV, MV and HV, the limit values of the current load of the cables and transformers, limits of economical use of transformers.

The calculation of the ratios in the power grid for 24 hours follows. The results then show the values of the quantities described in the introduction, including the number of states exceedences of limit values. Part of the results is a graphical processing result using graphs. Calculation continues for other possible network topology and the results of the exceeded limit for each network topology are entered into the table, which is stored to file of .txt in end of calculation.

The user can then select the optimal network topology according to the amount of warning messages according to the priority which attaches each criterion.

3 Mathematical model of analyzed power system

The mathematical model of the power system presents a real power system in three voltage levels (110 kV, 22 kV and 0.4 kV). Superior system is a system of 110 kV, the short-circuit power of 700 kV·A. Devoted power system operates through two transformers 110/22 kV to 22 kV power system, which includes 12 distribution transformers 22/0.4 kV and 34 pcs line, see Fig. 1.

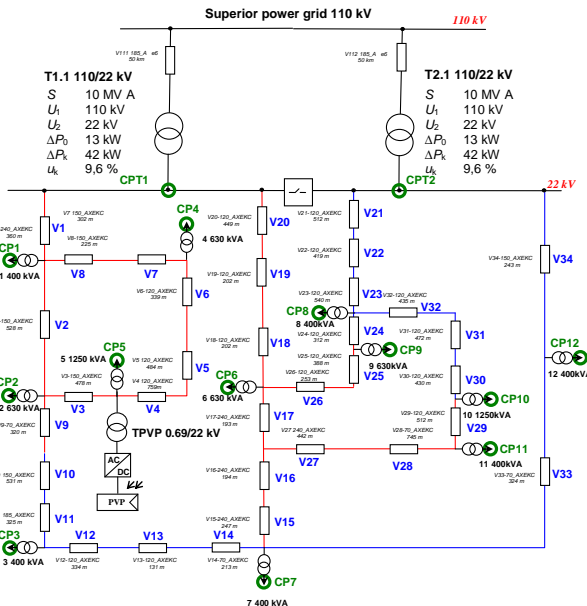


Fig. 1. Block diagram of analyzed power grid

3.1 Matematic model of photovoltaic power plant of 4 MWp.

At this moment the (fictional) photovoltaic power station of nominal output power of 4 MWp is situated in analyzed power system. It is representative of only one renewable source for now. This plant was selected for a rich data basis of measurement applied to the actual electricity unit of the same power.

PV plant model is created in ATPDraw to meet the requirement of pulse width modulation, and also to be created using a current source. Request of pulse width modulation was achieved by interactions of sine and saw general sources. Pulse width general signal is generated by mathematical combination of these two sources. Current source function is then assigned to him from the menu TACS (Transient Analysis Control Signals). One source is then assigned to each phase. They differ from each other then the phase shift. It is related to the moment of the start of the first phase sources (always shifted about 120 °).

3.2 Consumption part of analyzed power grid

Because at this time the prediction system of power consumption is under development, this information is taken for now from the standardized daily load diagrams, which are downloadable on web site of Electricity Market Operator. Therefore, the user inserts the selected daily load diagram for 24 hours, which is expected for the period and region.

The user then selects the maximum percentage loading of each distribution transformer 22/0.4 kV. Daily consumption is then calculated from the following equation:

$$P = S_{ni} \cdot h \cdot P_p \cdot \cos \varphi \quad (1)$$

Where P is the actual power consumption of the transformer, h is the percentage loads of transformers, Pp is the relative power value of the TDD in a given hour, Sni nominal power of i-th transformer. The ratio of active and reactive power component gives constant power factor = 0.95 inductive.

4 Example of calculation

It is expected one day, to be executed calculation and assessment of the appropriateness variant of topological network connections. The file with information about the curve of solar irradiation during the day was downloaded from the prediction system, where the period of one day was separated. Concretely 30.3.2014. The resulting curve is shown in the Fig. 2.

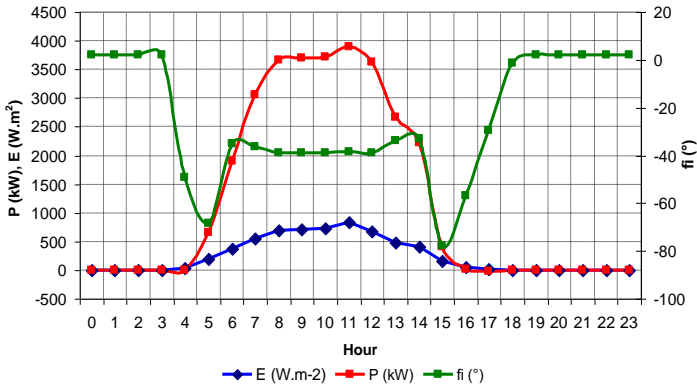


Fig. 2. The values for the calculation of parameters

Also, file of daily load diagram for the user without the use of heat energy was chosen, i.e. DLD 4.

The 10 variants of operating system configurations have been determined in advance, which are stored in the user interface of Solver platform. The limit values of all type of elements were also set. The criterions and resulted values are presented in Table 1.

Table 1. of calculated parameters of decision criteria

<i>Variant</i>	<i>Number of handling operation</i>	<i>Overvolt tage LV tr.</i>	<i>Overvolt age MV cables</i>	<i>Overvolt age HV line</i>	<i>Undervolt age LV tr.</i>
1	0	6	26	1	8
2	2	6	41	1	8
3	2	6	31	1	10
4	2	35	35	1	10
5	3	35	88	1	10
6	3	35	88	1	10
7	3	19	96	1	8
8	4	17	104	1	8
9	6	14	83	1	10
10	6	8	21	1	20
<i>Variant</i>	<i>Undervolt tage MV cables</i>	<i>Undervoltage HV lines</i>	<i>Overcurrent MV cables</i>	<i>Uneconomical operation LV tr.</i>	<i>Uneconomical operation MV tr.</i>
1	26	1	21	6	4
2	62	1	35	6	4
3	69	1	29	6	4
4	116	1	18	6	4
5	116	1	17	6	4
6	116	1	18	6	4
7	126	1	63	6	4
8	115	1	62	6	4
9	122	1	49	6	4
10	228	1	90	6	4

It is understood that the above parameters become different criteria for selecting the correct topology of analyzed power system.

Table 1 show that in terms of Overvoltage LV transformers are among the best topological connections variants 1, 2, 3, where exceeding the specified values in 6 cases. However, in terms of Overvoltage MV cables may be the best option 10, where the voltage limit values are exceeded in 21 cases. In terms of Undervoltage LV transformers are the best choice 1, 2, 7, 8 variants with the same number of alarms of 8. In terms of undervoltage of MV cables is the best option connection 1 with the number of exceedences of limit values 6. In terms of Overcurrent MV cables is the best option 5 with 17 limits of operation exceeding.

Some parameters stored in the table are not changed numerically. It is, therefore, that all possible combinations of connections of power system were not calculated (overvoltage and undervoltage of HV lines), but also because these parameters are not dependent on the grid topology changes (uneconomical operation of LV transformers).

The last criterion for choosing the appropriate power system topology is the Number of handling operation. This is the number of manipulations of switches for changing the network configuration related to the reference version (in this case, option 1).

5 Conclusion

This paper describes a optimization tool for a distribution grid operation. It is software, which mathematically calculate a behaviour of distribution grid under different conditions of power supply from renewable source and energy consumption in each consumption point.

It is possible to follow from results table user can choose optimal distribution grid topology according some criterion. It can be said solver platform can be a very important software for power engineer in proces of revitalization current distribution grids to Smart grids.

Future work will be an upgrade solwer platfom vizualization and implementation of implementation choosing the best variant of grid operation using the priorities assigned to each criterion.

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.: Influence of renewable resources operating to the Smart Grid power flow Elektroenergetika 2013 : proceedings of the 7th International Scientific Symposium on Electrical Power Engineering : Stará Lesná, High Tatras, September 18-20, 2013
4. Uher, M., Mišák, S., Analysis of feedback effect of distribution grid with renewable resources 2013 13th International Conference on Environment and Electrical Engineering : 1-3 November 2013, Wroclaw, Poland : conference proceedings

Hydrogen as a storage of electrical energy

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. This paper deals with accumulation of electric power by means of hydrogen technology and reverse transformation of such power stored in hydrogen into electric power again. The source of renewable energy in this system is represented by a small photovoltaic power plant feeding the hydrogen generator - electrolyzer as the main appliance for reverse chemical transformation together with hydrogen fuel cells. Both of these appliances operate at low temperature, they are based on PEM (Proton Exchange Membranes) technology. The electrolyzer ensures decomposition of purified water to oxygen and hydrogen by action of direct electric current. The fuel cell houses a reverse process, when chemical reactions between hydrogen and oxygen are electric power of water result in production of electric power and water. The crucial part of this paper comprises calculation of efficiency achieved by accumulation of power into hydrogen, including the efficiency of individual components forming the accumulation (storage) system located in the fuel cells laboratory of VŠB - TU Ostrava. The electrolyzer ensures decomposition of purified water to oxygen and hydrogen by action of direct electric current. The fuel cell houses a reverse process, when chemical reactions between hydrogen and oxygen are electric power of water result in production of electric power and water.

1 Introduction

Accumulation of electric power can be seen as an intermediate stage between the power produced and the power consumed. Accumulation is a process of storage of power in a certain form, which can be converted back into electric power from the electric engineering point of view. The accumulation itself can be divided into long- and short-term respectively. Appliances for short-term accumulation can include flywheels, super capacitors, super conductive batteries etc.



Fig. 1. Small photovoltaic power plant on the roof of laboratory

These accumulation (storage) systems are significant for industry and other fields of power engineering, yet they are unable to secure long-term accumulation that is vital for electric power engineering. The systems for long-term accumulation can include lead batteries, flow batteries (Redox flow batteries) [1], the compressed air energy system (CAES) [2] and the last but not the least item would be a system based on hydrogen technology, which actually still struggles with lower efficiency now. Raising its level to competitive figures with respect to other accumulation technologies is the goal pursued by numerous laboratories all over the world.

2 Description of Hydrogen storage system

Power supply to this island system is provided by twelve PV polycrystalline panels located on the lab roof inclined to 45° (see Fig. 1). The installed capacity of these panels connected into serial-parallel configuration equals to 1,980 Wp. Surplus electric power produced by these panels is stored into the power storage bank with priority and once it has been fully charged, the system starts to produce gaseous hydrogen via the proton exchange membrane and initiates its storage into a suitable magazine.

A module of fuel cells represents a reverse technology to the electrolyzer, PEM type once again, with the nominal capacity of 1,200 W and the voltage of 26 V that serves for generating of electric power from the previously produced hydrogen. The module is initiated only at the moment, when the demand for electric power can be no longer covered by the power storage bank. That serves as "a buffer" for moments of weak supply of power from the photovoltaic system. Selected parameters of PV panels have been stated in table 1.

Table 1. Parameters of one PV panel (SCHOTT POLY 165 TYPE)

nominal capacity (Wp)	≥165	nominal capacity (Wp)
voltage at nominal capacity (V)	35.10	voltage at nominal capacity (V)
current at nominal capacity (A)	4.70	current at nominal capacity (A)
no-load voltage (V)	43.60	no-load voltage (V)
short-circuit current (A)	5.27	short-circuit current (A)

2.1 Storage of Hydrogen

The electrolyzer keeps hydrogen pressurised at high pressure and stored into pressure vessels in gaseous form and at the maximum feed pressure, which is 13.8 bar in our case. This is the maximum output pressure of the electrolyzer possible. Further options involve storage on suitable carriers, which are metal hydrides, where 1 cm³ of which can absorb one litre of hydrogen. Bonding is performed at convenient pressure and temperature. Bonded hydrogen is in gaseous form. Filling is associated with exothermic reaction and the vessels (reservoir) requires cooling. On the contrary, the vessel must be heated during the emptying process. [3]

3 Operation of Hydrogen storage system

The storage system illustrated by block diagram in Figure 2 operates as follows: The output of PV panels rises from the beginning of the day and charging of batteries is in progress. The AC load is supplied from NEXA fuel cells module, consuming the hydrogen previously produced. As the output from PV panels reaches a sufficient level (assessment conducted by the control system), the NEXA module is switched off and the supply for AC load will be taken over by PV panels. Potential deviations of power supplied are then covered by batteries. Once they are fully charged, the control system will initiate the electrolyzer and start with production of hydrogen. PV panels will supply power to the AC load and electrolyzer under such circumstances. The electrolyzer remains in operation until the drop of solar radiation intensity, i.e. output from PV panels drops below the specified level. During night time, the AC load is supplied from batteries first, followed by involvement of the fuel cells module at the moment, when the battery voltage drops to a specific level. The module processes hydrogen from storage to produce electric power for supply into the AC load via converters. [4]

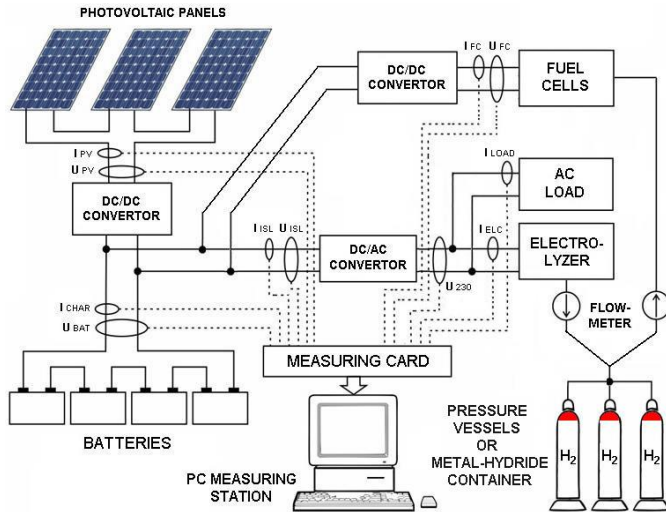


Fig. 2. Simplified diagram of the hydrogen storage system

4 Measurements of Hydrogen storage system

The aim was to determine energetic balance within the entire hydrogen storage system and to calculate its efficiency. Efficiency calculations were also made for individual components of the system, including the hydrogen technology devices, i.e. the electrolyzer Hogen GC600 and the fuel cells module NEXA, as well as the semiconductor equipment, i.e. converters used:

- Sunny Island Charger 40
- Sunny Island 4282
- SD-1000L-48

DC converters respond to arrival of energy into the DC bus, which further supplies power to terminals of the DC/AC converter. There is one DC/DC converter to transform the electric power from PV panels to a suitable level and to operate as battery charging controller at the same time. Another DC/DC converter serves for stabilising of output voltage from the fuel cells module. The input at DC/AC converter is connected to either batteries or the fuel cells module. Its power outlet is used to supply the AC load and electrolyzer Hogen GC600. Assurance of efficiency of the hydrogen storage system was pursued by means of several days of measurements during August and September 2013, where the values obtained on a specific selected day are shown in fig. 3

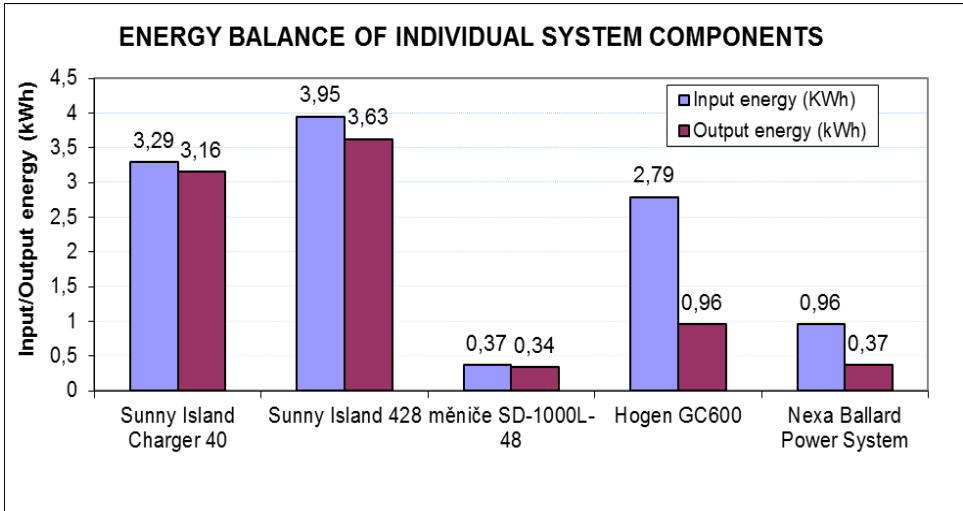


Fig. 3. Energy balance of individual system components

Table 2. Energy balance of individual system components

	(%)
Sunny Island Charger 40	95,99
Sunny Island 4282	91,82
Fuel cells module NEXA Ballard Power System	38,36
DC/DC converter SD-1000L-48	93,6
Electrolyser (Hogen GC600)	34,35
Overall Efficiency of Hydrogen Storage System	10,87

5 Conclusion

The laboratory hydrogen storage system described in this paper has been designed to ensure uninterrupted power supply to a specific appliance - the AC load with capacity of 250 W. The entire system has been set up in an island configuration with operation independent to the surrounding grid and able to balance immediate surplus or lack of electric power produced by a renewable energy resource. Long-term measurements helped to assess efficiency of individual components of this system. The overall efficiency of system for accumulation of electric power from photovoltaic source approximates the value of 11 %. The system efficiency figure does not include the efficiency of photovoltaic power plant, as it is a source of renewable energy only, with no effect on efficiency of accumulation. The achievable level of hydrogen storage system should be up to 30 %, considering the theoretical efficiency of electrolyzers around 70 % [5] and the level of 60 % achieved by fuel cell. The said values are relevant to high-temperature equipment only and not our hydrogen storage system.

In spite of the low efficiency established, the importance of accumulating renewable energy into hydrogen will keep growing in the near future. Hydrogen storage systems can be used to solve operating issues with both systems based on renewable resources connected to the grid as well as island systems. Hydrogen systems also help with accumulation of renewable energy to a great extent of output. Even though hydrogen serves as a mere carrier of electric power rather than its source in this case, its application is very convenient with respect to the ratio of chemical energy and weight.

One of the options for increase in efficiency of the hydrogen storage system is to ensure supply of pure oxygen for fuel cells. Further areas for improvement span even with respect to semi-conductor converters required for transformation of current systems and stabilisation of voltage.

Acknowledgements

This work was supported by the SGS grant from VŠB - Technical University of Ostrava (No. SP2014/187) and by the project ENET (No. CZ.1.05/2.1.00/03.0069).

References

1. P. Alotto, M. Guarnieri, F. Moro, "Redox flow batteries for the storage of renewable energy: A review," *Renewable and Sustainable Energy Reviews*, Vol. 29, January 2014.
2. S. Karellas, N. Tzouganatos, "Comparison of the performance of compressed-air and hydrogen energy storage systems: Karpathos island case study," *Renewable and Sustainable Energy Reviews*, Vol. 29, January 2014.
3. P. Moldrik, D. Minarik, R. Chvalek, J. Vaculik, Z. Hradilek, "Metal Hydrides as One of Hydrogen Storage Options," 13th International Scientific Conference Electric Power Engineering, Ed. J. Drapela, J. Machacek, Brno:Brno University of Technology, Vol. 13., Issue 1/2, p. 457-461, May 2012.
4. P. Moldrik, D. Minarik, Z. Hradilek, "Energy storage system in connection with a PV power plant," *Przeglad Elektrotechniczny*, Vol. 89., Issue 11, p. 65-68, November 2013.
5. S. Geitmann, "Wasserstoff & Brennstoffzellen - Die Technik von Morgn," Hydrogeit Verlag, ISBN 3-937863-04-4, 300 s. : il., Kremmen, 2004

Adaptive relay protection system for Off-Grid systems

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 article presents an adaptive protection system developed for the failure state and abnormal operation states, such as short-circuits and overloads in an autonomous 1-phase Off-Grid energy system. The Off-Grid system is built in the area VSB and is in active operation for two years. This adaptive protection system intended for the Off-Grid system, which contains several independent sources of short-circuit power, is able to change and configure itself for various operation states in the Off-Grid. The adaptive protection system is based on central control and real-time analysis of data acquired from electronic measuring devices.

Keywords: Off-Grid system, short-circuit current, protection

1 Introduction

VSB–Technical University of Ostrava has designed a power supplying concept which complies with the requirements for the SMART Grid Systems used for supplying an energy self-sustaining residential home in an Off-Grid operation mode. The energy sources and energy storage system were selected on the basis of the power consumption of a conventional residential home as well as on the size of the initial investment. [1] The Off-Grid system, in this instance, is defined as an energy unit independent from the electric power supplied from a superior grid. This system primarily integrates the parallel cooperation of a wind power plant and a photovoltaic power plant, where the power of the individual sources may be used to cover the consumption of electric energy in a separate low-voltage system and the surplus of power is used for charging the energy storage system. In the majority of cases, the general benefits and pro-active developmental trend of these concepts are mentioned, but as with any other product, issues related to the long-term operations of these power units and specific features tend to be neglected particularly when speaking of the mode of protection in case of failure or abnormal operation states. [1], [2]

An important task in designing the Off-Grid system for the power supply for the energy self-sustaining building is the design and implementation of the protection concept concerning the short-circuit conditions and the overload in the autonomous Off-Grid energy system. The short-circuit power in the Off-Grid system is determined by the Off-Grid inverter, the wind power plant and the photovoltaic power plant along with

the performance electronics. Electric power generation from renewable sources is considerably stochastically variable. This problem has a negative effect on change in the short-circuit power in the Off-Grid system as compared with a conventional low voltage distribution line (On-Grid). The short-circuit current in the Off-Grid system reaches much lower values in comparison with a conventional On-Grid system. In order to ensure selective and hundred-percent protection against short circuit and overload in the Off-Grid system, there is a need to modify the standard protection concept used in low-voltage networks with a considerably higher short-circuit power than in the Off-Grid system. For this reason, VŠB–Technical University of Ostrava proposed an adaptive protection system which uses information about the current operation states in the Off-Grid system. Based on the proposed algorithm, it evaluates the current values in individual current circuits and makes further decisions based on the values read. This protection concept makes use of conventional safeguarding elements, such as a breaker with an under-voltage trigger. With the adjustable threshold values adjusted in the adaptive protection system concept, it will be possible to apply the system to various power levels, from residential homes to administrative buildings. The adaptive protection concept is based on the regulations and standards intended for the basic requirements for the protective measures which need to be implemented in electrical wiring and installations with a voltage level of up to 1000 V AC. [3]

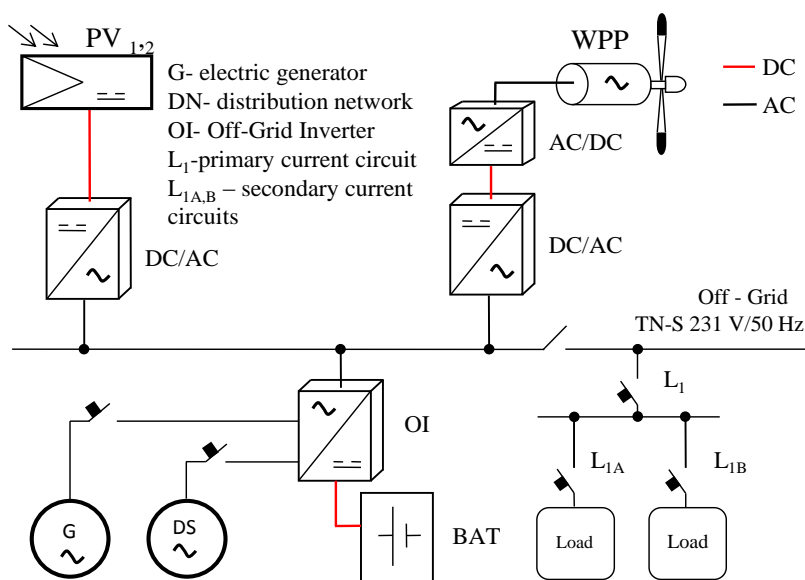


Fig. 1 Simplified wiring diagram of the Off-Grid system (the “By-Pass” topology)

The physical platform of the Off-Grid system built on the premises of VSB–Technical University of Ostrava is based on the By-Pass topology. For a simplified wiring diagram of the By-Pass concept, see Fig.1. The physical model of the electric unit testing platform was used as the basic element in order to create the protection

concept and active control system for the Off-Grid system. [1] The active control system, along with the protection concept, is essential for the long-term and stable operations of the power supply system. The source with the highest rated output is WPP, with an installed power of $8\text{kV}\cdot\text{A}$. It is a wind power plant with a 20-pole synchronous generator equipped with permanent magnets, whose output voltage is first rectified and adjusted to the optimal voltage value at the inverter input. The output power is subsequently led through the DC/AC inverter (with a maximum power output of 5.5 kW) to the AC 1-phase power system. The hybrid source of the Off-Grid system is also completed with two photovoltaic power stations (PV1, PV2). The maximum output of each PV is 2 kWp. PV1 uses polycrystalline panels and PV2 uses monocrystalline panels. The PV 1, 2 output is led into the DC/AC inverter, with a maximum output of 2.1 kW. The Off-Grid inverter, used in the testing platform of the Off-Grid system, creates conditions defined for a standard AC 1-phase network which can integrate individual electrical appliances. The Off-Grid inverter is able to provide continuous consumption of 4 kW, with a maximum of 5 kW for a period of 1 minute. Due to its sophisticated control of the accumulators, the Off-Grid inverter always knows the charging status of the accumulators, and makes further decisions based on its control function in the system. It also determines the optimal strategy for charging the accumulators. If the input DC falls below a predefined limit, the system will shut down automatically. The accumulator charging uses the characteristics of IUoU Active Inverter Technology. [20] To accumulate electric power in the testing platform of the Off-Grid system, Ni-Cd batteries are used in a series-parallel configuration, with a nominal voltage of 24 V. Owing to the testing operations of the entire system, the capacity of accumulators was selected at 750 A·h. [1]

2 Analysis of failure states in the Off-Grid system

The analysis is specified for two failure states. As regards the rate of 1-pole failures which may occur in the electric circuit of the residential home, the most frequent is an overload and 1-phase short circuit. According to the manufacturer and Standards EN 61000-6-1, the Off-Grid inverter has its own protection against over-heating and the output AC is monitored for short circuit and overload. In case of a short-term overload (motor starting current), the Off-Grid inverter of the testing platform is able to provide 50 A for 100 ms. Once the period defined for a short-term overload expires and the maximum current consumption is detected as exceeded, which occurs during the 1-phase short circuit and overload, the Off-Grid inverter of the testing platform will disconnect the AC end of the Off-Grid system. Using a contactor controlled by the Off-Grid inverter, it disconnects the individual current circuit as well as the individual appliances, as well as the sources connected to the primary circuit. The inverter disconnects the power supply to the AC side of the Off-Grid system when the failure duration exceeds 6 seconds. This period for shut-down is intended for the protection of the Off-Grid inverter, but does not correspond to the time limit for a failure shut-down, e.g. upon the 1-phase short circuit in low voltage networks (0.4 s for TN networks).[3]

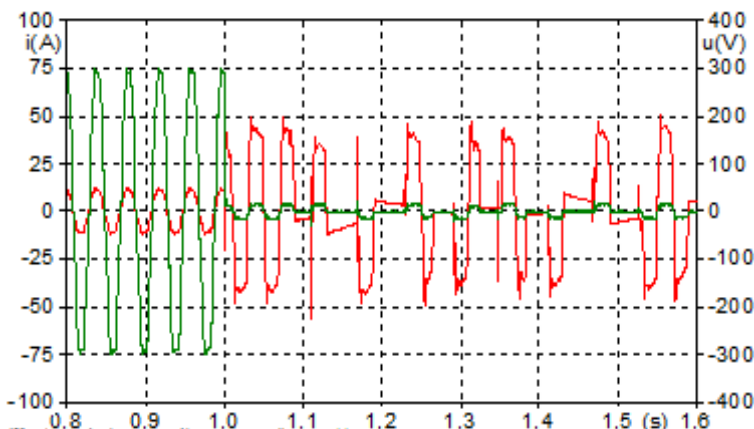


Fig. 2 Voltage and Current waveform during the short circuit in the Off-Grid system

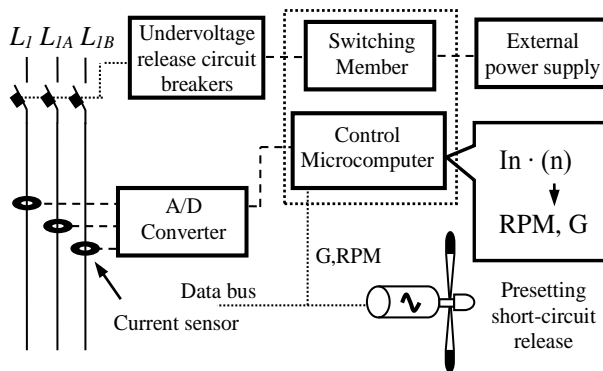
Fig.2 illustrates the curve of the short-circuit current and the voltage at the output end of the Off-Grid inverter. Prior to the simulated short circuit, the Off-Grid inverter is loaded at 45% of its maximum possible load. The voltage value is 231 V and the Off-Grid inverter supplies the Off-Grid network with 8.2 A. Upon short circuit and after the stabilization of the transitional events, the stabilized effective value of the short-circuit current is 34 A, whereas the effective value of voltage decreases to 10 V. The short circuit does not occur in the situation when the voltage passes through zero (it only contains the symmetrical short-circuit current), the short-circuit current will gradually follow the current flowing through the circuit prior to the short circuit. Automatic disconnection from the source is used (breaker, fuse) as a conventional protective measure. The protection will ensure automatic disconnection if the conditions indicated in formula (1) are met. If this requirement is met, the current which flows through the failure loop (depending on the value of the loop impedance) is higher than the release current I_a (in the short-circuit protection) The protection can consequently be shut down in the required time limit (the prescribed releasing time is 0.4 s; this value is established for TN networks). If conventional protection, such as a circuit breaker B, $I_n = 16A$, factor 5, is used, the release current of the breaker is $I_a = 80 A$.

$$I_p = \frac{U_0}{Z_p} (A) \quad Z_p = 0,8 \leq \frac{U_0}{Z_p} = \frac{231}{80} = 2,88(\Omega) \quad (1)$$

Z_p - impedance of the failure short-circuit loop (in the Off-Grid network, at the point of short circuit 0.8 Ω), U_0 - voltage before the short circuit, I_p - fault current

The aforementioned operation state implies that the use of the conventional concept of electric protection (without limiting the maximum power input of the appliances) for safeguarding the low-voltage lines will not ensure the safe shut-down of the failure at the AC side of the Off-Grid network in the first operation state during the determined

time limit. The value of the short-circuit current is insufficient for installing the safeguarding element, such as the B16, B10 breakers ($34 \text{ A} < 80, 50 \text{ A}$), in a determined time limit, a limited short-circuit current flows through the Off-Grid inverter and the point of short circuit. This current cannot be broken by the standard protection concept, therefore a failure state emerges in the Off-Grid system with this state endangering the reliability of the system. Another operation state emerges when the source of electric power and the short-circuit power for the Off-Grid network is the Off-Grid inverter supplied from the batteries as well as from the WPP and PV generator and the power electronics. In the first operation state, the short-circuit current measured at the Off-Grid inverter was 34 A. The maximum contribution of PV, along with the limitation ability of the PV inverter, to the total short-circuit power of the Off-Grid system is just beyond the limit of the PV maximum output (MPP- maximum power point). MPP is the point when the PV inverter sets up so that it corresponds to the present point of the maximum output of the used photovoltaic panels. The maximum current of PV is determined by the short-circuit current which is the maximum current which may be supplied by the solar element at a given sunlight intensity, the maximum contribution of the overall short-circuit current at the point of the short circuit is approximately 10 A. If the short circuit is electrically close to the synchronous generator WPP, the curve of the short-circuit current is more complicated and the change in the electrical parameters of the machine during the short circuit. The reactance of the generator changes gradually, from the impact to the transitional and synchronous value, the maximum contribution of the overall short-circuit current at the point of the short circuit is approximately 118 A.



G-global radiation, *RPM* - rotor speed WPP, *L₁* - primary current circuit, *L_{1AB}* - secondary current circuits

In order to ensure a safe selective protection against the short circuit and overload in all operation modes of the Off-Grid system, there is a need to identify the dynamically changing value of the short-circuit trigger in the electrical protection. In the first operation state (with the short-circuit current source being only the Off-Grid inverter), the value of the short-circuit current is only 34 A. The impulsive value for the short-circuit trigger must be set below this limit $In \cdot n < 34 \text{ A}$. If this limit remains set for the second operation state as well, the maximum transmitted current reaching a value of 50 A would be limited in the By-Pass system concept. The source with the greatest output

and short-circuit current is the WPP with its inverter. In order to avoid the adverse activation of the short-circuit degree by exceeding the impulsive value, this value must dynamically change in real time as a function of the present output of WPP. As regards the WPP generator parameters, the wind run-up speed at WPP is $2,5 \text{ m}\cdot\text{s}^{-1}$, which corresponds to 55 rpm. The maximum wind speed is $12 \text{ m}\cdot\text{s}^{-1}$, which corresponds to 180 rpm. The adaptive protection system, designed by VŠB–Technical University of Ostrava on the basis of measurements, evaluates the RPM of the WPP rotor and the values of global radiation in real time. Based on the measured and evaluated values, the system adjusts the impulsive value of the short-circuit degree for the primary circuit L1. Based on the current RPM of the WPP rotor, which are directly proportional to the wind speed, the system increases or decreases the impulsive value of the short-circuit degree. If the preset impulsive value is exceeded, the independent instantaneous trigger will be released. Time of weaning failure (1-phase short circuit) is less than 0.4 s. Fig.3 depicts the wiring diagram of the adaptive protection system. The adaptive protection system is similar in its activity to the 3rd generation electronic breakers. The basic circuit of this system is formed by the converter of the analogue signal from current probes to the digital format, and the controlling microcomputer which evaluates the flowing current, pre-sets the values of the impulsive component according to the current RPM of the WPP rotor and values of global radiation, and activates the tripping relay at the moment of the breaker switching off.

3 Conclusion

Modern information technologies intertwined with conventional safeguarding devices offer a number of options for designing modern electric protections with the dynamically changing impulsive values of the short-circuit degree. The final result is a highly variable, reliable and economical system for electric power supply at the level of low voltage, for supplying energy self-sustaining houses and buildings.

ACKNOWLEDGEMENTS

This paper was conducted within the framework 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. SP2014/49, project LE13011.

4 References

- [1] SGS Students Grand Competition: Hybrid Energy systems . VSB - TUO. [online]. [cit. 2012-02-02]. Available from: <http://hybridni-system.vsb.cz/vizualizace/>
- [2] Smart Grids. In: Smart Grids Czech technology platform [online]. [cit. 2014-02-02]. Available from: <http://www.smartgridcz.eu>
- [3] EN 61140. Protection against electric shock. Common aspects for installation and equipment. 2001
- [4] N. Miller, R. Zrebiec, R. Delmerico, Battery energy storage systems for electric utility, Conference on Applications and Advances, January 9–12, 1996, pp. 235–240.

Identification of System Parameters using Genetic Algorithms

Jiri Hajovsky

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

Abstract. This article is focused on genetic algorithms and their using for identification of unknown parameters in equations expressing the behavior of a physical system. In the last section of this document is presented experimental result of parameter identification of simple electric circuit with unknown parameters. Measured data and fitness function are inserted into LabView application. Genetic algorithms are used to search the right parameters of the system.

Keywords: Genetic Algorithms, optimization problems, LabView

1 Introduction

There are a lot of new approaches and principles which are based on a relatively new scientific discipline which is called Soft Computing. The main feature of Soft Computing is a departure from traditional modeling based on Boolean logic, analytical models, sharp classification and deterministic search. In the title the word "soft", expressing "softness, gentleness" here means "soft" requirements on the accuracy of the described phenomena. Among the main representative of Soft Computing we include fuzzy logic, neural networks and genetic algorithms - these will be prioritized for attention in the following paper.

2 Genetic Algorithms

Genetic algorithms are stochastic optimization methods used for finding solutions of searched problem. Genetic algorithms try to simulate evolution in the way that occurs in nature. As a theoretical basis or model they take Mandel's theory of genetics and Darwin's theory of natural selection. Darwin's theory basically says that faster, stronger and more intelligent individuals have more chance to survive in a dynamic and constantly changing environment than other individuals - slower, weaker and dumber. These "inferior" individuals in their environment survive rather because they are lucky, than due to its properties. Therefore, reproduction (creating the next generation) also participate, but to a lesser extent than individuals "better". Georg Mendel

has discovered transfer mechanism of character traits from parents to offspring. Later was shown that for the individual properties are responsible gens which are arranged linearly in the chromosomes.

Genetic algorithms are a scanning method for general use (not dependent on the type of the problem), which can be adjusted by setting the parameters a balance between the focus on promising areas and searching the largest part of the state space. These properties of genetic algorithms are one of the biggest advantages:

- Genetic algorithms are robust in the sense that they are applicable to very diverse tasks, taking them only need to make minor adjustments
- Genetic algorithms can work with all kinds of state spaces, including nonsmooth, multimodal and discontinuous
- Genetic algorithms can find solutions in terms of multiple criteria, it is not necessary to explicitly define a common evaluation function
- Genetic algorithms can find more optimal solutions or near optimal solutions
- Genetic algorithms can be used for dynamic optimization

Genetic algorithms try to imitate evolution in nature. The most important part of the genetic algorithm is the fitness function - an equation describing searched problem. First population is generated randomly from specified range of values. According to the results of the individual in the fitness function for each individual is assigned a fitness value that reflects how strong this individual is. Strongest individuals continue in the algorithm and the weaker die.

Next is applied operations of mutation and crossover of most powerful individuals. During the crossing parents remain and the population grows by the new offspring resulting from accidental crossing of chromosomes randomly selected individuals - parents. The newly formed individual may be stronger than parents or weaker and dies in the next generation. The next step is the implementation of a mutation. This creates new individuals by mutation of individuals randomly selected from the population. In the developed algorithm is mutation implemented by multiplication chromosomes and random value close to one. In the next step of the algorithm are randomly generated new individuals. If not to create new random individuals, the space would be poorly scanned and the algorithm would get stuck in a local extreme of fitness function. Add new randomly generated individuals ensures a random search of the entire area, allowing the algorithm to jump out of the field of local extreme. The population was thus extended by new individuals resulting from the crossing and mutation of new individuals, but also randomly generated individuals. The whole population progresses to the next generation, where the process is repeated again - determining of fitness values - selection - crossover - mutation - the emergence of new randomly generated individuals. This process is repeated continuously till not reached the desired number of generations or until the individual is sufficiently mature – high enough fitness value.

The measured points and the function which describe the physical phenomenon are loaded to the application. Chromosomes of the individual are function parameters and function values are calculated in measured points. Differences between measured and calculated values are calculated and summed. Total sum of the deviations then indi-

cates how the individual is strong. The lower sum of differences means that the individual is stronger. The aim therefore is to find the parameters of the function to function as close as possible to the measurement points.

Summation of differences has the consequence that in this case the behavior of the function exactly intersects some points and some points are disproportionately away. Sum of differences is the lowest in this case. Even during the fitting function all points can provide the sum of squared differences. The sum of squared deviations is smaller number than the large square of one difference.

The sum of the differences is exactly interleaves the most points and ignores points outside the course of function. The sum of the deviations quadrates create graphs of functions uniformly interleaved all points.

All measurements generally are flawed and because of this error sometimes it is not possible to achieve sufficient results by substituting into the equation, or by inserting of the measured points in to the system of equations and solve the equation parameters. In case of an inaccurate measurement should be measured points evenly interspersed by curve.

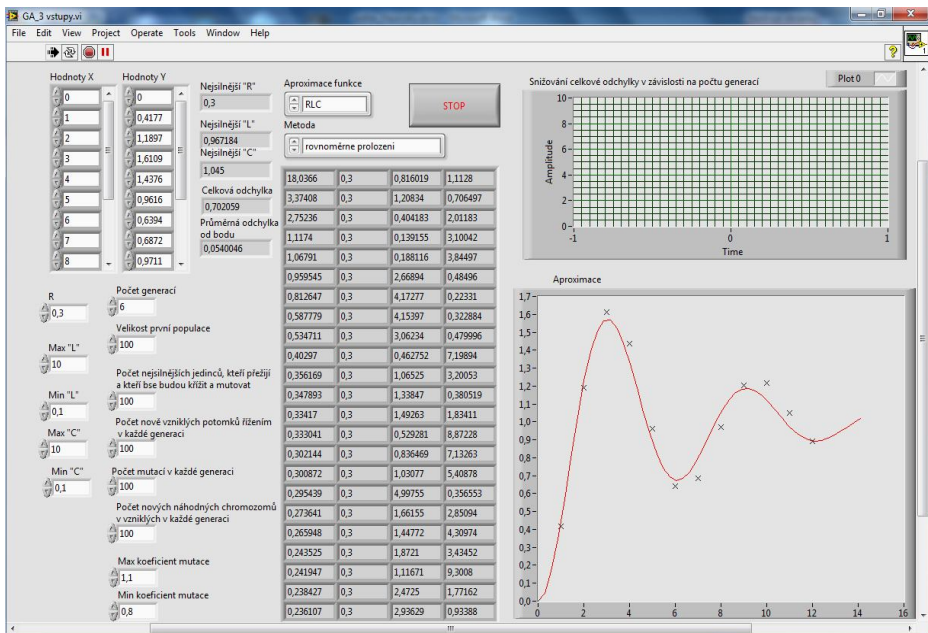


Fig. 1. LabView application

Figure Fig. 1. LabView application shows main screen of developed program. At the left can be set to different parameters such as the number of generations, population size, the number of new individuals per generation, and so on. The table in the middle shows a whole generation. The graph on the right below graphically shows the behavior of the fitness function with the best individual in the generation - the chromosomes of the strongest individual in a generation. The top right graph shows the de-

creasing sum of the deviations depending on the generation of the first ten individuals.

3 Experimental results

Genetic algorithms solve optimization problems in a number of industries. In economics can be used genetic algorithms to calculate the best investments in the industry to calculate optimization. In the field of electronics genetic algorithms are used to identify the parameters of electronic systems, optimizing controller settings and optimizing adjustment of adaptive controllers. Adaptive controller’s parameters need to be changed depending on the operation of the entire system. While the engine is running the engine warms up and there are changes in the electrical parameters and the adaptive controller that must to react.

I used a genetic algorithm to identify the unknown parameters of the electrical circuit from the step response. As electrical circuit I chose passive second order low pass filter. Figure 2 shows scheme of circuit and transfer function describing this circuit is given by equation (1), where ω_0 and Q is given by (2) and (3), respectively. Equation (4) shows solution of Step response in Laplace Domain and equation (5) shows “converting” from Laplace Domain to Time Domain.

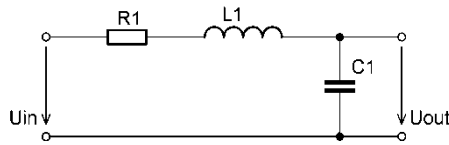


Fig. 2. Passive low pass filter

$$G(s) = \frac{U_{out}}{U_{in}} = \frac{\omega_0^2}{s^2 + \frac{s \cdot \omega_0}{Q} + \omega_0^2} \quad (1)$$

where:

$$\omega_0^2 = \frac{1}{L \cdot C} \quad (2)$$

$$Q = \frac{L \cdot \omega_0}{R} \quad (3)$$

Step response in Laplace Domain:

$$H(s) = \frac{1}{s} G(s) \quad (4)$$

Step response in Time Domain:

$$h(t) = L^{-1}\{H(s)\} \quad (5)$$

Solution in Matlab:

```

syms s t R L C;
G = (1/(L*C)) / (s^2+((R/L)*s)+(1/(L*C)));
H = G*(1/s);
h = ilaplace(H,s,t);
pretty(simplify(h))
    
```

$$1 - \exp\left(-\frac{R t}{2 L}\right) \frac{\cosh\left(\frac{\sqrt{4 C R - L^2} t}{4}\right) + \frac{C R \sinh\left(\frac{\sqrt{4 C R - L^2} t}{4}\right)}{2 \sqrt{4 C R - L^2}}}{\sqrt{4 C R - L^2}} + \frac{\exp\left(-\frac{R t}{2 L}\right) \left(\frac{C R}{2 \sqrt{4 C R - L^2}} \cosh\left(\frac{\sqrt{4 C R - L^2} t}{4}\right) + \frac{1}{2} \sinh\left(\frac{\sqrt{4 C R - L^2} t}{4}\right)\right)}{\sqrt{4 C R - L^2}}$$

Fig. 3. Step response in time domain solved by Matlab

Figure 3 shows equation of step response of second order low pass filter in time domain solved by Matlab. Values of the electrical components were chosen: $R = 0,3 \Omega$, $L = 1 \text{ H}$, $C = 1 \text{ F}$. Parameters Function values were calculated for each second and these values were inserted into LabView application. It was also necessary to insert the value of the resistance R , as otherwise there would be infinitely many solutions. The population was chosen size of 50 individuals.

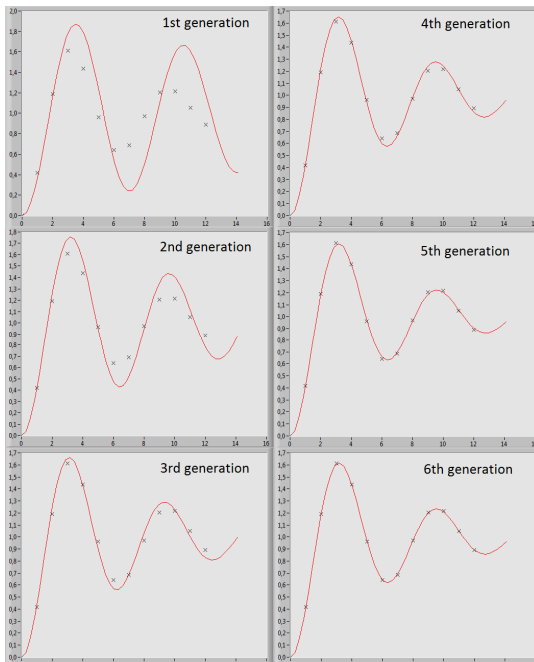


Fig. 4. Development of the fittest individuals in each generation

Conclusion

The figure 4. shows the improvement of the graph of function with the best individual in each generation. In each generation the sum of deviations is much smaller and the curve follows the given points. Applications typically finds parameters L and C with an error smaller than 0,1% in six generation with the first population size of fifty individuals. This application can be used for almost every optimization problem and in future it will be used for determine the parameters of the induction motor.

Acknowledgment

In the paper there are the results of the project SP2014/119 which was supported by Student Grant Competition of VSB - Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, CSc.

References

1. MARIÁN, Jančovič. *Využitie genetických algoritmov v striedavých servo-systémoch*. Bratislava, 2008. Disertační práce. Slovenská technická univerzita v Bratislavě.
2. HAUPT, Randy L. a Sue Ellen HAUPT. *PRACTICAL GENETIC ALGORITHMS*. Hoboken, New Jersey: A John Wiley & Sons, Inc., Publication, 2004. 2nd edition. ISBN 0-471-45565-2.
3. RADÍČOVÁ, Tatiana. *Moderné prístupy číslicového riadenia polohových servo-pohonov s LMPM*. Bratislava, 2012. Disertační práce. Slovenská technická univerzita v Bratislavě.
4. URSEM, Rasmus K. *Models for Evolutionary Algorithms and Their Applications in System Identification and Control Optimization*. Aarhus, Denmark, 2003. ISBN 1396-7002.
5. VOLNÁ, Eva. *Základy softcomputingu*. Ostrava, 2012.

The Contribution of 3-Phase Induction Motor Parameters Identification Using the Equivalent Circuit

Jan Otýpka

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.otypka@vsb.cz

Abstract. This paper is focused on the parameters identification 3 – phase induction motor with squirrel cage. The identification method is introduced for equivalent circuit shaped Γ -network. The direct application of identification of parameters is introduced for asynchronous motor type: AOM090L02-016, with power rating 2.2 kW, $2p = 2$.

Keywords: Induction motor, power factor, power quality, efficiency, parameter identification

1 Introduction

The three phase squirrel cage induction motors (IM) are widely used in a large scale of industry applications because of their simple construction, reliability, low maintenance, wide range of power ration and a relatively low cost.

The mathematical model of the one phase IM in form of the equivalent circuit is used for its properties and behavior the description. This equivalent circuit contains several circuit elements (parameters). These parameters shall be determined very precisely. The equivalent circuit of IM may be used in the Γ -network, inverse Γ -network or the classical T-network form.

This paper is especially focused on the parameters identification with using the Γ -network. The principle of this method is performed by the parameters identification for the Γ -network in the operation area of the IM torque characteristic. The paper deals with the problem of this method, because the different solutions are obtained by the mutual combinations of the torque characteristic points.

The identification of parameters is performed for motor rated at 2.2kW, type: AOM090L02-016, $2p = 2$.

2 The parameters identification of induction motor by method of adjacent points

The parameters of IM are required to identify for the purposes of analysis and simulations. The mathematical model of one-phase IM is most often used as the substitute circuit in T-network form and $(L_{\sigma 1}, L'_{\sigma 2}, L_{1m}, R)$. The Γ -network is completely sufficient for the purposes of identification, because it consists of only three parameters $(L'_{\sigma 2}, L_{1m}, R)$, which are required to identify in comparison to T-network, which has four parameters $(L_{\sigma 1}, L'_{\sigma 2}, L_{1m}, R)$. The advantages of using the Γ -network are described in [1].

The method of adjacent points is used for the purpose of IM identification parameters. This method is derived for Γ -network and it is introduced in [1], [2] and [3]. The motor input impedance is intended in both points A and B, which are situated on the torque/mechanical characteristic in the vicinity of the nominal point N (the nominal point N should be between points A and B or it should be one of them). The input impedance is given by the formula:

$$\underline{Z}_{1,A}(\underline{Z}_{1,B}) = R_{1,A}(R_{1,B}) + j\omega \cdot L_{1,A}(L_{1,B}). \tag{1}$$

The equivalent circuit parameters R_1 (resistance of stator winding), R_{Fe} (equivalent resistance for core loos), L_{1m} (magnetizing inductance), $L'_{\sigma 2}$ (rotor inductance brought to stator), R (equivalent resistance for mechanical load of IM and rotor winding resistance brought to stator) are funded for the measured values R_A, L_A in point A for slip s_A and R_B, L_B in point B for slip s_B . The Γ -network equivalent circuit input impedance on Fig. 1b) must be equal the experimental measured impedance on Fig. 1a).

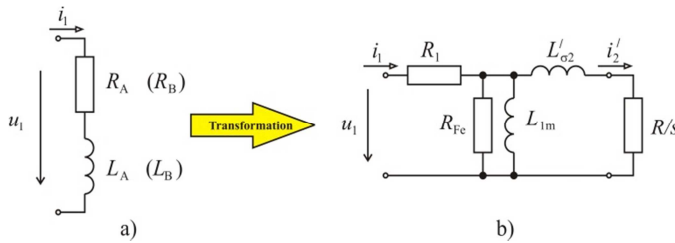


Fig. 1. Input impedance $\underline{Z}_{1,A}(\underline{Z}_{1,B})$ a) experimental measured impedance and equivalent circuit in Γ -network impedance b) (taken from [2])

The Γ -network equivalent circuit input impedance is given by the formula:

$$\underline{Z}_{1,\Gamma} = R_1 + \frac{j \cdot \omega L_1 \cdot R_{Fe} \cdot \left(j \cdot \omega L'_{\sigma 2} + \frac{R}{s} \right)}{j \cdot \omega L_1 \cdot R_{Fe} + j \cdot \omega L_1 \cdot \left(j \cdot \omega L'_{\sigma 2} + \frac{R}{s} \right) + R_{Fe} \cdot \left(j \cdot \omega L'_{\sigma 2} + \frac{R}{s} \right)}, \tag{2}$$

where s is the slip. The comparison of the experimental measured impedances $\underline{Z}_{1,A}$ and $\underline{Z}_{1,B}$ (it can be seen in Fig. 2) with the algebraic derived formula for the impedance of Γ -network $\underline{Z}_{1,\Gamma}$ is essential for this method.

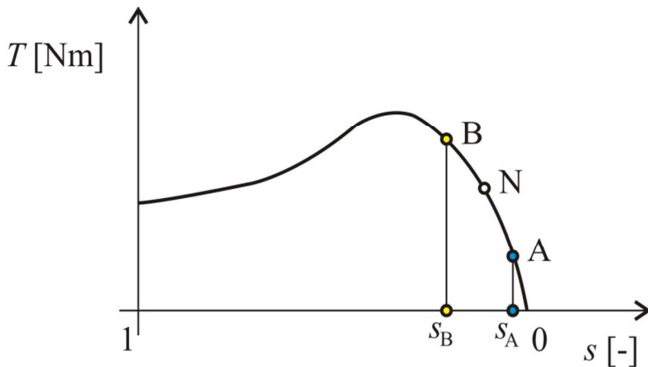


Fig. 2. The method principle of the adjacent points on the torque/mechanical characteristic.

The set of four equations with four variables are obtained from the comparison of the real and imaginary sides of the impedances $\underline{Z}_{1,A} = \underline{Z}_{1,\Gamma-A}$ for point A and $\underline{Z}_{1,B} = \underline{Z}_{1,\Gamma-B}$ for point B. The parameters R , $L'_{\sigma 2}$ and L_{1m} are obtained from these equations, which are introduced in [1]. The parameter R_{Fe} is difficult to determine, because we must deal with the 14th - level equation. Therefore, the parameter R_{Fe} must be a known value, which is obtained from the no-load measuring. The resistance R_1 is measured by the Ω -meter on the IM terminal. These parameters R , $L'_{\sigma 2}$ and L_{1m} are given by the formulas:

$$L_{1m} = \frac{K_{10} \pm \sqrt{K_{10}^2 - K_9 K_{11}}}{K_9}, \tag{3}$$

$$L'_{\sigma 2} = \frac{L_{1m}^2 K_4 - L_{1m} K_3}{L_{1m}^2 K_2 - L_{1m} \cdot 2 \cdot K_4 + K_3}, \tag{4}$$

$$R = \frac{\omega^2 L_A L_{1m} L'_{\sigma 2} s_A - L_{1m} K_1 - L'_{\sigma 2} K_1}{L_A R_{Fe} + L_{1m} R_X - L_{1m} R_{Fe}}, \tag{5}$$

where K_1 till K_{11} are the substitution constant, which are given by the algebraic combinations of the parameters R_1 , R_{Fe} , R_A , R_B , L_A , L_B , s_A , s_B and ω . These constants are given by formulas:

$$K_1 = R_{Fe} R_X s_A, \tag{6}$$

$$R_X = R_A - R_1 \tag{7}$$

$$K_2 = \omega^2 R_X^2 - 2\omega^2 R_{Fe} R_X + \omega^4 L_A^2 + \omega^2 R_{Fe}^2 \tag{8}$$

$$K_3 = \omega^2 L_A^2 R_{Fe}^2 + R_{Fe}^2 R_X^2 \quad (9)$$

$$K_4 = \omega^2 L_A R_{Fe}^2 \quad (10)$$

$$K_5 = R_{Fe} R_X R_Y \left(1 - \frac{s_A}{s_B} \right) + R_{Fe}^2 R_X \frac{s_A}{s_B} - R_{Fe}^2 R_Y \quad (11)$$

$$R_Y = R_B - R_1 \quad (12)$$

$$K_6 = L_A R_{Fe}^2 R_Y - L_B R_{Fe}^2 R_X \frac{s_A}{s_B} \quad (13)$$

$$K_7 = K_5 - \omega^2 L_A L_B R_{Fe} \left(1 - \frac{s_A}{s_B} \right) \quad (14)$$

$$K_8 = \omega^2 L_A \frac{s_A}{s_B} (R_Y - R_{Fe}) + \omega^2 L_B (R_{Fe} - R_X) \quad (15)$$

$$K_9 = K_2 K_5 + K_4 K_8 \quad (16)$$

$$K_{10} = \frac{1}{2} (K_3 K_8 + 2K_4 K_5 - K_2 K_6 - K_4 K_7) \quad (17)$$

$$K_{11} = K_3 K_5 - K_4 K_6 - K_3 K_7 \quad (18)$$

The both different cases could be obtained for the solution of this method, when both point A and B are deliberately changed as it is shown in Fig. 3 a) and b). This change has a great impact to the Γ -network calculation solution. Respectively, the both procedures (method of points A + B and method of points B + A – it is defined from the slip 0 to 1) are generated by this change. Each procedure generates the different mathematical correct solution. These solutions could be physically feasible or not. It is given by the formula (3) for L_{1m} , which is the solution of the quadratic equation.

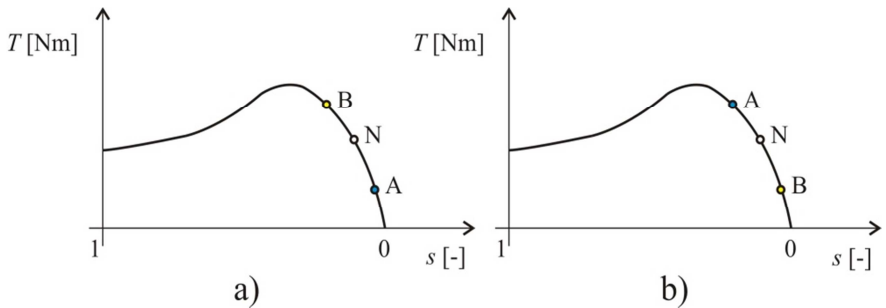


Fig. 3. The adjacent points method principles a) A + B and b) B + A.

3 The parameters identification from the measuring on the induction motor

The parameters identification was realized for IM type: AOM090L02-016, with power rating 2.2kW, $2p = 2$, 400V and frequency 50Hz. The analysis was made by the torque characteristic, which is introduced in Tab. 1. The motor was measured in the nine operating points for the different load and the same voltage (3x400V). The measured values were the motor slip s (speed), the torque T , the active P and reactive Q power from the main, the effective values of voltage U_{rms} and the current I_{rms} in the each phase. The calculated values are the apparent power and $\cos\phi$ from the measuring. The input impedances $\underline{Z}_{1,A}$ ($\underline{Z}_{1,B}$) were calculated from these values, it is shown in tab. 2. The point A is selected for torque 4Nm (it is marked by the blue color) and point B is selected for torque 8Nm (it is marked by the yellow color) for method A + B, these points are chosen for method B + A.

Table 1. The table of experimental measured and calculated values of active power P, Q a S for the different loading torque of IM type: AOM090L02-016

	T [Nm]	n [rpm]	U_{rms} [V]	I_{rms} [A]	P [W]	Q [VAr]	S [VA]	$\cos\phi$
	2	2972	231.87	2.320	838.12	1378.25	1621.03	0.5194
A	4	2947	231.75	2.937	1490.26	1394.49	2041.40	0.7300
	5	2934	231.22	3.333	1821.32	1421.69	2311.16	0.7881
	6	2918	231.02	3.723	2133.16	1451.65	2580.98	0.8265
	7,3	2898	230.96	4.307	2574.92	1511.47	2984.18	0.8623
B	8	2886	230.90	4.640	2814.46	1554.09	3220.00	0.8753
	9	2866	230.80	5.137	3160.00	1616.86	3560.00	0.8906
	10	2844	230.86	5.677	3550.00	1694.88	3930.00	0.9022
	12	2800	230.62	6.750	4270.00	1881.43	4670.00	0.9153

Table 2. Table of values slip s and input impedance for this IM

$s_A, (s_B)$ [-]	$R_A, (R_B)$ [Ω]	$L_A, (L_B)$ [H]	$\underline{Z}_{1,A}, (\underline{Z}_{1,B})$ [Ω]
0.0093	51.9050	0.2717	51.91+85.36i
0.0177	57.6012	0.1716	57.60+53.90i
0.0220	54.6396	0.1358	54.64+42.65i
0.0273	51.2907	0.1111	51.29+34.90i
0.0340	46.2764	0.0865	46.28+27.16i
0.0380	43.5750	0.0766	43.58+24.06i
0.0447	39.9212	0.0650	39.92+20.43i
0.0520	36.7215	0.0558	36.72+17.53i
0.0667	31.2391	0.0438	31.24+13.76i

The solutions of this identification are introduced in Tab.3. The combinations of both points A and B are obtained from four solutions, but only once it is physical feasible (it is marked by the green color).

Table 3. The identified the parameters of IM: AOM090L02-016 from elected values of torque $T_A = 4\text{Nm}$ a $T_B = 8\text{Nm}$ with method A + B and B + A.

		Metod A + B		Metod B + A	
		Resistance of stator winding	R_1 [Ω]	3.1	3.1
Equivalent resistance for core loos	R_{Fe} [Ω]	768.52	768.52	768.52	768.52
Equivalent resistance for mechanical load and rotor resistance brought to stator	R [Ω]	1.1012	2.2091	2.3665	0.4707
Magnetizing inductance	L_{1m} [H]	0.1847	0.3420	0.3401	0.0455
Rotor inductance brought to stator	$L'_{\sigma 2}$ [H]	-0.1990	-0.0047	0.0193	-0.0459

4 Conclusion

This paper deals with the procedure for the parameter identification equivalent circuit in the Γ -network form for the IM with the usage of the method of adjacent points, which is appropriately used for the identification on the torque/mechanical characteristic. This method could be declared as accurate, because this analysis is performed in the operating area of the torque/mechanical characteristic, i.e. in stable area of shaft speed. The different mathematical solutions are generated by this method, when some solutions are physical feasible and some are not, as shown in Tab. 3.

Acknowledgment

This paper has been supported by project SGS №. SP2014/45. This research was supervised by assoc. prof. Vítězslav Stýskala, Ph. D. from Department of Electrical Engineering, VŠB - TU Ostrava.

References

1. BĚLOUŠEK, J., PATOČKA, M.: Identifikace parametrů asynchronního motoru, EPVE 2006. Brno, VUT Brno, FEKT, UVEE, 2006. s. 179 (s.), ISBN: 80-214-3286- 1.
2. OTÝPKA, J.: Identifikace parametrů náhradního zapojení asynchronního motoru. Brno, Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008, 68 p.
3. NOVOTNÝ D. W., LIPO T. A.: Vector Control and dynamics of AC Drives. Oxford university Press Inc., New York, 1996.

Design of Luenberger Observer for Sensorless Control of Induction Motor

Ondřej Petrtýl

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

Abstract. The aim of this paper is using a Luenberger state observer to estimate a rotor speed of the sensorless vector controlled induction motor. As a control structure we used the commonly known structure of the vector control for the IM supplemented by a block for an estimation of the rotor speed. The paper also deals with a calculation of poles of the IM and their placement, because if we have knowledge about the IM, we can better design poles of observer.

Keywords: sensorless speed control, induction motor, Luenberger observer, poles placement

1 Introduction

The induction motors (IM) are more and more popular for their attributes. They are relatively cheap, simply and reliably constructed with low requirements for the maintenance. The control techniques for the IM use information of the rotor speed which comes from the speed sensor. This sensor is a source of problems, by removing this sensor from controlled structure we receive the sensorless control structure, which still contains other sensors e.g. the current or voltage sensors.

The sensorless drives have lower requirements for a maintenance, simpler construction, higher reliability and mechanical robustness. They also can be used in applications, where the speed sensor can't be used. Main purpose in sensorless control is estimation of position or speed.

The speed estimators, operated on the motor model, can be divided to these three groups – open loop estimators, MRAS (Model Reference Adaptive System) and observers (Kalman, Luenberger, Gopinath etc.). [1]

2 Induction Motor

The current-flux mathematical model of the IM in the stator coordinate system [2] is shown by equation (1).

$$\frac{d}{dt} \begin{bmatrix} \mathbf{i}_s^S \\ \boldsymbol{\psi}_r^S \end{bmatrix} = \begin{bmatrix} -\frac{L_m^2 R_r + L_r^2 R_s}{\sigma L_s L_r^2} & \frac{L_m R_r}{\sigma L_s L_r^2} - \frac{L_m}{\sigma L_s L_r} j\omega_r \\ \frac{L_m R_r}{L_r} & -\frac{R_r}{L_r} + j\omega_r \end{bmatrix} \begin{bmatrix} \mathbf{i}_s^S \\ \boldsymbol{\psi}_r^S \end{bmatrix} + \begin{bmatrix} \frac{1}{\sigma L_s} \\ 0 \end{bmatrix} \mathbf{u}_s^S \quad (1)$$

$$\sigma = 1 - \frac{L_m^2}{L_s L_r} \quad (2)$$

Where:

\mathbf{i}_s^S	Stator current space vector with components $i_{s\alpha}, i_{s\beta}$
\mathbf{u}_s^S	Stator voltage space vector with components $u_{s\alpha}, u_{s\beta}$
$\boldsymbol{\psi}_r^S$	Rotor magnetic flux space vector with components $\psi_{r\alpha}, \psi_{r\beta}$
L_m	Magnetizing inductance
L_s, L_r	Stator and rotor inductance
R_s, R_r	Stator and rotor resistance
ω_r	Electrical angular speed
σ	Leakage factor

In our case we are using these space equations (3), (4), where respective matrices are (5) – (7) [2].

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad (3)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \quad (4)$$

$$\mathbf{x} = [i_{s\alpha} \quad i_{s\beta} \quad \psi_{r\alpha} \quad \psi_{r\beta}]^T, \quad \mathbf{u} = [u_{s\alpha} \quad u_{s\beta}]^T \quad (5)$$

$$\mathbf{A} = \begin{bmatrix} -\frac{L_m^2 R_r + L_r^2 R_s}{\sigma L_s L_r^2} & 0 & \frac{L_m R_r}{\sigma L_s L_r^2} & \frac{L_m}{\sigma L_s L_r} \omega_r \\ 0 & -\frac{L_m^2 R_r + L_r^2 R_s}{\sigma L_s L_r^2} & -\frac{L_m}{\sigma L_s L_r} \omega_r & \frac{L_m R_r}{\sigma L_s L_r^2} \\ \frac{L_m R_r}{L_r} & 0 & -\frac{R_r}{L_r} & -\omega_r \\ 0 & \frac{L_m R_r}{L_r} & \omega_r & -\frac{R_r}{L_r} \end{bmatrix} \quad (6)$$

$$\mathbf{B} = \frac{1}{\sigma L_s} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (7)$$

3 Luenberger Observer

The Luenberger observer (LO) is a typical representative of the close loop estimators, so-called observers, namely of the deterministic observers. The expressions (8), (9) describe the basic Luenberger observer, which can be used for state estimation of a time invariant system. [3]

$$\dot{\hat{x}} = \hat{A}\hat{x} + Bu + G\tilde{y} = \hat{A}\hat{x} + Bu + G[y - \hat{y}] \tag{8}$$

$$\hat{y} = C\hat{x} \tag{9}$$

$$\hat{A} = A|_{\omega_r = \hat{\omega}_r} \tag{10}$$

Equations (8) and (9) are the base for design a block diagram (Fig. 1) [2]. The observer state matrix \hat{A} is a transformed state matrix A by eqn (10). The input matrix B and output matrix C are same as in the IM description (7).

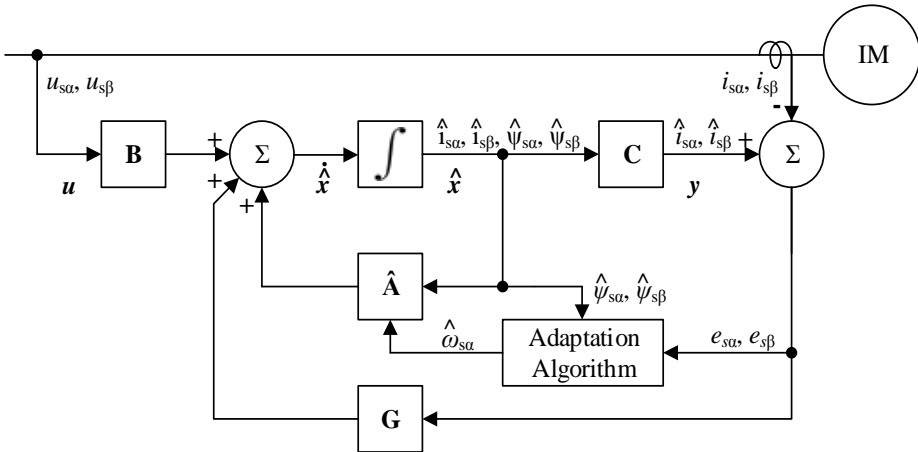


Fig. 1. Block diagram for the Basic Luenberger Observer

The observer matrix G can be written as (11), and their components are described by (12) – (15). [4]

$$G = \begin{bmatrix} g_1 & -g_2 \\ g_2 & g_1 \\ g_3 & -g_4 \\ g_4 & g_3 \end{bmatrix} \tag{11}$$

Design of the observer matrix is based on the rule that the Luenberger observer must be stable and faster than observed system, in our case the IM. [1]

$$g_1 = (k - 1)(a_{11} + a_{33}) \quad (12)$$

$$g_2 = (k - 1)\hat{\omega}_r \quad (13)$$

$$g_3 = (k^2 - 1)(ca_{11} + a_{33}) - c(k - 1)(a_{11} + a_{33}) \quad (14)$$

$$g_4 = -c(k - 1)\hat{\omega}_r \quad (15)$$

$$c = -\frac{\sigma L_s L_r}{L_m} \quad (16)$$

Where a_{11} , a_{31} , a_{33} are components of matrix \mathbf{A} from (6).

The estimated speed $\hat{\omega}_r$ can be given by an adaptation algorithm by two different expressions. First procedure is a simple integration (17), second one is integration with a proportional component (18). The correction term z is defined by term (19). [1]

$$\hat{\omega}_r = K_I \int z \, dz \quad (17)$$

$$\hat{\omega}_r = K_P z + K_I \int z \, dz \quad (18)$$

$$z = e_{is\alpha}\hat{\psi}_{r\beta} - e_{is\beta}\hat{\psi}_{r\alpha} = (i_{s\alpha} - \hat{i}_{s\alpha})\hat{\psi}_{r\beta} - (i_{s\beta} - \hat{i}_{s\beta})\hat{\psi}_{r\alpha} \quad (19)$$

4 Poles placement

We can compare poles placement of the IM and the LO for better design of the observer's parameters. For finding the poles can be used (20), where matrix \mathbf{A} is the same as in (6).

$$s\mathbf{I} - \mathbf{A} \quad (20)$$

Next step is to find determinant of (20), which gives us characteristic equation $N(s)$ (21). Form $N(s)$ we can calculate roots, namely real and complex.

$$\det(s\mathbf{I} - \mathbf{A}) = a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0 = N(s) \quad (21)$$

For a calculation of poles for the LO we need another expression (22), where \mathbf{A} and \mathbf{C} are same as in (6) and (7). Next step is calculation roots of equation $N_g(s)$. [4]

$$\det(s\mathbf{I} - (\mathbf{A} + \mathbf{GC})) = a_{gn} s^n + \dots + a_{g1} s + a_{g0} = N_g(s) \quad (22)$$

5 Simulation results

I made simulation a of poles of the IM and the LO in MATLAB. First step to set settings parameters of the IM. Namely $R_s = 2,1 \Omega$, $R_r = 2,51 \Omega$, $L_s = 137 \text{ mH}$, $L_r = 137 \text{ mH}$ and $L_m = 129 \text{ mH}$.

In second step I wrote down components of the state matrix \mathbf{A} (6) and the observer matrix \mathbf{G} (13) – (16) and other necessary equations. I calculated results in loop by using command `eig()` to obtain eigenvalues (eigenvector) and command `real()`, `imag()` to obtain real and imaginary part of solution, respectively.

```
lambdaA=eig(A); lambdaG=eig(A+GC);
xA=real(lambdaA); yA=imag(lambdaA);
xG=real(lambdaG); yG=imag(lambdaG);
```

[Example of code for calculating poles of the IM and the LO in MATLAB]

Last step was plotting a graph (Fig 2.), where black curve is for poles of the IM and red one is for poles of the LO. I tuned parameters of the LO via coefficient k . I obtained optimal result if $k = 1,05$.

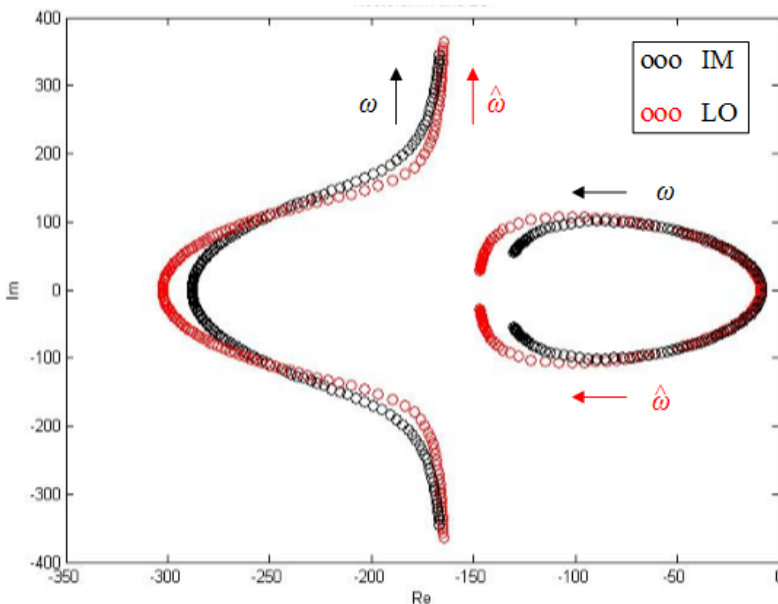


Fig. 2. Placement of poles of the IM and LO

6 Conclusions

The aim of this paper was the design of the Luenberger observer as estimation technique of speed for the induction motor. It was simulated in program MATLAB, based on theory presented aforesaid. I described mathematical model of the IM, equations necessary to designed the LO. The speed adaptive algorithm for estimating speed from a difference between measured and estimated currents components. I also did calculation of poles for the IM and the LO shown by graph obtained from simulation results.

Follow-up in this research will be application of the LO to the vector controlled IM, at first in simulations, followed by experimental measurements. I will also have interest in further observers e.g. the Extended LO or the Gopinath observer.

7 Acknowledgement

In the paper there are the results of the project SP2014/119 which was supported by Student Grant Competition of VSB - Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, CSc.

8 References

1. Brandštetter, P. - Friedrich, J. Speed Estimation of Induction Motor Using Luenberger Observer. In Proceedings of the Acta Technica CSAV. vol. 57, pp. 113-126, 2012. ISSN 0001-7043.
2. Bose, B. K. Modern power electronics and AC drives. Upper Saddle River, NJ : Prentice Hall PTR, 2002, 711 pp. ISBN 0-13-016743-6.
3. Luenberger, D. G. An introduction to observers. Automatic Control, IEEE Transactions on, 1971, vol. 16, no. 6, pp. 596–602. ISSN 0018-9286.
4. Kojabadi, H. M. – Chang, L. Comparative Study of Pole Placement Methods in Adaptive Flux Observer, Control Engineering Practice 13, pp.749-757, 2005. ISSN 0967-0661.
5. Žalman, M. - Gacho, J. IM Based Speed Servodrive eith Luenberger Observer, Journal of ELECTRICAL ENGINEERING, VOL. 61, NO. 3, pp. 149–156, 2010. ISSN 1335-3632.

Induction Generator Optimized Operation for Small-Scale Wind Power in Program LabVIEW

Stanislav Rusnok, Pavel Sobota, and Michal Kelnar

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{stanislav.rusnok, pavel.sobota, michal.kelnar}@vsb.cz

Abstract. This article discusses a method of determining the optimal operation of the induction machine in generator mode in off-grid. This method of determining the optimum is processed into a program that was created in LabVIEW. This program is able to evaluate the optimal operation of the induction generator in island mode from the calculated parameters of machine and measured characteristics of the machine.

1 Introduction

Wind power has its origins in the incident solar radiation, the solar energy heats the air near the earth's surface. Due to different warming in various fields, there are significant differences in temperature of air zones. The result is a horizontal air flow, known as wind. Small-scale wind power is the name given to wind generation systems with the capacity to produce up to 10 kW of electrical power. Off-grid that may otherwise rely on diesel generators may use wind turbines as an alternative. Individuals may purchase these systems to reduce or eliminate their dependence on grid electricity for economic reasons, or to reduce their carbon footprint. Wind turbines have been used for household electricity generation in conjunction with battery storage over many decades in remote areas.

2 Theoretical Analysis

A. Small-scale wind power

S-SWPs are usually mounted with induction generator with squirrel cage rotor. To a lesser extent, are mounted with synchronous generators with permanent magnets or DC motors. Induction motor needs reactive power for creating a magnetic field source machine. According to the operating conditions, it may be a capacitor battery or capacity of connected network. From the viewpoint of construction is not the difference between induction motor and induction generator. Therefore, it is common that S-SWPs are mounted mass-produced induction motors operating in generator mode. Installed wind power plants equipped with mass-produced induction motors, have low power factor at nominal electrical power therefore the induction machine is very current overload. Based on this experience, we perform a usability analysis of an induction motor acting as a generator.

The analysis is based on change of energy flow, so it's substitution of electric power and mechanical power engine mechanical power and electric power generator. Power induction motor P_{OUT} is the mechanical power P_H on the machine shaft according to (1):

$$P_H = \omega \cdot M = \frac{\pi \cdot n \cdot M}{30} \quad (1)$$

Where P_H is the power of the induction motor, ω is the angular velocity of the shaft, and M is the torque on the shaft. Power of induction motor is given by electrical input P_{EL} on motor terminal's according to the equation (2):

$$P_{EL} = \sqrt{3} \cdot U \cdot I \cdot \cos \varphi \quad (2)$$

Where is P_{EL} electrical input power, U is the voltage, I is the current and $\cos \varphi$ is the power factor. Losses of machine P_Z are the difference between electrical input power P_{INP} and electrical power P_{OUT} of machine are given by (3)

$$P_Z = P_{INP} - P_{OUT} \quad (3)$$

Power of induction generator P_{OUT} means the electric power P_{EL} on machine terminal according to (2) and P_{INP} is mechanical input power P_H are given by (1).

On Fig. 1 is a replacement circuit diagram a) induction motor b) asynchronous generator. A fundamental change is reflected on the size of the internal induced voltage U_i .

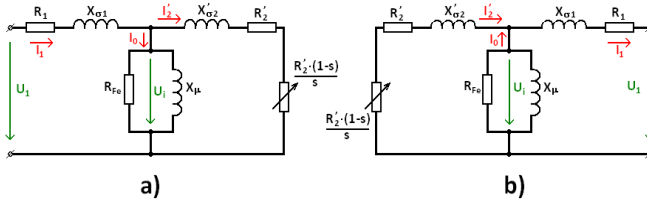


Fig. 1 Equivalent circuit

The size of the induced voltage of induction motor is derived from the (4)

$$U_i = U_1 - Z_1 \cdot I_1 \tag{4}$$

and the induced voltage of the generator is calculated according (5)

$$U_i = U_1 + Z_1 \cdot I_1 \tag{5}$$

Where U_1 is the voltage on terminal of the machine and $Z_1 I_1$ is the voltage drop on the stator. When an induction generator will operate on identical terminal voltage as an induction motor, the size of the induced voltage of machine will be increased by twice the value of the voltage drop on the stator. This is especially increasing the magnetic flux machines, higher induction in the air space, reducing power factor of the machine, increasing the total losses and danger of overheating machine. Reducing the power factor of machines decreases quality supplied electricity and increasing losses, decreasing of maximum machine utilization. Thus operated machine is able to supply to the network 50 to 70 percent of the rated electrical power.

B. Induction Generator in Off-Grid

Induction generator needs reactive energy for to create a magnetic field. Capacitors in parallel connected to the stator winding of the generator are the most commonly used for excitation. To determine the capacity of the capacitor it is necessary to know the magnetizing current of the induction motor obtained from no-load measurements. We proceed from the equality of induction reactance X_L with capacitor reactance X_C . Now we can determine the size of the capacitor reactance X_C

$$X_C = \frac{U_1}{I_0} \tag{6}$$

Where U_1 is the phase voltage of machine and I_0 is the no-load current. The value of the required capacitance of the capacitor C will be then

$$C = \frac{1}{2 \cdot \pi \cdot f \cdot X_C} \tag{7}$$

C. Measurement induction generator

Measurements were carried out on induction machine Siemens 1LA7106-0AB (4 poles / YY stator connection). Capacity of the capacitor was $46\mu\text{F} \pm 20\%$. We set rpm by dynamometer at a given value. After that we connected load and gradually burdening the generator. Measurement was carried out for $n = \text{const.}$ Schema of measurement is shown on Fig. 2. Voltages and currents were measured by voltmeter and ampere meter in each phase. Load with purely ohmic character, we have connected to the motor terminals after the motor has excited.

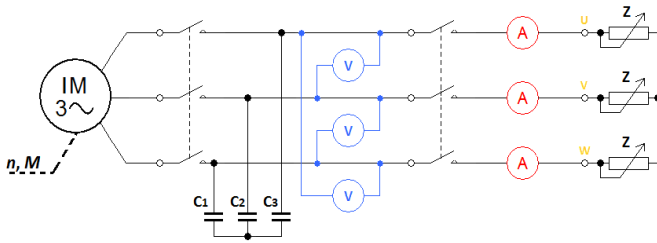


Fig. 2 Scheme of measurements

We have drawn output waveforms of induction generator from measured data (Fig. 3). We can determine from these waveforms the optimal operation of induction generator by LabVIEW.

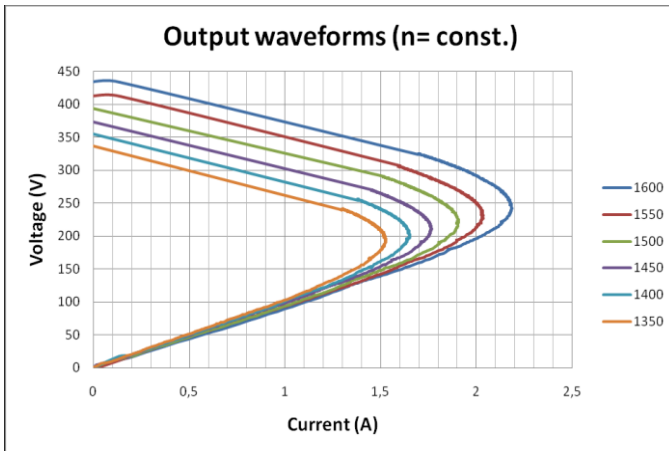


Fig. 3 Measured waveform of the generator

3 Program to determine the optimal operation of induction generator

Model to determine the optimized operation of the induction generator is made in the program LabVIEW 7. Dialog box for setting input parameters of induction generator is displayed after starting the program (Fig. 4). There is also performed calculation of a recommended value capacity of the capacitor (7).

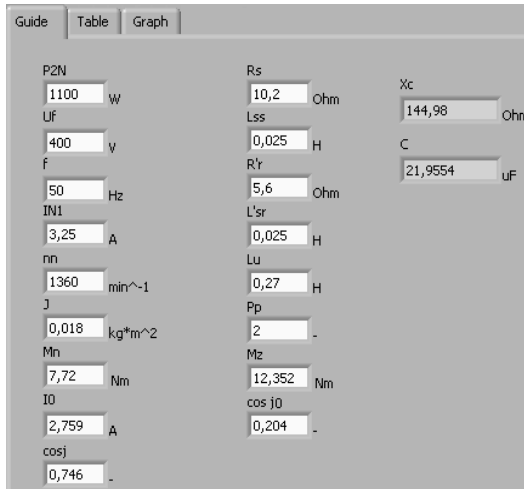


Fig. 4 Dialog box for setting input parameters

Data with which the program works are displayed in the left table labeled “Data table”. In right table labeled "Table of optimal operation" is the data for recommended to optimal operation of induction generator (Fig. 5).

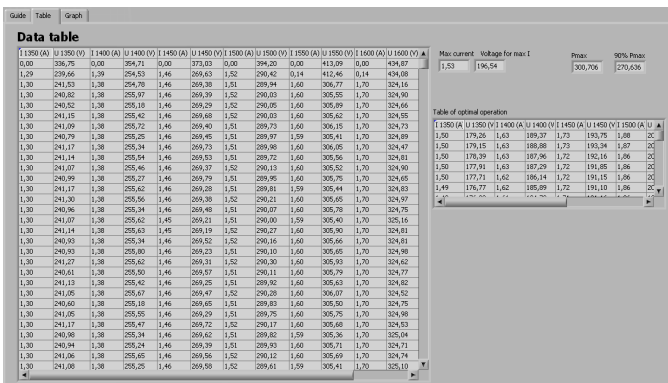


Fig. 5 Front panel for evaluating the optimum

Plotted the waveforms are a necessary part for clarity. The Fig. 6 shows the waveforms of the measured values for speed 1350, 1400, 1450, 1500, 1550 and 1600 rpm. C-V resistor characteristics of the values $R = 100, 150, 300 \Omega$ are also plotted. Pro-

gram also calculated so called optimum for operation (above this curve is recommended to operate the induction generator). This optimum is highlighted part of the curve in the graph.

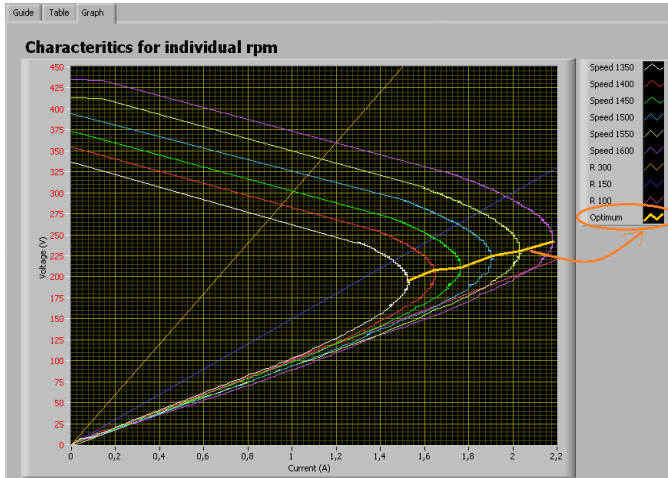


Fig. 6 Plotted waveforms in LabVIEW

4 Conclusion

Energy generated in asynchronous machines, is associate with increasing development of the construction of small-scale wind power. Induction machines in wind power are very often used for its simplicity, low maintenance as well as for low acquisition price. Induction machine with variable frequency drive is often used at present. However variable frequency drive considerably increase the acquisition price this wind power plant. Low acquisition costs, simplicity and above all the high efficiency are the main requirements of users. Method of operation, which there is solved, satisfies these requirements. The program is capable for actual speed of shaft of the generator to evaluate optimal operation of induction generator. This means that the program is able to determine the ideal size of the load. The ability to automatically control the size of the connected load will be the next phase of the program. The program itself will thus connect and disconnect the load on the terminals of the induction generator operating in off-grid so that the power that is at this time supplied to load was always in the optimum.

References

- [1] S. Rusnok, P. Sobota, R. Jílek, T. Mičák, "Induction Generator Optimized Operation for Small-Scale Wind Power in Program LabVIEW" in EPE 2014 Brno

Identification and Optimization of Parameters of Synchronous Machines with PM in axis D and Q

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.

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.

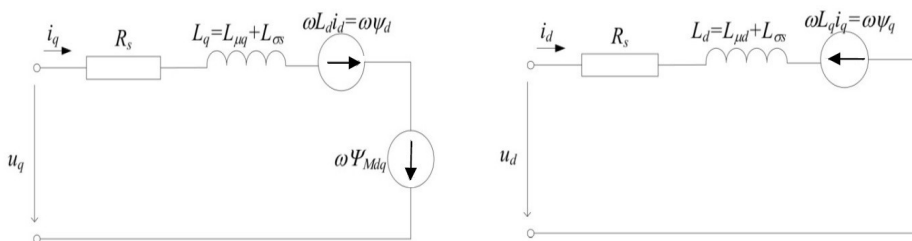


Fig. 1. Equivalent circuit IPMS with fully internally stored magnets on the axis d and q

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 $t = 28^\circ C$. The median value of stator resistance R_s converted to $20^\circ C$ is $R_{s20} = 3,93 \Omega$ and at $75^\circ C$ 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).

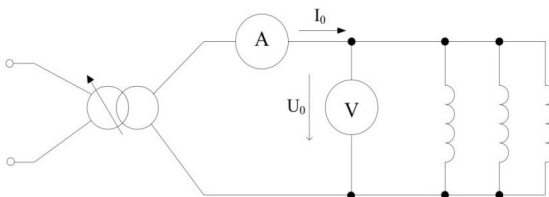


Fig. 2. Circuit diagram for measuring non-rotating reactance IPMS

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.

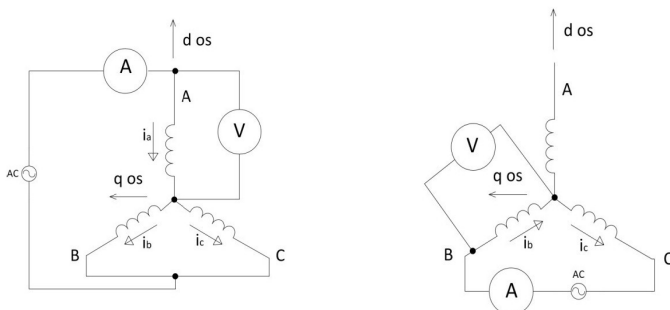


Fig. 3. Schematic diagram for measuring $L_{\mu d}$ and $L_{\mu q}$

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

$$L_{\mu d} = \frac{2}{3} \cdot \left(\frac{X_d}{2 \cdot \pi \cdot f} - L_{\sigma s} \right) = \frac{2}{3} \cdot \left(\frac{19,62}{2 \cdot \pi \cdot 50} - 0,02897 \right) = 0,02239 H$$

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

$$L_{\mu q} = \frac{2}{3} \cdot \left(\frac{X_q}{2 \cdot \pi \cdot f} - L_{\sigma s} \right) = \frac{2}{3} \cdot \left(\frac{32,177}{2 \cdot \pi \cdot 50} - 0,02897 \right) = 0,04905H$$

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.

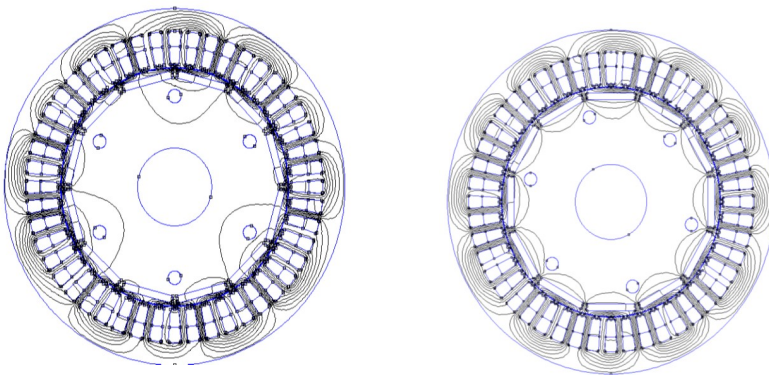


Fig. 4. Cross-section of synchronous machines with PM on the axis d and q

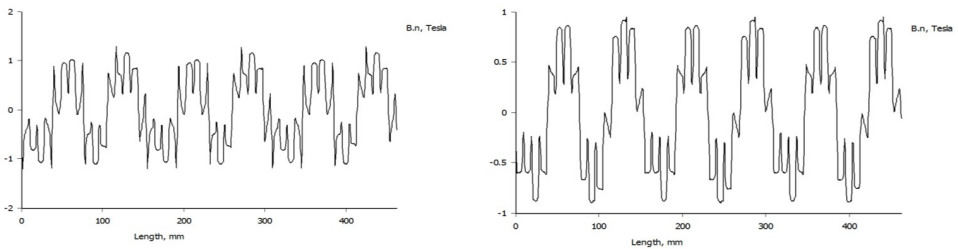


Fig. 5. Spatial distribution of magnetic induction in the middle of the air gap on the axis d and q

Subsequently, on the program Ansys Workbench was constructed 3D model and solved in idle state for the purpose of analysis of a magnetic circuit.

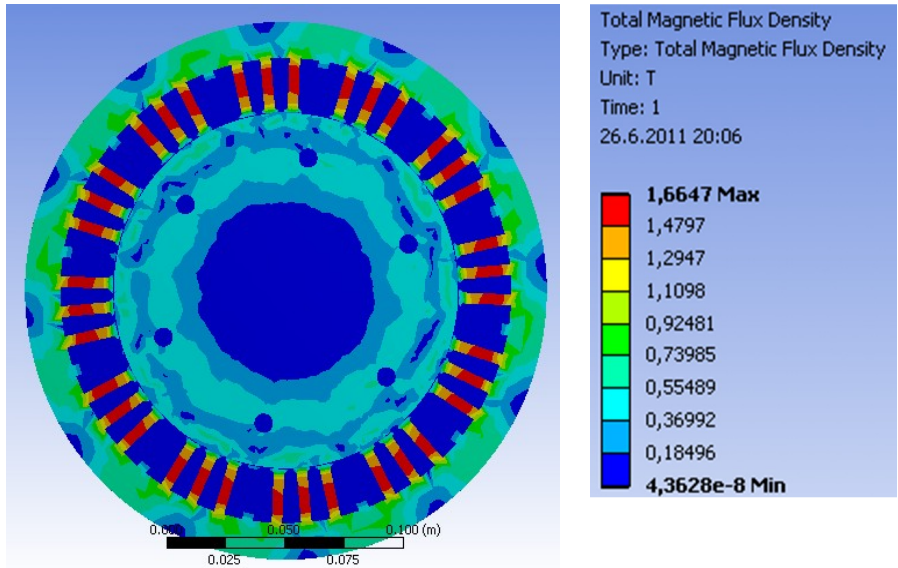


Fig. 6. Status idle (3D model)

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.

Control of Induction Motor Using Artificial Neural Networks

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). The induction motor drive is a dynamic nonlinear system with uncertainty in the machine parameters. The main aim is to improve tracking performance of the induction motor drive etc. A method for controlling induction motor drive is presented with Artificial Neural Networks (ANNs). Artificial Neural Networks (ANN) is successfully used in many areas such as fault detection, control and signal processing in our daily technology. Artificial Neural Networks have nonlinear structure and this is an effective feature that it approaches to the results of learning phase. Then, it gives results in test phase in short time (the degree about 3-10 second). It is a very preferable according to the other approaching methods. Matlab-Simulink is software which used to develop a three phase induction motor model and we can use to the creation of simulation models for the sensorless control and another methods for motor drives. The ANN is trained so that the speed of the drive tracks the reference speed etc.

1 Introduction

By using modern resources used in the regulation of AC drives can be considerable savings and improvements in the control of the drives. ANN models are composed of many linear or nonlinear computational elements (neurons or nodes) operating in parallel. Parallelism, robustness, and learning ability are among the main features, which determined wide applications for ANN to control of industrial processes. The special interest from this point of view has Multilayer Neural Network (AMNN) with backpropagation learning algorithm. Artificial neural networks have several important characteristics, which are of interest to control, like: Artificial Neural Networks, by their nature, have many inputs and many outputs and so can be easy applied to multivariable systems, ANN have parallel structure and this feature implies very fast parallel processing, fault tolerance and robustness. Artificial neural networks offer the advantage of performance improvement through learning using parallel and distributed processing. These networks are implemented using massive connections among processing units with variable strengths, and they are attractive for applications in system identification and control.

2 Basic of artificial neural networks (ANN)

Artificial neural networks (ANNs) are computing systems whose structures are inspired by a simplified model of the human brain. Many different types of ANNs have been proposed. The prime examples is biological neural networks, especially the human brain. An artificial neural network is a mathematical or computational model for information processing based on a connectionist approach to computation. The original inspiration for the technique was from examination of bioelectrical networks in the brain formed by neurons and their synapses. In a neural network model, simple nodes (or "neurons", or "units") are connected together to form a network of nodes hence the term "neural network". One type of network sees the nodes as 'artificial neurons'. These are called artificial neural networks (ANNs). An artificial neuron is a computational model inspired in the natural neurons. Natural neurons receive signals through synapses located on the dendrites or membrane of the neuron. When the signals received are strong enough (surpass a certain threshold), the neuron is activated and emits a signal though the axon. This signal might be sent to another synapse, and might activate other neurons.

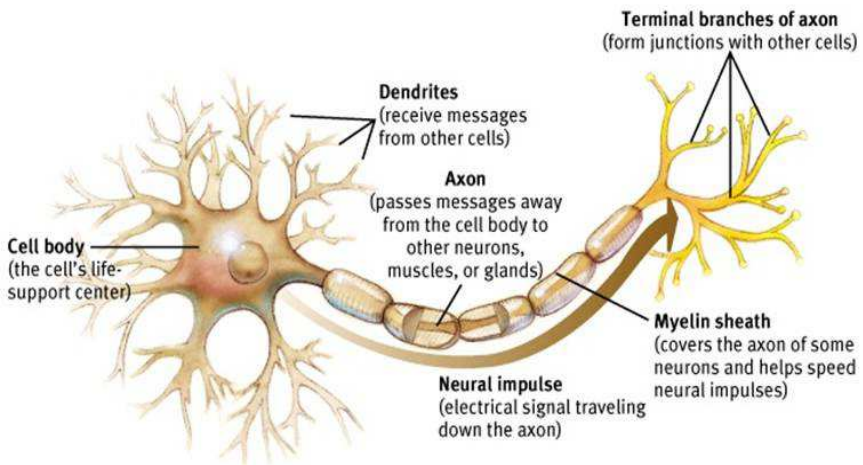


Fig.1. Natural neuron [7]

The complexity of real neurons is highly abstracted when modelling artificial neurons. These basically consist of inputs (like synapses), which are multiplied by weights (strength of the respective signals), and then computed by a mathematical function which determines the activation of the neuron. Another function (which may be the identity) computes the output of the artificial neuron (sometimes in dependence of a certain threshold). ANNs combine artificial neurons in order to process information.

Because of matters of space, we will present only an ANN which learns using the backpropagation algorithm (Rumelhart and McClelland, 1986) and feedforward methods for learning the appropriate weights, since it is one of the most common models used in ANNs, and many others are based on it.

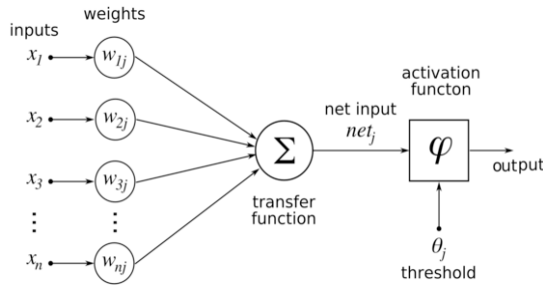


Fig.2. Artificial neuron model [7]

2.1 Multilayer perceptron network

Multilayer perceptron is class of networks consists of multiple layers of artificial neurons, usually connected in a feedforward way. Each neuron in layer has directed connections to the neuron of the following layer. Especially in non-linear applications the neurons of these networks apply a sigmoid function as a signal function.

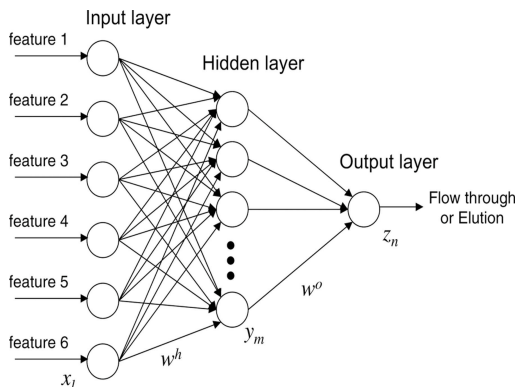


Fig.3. Multi-layer perceptron neural network [3]

3 Artificial neural network learning

We can categorize the learning situations in two distinct sorts:

- Supervised learning or associative learning in which the network is trained by providing it with input and matching output patterns. These input-output pairs can be provided by an external teacher, or by the system which contains the network self-supervised).
- Unsupervised learning or self-organization in which an output unit is trained to respond to clusters of pattern within the input. In this paradigm, the system is supposed to discover statistically salient features of input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified, rather the system must develop its own representation of the input stimuli. [1,6]

4 Sensorless control of induction motor using artificial neural networks

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. With developments in DSPs and power electronics, induction motors can now be used in high-performance variable-speed drives. Some of these drives are based on the indirect rotor field-oriented method.

Advantages of the sensorless control:

- reduction of hardware complexity and price of the system,
- increased mechanical robustness and overall stability,
- higher reliability,
- reduced maintenance requirements,
- increase noise immunity,
- improvement of machine vibration,
- elimination of cabling for sensors, etc.

The goal is to achieve speed-sensorless operation of the drive, e.g. the ANN replaces the digital encoder [3-5]. Simulation-based data or measured data can be used for training. Once the ANN is trained and tested, it replaces the digital encoder. To obtain good estimation accuracy, the inputs to the network are the present and past values of the stator voltage and current components in the stationary reference frame. The final structure of the neural network used is a multilayer net with three layers. The training algorithm of the neural network speed observer is as follows [3]:

1. Initially randomize the weights,
2. Obtain the stator currents and voltages,
3. Calculate the error between real and observed speeds,
4. Adjust the weights of the neural network,
5. Calculate the output of the neural network,
6. Go to 2nd step until the stipulated error is reached.

4.1 Backpropagation algorithm

The backpropagation algorithm (Rumelhart and McClelland, 1986) is used in layered feed-forward ANNs. This means that the artificial neurons are organized in layers, and send their signals “forward”, and then the errors are propagated backwards. The idea of the backpropagation algorithm is to reduce this error, until the ANN *learns* the training data.

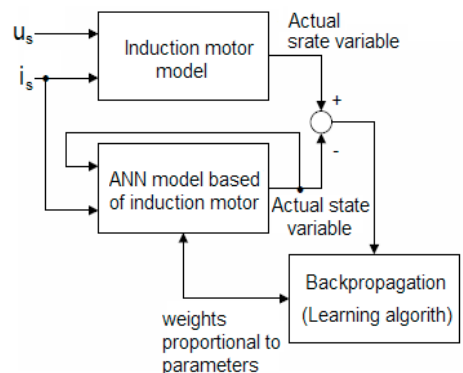


Fig.4. ANN identification parameters (speed)

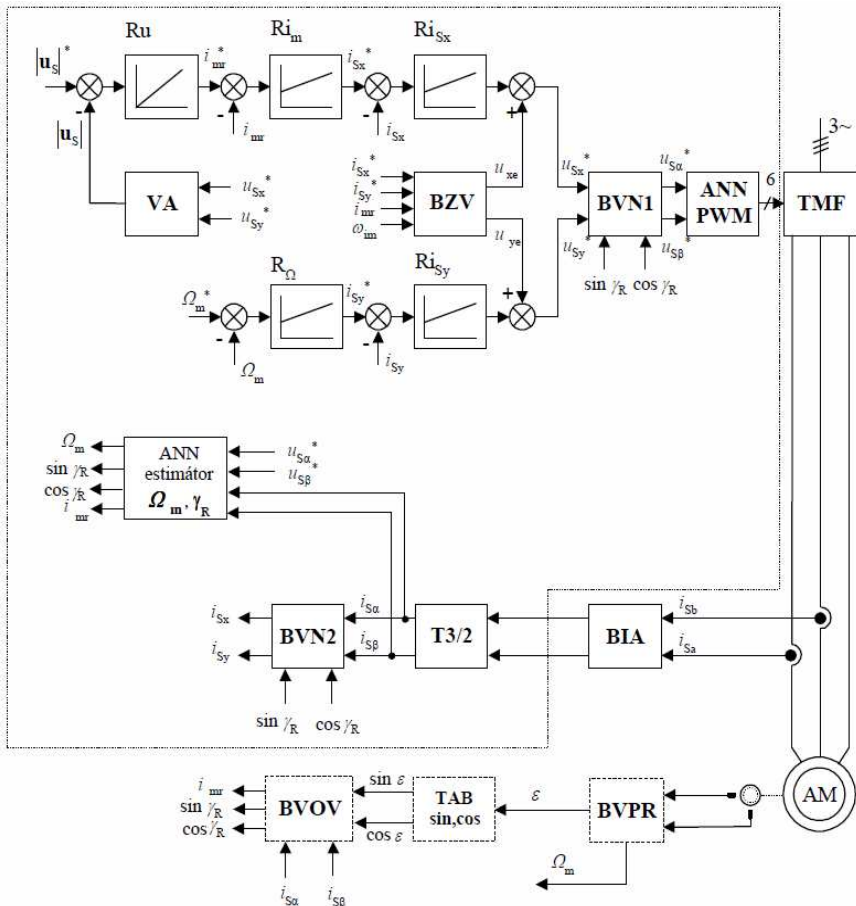


Fig.5. Control structure of sensorless vector controlled induction motor drive with using artificial neural networks [1]

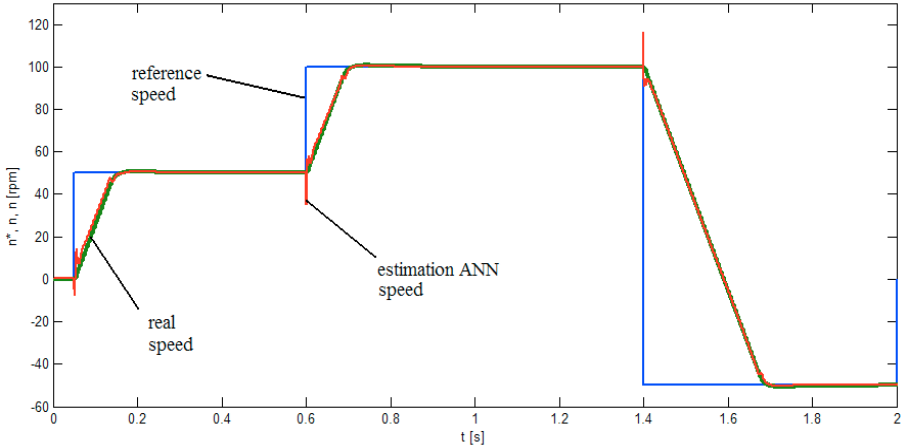


Fig.6. Real, ANN and reference speed of ANN training feedforward methods

Figure 6 shows the ANN performance for step changes in speed reference. The ANN responds well to these changes in speed reference as can be seen from the plot, but the network has a small problem with the change during step change, which is reflected as an instability in a network. But even with these small problems, the application of neural networks has very good properties.

5 Conclusion

This paper has described the use of artificial neural networks to identify and control the induction machine. There are many methods and approaches to train, learn and use ANN that can't be described here, but this is a basic view of the control using ANN. The network learning is based on the backpropagation algorithm, which is a relatively low cost computing method, and so, it is easy to implement in real time. The advantage of the neural network is to control the system without exact knowledge of its model. In conclusion, the proposed artificial neural network shows high performance and good control accuracy for the system.

Acknowledgement

In the paper, the results of the project SP2014/119, 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] MOUNA, B., LASSAÂD, S.: *Neural Networks for Controlled Speed Sensorless Direct Field Oriented Induction Motor Drives*. JEE 2008, Vol. 8, 88-99.
- [4] TSAJI, C.: *CMAC-Based Speed Estimation Method for Sensorless Vector Control of Induction Motor Drive*, *Electr. Power Compon.Syst.* 2006, Vol. 34, 1213–1230.
- [5] ZAKY, M., KHATER, M., YASIN, H., SHOKRALLA, S.: *Review of Different Speed Estimation Schemes for Sensorless Induction Motor Drives*, *JEE* 2008, Vol. 8, 100-138.
- [6] TOMAN, S.: *Hardware accelerated neural networks*. Diploma thesis, Ostrava, VŠB –TU Ostrava, 2011.(in Czech language)
- [7] SOZBUK, D. L.: *Application of ANN for control of PWM inverter fed induction motor drives*. Ph.D thesis, Warszawa, Warsaw Uni. of technology.

The converter for coupling to photovoltaics with accumulation

Katerina Baresova

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

Abstract. This paper deals with the design of converter for cooperation of photovoltaic strings with accumulator. During the design, the system optimization is focused on the dimensions of the DC/DC converter, which provides the link between photovoltaics and accumulator. Design optimization of the system is based on the topology, devices and material optimization of the suggested solution whereas ensuring the required functions. The critical point of design is relatively high battery voltage at the output side of the proposed DC/DC converter and the necessity of parallel cooperation of the modules to the battery. At the end of this paper, there is 3D design of suggested solution.

Keywords: accumulation; boost converter; photovoltaic; renewable source.

1 Introduction

Photovoltaic modules used for the construction of photovoltaic systems connected to the distribution network are used in this country roughly from the mid nineties. The photovoltaic has become, particularly in Europe, one of the fastest developing fields in electrical power engineering with an annual increase of installed power that in some countries exceeded 30%. The concept of connection of photovoltaic systems to the power network, so-called power network system (on-grid), brings many problems, particularly with the stability of the power network. At the same time the demand for island mode (off-grid) is rising, which is used everywhere where there is no power grid at the disposal, yet there is a need for alternating voltage.

Essential importance, when planning the use of solar energy, has distribution of sunlight intensity in individual seasons and particularly during a day. Accumulation appears as suitable solution for this case as acquired energy is saved for some time, thus accumulated. Disadvantage of this solution is that the more solar energy and for longer period of time is accumulated, the more losses and higher investment cost for building occur. Ultimately, the technical design optimization of converter must bring economic benefits to the entire solution. Increasing importance has requirements for high profit of electricity from sunlight, which are reflected both in the development of new methods MPPT (Maximum Power Point Tracking) and the col-

lapse of management of photovoltaic systems to smaller units on which the MPPT is performed individually. Our workplace has lots of experience with management optimization of power systems by modern methods, [1], [2], [5]. Algorithms and optimization methods MPPT are not part of this paper though.

Within the research of the Pre-seed project (Accumulation Technologies of Electric Energy for Renewable Sources) was therefore assigned to develop a modular design of converters with a common control system that would transfer energy from photovoltaic strings to accumulators with a nominal voltage of 750V. The basic module should be able to transfer a maximum power of 3kW, so parallel cooperation of outputs to a common accumulator is required. In the case of higher input power, the parallel connection of inputs (parallel modules cooperation) should be enabled. Module design should be dimensionally optimized to allow subsequent assembly of modules into standardized cabinet with possibility of air or liquid cooling. The design of cooling system was not an optimizing requirement, only power loss limit of all the modules was set.

2 Starting points of optimization

As was already mentioned, the energy conversion efficiency together with a power density are currently the main indicators of qualitative properties of DC/DC converters. Relative to the connections to photovoltaic systems, these parameters are required in a wide range of input voltage fluctuations. Due to the possibility of connecting various panel technologies and different number of panels in the string (maximum output voltage at the lowest temperature is typically 4x125Vdc or 6x96Vdc), the maximum working input voltage of the converter $U_{IN(max)}=600Vdc$ was selected. In this condition, a maximum output voltage of the unloaded string $U_{IN(0)MIN}=750Vdc$ is considered. The MPPT range is expected to be between 75-600Vdc. The solution must also respect the variation of the accumulator's output voltage U_{OUT} from 700Vdc to 925Vdc. From the perspective of voltage dimensioning of the whole converter, the unbreakable voltage of the system is considered to be $U_{MAX}=1000Vdc$. (e.g. critical insulating distance).

Under specified conditions it is difficult to define the transmitted power of one module. At a rough guess, the nominal power of the module is set to be $P_N=3.5kW$. However, for the design of one module are essential current values which are set to be $I_{N(AVG)}=7A$ and $I_{N(MAX)}=15A$.

These input terms were the basis for topology optimization, selection of the components, magnetic materials and design solutions, so as to fulfill the parameters of the power density and efficiency, which are corresponding to the present trends.

In about 10 years ago, the standard value of power density of DC/DC converters was approximately $0.3W/cm^3$, while at the present days this parameter ranges in values around $1.5W/cm^3$ with the assumption that over the next five years, the value will be around $1.8W/cm^3$. From this perspective, the cubic capacity of the designed solution should be in the volume of approximately $2000cm^3$ (e.g. 10 x 10 x 20cm). It

is clear that achieving this cubic capacity is only possible through the system optimization.

The efficiency of the converter is the second very important monitored parameter. The efficiency has impact on the environmental indicators (material and energy demands from the perspective of production facilities) in terms of energy (useful energy obtained per unit of sunlight) in the final analysis.

In the past, the efficiency was usually related to the nominal working point. This definition is however insufficient, especially for photovoltaic converters with strong fluctuations in power. Therefore many of international organizations set a different definition. The best known is the definition of efficiency with 20%, 50% and 100% of nominal power (Program 80 Plus). Values of efficiency in these working points are continually increasing (since 2010 by about 3%). At present, the commonly required efficiency is between 94 -96% in specified ranges of the power 20-100%. Part of dimensional optimization thus becomes the minimization of energy losses and thereby achievement of the highest effectiveness possible. Because the efficiency is associated with the reduction of converter's power losses, the economical cooling system affects also the power density of the converters.

3 Concept of converter topology without galvanic isolation

Due to the previous, it was decided to focus on the boost converter.

To achieve sufficient transmission of power, it was requested to create a converter in the form of three parallel modules with one-third of power, that would be able to transfer the summation power about 10kWp (total current $I_{INC}=20A$ at 550V of panel voltage). Concept of a converter for one module will be thereby derived from the diagram in the Fig. 1. To achieve a uniform distribution of current and improve the efficiency parameters, these modules are controlled from a central microprocessor controller with an offset of $T/3$ and with modern control methods, that have been verified at the department of electronics [1], [3], [4].

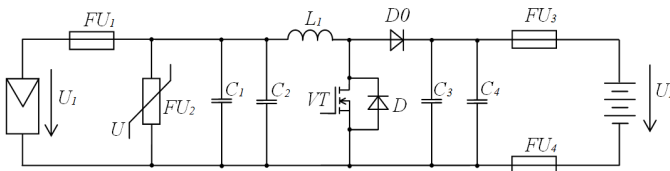


Fig. 1. Schematic diagram of boost converter module.

The operating principle of the converter is very simple and does not require to be analyzed in this paper. Essential for the realization is the favorable load of transistor VT, which must be dimensioned on the voltage $U_{DS} < U_2$ and the voltage on diodes $U_{RRM} < U_2$. When the voltages are set to the values mentioned above then these condi-

tions are with reserve easily achievable with 1200V class components. The choice of components is mentioned in the following section.

4 Selection of suitable components

The high switching frequency determines the use of unipolar MOSFET transistors. Disadvantages of high voltage type MOSFETs are large resistance $R_{DS(ON)}$ in switched state and high values of input and output capacity. Based on used structure of Trench MOSFET (business signage transistor CoolMOS™ or SupreMOS™), up to fivefold reduction in the resistance drift zone occurred at transistors with $U_{DS}=1000V$. Both parameters mentioned above have effect on the FOM parameter (Figure Of Merit), that was introduced by manufacturers for comparing the performance of each MOSFET transistors. In the simplest form it is calculated by:

$$FOM=Q_{GS} \cdot R_{DS(ON)} \tag{1}$$

Where Q_{GS} – is the charge of capacity between gate and source electrodes of the transistor and $R_{DS(ON)}$ is the resistance of the conductive channel in the switching state. Equation 1 is often supplemented by a member representing the loss of excitation circuit that can be however neglected for a single-transistor layout.

The FOM of several transistors were compared when choosing a particular one. Specifically it is the transistor type IPW90R120C3 by Infineon, the transistor type IPI90R340C3 by Infineon, the Silicon Carbide power MOSFET CMF20120D by CREE and IXKC13N80C by IXYS Corporation. The transistor type IXKC13N80C was selected based on the comparison in Table 1.

Those transistors are suitable for vast majority of applications. In case that the battery voltage reaches values around 900V, the utilization of 1200V class transistors is necessary. Therefore, the SiC transistor CMF20120D was also verified.

Table 1. Comparison transistors by figure of merit

Figure of merit for transistors	
<i>Type of transistor</i>	<i>FOM</i>
IPW90R120C3	3,84
IPI90R340C3	3,74
IXKC13N80C	2,25
CMF20120D	10,84

The characteristics of fast diodes can also be similarly evaluated. FOM factor of a diode can be calculated from the modified formula by:

$$FOM=Q_{RR} \cdot U_{TO} \tag{2}$$

Several diodes were similarly compared as transistors, where the diode MUR1100 from Taiwan Electronics Company was eventually selected.

In relation to the anticipated onset of components with a wide energy zone (especially SiC) and at the same time with low loss during operation at high frequencies, further increase of converter's power density is expected. Under the present terms turned out, that the better solution is based on the technology CoolMOS™.

In the conclusion of the optimization, the comparison of magnetic superlattices of converter's coil was performed with a view to the construction of the transformer. This analysis goes beyond the framework of this paper, therefore only the material that has been selected is given. In the mentioned application the Ni-Zn ferrite 3F3 was proved as best choice, because it showed best qualities for applications at frequencies from 200 to 500kHz.

5 Construction design of the converter

The transistor IXKC13N80C with technology CoolMOS™ from IXYS Corporation was selected on the dimensioning basis for the construction solution of the converter. It is a unipolar transistor for voltage $V_{DS}=900V$ and current $I_D=36A$. Based on the availability, the ultra-fast diode MUR1100 from Taiwan Electronics company was selected with parameters $V_{RRM}=1000V$, $I_{FSM}=30A$ and $t_{rr}=75ns$. SiC transistors are considered as a perspective solution for higher voltages, whereas the CoolMOS™ transistors are often used for the basic design with the limited voltage $U_{IN(max)}=800V$. SiC transistors have also a worse parameter of FOM. The control circuit, which was developed on the Department of Electronics, was used for the selected transistor. Structural solution of the module is shown in the Fig. 2.

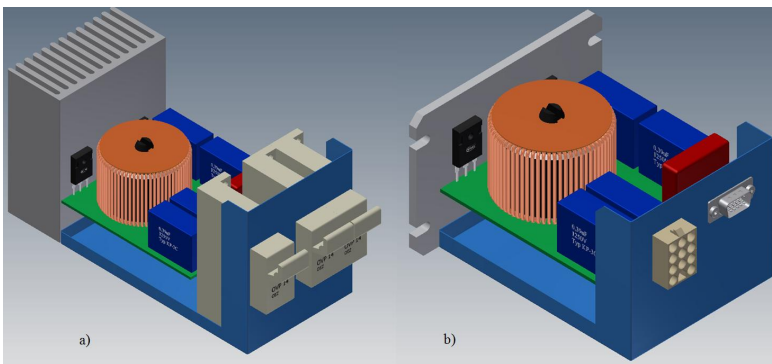


Fig. 2. Structural 3D model of the boost converter

Based on the previous designing, used to control voltage and current dimensioning of all used components including capacitors and fuses, the dimensions of heat sink with air cooling were calculated for installation in the common air channel around the converter, which is composed of total three modules.

6 Conclusion

This paper introduces an approach to the selection of the boost converter's components, with which we are trying to achieve the best specific power density. In the case of using the air cooling, the total size was limited by the length of the heat sink. That is the reason why the part of fuse disconnectors was included into the content of converter (Fig. 2a).

In the converter housing, for which the module was designed, is available to use a water cooling, therefore a variant of the converter module with water cooling and without fuse disconnectors was also designed (Fig. 2b). Fusing in this case creates a part of the entire converter. The volumetric power density achieved in the first case is $2,54\text{dm}^3$, in the latter case then $1,14\text{dm}^3$.

The possibility of using SiC transistors was also verified, but due to higher price and higher parameter of FOM, the solution utilizing SiC power parts was not considered as a perspective for further development.

Acknowledgment

The presented research was supervised by prof. Ing. Petr Chlebis, CSc. from Department of Electronics, VSB - TU Ostrava, Czech republic.

References

1. P. Brandstetter, P. Chlebis, P. Palacky, P. Simonik, "Unconventional Soft Switched Parallel Active Power Filter," In 10th International Scientific Conference Electric Power Engineering, Kouty nad Desnou, Czech Republic, 2009, pp. 347-351, ISBN 978-80-248-1947-1.
2. P. Brandstetter, M. Skotnica, "ANN Speed Controller for Induction Motor Drive with Vector Control," International Review of Electrical Engineering - IREE, Vol. 6, N. 7, pp. 2947-2954, 2011, ISSN 1827- 6660.
3. M. Cacciato, A. Consoli, V. Crisafulli, "A high voltage gain DC/DC converter for energy harvesting in single module photovoltaic applications," In Industrial Electronics (ISIE), 2010 IEEE International Symposium on. Bari: IEEE, 4-7 July 2010, ISBN 978-1-4244-6390-9, pp. 550 - 555.
4. A. Havel, P. Vaculik, M. Tvrdon, M. Hromjak, "The Design and Construction of a Reversible Voltage Inverter Prototype," 13th Electric Power Engineering - EPE 2012, May 23-25, Brno, Czech Republic, p. 1059-1062, ISBN 978-80-248-2393-5.
5. P. Palacky, P. Hudecek, A. Havel, "Real-time Estimation of Induction Motor Parameters Based on the Genetic Algorithm," 7th International Conference on Soft Computing Models in Industrial and Environmental Applications - SOCO 2012, September 5-8, Ostrava, Czech Republic, p. 401-409, ISBN 978-3-642-33017-9, ISSN 2194-5357.

The Properties of Boost Converter with SiC MOSFET

Martin Borski and Jiří Špička

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

Abstract. This paper describes the properties of boost converter with silicon carbide MOSFET. At the beginning of this article is shown the connection and description the boost converter. A several measurements are performed especially in relation to EMC. The base of this converter is SiC MOSFET and toroidal air core coil. These components enable the operation on high switching frequency up to 1MHz. The created model will be used to next research in the Department of Electronics.

Keywords: Boost converter; silicon-carbide MOSFET (SiC); air core coil; electromagnetic compatibility (EMC)

1 Introduction

Semiconductor converters created for cooperation of high power photovoltaic systems with accumulation systems form a special category of DC/DC converters. These converters have specific demands for relatively high transmitted power by relatively high voltage values. Typical requirements specified on the converter usually come from trends which determine the category of switching power supplies and similar devices. These devices operate in the line voltage range of which generally does not overcome 350V in the DC unit. In mentioned applications of converters used for coupling of photovoltaic systems with accumulation units is often worked with voltages of minimally twice the size. Due to the fact that the device is exposed to an outside influence, the insulation levels must be sufficiently resistant to various atmospheric over-voltages etc.

Increasing of the specific power per a volume unit of a converter and increasing of the efficiency values are contemporary trends used in switching power sources. Increasing of the switching frequency to values up to 1MHz is typical for reaching high values of above mentioned parameters in boost converters.

Values of inductances of coils and transformers are decreasing in the context of increasing switching frequencies. It was decided to go with the trend of application air coils and transformers for utilization in power DC/DC converters. An advantage of air coils is their linearity; the disadvantages are large dimensions and stray field. Already at the very beginning it can be said, that the large dimensions of the coils do not have to be a problem due to increasing insulation requirements imposed on the system.

The question of stray field can be solved by appropriate construction of the coils. Nevertheless, the considerable influence of interference can be expected in the question of EMC by the presented solution.

2 Design of Boost Converter with SiC Components

A circuit block diagram of designed the boost converter is shown in Fig. 1. This boost converter with 200 kHz – 1 MHz operating frequency and consist of a SiC parts will be the article of further research. There are used one SiC MOSFET 1200V/24 A 80 mΩ (CMF20120D) and one SiC schottky diode 1200V/15A (C4D15120A) in this configuration. The boost converter does not need any circuit for soft-switching. The control of the SiC MOSFET is provided by DSC 56F8037.

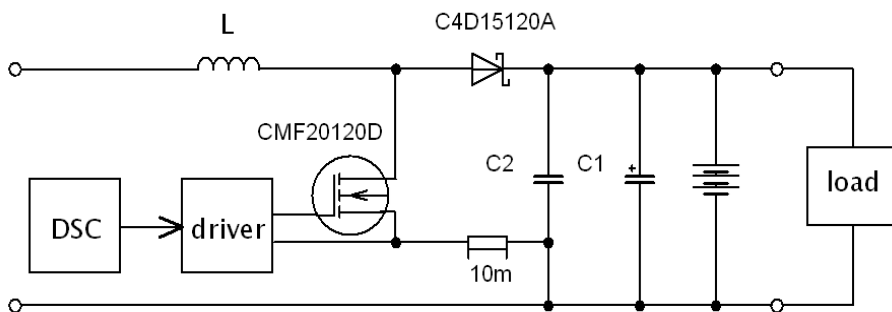


Fig. 1. Controlled voltage inverter block diagram

2.1 Description of a Block Diagram

There is a capacitor battery on the output of the converter. A total capacity of the capacitor battery is 1.1mF. The power supply of a driver is 12V. The switching is provided by an optocoupler and controlled by a PWM signal from the DSC. Voltage levels for switching MOSFET are -5V/25V. A driver also creates the galvanic insulation. The insulating voltage from the side of the optocoupler is 1kV and from the side of the power circuit of the converter it is about 5.2kV. A commutating circuit is made of a loop of SiC MOSFET, SiC Schottky diode and does not contain the inductive capacitor. [1], [2], [3], [4]

2.2 Toroid coil with an air core

The inductance calculation formulas are very complicated for air cored coils. They are based on equations, which are including correction factors coming from relative values of coil sizes. For inductance calculation is applicable the next formula:

$$L = \frac{\mu_0 N^2 h}{2\pi} \ln \frac{D_b}{D_a} \quad (1)$$

Parameter h represents the height, D_b is the outer diameter and D_a is the inner diameter of the core

3 Experimental results

One of the experimental measurements is shown in this sub-chapter with waveforms of following values: Transistor voltage UDS (red) and current in the air core coil (blue) are matching with the simulation waveforms. The current in the transistor is measured by the shunt resistor (yellow) where the voltage drop 10mV corresponds with current 1A. Voltage drop on the diode (green) indicates its conducting time.

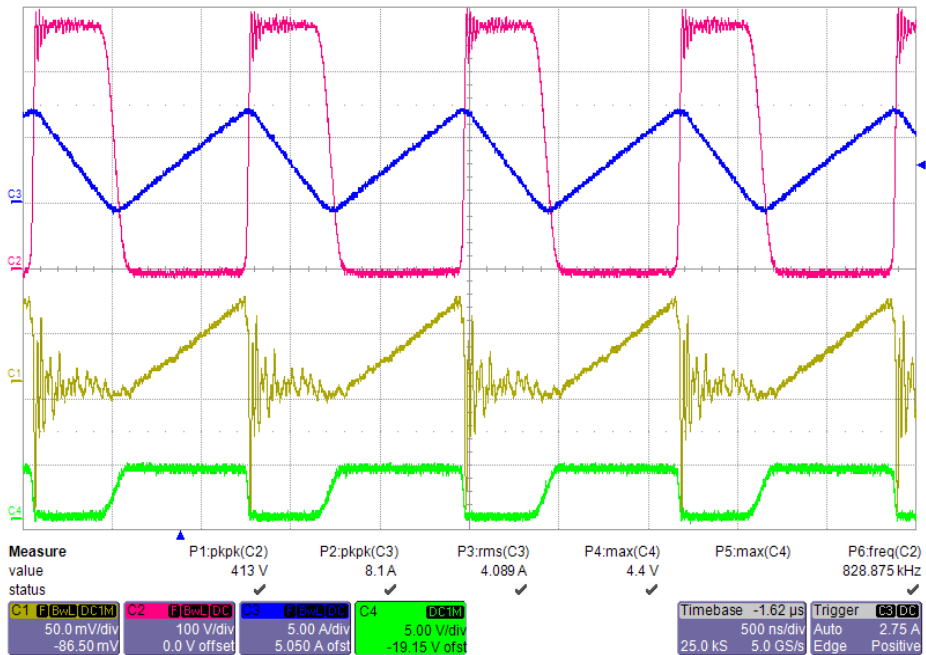


Fig. 2. Measurement of the electrical values

3.1 Check of the properties in relation to EMC

The designed boost converter is a source of high frequency interference. The properties of this converter were proved in relation to EMC. An analyzer Agilent CXA N9000A was used for this measurement. The measurements were performed in range from 150kHz to 30MHz because the converter operate in 1MHz. A standard CSN EN 55011 was set in the analyzer. A ring probe was used to the measurement for detection a near field due to finer analysis interference of electromagnetic fields. The interference was measured in center and edge of air core coil.

Due to standard CSN EN 55011 the measurement should by provide at a distance of one meter. In this distance the designed converter definitely meet required value of EMC.

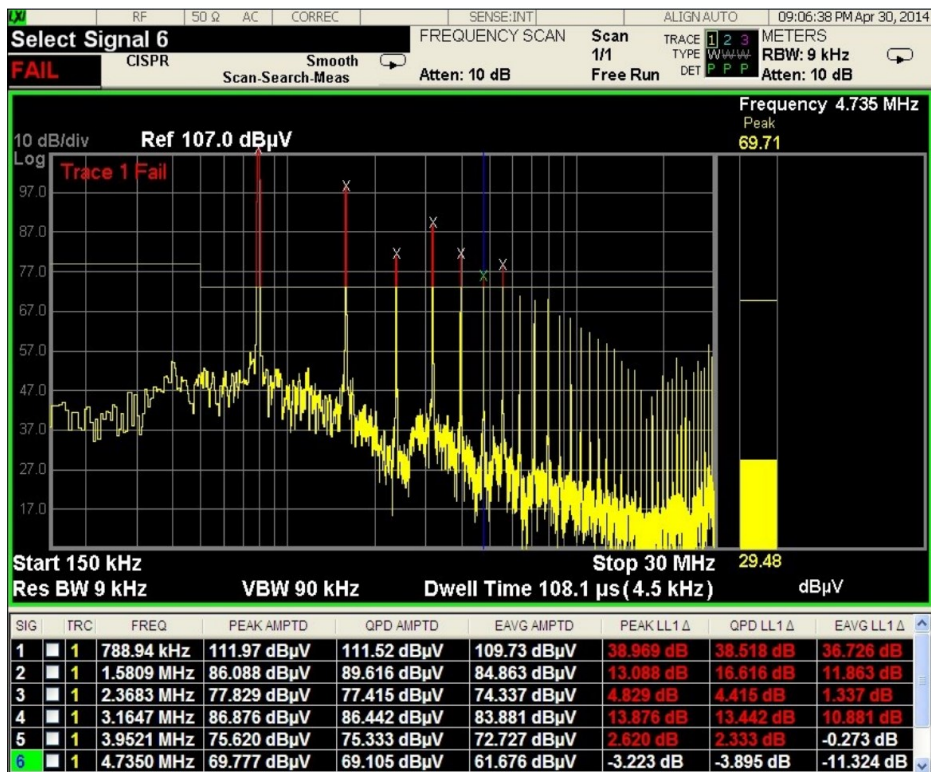


Fig. 3. Measured values on the edge of the coil

According to figure 3 the measured values are exceeded on seven frequencies in measured range. The greatest interference come in the frequency about 790kHz where a quasi-peak values (QP) is higher than 39dB. Interference on the center of the coil is much larger than on the edge of the coil according to measurement.

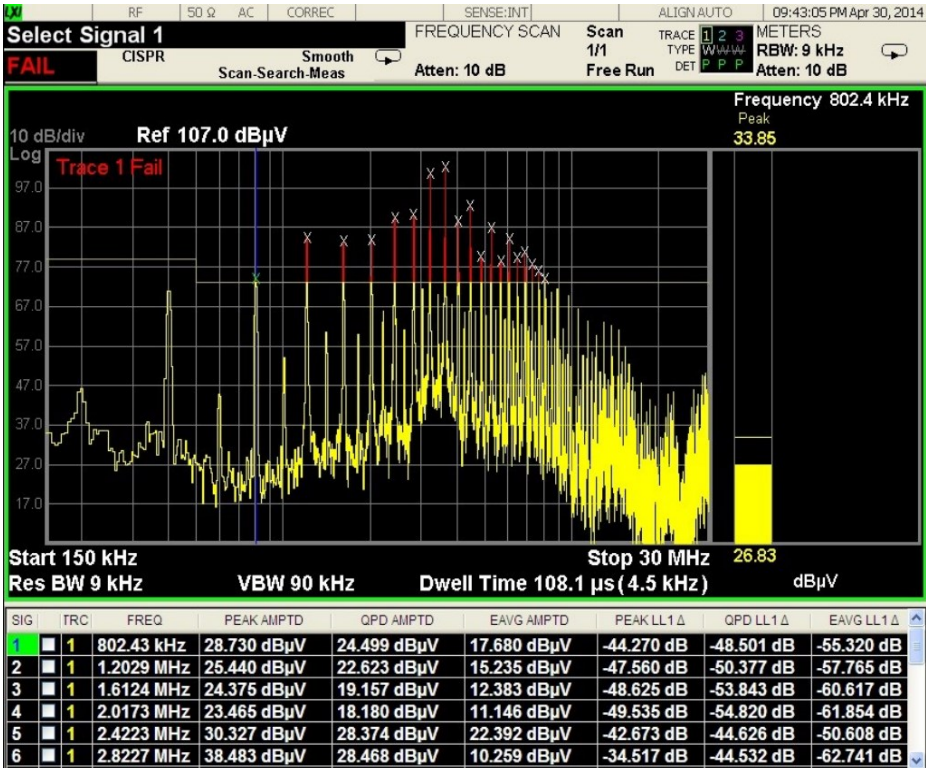


Fig. 4. Measured values on the center of the coil

4 Conclusion

The main object of this paper was analysis the boost converter which operates in the middle field of frequency and design of the air core coil. Based on the introductory analysis it was necessary to determine the application possibilities of an air coil in power circuits of pulse converters. The operation of converter was proved by measurement. The electromagnetic interference was detected its magnitude and influence to environment. This influence was within a standard under the given conditions. The toroid coil had not an impact on surrounding devices in test up to now. Using of air core coil proves as applicable conception and makes an area for next research.

5 Acknowledgment

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

- SGS project SP2014/203 – Modern design of semiconductor converters. This research was supervised by prof. Ing. Petr Chlebiš, CSc. from Department of Electronics, VSB - TU Ostrava.

6 References

1. Brandstetter P., Skotnica M.: ANN Speed Controller for Induction Motor Drive with Vector Control. *International Review of Electrical Engineering - IREE*, Vol. 6, N. 7, pp. 2947-2954, 2011, ISSN 1827- 6660.
2. Brandstetter P., Chlebis P., Palacky P., Skuta O.: Application of RBF Network in Rotor Time Constant Adaptation. *Elektronika Ir Elektrotechnika*, Issue 7, pp. 21-26, 2011, ISSN 1392-1215.
3. Brandstetter P., Chlebis P, Palacky P., Simonik P.: Unconventional Soft Switched Parallel Active Power Filter. In *10th International Scientific Conference Electric Power Engineering*, Kouty nad Desnou, Czech Republic, 2009, pp. 347-351, ISBN 978-80-248-1947-1.
4. P. Brandstetter, T. Krecek, "Speed and Current Control of Permanent Magnet Synchronous Motor Drive Using IMC Controllers," *Advances in Electrical and Computer Engineering*, Vol. 12, No. 4, pp. 3-10, 2012, ISSN 1582-7445.

Actuators for Intake Air Regulation of Modern Turbo Diesel Engines

Tomas Mrovec

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

Abstract. This paper describes fault symptoms and options of problematic actuators of modern turbocharged diesel engines. These engines are very popular in present time. Using these components is possible to regulate pressure and weight of intake air. There is currently no technical literature describing management of most modern control actuators such as valves (these components often have a relatively short life). For this reason was made a practical analysis of selected actuators and precise description of the nature their management.

Keywords: Actuator control; butterfly valve; PWM; turbocharger control; diesel engine; EGR

1 Introduction

Due to increased demands for environmental friendliness, economy of running and also control precision, proportional actuators replaced actuators, which previously only allowed two-stage control (on / off). They are used wherever it is necessary to continuously regulate the manipulated variable. In modern diesel engines, these actuators control systems such as exhaust gas recirculation (EGR) regulation of supercharging (control of Variable Geometry Turbocharger - VGT or bypass valve) and also, for example influencing fulfillment of cylinders using butterfly valves.

Modern turbo-diesel engines are controlled by so-called combined regulation. In this case is air entering the engine at first compressed by turbocharger and then cooled by intercooler. Density of intake air will be increased and temperature of circulation will be reduced. Power regulation is essentially qualitative, varies with filling air pressure, when an increase in mean effective pressure is increased filling mass for cylinder and this regulation have the qualitative characteristics of regulation. The modern concept of motor control is designed so that for management is responsible ECU which is set to using the optimal pre-injection. That according to the engine mode, regulates and controls the boost pressure and exhaust gas recirculation, the EGR valve.

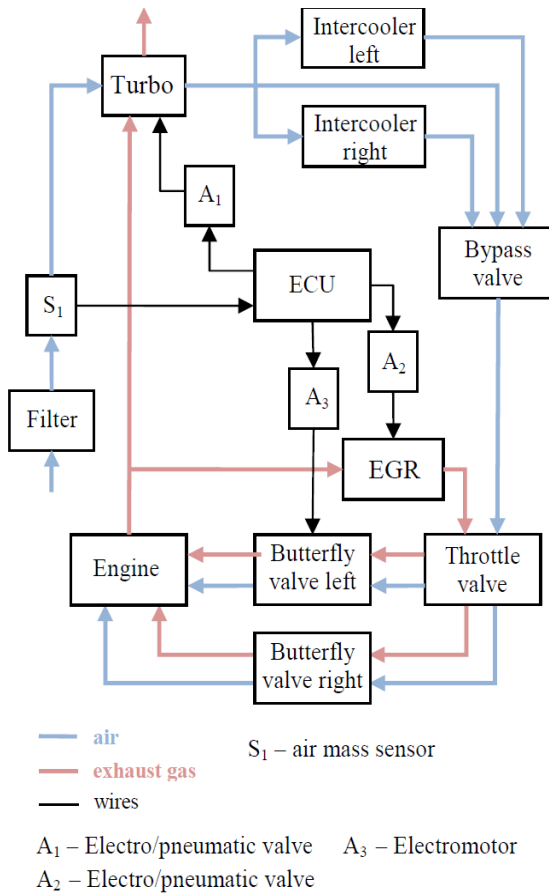


Fig. 1. Diagram of the intake and exhaust system of modern diesel engine

2 The EGR Electromagnetic Valve

Control of EGR valve is similar case as electromagnetic valve of turbocharger. It is controlled by changing duty cycle of pulse width modulation signal. EGR regulation is performed after the engine warms up to operating temperature, but in need for acceleration or deceleration is inactive. Analysis of the EGR valve was performed on the vehicle Hyundai i30. Condition for the functionality of the EGR system is achieving engine operating temperature.

In Fig. 2 are displayed waveforms of engine speed and EGR valve duty cycle. From these waveforms is obvious that the EGR system bypasses exhaust gas into the intake manifold, when the engine has reached idle speed. At the first the engine speed was twice rapidly increased up to 3000 rpm, then the engine speed was again increased above 3000 and then gradually reduced until the engine reach idle speed.

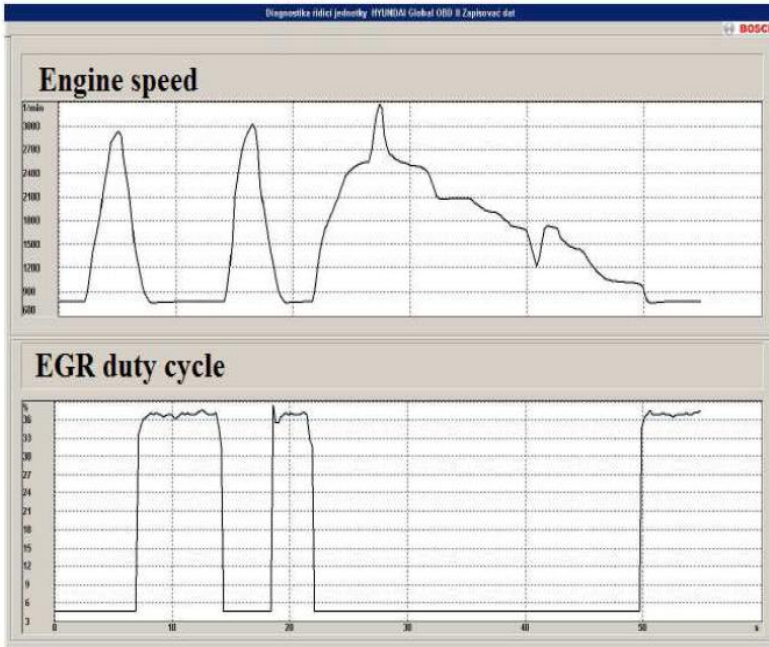


Fig. 2. Waveforms of engine speed and EGR valve duty cycle

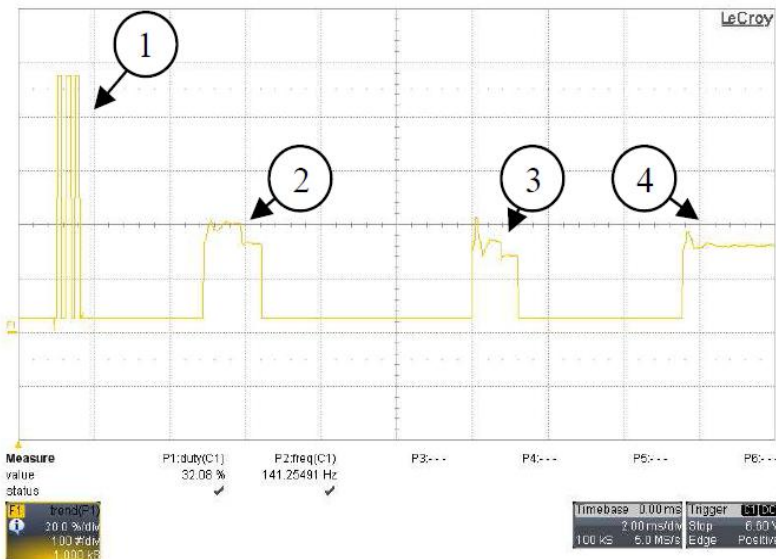


Fig. 3. Duty cycle of PWM control signal for the EGR valve

In Section 1 the ignition is switched on, the duty cycle is set to 5%. Then there is a short initialization setting of EGR system. The valve is three times fully open and

closed. Closed valve corresponds to duty cycle 5% and fully open to duty cycle 95%. Section 2 shows the moment when the engine is started. Next the engine is turning idle speed, duty cycle is 40%. Follows an increase of engine speed and EGR valve is closing. In Section 3 there was again a stabilization of engine speed to the idle speed, duty cycle of PWM signal is 32%. Then there was again engine speed increase. Section 4: The engine was left in idling. Duty cycle of signal is 32%.

3 Actuator of Butterfly Valves

Butterfly valves are used to direct the flow of air into the cylinders, depending on the current engine speed and load with respect to emissions, fuel consumption and torque. They are controlled by PWM signal with a certain frequency; their position is transmitted in the form of a digital signal. Location butterfly valves are controlled by a position sensor. Butterfly valves are open according to the requirements of operating status. From certain engine speed are open constantly. Position of butterfly valves has a decisive influence on the motion of fulfillment. Closing butterfly valves creates a circular motion filling the air around the vertical axis of the combustion chamber. Increased flow allows optimal production mix.

In Fig.4 is a photograph of actuators control electronics of butterfly valves. The electronics is placed directly in the housing of the actuator. Analysis of the activator of butterfly valves proceeded on the vehicle Audi A6 (model 2008). On the first pin is supply power from battery. On the second pin of connector is ground. The third pin receives a control signal from ECU. And fourth pin transmits position information of butterfly valves to the ECU. Position is detected using Hall probe.



Fig. 4. Photo of actuator control electronics of butterfly valves

In Fig.5 is displayed measurement on car Audi A6. ECU sends a PWM signal of size the on-board voltage at frequency 244 Hz. The position information is also represented by a PWM signal at frequency 255 Hz. Yellow color represents the duty cycle of control PWM signal and purple color represents current position of butterfly valves. Full closure corresponds the duty cycle 80% and complete opening the duty cycle 20%. With increasing engine speed the butterfly valves are beginning to open. In section 1 you can see the trend showing initially acceleration. In section 2 is the ignition switched on. In this case the butterfly valves are set to end positions.



Fig. 5. Measurement of butterfly valves on Audi A6 2.7 TDi

Immediately after the engine starts ECU sends a signal to the immediate closure of the valves (79.82 % duty cycle). This requirement is fulfilled with a slight inertia. When the engine idling and the valves are closed. After the engine speed is increased, the valves are open to their maximum opening. In section 4 follows engine shutdown.

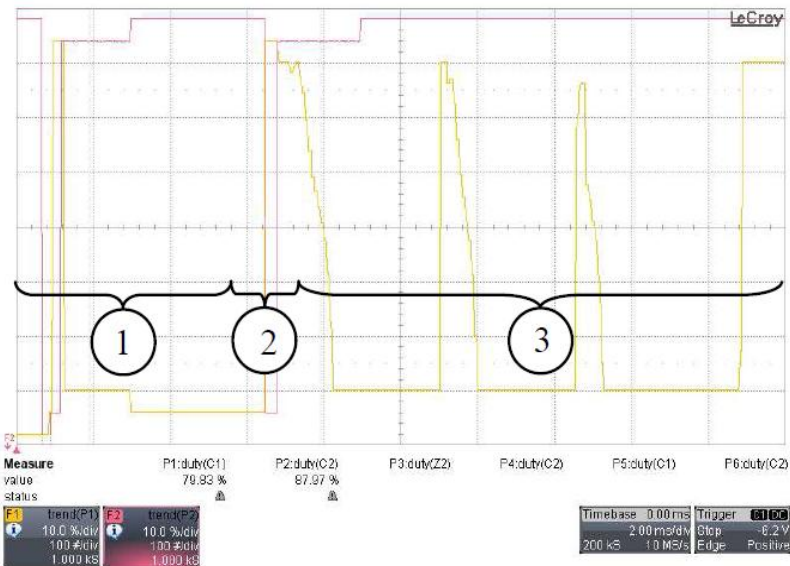


Fig. 6. Measurement of fault butterfly valves on Audi A6 2.7 TDi

Fig.6 shows the trend of duty cycle from defective valves, when the ignition is turned on with following start and acceleration. Section 1 shows the course of the valves setting to the reference position with the ignition on. Follows start of the vehicle. After launching vehicle the valves are permanently closed. Section 2 indicates a requirement of ECU for engine to the immediate closure valves at idle speed. Following section 3, which shows the waveform when the accelerator pedal was pushed 3 times in a row. Here you can see demand of engine control unit to open valves at high engine speeds. Responding to this requirement did not happen and valves are closed.

4 Conclusion

The paper provides review of theory and practical realization results of modern and real vehicle analysis. At present is the results of analysis applied for other modern actuators analyzing. The next stage of the reverse engineering will be also oriented to the cars with gasoline combustion engine. All applications are realized within the research activity in the branch of automotive electronics in the Department of Electronics.

Acknowledgment

Work is partially supported by Grant of SGS No. SP2014/119, VSB - Technical University of Ostrava, Czech Republic, for which author express their sincere gratitude.

This research was supervised by Ing. Petr Simonik, PhD. from Department of Electronics, VSB - TU Ostrava.

References

1. Brandstetter P., Chlebis P., Palacky P.: *Direct Torque Control of Induction Motor with Direct Calculation of Voltage Vector*. Advances in Electrical and Computer Engineering, 2010, Vol. 10, N. 4, pp. 17-22, ISSN 1582-7445.
2. Chlebis P., Vaculik P., Moravcik P.: *Direct Torque Control Methods for Three-level Voltage Inverter*. In: 10th International Scientific Conference Electric Power, Kouty nad Desnou, Czech Republic, May 12-14, 2009, pp: 352-356, 2009, ISBN 978-80-248-1947-1.
3. Kascak, S., Dobrucky, B., Prazenica, M.: *A New Approach for Estimation of Speed/Position of Two-Phase Induction Machine Using Virtual High Frequency Injection Method*, In: International Review of Electrical Engineering -IREE, Vol. 8, No.4, 2013, pp. 1156-1161, ISSN 1827-6660, simulations, In: Electrical Engineering - Archiv für Elektrotechnik, 2013, 5/95, ISSN 0948-7921, ElectrEng DOI 10.1007/s00202-013-0284-8, ADC, CC.
4. Brandstetter, P., Vanek, J., Pumr, J., Michalik, J.: *Microcomputer Control System for Industrial Electronics Applications*. Applied Electronics, Sept. 2009, ISSN 1803-7232.

Full – Wave Rectifier Using Functional Structure 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 operational rectifiers, in which new functional structure (OTA) is used. This article introduces precision full – wave rectifier with only commercially available current conveyor, which is the non – invert positive current II. generation conveyor (CCII+). These new functional structure is based on the essential idea that “every active element can be described by a set of controlled sources” - a voltage source controlled by voltage, a current source controlled by voltage, voltage source controlled by current, a current source controlled by current. A circuit provides precision rectification of small signal for frequencies up to 100kHz with small waveform distortion is presented. This new functional structure is consequently demonstrated in electronic circuit mentioned above.

Keywords. current conveyor, full-wave rectifier

1 Introduction

In 1966, the concept of an active element was introduced, which surpassed the characteristics of operational amplifiers. Three generation of current conveyors exists.. In these days, current conveyor is found as part of another complex structure. These elements are now used in electronic circuits with many advantages, such as: it extends frequency range of processed signals, it increases frequency, at which circuit can operate (up to several hundred MHz), it extends frequency range of processed signals, it significantly improves dynamics of the circuit (increases slew rate), it improves circuit's noise immunity at low supply voltage, and most importantly, it can operate at very low supply voltage.

Precision rectifiers are important for analog signal processing function and instrumentation [5,6]. Conventional voltage mode rectifiers based on diodes and operational amplifiers have a serious drawback. It's due to the fact that these circuits have to overcome the threshold voltage of the diodes. This problem prevents the rectification of signals below a voltage of about 0.6 V [7,8]. During transition of the diodes from their non-conduction state to their conduction state the operational amplifiers have to recover with a finite small signal, which leads to the significant distortion during the zero crossing of the input signal [9]. We can obtain wider bandwidth, improve the

signal to noise ratio, achieve higher precision of the output rectified signal and decrease energy consumption by using current active elements such as current conveyors, operational transconductance amplifiers, current followers.

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 current I_x is independent value which it is always conveyed from terminal X (current input) to the terminal Z by the current follower. The terminal Z is current output terminal of current conveyor [2]. We choose the voltage system (U_x, U_y, U_z) on the individual gates and we orient this voltage system to common node. The voltage U_y is transfer from gate Y (high – impedance non-inverting voltage input) to the gate X by voltage follower and this gate can also be considered as the voltage output gate.

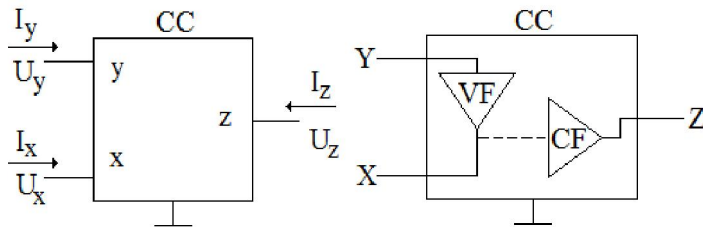


Fig. 1 Schematic symbol and block scheme of current conveyor

The behavior between input and output terminals of current conveyor [1]:

$$u_x = u_y, i_y = i_x, i_z = i_x \tag{1}$$

Another important element is the universal current conveyor (UCC) which it can represent every mention current conveyor any generation.

3 Basic structure with current conveyors (OTA)

The best description every active element is by a set of controlled sources, which can describe all the structures [3]. 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.

OTA is a differential voltage controlled current source where the output current is controlled by an applied input voltage signal. The OTA is similar to a standard operational amplifier in that it has a high impedance differential input stage and that it may be used with negative feedback.

The structure, which replaces a 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 added by voltage follower and resistor to ground [4]. This resistor correct voltage transmission of structure.

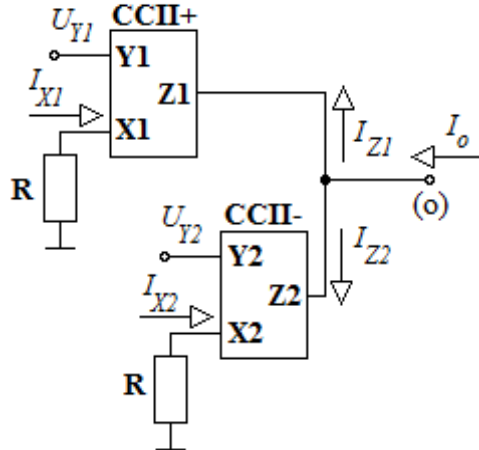


Fig. 2 Functional block structure (OTA) with current conveyors

4 Full-wave rectifier with OTA based on current conveyors

The connection of full – wave rectifier is shown on Fig.3. The functional structure replaces operational amplifier, which it provides the equivalent properties as a classical voltage operational amplifier. In the circuit, the structure was inserted without any further connections adjustments.

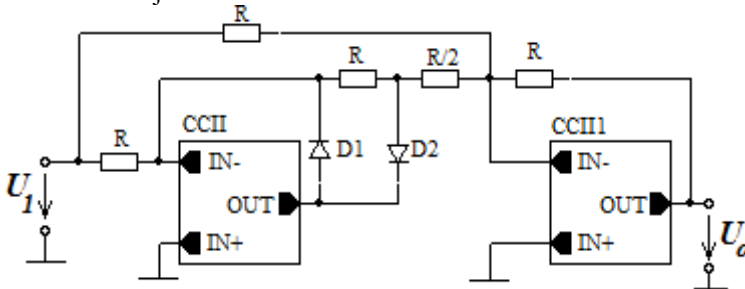


Fig.3 Full-wave rectifier

The time waveform of full – wave rectifier is shown in Fig.3 and Fig.4 The full – wave rectifier was simulated with the signal amplitude $U = 0.1V$, signal frequency $f = 1MHz$ and $f = 100kHz$.

Precision rectifiers are frequently dependent circuits. We can see this dependence on Fig.5 and Fig.6 which show waveforms at signal amplitude $U = 50mV$, signal frequency $f = 1MHz$ and $f = 100kHz$.

The peak increases and goes to negative values when the frequency is increased. This peak is caused by accumulated charge of the diodes which must be drained that the diode could restore own original operating characteristics. It is complicated on higher frequencies because diode must reach it in a shorter time.

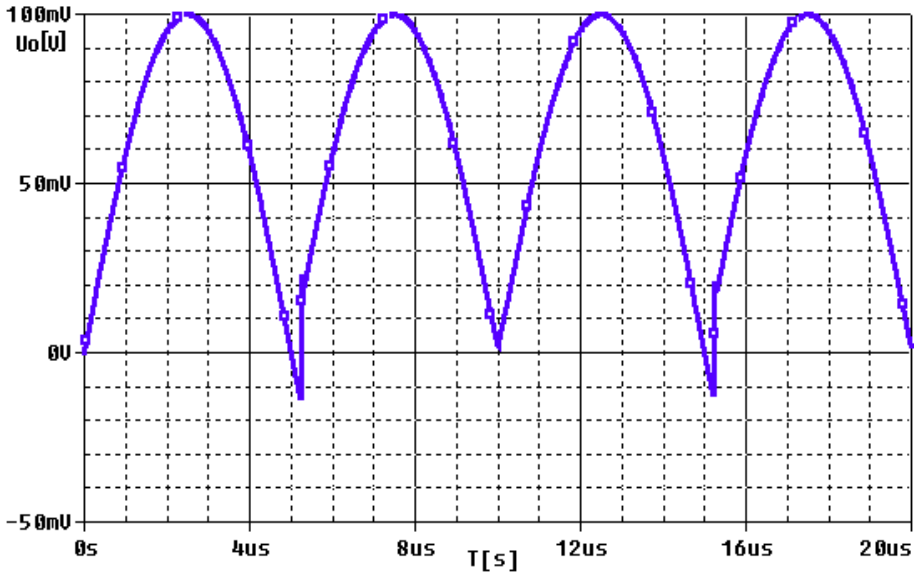


Fig. 3 Time waveform full-wave rectifier ($f = 100\text{kHz}$, $U = 100\text{mV}$)

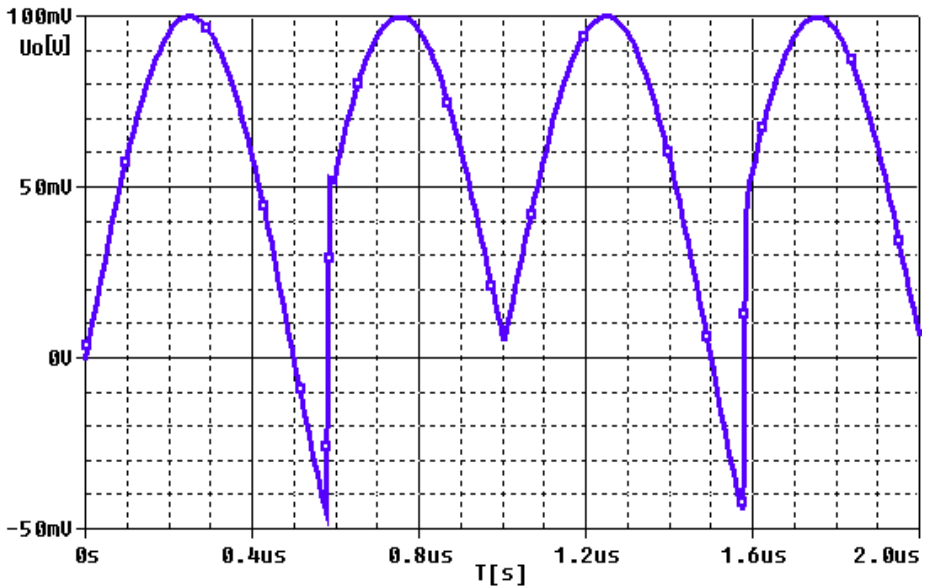


Fig. 4 Time waveform full-wave rectifier ($f = 1\text{MHz}$, $U = 100\text{mV}$)

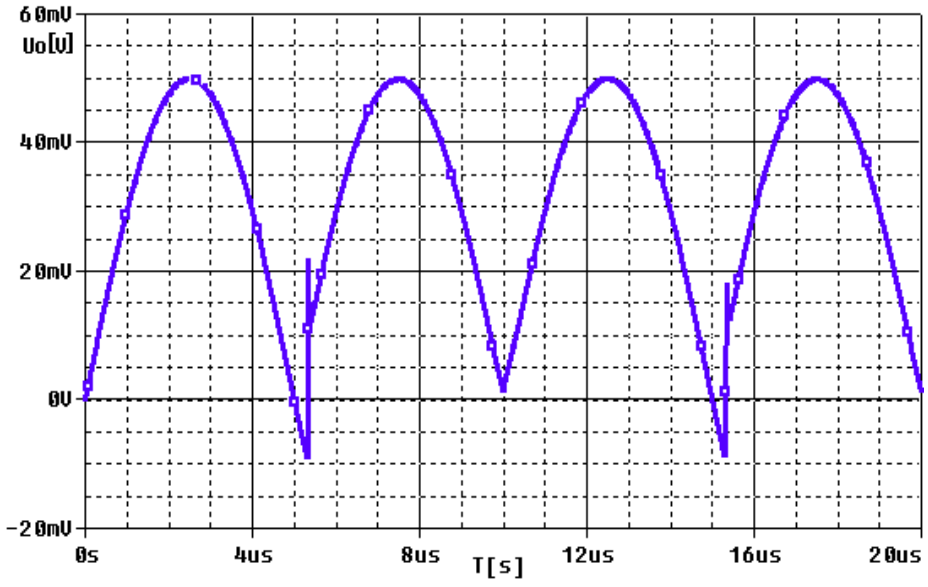


Fig. 5 Time waveform full-wave rectifier ($f=100\text{kHz}$, $U=50\text{mV}$)

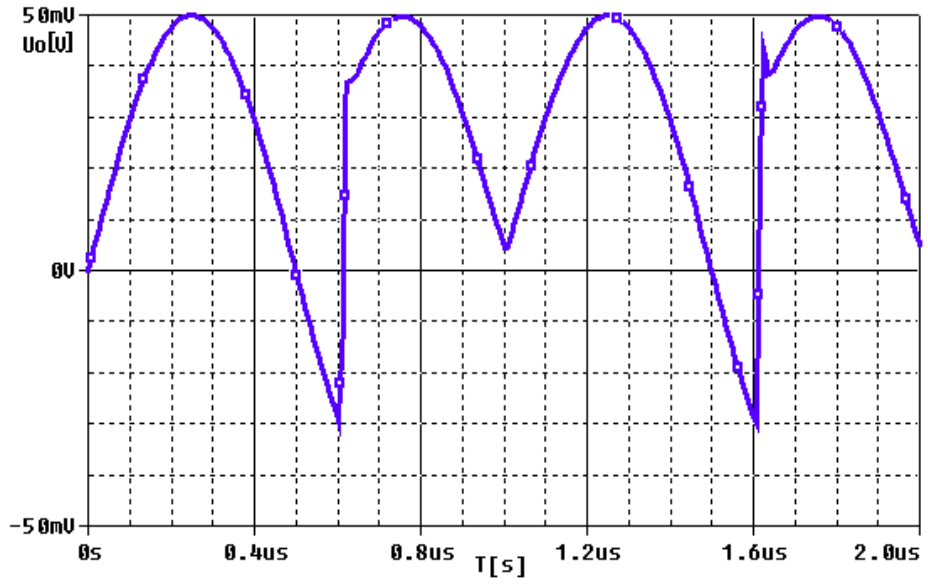


Fig. 6 Time waveform full-wave rectifier ($f=1\text{MHz}$, $U=50\text{mV}$)

5 Conclusion

The paper describes basic structure OTA based on current conveyors and its using in full – wave rectifier. From graph, we can see that waveform is deformed with decreasing input voltage and increasing frequency. Even at higher frequencies, operational full-wave rectifier preserves typical characteristics for rectifying with small distortions. It is confirmed that the current conveyor of II. generation can operate on very high frequencies. This can be also seen in the provided waveform graph. Especially in case of frequency dependent circuits.

Acknowledgement

In the paper there are the results of the project SP2014/119 which was supported by Student Grant Competition of VSB - Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, 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.
5. WILSON B., MANNAMA V.: *Current-mode rectifier with improved precision*. IEE 1995, Electronics Letters Online No: I9950185.
6. GIFT S. J. G., MAUNDY B.: *Versatile precision full-wave rectifiers for instrumentation and measurements*, Manuscript received December 7, 2005; revised May 3, 2007. Digital Object Identifier 10.1109/TIM.2007.904565.
7. Gift, S.J.G.: "A high - performance full - wave rectifier circuit," Int. J. Electron., vol. 87, no. 8, pp. 925 - 930, 2000
8. KOTON J., HERENCŠAR N., VRBA K., CICEKOGLU O.: *Versatile Precision Full-Wave Rectifier Using Current and Voltage Conveyor*, Latest Trends on Circuits, Systems and Signals, ISSN: 1792-4324
9. TOUMAZOU C., LIDGEY F. J., CHATTONG S : *High frequency current conveyor precision full-wave rectifier*. IEE 1994, Electronics Letters Online No: 19940539.

Design of a controller for the motor speed

Tomáš Dočekal and Kristýna Friedrischková

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{doc0021, kristyna.friedrischkova}@vsb.cz

Abstract. This document focuses on basic controller and especially the implementation of it. As the model for this purpose, a small mobile robot with the 32bit microprocessor was chosen. The experimental identification of system from its step response was found as suitable. All calculations were performed in Matlab.

Keywords: controller of motor speed, identification of system, controller implementation.

1 Introduction

In these days, we can find various controllers in many facilities, from simple electronic products used daily to complex and large units in the industry. One of the simpler applications relates to the regulation of electric motor revolutions. The utilization exists also in mobile robots of varied designs, sizes and use, in travels, electric cars, or in industrial facilities, milling machines and drills.

Many people deal with this issue when they need to control revolutions of motors in wide field of application, for example in the construction of mobile robots designed into the maze 2,3.

1.1 Feedback control system

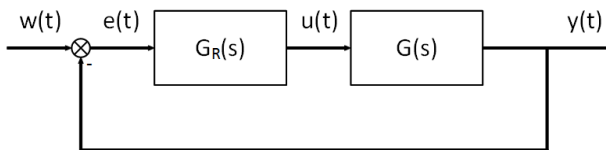


Fig. 1 Block scheme of the control loop

The Fig. 1 illustrates the basic block scheme of the feedback control system. It consists of two suitably interconnected systems. The first one is an independent system described by the transmission $G(s)$. Within this specific issue, it is the motor itself, together with the extending mechanical part and the complete load. The second one is the regulator $G_R(s)$, which is, in practice, realized by a microcomputer, programmable logical controller or by some other similar facility. A high performance

circuit, which corresponds to the engine type, is considered as its part. Its task is the generating of control action $u(t)$. The motor revolutions $y(t)$ are measured and taken, within the feedback, to a differential unit where the error $e(t)$ entering the regulator is created by the deduction from the required quantity $w(t)$ - the required speed.¹

1.2 Model system

The next part describes the controller's design by an experimental way using the system identification with the aid of the deterministic signal. For this purpose, there was a simple mobile robot determined for the line following.

The robot utilises two DC brushed motors Faulhaber 1717. Each of them is equipped with an incremental quadrature encoder having the resolution of 512 ticks per revolution. The encoders have been determined for the chance to implement the feedback. A microcontroller ARM Cortex M3 by STMicroelectronic, specifically STM32F100, was used as the controlling unit for the controller. The microcontroller includes peripherals, which are able to process the quadrature signal without the need to load the main core itself. It can generate PWM (pulse width modulation) signal controlling the motors and it has been performing well enough for the creation of the controller with a short cycle period. The output part has been composed by h-bridges consisting of unipolar transistors and their drivers.

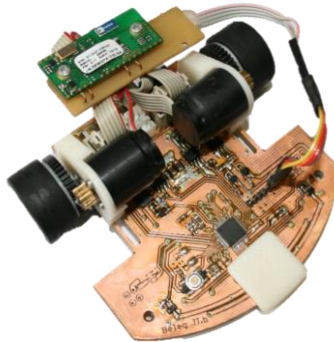


Fig. 2 Mobile robot used for the regulator's design

2 Implementation

The whole process designing the controller consists of several parts. Firstly, we need to get the description of the system which is supposed to be controlled. Then, the controller's design is prepared and, consequently, it must be implemented into the microcontroller.

2.1 The system identification

For this purpose, we opted for the experimental system identification, thanks to the step course. The first step was the measuring of the step response. There was the

PWM signal of a high frequency 27% duty cycle in steps to the system because higher values caused the lifting of the front robot part and affecting the measurement. The motor revolution course was recorded in periods of 1 ms.

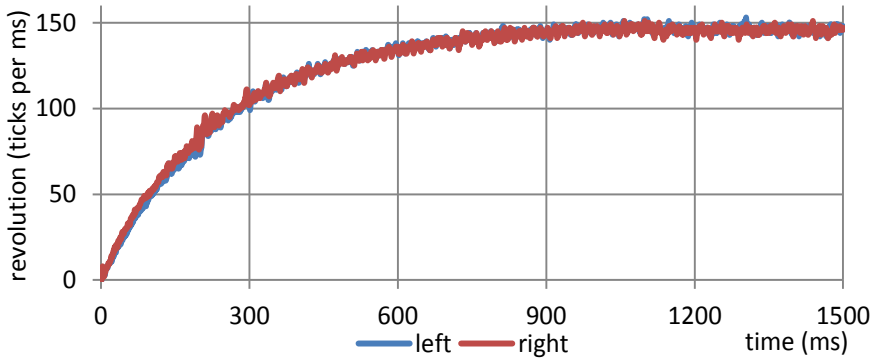


Fig. 3 The engines’ response to the step changes in PWM

As the step response of the right and the left engines were identical, we used consequently only one characteristic. The measured data were imported to the program Matlab. A second vector of the same length was created to this vector. It represented the input signal. Its height was selected to correspond to the PWM duty cycle used on the microcontroller.

The System Identification Toolbox has been used for the creation of the system description in the form of transfer function. When importing the input and output signals, from which the transfer function is then determined, the entering the correct period of sampling was very important. Consequently, there were several transfer functions variants tested, the first to the third order systems with various variants of roots, including zeros. The resulting transfer function were, however, almost identical and we have thus selected the simplest situation – the system description with the aid of the first order transfer function without zeroes.

$$G_{s\ ident}(z) = \frac{2.115}{0.2478s+1} \tag{1}$$

2.2 Design of the regulator

A proportionally integrated controller – more precisely its discrete PS variant, was selected as suitable and satisfactory for this task. A model of the control loop was created with the Simulink tool to determine the necessary constants.

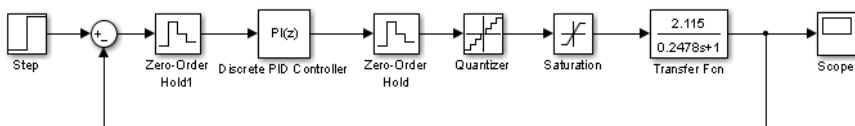


Fig. 4 Block model of the regulation circuit with non linearities

When creating it, it was important to consider several limits given by the consequent implementation. All components had to have the determined sampling period of 1 ms. The value leaving the controller, corresponding to the PWM duty cycle, was limited by the maximal value and it can be expressed only by an integer. For that reason, there were nonlinear saturation and quantizer items included into the model.

The following formula shows the general form of the controller's transfer function:

$$G_R(z) = P + I \cdot T_s \frac{1}{z-1} \quad (2)$$

The sampling period value had been already selected and it was consequently necessary to determine the constants P and I. To make it simple, we selected the PID Tuner tool for it as it allowed for the adjustment of the step response of a closed regulation circuit and, on its basis, to calculate the constants.

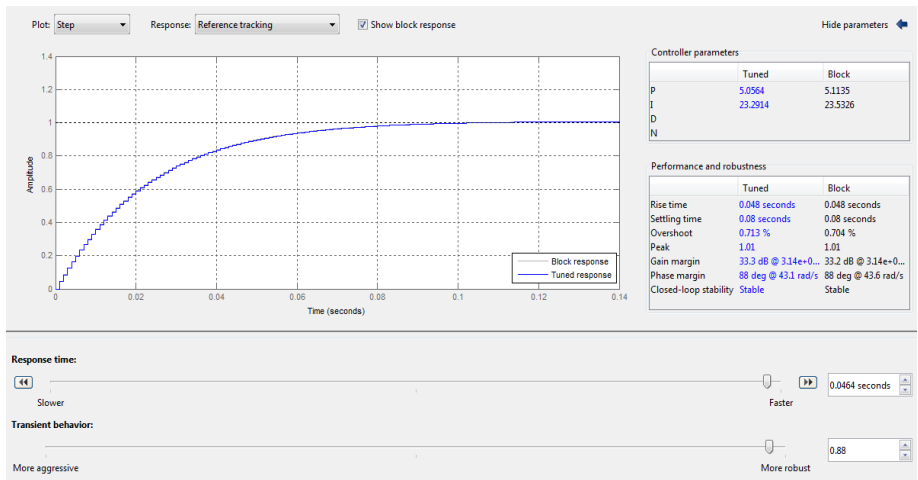


Fig. 5 PID Tuner tool for the regulator set up

The resulting controller's transfer function has thus got the following form:

$$G_R(z) = 5,1135 + 23,5326 \cdot 0,001 \frac{1}{z-1} \quad (3)$$

This form must be transferred, with the aid of Z inversion transformation, into the time domain for the implementation into the microcontroller. The Z transformation dictionary was used for that:

$$G_R(z) = P + I \cdot T_s \frac{1}{z-1} = \frac{U(z)}{E(z)}$$

$$U(z) = P \cdot E(z) + I \cdot T_s \frac{E(z)}{z-1}$$

$$u[k] = P \cdot e[k] + I \cdot T_s \cdot \sum_{j=0}^{k-1} e[j]$$

$$u[k] = 5,1135 \cdot e[k] + 23,5326 \cdot 0,001 \cdot \sum_{j=0}^{k-1} e[j] \quad (4)$$

To make all calculations related to the microcontroller satisfactorily fast, the formula, calculating the control action, was adjusted to utilize only the integer data types. The controller's implementation was done in the C language.

```
eL = (cppLdem - cppL);
eR = (cppRdem - cppR);

eL_sum += eL;
if (eL_sum > 30000)
    eL_sum = 30000;
elseif (eL_sum < -30000)
    eL_sum = -30000;

eR_sum += eR;
if (eR_sum > 30000)
    eR_sum = 30000;
elseif (eR_sum < -30000)
    eR_sum = -30000;

l_value = (int16_t) (((int32_t) 511 * eL + 2 * eL_sum) /
100);
r_value = (int16_t) (((int32_t) 511 * eR + 2 * eR_sum) /
100);

eL0 = eL;
eR0 = eR;
```

The variables `cppLdem` and `cppRdem` mean the required speed value (count per period L/R demand). Following the same logics, the variables `cppL` and `cppR` then include the data about the measured speed.

2.3 Verification of the functionality

The implemented controller was consequently tested. The speed requirement of the mobile robot was changed in step to the value 50 and the responses were measured.

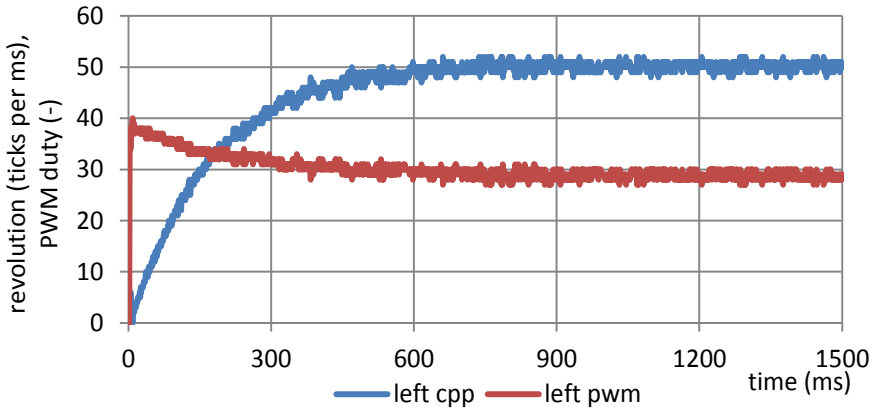


Fig. 6 Measured responses in the closed regulation circuit

3 Conclusions

Within the study, we firstly identified the system driving the mobile robot in an experimental way from the step response. A controller of the motor speed was designed and then implemented into the microcontroller on the basis of its description. Then, the response of the closed feedback control loop on the step changes was measured and also the response of the control action, in the form of the PWM signals duty cycle. The first mentioned course indicates that the response is not of the oscillating character, which would be undesirable. The setting time of the controlled value has been also relatively short – the hundreds of milliseconds and the error in the steady state has been zero.

We might notice a narrow peak at the beginning of the measured response. Probably, it has been caused by tolerances in gears and that is thus a failure quantity inserted into the control loop. If this phenomenon did not occur, the starting values of PWM signal duty cycle would be slightly higher.

4 Sources

1. OŽANA, Štěpán. *Navrhování a realizace regulátorů (Designing and implementations of regulators)* [online]. Edition 1. Ostrava: VŠB – Technical University, 2012 [cit. 2014-07-31]. 2 DVD-ROM. ISBN 978-80-248-2605-9. Accessible at: <http://www.person.vsb.cz/archivcd/FEI/NRR/Navrhovani%20a%20realizace%20regulatoru.pdf>. Textbook. VŠB-TUO.
2. Designing a phase lead controller. HARRISON, Peter. *Micromouse online* [online]. 2011 [cit. 2014-08-1]. Accessible at: <http://www.micromouseonline.com/2011/05/16/designing-a-phase-lead-controller/>
3. Designing the motor controller. HARRISON, Peter. *Micromouse online* [online]. 2011 [cit. 2014-08-1]. Accessible at: <http://www.micromouseonline.com/2011/05/15/designing-the-motor-controller/>

Software diagnosis and drivers for photovoltaic power station

Petr Drábek

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

Abstract. Renewable sources of energy are more often mentioned terms. For optimum utilization of these sources of energy is most important monitoring output and other operation values. The goal of this article is describing how diagnostic application finds fotovoltaic invertors and how drivers work with fotovoltaic invertors. In the next case is explained how data are processed, transmitted and displaed to the end users. In other important case this article deals uniform format of output data about status of fotovoltaic power station.

Keywords: Fotovoltaic power plant, Fotovoltaic invertor, RS485, Ethernet, XML, solar energy.

1 Introduction

This article deals with the development of two closely-related programs. The first software can automatically diagnose what photovoltaic inverters are connected in photovoltaic power and bus settings. The output of this software is the configuration XML file, which is required to run the other software – driver. Driver reads the values of photovoltaic inverters (FVIs). The problem with monitoring the photovoltaic power can be difference FVIs types and typical monitoring interface which is impractical because the interface from another producer doesn't read data from FVIs other producers. Developed software solves this problem. Transparency software is its ability to read data from multiple inverters connected to the bus used but especially in the ability of configurability ssoftware for various types of buses and various types of inverters. This solution has many advantages such as: low cost compared to conventional PC or PLC, then the other advantages are low power consumption, small size, the passive cooling and the possibility of extending application. Above solution much more advantageous than data logger, because it doesn't have a unified output format. Each data logger from different manufacturers have different data format and its price is also higher.

1.1 Remote configuration of the monitoring system

For a detailed description of the complete system is displayed photovoltaic power plant (FVPP) before the installation of the monitoring chain in the picture Fig. 1. Photovoltaic power plant consists of a field of photovoltaic panels, solar inverters and miniPC with a static IP address. MiniPC is connected to the Internet. Before starting the monitoring operation, the operator connects from a remote PC and runs miniPC diagnostic applications. After completing the diagnostic application is created XML configuration file and it can be started monitoring system.

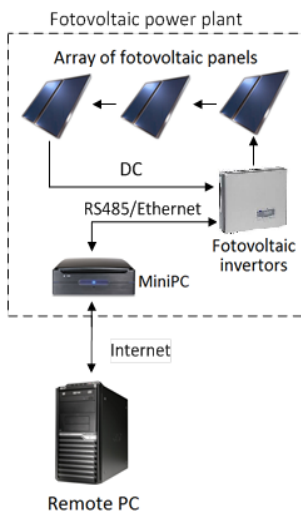


Fig. 1 Block scheme of monitoring chain photovoltaic power plant

2 Diagnostic applications for photovoltaic inverters

Before starting the monitoring of photovoltaic power plants is necessary to determine the number of FVIs connected to the communication bus and bus configuration parameter settings such as for RS485 communication speed, parity, port RS485 converter, addresses FVIs. In case of using ethernet communication bus, the configuration data FVIs IP address and virtual port on which it communicates. Currently, diagnostic application is developed for different types FVIs 4 – Vacon, SMA sunnyboy, RE-FUsoI and OK4E. The development is still ongoing and the applicability of SW will expand to another type FVIs.

2.1 The configuration XML file.

The configuration XML file contains data that are loaded at boot driver and driver is controlled by the loaded data. For each type FVIs XML configuration file is different,

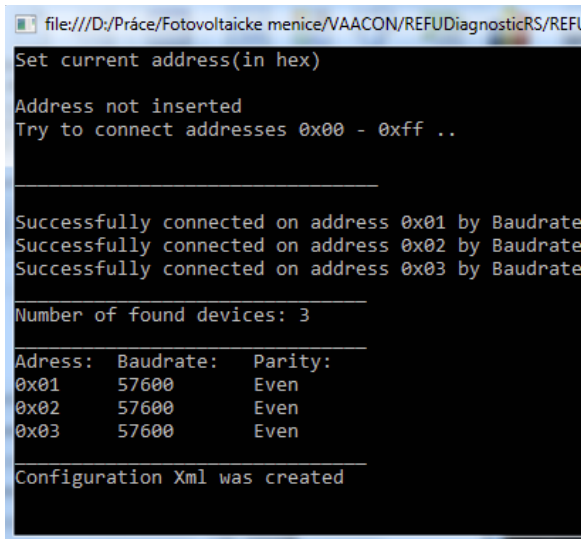
just for the type of used bus and its settings. For example, for communication via RS485 are the important data such as communication baudrate, parity, port of the RS485 converter, addresses of FVIs, name of the FVPP and the path where will be stored measured data. In the case of communication via ethernet XML file contains the IP address of each FVIs, virtual port, IP address of the FVPP, and the name of the FVPP. Example of the configuration XML file is in Fig. 2.

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
<StartInf ActDev="2,3,4" XmlVer="1" FVE_ID="8"
  Parity="Even" Baudrate="115200"
  Port="/dev/ttyUSB0" Name="FVE Prerov"
  IP="195.22.51.108"/>
```

Fig. 2 Example of the configuration XML

2.2 Diagnostic applications for FVIs OK4E

FVI OK4E is equipped with RS485 communication bus. Diagnostic application is designed for this interface. At the start of diagnostic applications via RS485 user can either enter the address of FVI, then the user can set the port or all ports tested sequentially. For each address is sent on the bus contact data packet. If it is received the correct answer, then is captured configuration (port and address of the FVIs). Communication is sequentially tested all communication speed and parity options. In the second case is checked if aren't more devices with same address on the bus. Results are shown and it is created XML configuration file Fig. 3.



```
file:///D:/Práce/Fotovoltaicke menice/VAACON/REFUDiagnosticRS/REFU
Set current address(in hex)
Address not inserted
Try to connect addresses 0x00 - 0xff ..

Successfully connected on address 0x01 by Baudrate
Successfully connected on address 0x02 by Baudrate
Successfully connected on address 0x03 by Baudrate

Number of found devices: 3

Address: Baudrate: Parity:
0x01 57600 Even
0x02 57600 Even
0x03 57600 Even

Configuration Xml was created
```

Fig. 3 Output from diagnostic program

3 Driver for fotovoltaic invertors and target device

Because each FVI communicates different protocol then must be used a specific driver for a specific FVI, then it's necessary to run the driver to loaded data from a XML configuration file. When OS starts the program are first loaded FVIs addresses, physical port, which is connected transmitter. According to the loaded data from configuration XML file packets are generated with specific commands FVIs. After generating all commands array, the packets are send one after the other on the collection, in the event that a response is received, the packet is sent repeatedly to a maximum of three time. After sending all packets algorithm generates packets for other FVM and the cycle is repeated until the data obtained from all the FVM. Furthermore the data are decoded and is creating output XML file, as seen from the flowchart (Fig. 4).

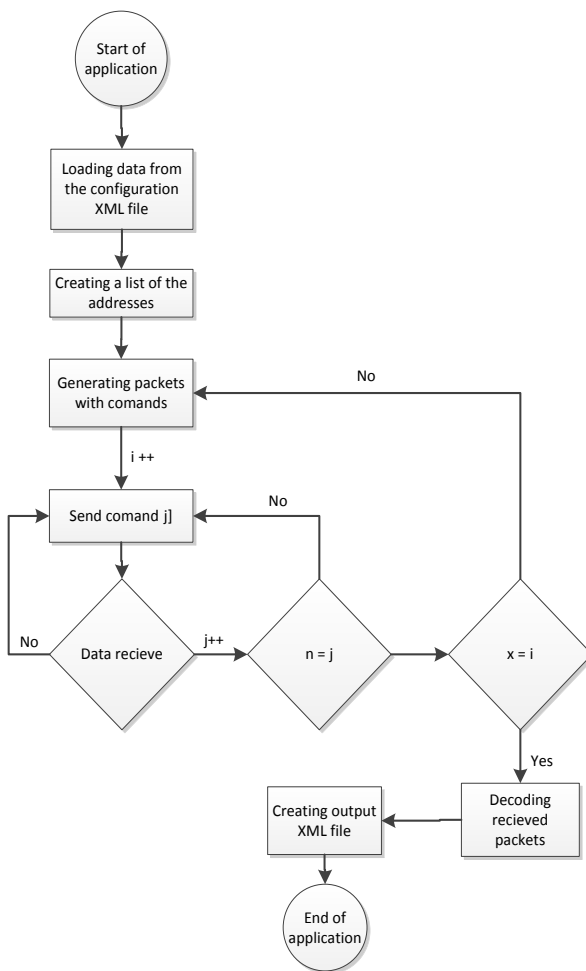


Fig. 4 Flowchart for driver of FVIs OK4E

4 Communication with the monitoring system

The driver that is cyclically executed at FVPP, generates output XML files that contain information on measured variables on FVIs. The format and structure of these XML files is completely unified. The fact that output XML file has a completely unified format allows the creation of the common monitoring system for all types of FVIs on which to run the above driver. Output Xml file is included by the name of file consists from the identification number of the FVPP, the date and time, when XML file was created XML file. This method of assignment names to files allows of the classification of data from both data sources and in terms of time inclusion.

5 Conclusion

Currently, it was developed 5 diagnostic applications and 5 drivers for FVIs, which was long – term tested in practise and real FVPP. Testing is passed in getting data, parse data, unified data format and user monitoring interface. Testing passed without errors or faults. In connection with the requirement for a common monitoring system was designed uniform format for all drivers, unified output XML file. Advantage of the common output format is portability in developing of the monitoring system. This application is still developing. There are currently developed new drivers. Results of all monitoring chain are represented in the web-user interface in the Fig. 5.



Fig. 5 Screen of the processed data in user interface (Power)

6 References

[1] MESSENGER, Rogger A. a Jerry VENTRE. Photovoltaic systems engineering. 3rd ed. Boca Raton, FL: CRC Press/Taylor, c2010. xxiii, 503 p. ISBN 14-398-0292-0.

[2] DRÁBEK, Petr. Softwarová implementace ovladačů pro fotovoltaické měniče. 2013. Diplomová práce. Vysoká škola báňská - Technická univerzita Ostrava. Vedoucí práce Ing. Zdeněk Slanina, Ph.D.

Management of energy systems combined with renewable or alternative energy resources

Kristýna Friedrischková, Bohumil Horák, and Tomáš Dočekal

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{kristyna.friedrischkova, bohumil.horak, tomas.docekal.st}@vsb.cz

Abstract. The management of energy systems combined with renewable or alternative energy resources has been currently a very intensively studied issue. The often not continual and in time determined electric power production from the current alternative resources requires research of correlation between its production and its consumption. The systems can be managed by changes in time/side demand management. In most cases, there is, however, an accumulation component needed. It must allow for the full production implementation and the electric power consumption together with the minimising of deliveries from the public grid. The presented article describes the experimental utilisation of the traction battery of an electric-powered car for the delivery of power to the family house system.

Keywords: electricity, electro-mobility, demand, production, consumption, accumulation, control.

1 Introduction

The development of modern technologies is accompanied by the increased demands on the delivery of high quality electric power. The current consumer considers the quality not only as its technical and in time and side demand management parameters, but also its price. The development trend related to small distributed alternative or renewable energy resources supplying power into the public grid requires higher demands on its management.

The primary motivation, leading unfortunately to the currently bad situation in both technology and legislature, has been the looking for energy alternatives to the limited resources of crude oil, coal and natural gas on the Earth.

The speedy development of production technologies and production capacities in the recent ten-year period has allowed the big increase in the number of power resources motivated not by the alternative electric power production and decreased consumption, but mostly by its selling to the public grid under not completely market

terms and conditions, thanks to the market globalization (including the electric power market), the market monetary-oriented economics and the lacking legislature.

The management of energy systems combined with renewable or alternative energy resources is thus currently very important. It might allow for the optimising of power deliveries and consumption, for the reduced dependency on the power deliveries or for the reduction of power costs.

2 Household consumption

The household power consumption is not balanced. We can find daily, weekly and annual periodicities. Currently, the most power is consumed for heating and the heating of hot utility water. The consumption of electric power in the period 2000 - 2012 increased in households by 12% and by 15% in the industry. The consumption growth in households has not been caused only by non economic utilisation of electric appliances, but also by the use of new appliances like, for example, dishwashers, cooking plates, coffee-makers, communication, presentation and leisure electronics, and other facilities without which modern people cannot imagine their life.

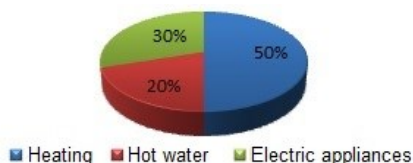


Fig. 1. Graphic distribution of the electric power consumption in households

The referential scenario suggests the increase in the total measured power consumption in households by 29% in the period from 2009 to 2040 and the following changes in its distribution:

- Reduction in the measured power consumption for heating by 22%
- Reduction in the measured power consumption for the heating of hot water by 18%
- Reduction in the energy demands by other consumption by 7%

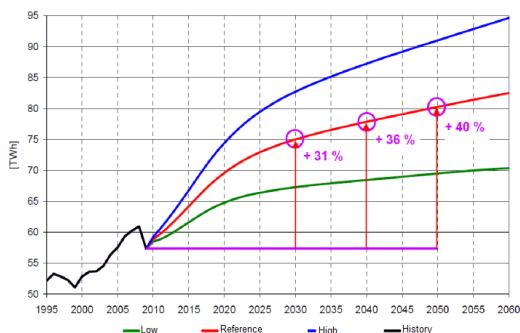


Fig. 2. Prediction of the household electric power consumption in the Czech Republic

3 Electric mobility

The electric mobility era historically started in the period when it was not yet clear which energy medium will be the best for transport. Works by H. Ford highly influenced the direction of the development at the time when he was employed at T.A. Edison and he worked on improvements of the conception of an economic petrol combusting engine. At that time, there were more electric-cars manufactured than those equipped with combustion engines. The development of electric mobility, however, ran up and it is still running against electric power accumulation problems in spite of important positives in the design. The pressure put on alternatives in transport, the efforts related to the emission reduction in transport means, changes in the structure of settlement in large agglomerations, the increased mobility needs, and at also last but not least the lowering of manufacturing costs by car manufacturers have indicated the re-emergence of the electric car conception as the transport equipment used in cities and in their outskirts and featuring minimal noise, no emissions in areas of operations, and lower costs of driving unit manufacturing.

The development of the electric mobility in cities however shows also its negative impacts and significant limitations. The public distribution grid has not been prepared for such a demand – the delivery of electric power for the electric car charging. The mass spreading of the electric cars in current cities is thus very problematic. It is different when it comes to city expansions or their spreading into outskirts, which are newly designed. Their plans might cover the satisfactory energy infrastructures.

One of the immediate solutions of the occurring situation relates to the charging of electric cars at the time of lower consumption – at night. The delivery tariff D27d for electric cars currently relates to this situation. It has been created on the basis of negotiations by E.ON, PRE, ČEZ with the Energy Regulating Authority and the Association of Electric-car Industry (ASEP).

4 Experiment

The implemented experiment counts on the future user as the person living in an intelligent family house utilising alternative energy sources and connected with the public electric power and natural gas grids. In transport from home to work, he or she will use an electric car charged alternatively at night hours from the public grid and, during days, at their workplaces or in their family houses (e.g. on Saturdays and Sundays). That will allow for the charging, for example, from a photovoltaic solar system or from the energy surplus in the cogeneration unit.

The control system in a family house reacts autonomously to users' needs and uses one of the pre selected management algorithms.

Within this model, the electric car is only an appliance running on power and it allows for the mobility of its user. However, the vehicle is standing for the most of time.

The current electric power accumulators in electric car traction batteries allow for tens of thousands of charging cycles. The vehicle reach is about hundreds of kilome-

tres. The average daily travel by a user is only between 50 and 100 Km. Great energy part is thus utilised. The experiment thus offers to verify the possibility of communication between the family house controlling systems and the vehicle and to utilise the vehicle traction battery capacity for the delivery of power to the house, under the condition of the guaranteed user’s mobility.

Based on these ideas, we verified the conception by the experimental part utilising the technology of the electric car prototype KAIPAN VoltAge. The prototype “K0” was set up at the end of 2009 and it achieved the reach of 107 Km per charge on the Hyundai circuit in Nošovice at the beginning of 2010. The vehicle driving unit had been designed for operations close to cities and the vehicle reaches the maximal speed of 65 Km/h. The traction battery system with 100 cells, the LiFeYPO4 accumulators, the voltage 320V, and the capacity 40 Ah suggests the life-span “the reach” of more than 300,000 Km. The system is charged from the grid 400V/32A and the charging time, at the charging current 32A, does not extend 2 hours (it might be charged or discharged by currents up to 120 A within the experiment).

Within the experiment, we measured the time duration of the charging and discharging current and performed the synthesis of the solution simulating the inclusion of the electric car into the house energy system. The considered energy system had several sources with installed photovoltaic system having the output of 3kWp and with a micro cogeneration unit with the output of 1kWe. At the same time, we assumed that in the case of enough energy from alternative sources, there was no power from the public grid taken.

Within the implementation experiment stage, we linked the electric car traction battery as indicated in the block scheme in Fig. 3.

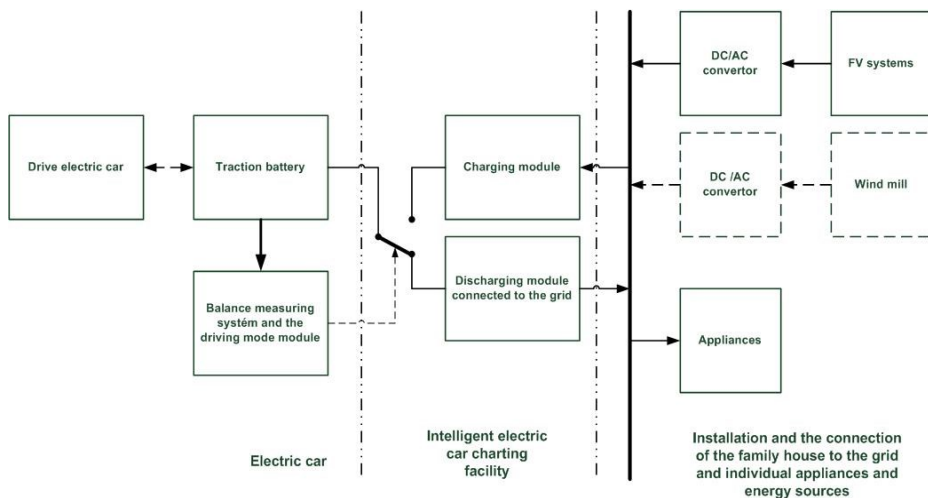


Fig. 3. Block scheme showing the linking of the electric car with the house energy system

4.1 Charging the electric car traction battery

Electric car traction batteries have got prescribed traction battery charging and discharging algorithms. These are monitored by the balance system. This measuring and action sub system protects the traction battery cells against the complete discharging in the discharging mode and against overcharging in the charging mode.

In the case of the experiment, we opted for charging AC/DC converter TCCH-H389-6. This charging device can be connected and fed from a single phase power network 230V and allows for the electric car traction battery charging with the current of up to 11A at the voltage 389V. The communication interconnection with the vehicle balance system ensures the use of the charging algorithm.

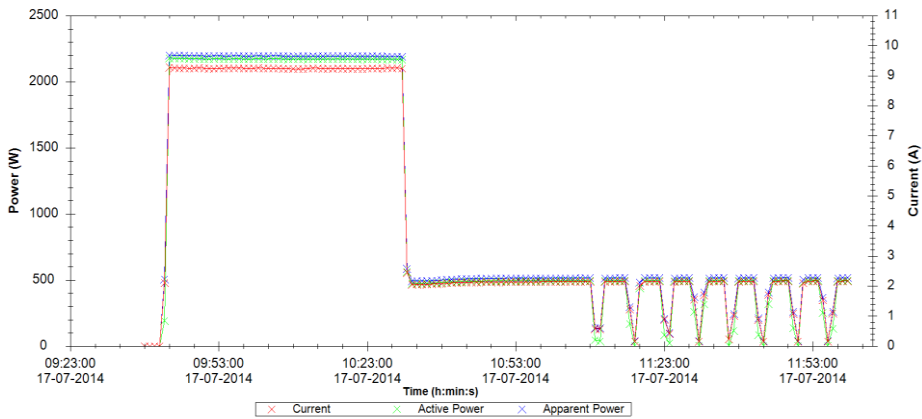


Fig. 4. The course of the electric car Kaipan VoltAge K0's traction battery charging system by the charging device TCCH-H389-6

The Figure 4 shows the course of the electric car charging. The graph illustrates the “plateau” at the beginning of the charging by the current of about 9A. The charging current lowers down to 2.5A after the balance level is achieved.

4.2 Discharging traction battery the electric car

Within the verification of the discharging conception and thus the strengthening of the house power network (the reduction of deliveries from the public grid), we organised the experiment with the direct output from the electric car to the alternating network by connecting the vehicle via the converter SMA 3000. It was linked to the control system through the communication interface RS485. The measuring focussed on varied converted set ups and on the possibility of the electric car traction battery discharging below the determined level. In the course of the taking verifying measurements, we set up the minimal voltage of the traction battery to 319V and the maximal output to the network to 2kW.

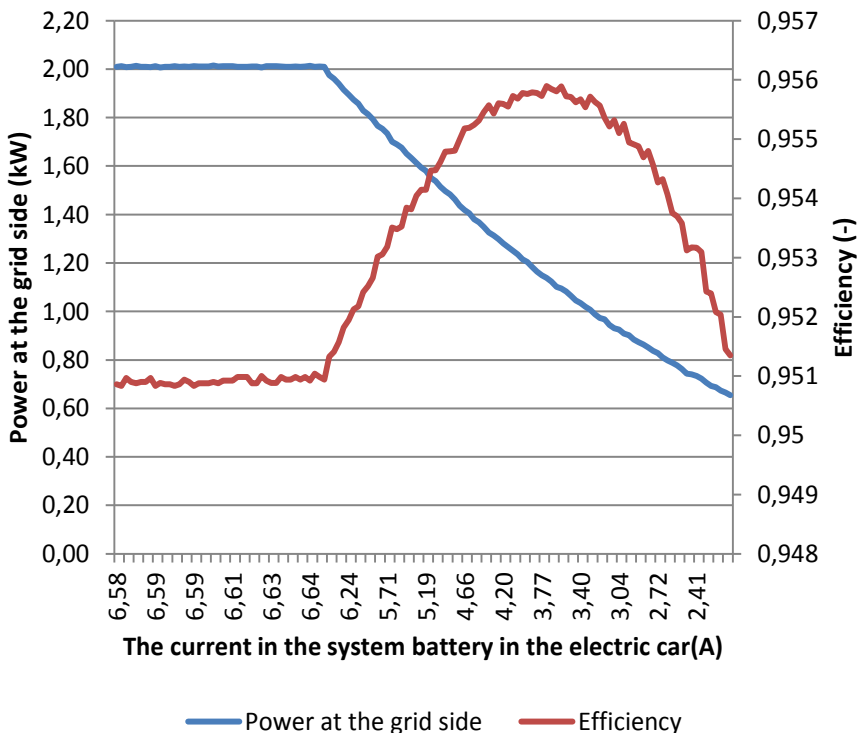


Fig. 5. The course of the output reduction dependent on the battery status

5 Conclusion

The measured values and their graphic presentation in Fig. 4. The course of the electric car Kaipan VoltAge KO's traction battery charging system by the charging device TCCH-H389-6 indicate the instability of the electric car balance system caused by the unsatisfactory specifications of the balancers' cooling (the course of the electric car charging with the lowering charging current down to 2.5A and the periodic repetition of current increases and decreases before the system shutting up).

Fig. 5 presents the course of the output reduction dependent on the battery status (the battery charging and its internal resistance). The efficiency of the converter extended 95% during the complete duration.

We have proved on the basis of this simple experiment that the electric car traction battery might be in principle utilised as the accumulation source in an intelligent house.

The current technology allows the connection of this direct source with the alternating power network, the phase connection and the power delivery on the basis of information shared with the house and the electric car controlling systems.

Further research and the technological development will focus on the area of expert management which should, on the basis of information about the system operational

history and the information about the planned future operations as well as some other information, predict the energy need guaranteeing the mobility and thus determine the amount of energy for the disposal within the intelligent house.

6 Acknowledgement

The experimental project part has been utilising and extends the results of the project MPO (MIT) TIP FR TI 1/223. The experiment results will be used as the support of the research issue “Multi source energy units” within the project Preseed II.

7 Bibliography

1. BECHNÍK, Bronislav. Národní akční plán České republiky pro energii z obnovitelných zdrojů. TZB info: Technická zařízení budov [online]. 13.3.2012, č. 3 [cit. 2014-07-15]. Dostupné z: <http://www.tzb-info.cz/8364-narodni-akcni-plan-ceske-republiky-pro-energii-z-obnovitelných-zdroju>
2. HALUZA Miroslav, MACHÁČEK Jan. Spotřeba elektrické energie domácností, predikce a potenciální úspory pomocí BACS. TZB info: Technická zařízení budov [online]. 7.5.2012, [cit. 2014-07-15]. Dostupné z: <http://elektro.tzb-info.cz/8570-spotreba-elektrické-energie-domácnosti-predikce-a-potencialni-úspory-pomoci-bacs>
3. E.ON přináší na trh novinku – distribuční tarif pro elektromobily: Tiskové zprávy. E-ON. Webové stránky společnosti e-on [online]. 22. 1. 2013 [cit. 2014-07-21]. Dostupné z: <http://www.eon.cz/cs/o-spolecnosti/media/tiskove-zpravy/eon-prinasi-na-trh-novinku-distribucni-tarif-pro-elektromobily-140.shtml>
4. SMD INC. Yazaki vehicle power and data related connectors and components. [online]. neuveveno [cit. 2014-07-22]. Dostupné z: <http://www.smdinc.com/j1772.php>

8 Publications

1. FRIEDRISCHKOVÁ Kristýna, HORÁK Bohumil. Use of electric vehicle for power consumption optimizing in household. RSYIP. Písek u Jablunkova. VŠB TU OSTRAVA. 12. - 13. září 2013, [cit. 2014-07-15].
2. FRIEDRISCHKOVÁ Kristýna, HORÁK Bohumil, MINAŘÍK Daniel. „Kaipan VoltAge and his serial hybridisation“. 10th International Hydrogen and Fuel Cell Conference. Birmingham, UK. 26 - 27 March 2014, [cit. 2014-07-15].
3. FRIEDRISCHKOVÁ Kristýna, HORÁK Bohumil. Využití elektromobilu jako součást akumulací soustavy elektrické energie energeticky nezávislého rodinného domu. Konference alternativní zdroje energie 2014. 1-3.7.2014, [cit. 2014-07-15]. Kroměříž.
4. FRIEDRISCHKOVÁ Kristýna, HORÁK Bohumil. Electric car traction system and his secondary exploitation. Workshop nekonvenční zdroje elektrické energie a moderní akumulátory. 21-23.5.2014, [cit. 2014-07-15]. hotel Panorama, Blansko

Autonomous mode generator unit start choke control

Kristýna Friedrischková¹, Tiantian Yu², Boris Vukojević³, Jakub Rawicki⁴, and Shen Junwen⁵

¹ VSB-Technical University of Ostrava, Ostrava, Czech Republic
kristyna.friedrischkova@vsb.cz

² Shanghai Maritime University, Shanghai, China
yttsweet@126.com

³ University of Osijek, Osijek, Croatia
boris.vukojevic@hotmail.com

⁴ Silesian University of Technology, Katowice, Poland
jakuraw801@student.polsl.pl

⁵ Shanghai University, Shanghai, China
shenjw92@gmail.com

Abstract. This paper describes the design and realization of autonomous choke control of fuel of the electricity generator. Before starting the generator, the choke needs to be pulled and after start, it needs to be pushed back. The DC driver assembly includes slide block, DC driver, limit switch controlled by the embedded pi. The DC driver assembly in the control unit works when it receives the sign from the computer and stop working when it touches the limit switch. The created model will be used for pull and push the choke in the generator automatically.

Keywords: generator, autonomous mode, start choke, control

1 Introduction

With the development of modern technology, the electrical devices need to work in a smart way. Production and consumption of electrical power can be constantly monitored by computer. If needed, computer can adjust amount of produced energy. For this, electricity generator has to be able to start on demand.

In electricity generation, an electric generator [3, 4] is a device that converts mechanical energy to electrical energy. A generator forces electric current to flow through an external circuit.

Generator used in this project is shown in Fig 1. To start the generator specific routine has to be followed. Routine is:

- Pull choke to on
- Pull rope of starter
- Push choke to off

In this paper device for operating choke is described.



Fig. 1. Electricity generator chosen for the project

2 Manual control of the choke

Generator choke has two positions, on and off. Device for operating the choke needs to be able to pull and push choke on generator into these positions. Functions of the device are:

- put lever to ON position (Figure 2 right)
- put lever to OFF position (Figure 2 left)



Fig. 2. Combustion engine “choke” in “on” position

3 Measured mechanical properties

3.1 Choke movement

To push or pull the choke, the length of the choke is needed. This length dictates the distance between the slide block and the limit switch.

Measure method

Pull out the choke and use tape to measure the length of the choke. Then push the choke and measure the length.

Table 1. Length of choke

	1	2	3	4	5	6	Average
Longer/mm	2.61	2.62	2.59	2.58	2.60	2.62	2.60
Shorter/mm	0.18	0.19	0.21	0.20	0.21	0.19	0.20

$2.60\text{cm}-0.20\text{cm}=2.40\text{cm}$.

Distance between two limit switches is 2.40cm.

3.2 Maximum device movement**Measure method**

Start the DC driver and let the slide block move to the limit switch. Make a mark at the position of the up circle. Then change direction of voltage to make the DC driver drive the slide block to another limit switch, then mark the position of the up circle. Use micrometer to measure the distance between the two marks.

Table 2. Get the length of distance

	1	2	3	4	5	6	Average
On side/mm	37.18	37.84	37.39	37.69	37.35	37.28	37.46
Another side/mm	77.58	77.46	77.67	77.35	77.58	77.40	77.51

$77.51\text{mm}-37.46\text{mm}=40.05\text{mm}$

The distance between the two limit switch equals to 40.05mm.

Force needed for choke movement

Use force gauge to Measure the force for push and pull choke is used.

Table 3. Results

Force/N	49.6	57.8	51.2	51.6	46.4	50.2	53.8
---------	------	------	------	------	------	------	------

Max = 57.7 Min=46.4

Average= $(49.6+51.2+51.6+50.2+53.8)/5=51.28\text{N}$

4 Proposed automatization of the choke lever

Servomechanism – an automatic device which enables precise steering of diverse kind of objects, usually called servomotors or stepper motors. Users of these types of

mechanisms are entirely informed about current position of operated object. The output signal can be as well data concerning variables like speed or acceleration. Block schematic of system for operating choke is shown in Fig 3.

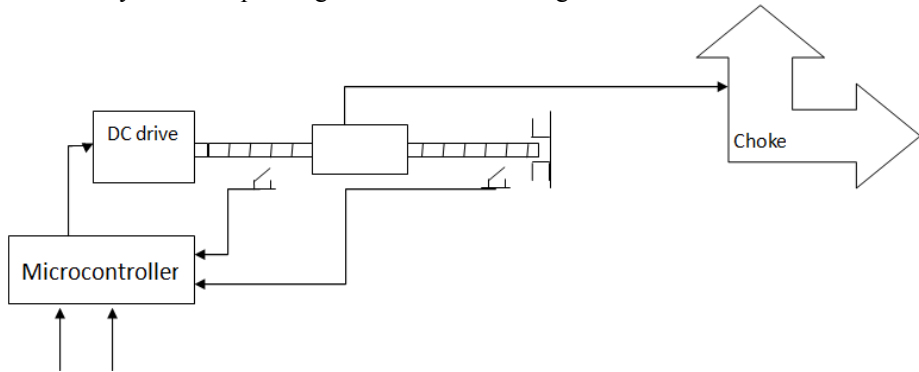


Fig. 3. Block schematic of choke operating device

In this system feedback is provided with two switches used for detecting extreme positions. Operator in this case is a microcontroller. Interface between microcontroller and DC drive is provided by L298N component [2].

L298N is a kind of h-bridge, which can control forward rotation and reverse rotation of the DC motor. The microcontroller send signals to the h-bridge drive the DC driver, when the slide block touch the limit switch, the DC driver will stop moving.

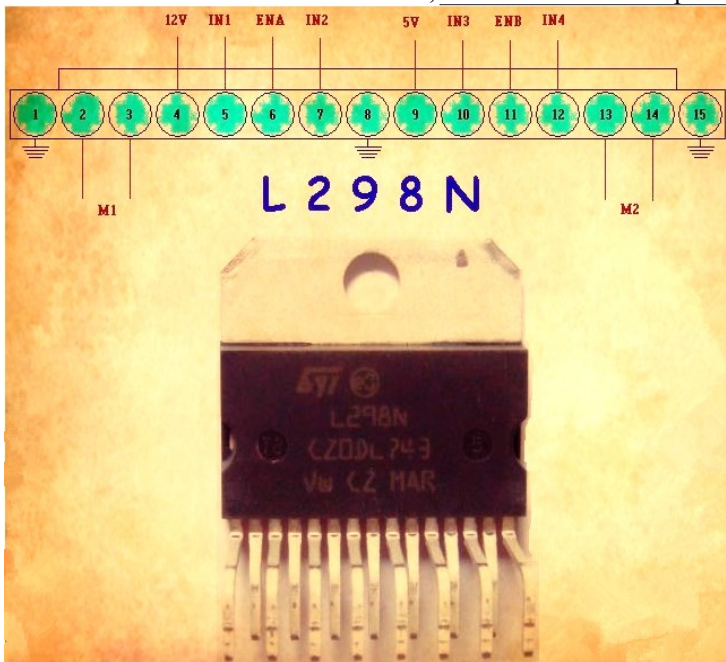


Fig. 4. L298N H-bridge with pin functions

5 Description of the control algorithms

Device for pushing and pulling the choke can be set at any position in between two ends. Being able to set the choke just in extreme positions is sufficient to operate it. Switches at extreme positions are used for stopping the movement and determining position.

Algorithm for moving choke:

```

Disable motor
Set direction of rotation
While( not_in_extream_position )
    Enable motor
Disable motor

```

Motor is controlled via L298N component. To implement shown algorithm three pins of that component are used:

- motor1_enable
- motor1_direction1
- motor1_direction2

Motor is enabled by high state at pin motor1_enable, and disabled by setting low state on the same pin. To set direction motor1_direction1 and motor1_direction2 have to be used. They have to have opposite states for motor to move.

Table 4. Settings for pins

Pin motor1_direction1	Pin motor1_direction2	Function
Low	Low	motor stopped
Low	High	Motor moving one direction
High	Low	Motor moving other direction
High	High	Motor stopped

6 Conclusion

In this paper problem of automating generator choke operation is addressed. Problem is described and presented. Problem parameters are measured. System for achieving automation is presented. System consists of mechanical parts and controller. Algorithm for controller is given. System still needs to be tested. Measurements of speed and power consumption need to be performed. The control of the choke must be synchronized with other mechanisms attached to generator. The logic part of steering the

choke should be able to stop the process if received measurements crossed the set limits.

7 Acknowledgement

This work was supported by projects “Preseed II” and SP2014/188 “Control of technological systems with OAZE providing an independent sustainable development of complex system 2014” within the student’s grant competition (VŠB - TU Ostrava).

8 Reference

1. me.queensu.ca [online]. 2014. [cit. 2014-07-31]. Available WWW: <http://me.queensu.ca/Courses/452/files/ghm01motordatasheet.pdf>
2. www.sparkfun.com [online]. 2014. [cit. 2014-07-30]. Available WWW: https://www.sparkfun.com/datasheets/Robotics/L298_H_Bridge.pdf
3. en.wikipedia.org [online]. 2014. [cit. 2014-07-30]. Available WWW: http://en.wikipedia.org/wiki/Electric_generator
4. www.gongchang.com [online]. 2014. [cit. 2014-07-30]. Available WWW: http://www.gongchang.com/Electricity_Generator-dp2343491/

Image Signal Processing and Face Detection

Jiri Haska*

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

Abstract. This paper describes an algorithm, which searches for the center point of a human face in front of the video cam and marks the position of the face with the rectangle. The hardware, which is responsible for sensing and face recognition, is placed inside the robotic head. The head is capable to distinguish colors, shapes, position of the objects and it can also count the number of objects in the image. The processed information are sent via bus to processing unit which is driving the actuators. Actuators then make the head turn. The algorithm saves the image as a matrix, where it searches the particular color and assigns one to the occurrences of the color in the image and zero to another colors. Finally it searches the center points or shapes of the objects.

Keywords: Face detection, Colors recognition, Robotic vision

1 Introduction

The vision of the grant task at the Department of Cybernetics and Biomedical Engineering was to create a robotic head which would be located in the lobby , see Fig. 1. This head will react to the surrounding stimuli. The task of the head is to scan its surroundings areas when in the presence of a person in front of this head , the head detects the person and looks at the person. If this robotic head is asked a question , the head reacts to it and accordingly responses.

This article was created in connection the the system for the recognition of signs as described in the article [5]. The application is being created in the team of four people when one deals with the mechanics of robotic head. The next person is in charge of detecting the voice, the recognition of the question and its following response. The third person creates a decision-making protocol to connect all the parts. This controls the communication between individual parts of the robot. The last task is the visual detection of human by robotic head which this article deals with.

* This work was supported by project SP2014/156, "Microprocessor based systems for control and measurement applications." of Student Grant System, VSB-TU Ostrava.



Fig. 1. The robotic model of a head is being developed at the Department of Cybernetics and Biomedical Engineering.

2 PROBLEM DEFINITION

The basic task is to convert the unloaded image to the matrix. To reduce the computing power only four numbers are stored in the matrix. The read value of the pixel is represented as RGB. After comparing whether the color is sought, these numbers are inserted into the matrix. They are listed in the Tab. 1.

Table 1. The numbers of colors for individual parts of human face.

0	Not seeking color
1	Face
2	Eyes
3	Lips

In this way filled matrix the search is performed. Searching is done by function similar to scanning, so always only one column in the matrix is being searched for.

3 Face recognition

3.1 The face detection algorithm

To the decision-making algorithm see Fig. 2 the intervals that define the colors appearing on the human face have been created.

For the purposes of the final application the interval of the colors of the face is set to the color of the face of the average central European man. If the comparison passes the condition, the number 1 will be saved into the matrix under the same index position.

```
//face
if ((c.R <= 184 && c.G <= 134 && c.B <= 107)...
    && (c.R >= 113 && c.G >= 71 && c.B >= 49))
```

The color of eyes that appeared on the snapshot was rather grey than white. Therefore the color of eyes was set more likely to light grey. If the color undergoes the condition, the number 2 is saved in the matrix.

```
//eyes
if ((c.R <= 100 && c.G <= 85 && c.B <= 70)...
    && (c.R >= 80 && c.G >= 75 && c.B >= 60))
```

The color of the lips was set from red to dark red up to brown. If the color undergoes this condition, the number 3 is saved to the matrix.

```
//lips
if ((c.R <= 125 && c.G <= 80 && c.B <= 70)...
    && (c.R >= 110 && c.G >= 65 && c.B >= 55))
```

After unloading the image from the camera to the matrix, the searched algorithm see Fig. 2 undergoes. Searching proceeds that way that one pixel is tested. If one of the conditions is met, its value in the matrix changes to the new one. If no new color is observed it changes to zero see Fig. 5, where zero is represented as a white color. If the value was successfully changed in the matrix, this would be repeated in the following pixel.

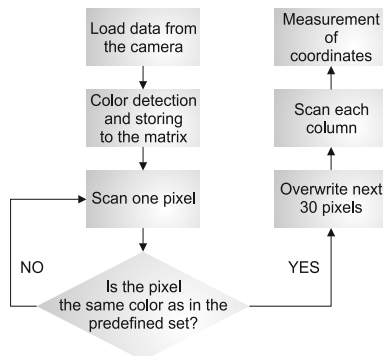


Fig. 2. Simplified flow chart of the face detection algorithm.

3.2 A modified matrix

The original image which is being captured by the camera is shown in Fig. 3. In this sample of the image the colors which algorithm was described in the chapter 3.1 are searched for.



Fig. 3. Original picture.

Fig. 4 shows the captured image which has been converted into the resulting matrix. For the idea there is the number 1 represented by an orange color, the number 2 by a yellow color and the number 3 by a red color.



Fig. 4. The detection of the sought colors and the conversion into the matrix.

In this customized matrix the distortion of detection happens that poorly enlightened places may not be correctly discovered on the face and therefore the blank places appears in the image. Therefore at all found pixels the following 30 pixels to the right are overwritten with the same number of the color as the flowchart describes in Fig. 2. This adjustment eases the detection of the centre of the face. A modified matrix is in Fig. 5.

In this customized matrix the search algorithm that compares each column whether there are all the colors used. If these colors are discovered then the counting of how many columns has been discovered this way can begin, see Fig. 6. Subsequently the number of pixels from the detection of the eyes to the appearance of the lips in the column is always counted up.

From the first sum the value of 30 is deducted from the resulting value. This value was added in the calculation by overwriting 30 pixels. In the second sum the center between the eyes and the lips is calculated. This way the axis X and Y are calculated which represent the center of the face.



Fig. 5. The modified matrix after overwriting of 30 pixels.

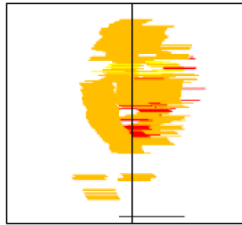


Fig. 6. Finding the position of the colors in the matrix.

4 Verification of face recognition method

In order to test the proposed algorithm the test application in Visual C# has been created. The following outputs are then exported from this test application. This testing foreruns the final implementation of the proposed algorithm to the firmware of the microcontroller.

4.1 Test

The first measurement was done in a well lit room where the door serves as the background and face will be read from the front, as seen in Fig. 7.

Lightening: The yellow bulb
 Background: The door
 View: From the front

The algorithm evaluated the axis of the center of the face:

X – 337
 Y – 212

The result is very precise which can be seen from Fig. 7.

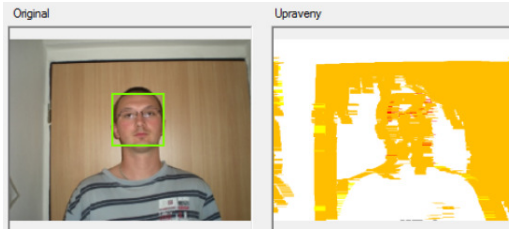


Fig. 7. The face detection algorithm test when viewed from the front

5 CONCLUSION

When testing it was obvious that if the tested object was in an environment where not many colors as background appeared on the face, the detection was quite accurate. The best results were found in the countryside and in the cities. On the contrary the worst results were detected when behind the searched object there was a red yellow white background, so the background consisted of colors that appeared in the face of the average central European man.

In comparison with the cameras allowing the detection of the face, the distortion of the result happens only with unsuitable background, however it has a distinct advantage of independence on the computing power and the size of the flash memory for an individual code.

The aim was to create a different vision of what is today commonly used in digital cameras. The emphasis was placed on the little computing power and the ability to deploy this solution on the less powerful processors. This algorithm was applied to MCF51AC256 Freescale microcontroller and the image was adopted from CMOS sensor Aptina MT9V131.

In the future work the algorithm will be modified for the automatic modification of the parameters, so that it will be able to work for other skin colors. This action will lead to the creation of the next iteration of the program when the algorithm will work with predefined kinds of the skin colors.

References

1. R. Slaby, R. Hercik, Z. Machacek "Compression methods for image processing implementation into the low capacity device," *Technical Gazette* Vol. 20, No.6, pp. 1087-1090, 2013.
2. R. C. Baker and B. Charlie, "Nonlinear unstable systems," *International Journal of Control, Automation, and Systems*, Vol. 23, No. 4, pp. 123-145, 1989.
3. M. Young, *The Technical Writer's Handbook*, Mill Valley, Seoul, 1989.
4. M. Young, *The Technical Writer's Handbook*, Mill Valley, Seoul, 1989.
5. *Correlation methods of OCR algorithm for traffic sign detection implementable in microcontrollers*, Hercik, R., Slaby, R., MacHacek, Z., Koziorek, J., 2013.

Advanced Method for Selective Extraction of Lutein and Zeaxanthin from Plant Material

Barbora Hrvolová^{1,2} and Jiří Kalina¹

¹University of Ostrava, Faculty of Science, Dvořákova 7, 701 03 Ostrava, Czech Republic

²Department of Cybernetics and Biomedical Engineering, FEECS, VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
barbora.hrvolova@vsb.cz, jiri.kalina@osu.cz

Abstract. In the Czech Republic there are a lot of environmental problems caused by knotweed (*Reynoutria spp.*). Millions of Czech crowns are spent for liquidation of that plant. Knotweed is invasive plant that performs really huge environmental problem all over the world. In the Czech Republic there is no usage for this plant. For this purpose we tried to find some possible usage for knotweed. In our research we found out high concentration of lutein a zeaxanthin (important antioxidants) in knotweed. Therefore we proposed advanced and simple method for selective extraction of lutein and zeaxanthin from knotweed. Proposed method is based on physical-chemical properties of individual pigments. Individual pigments were identified by UV/VIS absorption spectrophotometry and by liquid chromatography equipped with DAD detection. Designed method is simple, fast, and economic.

Keywords: lutein, zeaxanthin, extraction, HPLC

1 Introduction

Carotenoids are yellow, orange and violet pigments which can be found in plant and animal tissues. Carotenoids can be divided into carotenes and xanthophylls. Lutein and zeaxanthin belong to carotenoids xanthophylls. Due to their properties lutein and zeaxanthin are really popular [3, 6]. They can play many roles in biological organisms. For example, in light harvesting systems lutein has absorption function. Probably the most important role is their preventive effect against cataracts and age related macular degeneration (AMD) [3]. These pigments are essential for human health, but their abundance in the human body is entirely dependent on the dietary intake [3, 4]. They help to keep eyes safe from oxidative stress and high-energy photons of blue light. In addition, lutein and zeaxanthin perform many other important functions such as cardiovascular disease prevention, cancer prevention, and protection against oxidant-induced cell damage [1]. Due to their potential therapeutic functions, the presence of lutein and zeaxanthin in human diets has become of considerable interest. Moreover, many nutritional supplements enriched with lutein and zeaxanthin have been put on the market to enhance intake of these antioxidants. Lutein and zeaxanthin are found in green leafy vegetables such as kale, spinach or cabbage. These foods are

consumed in small amounts. While zeaxanthin is also found in dark green leafy vegetables, it is perhaps 20 times less abundant than lutein. In some plants there is also very high concentration of lutein and zeaxanthin. For example, marigold flowers are the main commercial source of these pigments especially of lutein [2]. The highest natural source of zeaxanthin was recently established to be Chinese wolfberry (Goji berries). In our research we found that the knotweed (*Reynoutria* spp.) contains really high-concentration of lutein and zeaxanthin. The knotweed is invasive plant that comes from Japan. The knotweed can be found all over the Moravian Silesian region, where it makes really big environmental problem [5]. Millions of Czech crowns are spent for liquidation of knotweed [5]. In the Czech Republic there is no usage for knotweed.

Many complicated extraction methods have been developed for dietary supplement industry [6, 7]. These complicated instrumental methods are very expensive and it is the reason why the dietary supplements enriched with lutein or zeaxanthin are so expensive. For this purpose a simple method for selective extraction of lutein and zeaxanthin from knotweed was developed. The second reason why this method was suggested was to find a use for knotweed.

2 Material and Methods

High concentration of lutein and zeaxanthin was found in knotweed by chromatographic analysis of extract of its pigments (Fig.1).

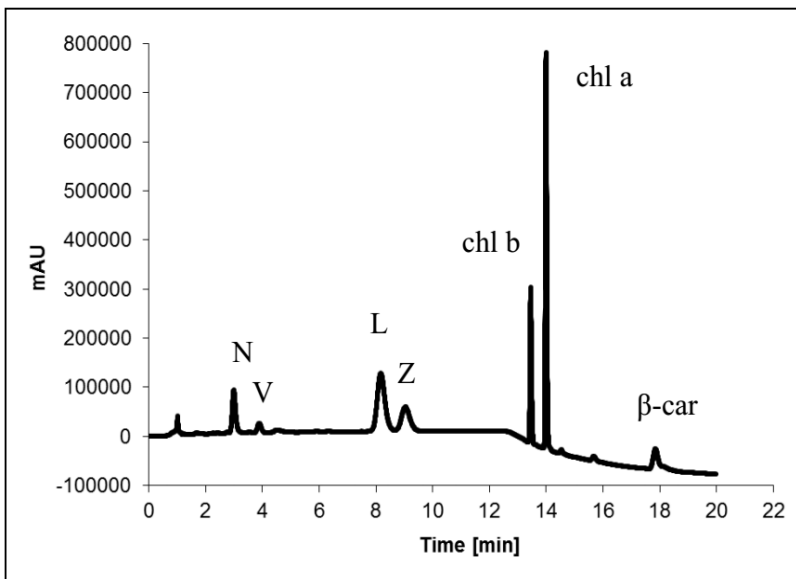


Fig. 1. Chromatogram of 80 % acetone extract of pigments from knotweed (*Reynoutria* spp.) (N - neoxanthin, V - violaxanthin, L - lutein, Z - zeaxanthin, chl b - chlorophyll b, chl a - chlorophyll a, β-car - β - carotene).

The quantity of lutein was 800 µg/g f.w. and the quantity of zeaxanthin was 300 µg/g f.w. Due to high concentration of these antioxidants the knotweed has become of considerable interest.

In our research middle segments of Knotweed leaves were used. For development of simple method for selective extraction of lutein and zeaxanthin the method for isolation of carotenoids from mixture of knotweed pigments had to be proposed.

2.1 Isolation of carotenoids

The method for isolation of carotenoids from mixture of knotweed pigments had four parts. The first three parts of suggested method were preparatory. The fourth part was the separation. For the purpose of elimination of phenolic compounds, extraction of pigments and saponification of green pigments the first three parts were realized. Last part of this method was separation carotenoids from that mixture. We used ultrasound (37 - 42 kHz, 130 W, UC006DM1, TESLA, Czechoslovakia) for selection of individual compounds.

The weight of knotweed samples was between 400 and 500 mg. Elimination of phenolic compounds was realized by ultrasound extraction in water for 1 minute. After the removal of phenolic compounds the sample was transferred to 10 ml of methanol. The removal of green pigments was realized by ultrasound extraction in 10 ml of 100 % methanol for 10 minutes. Last preparatory part was saponification. Saponification was performed by addition of 1 g of KOH. Selection of carotenoids from mixture was performed by addition of 5 ml of n-hexane. Carotenoids in mixture of pigments were transferred to n-hexane by shaking. n-hexane fraction with carotenoids was transferred to other glass test-tube.

2.2 Isolation of lutein and zeaxanthin

The mixture of carotenoids contained lutein, zeaxanthin, β – carotene, and small amount of chlorophylls. Selective extraction of lutein and zeaxanthin had two parts. First part was performed by addition of 5 ml of 100 % methanol to test-tube with mixture of carotenoids. Further the methanol fraction was three times washed by 100 % n-hexane. Washed methanol fraction contained only lutein a zeaxanthin.

2.3 Identification of pigments

All samples were analyzed by spectrophotometer UV550 (Unicam, Great Britain) and by HPLC (TSP Analytical, USA). Identification of individual pigments was based on description of their absorption spectra and retention times.

3 Results and Discussion

Proposed method for selective extraction of lutein and zeaxanthin was based on their physical-chemical properties. The main physical-chemical property was a polarity of

individual pigments. Due to the knowledge of these properties we could design a really simple and effective method for extraction of lutein and zeaxanthin from plant material especially from knotweed. The process of extraction consists of isolation of the mixture of carotenoids and selective isolation of lutein and zeaxanthin from that mixture. Final methanol fraction contained only lutein a zeaxanthin (Fig. 2).

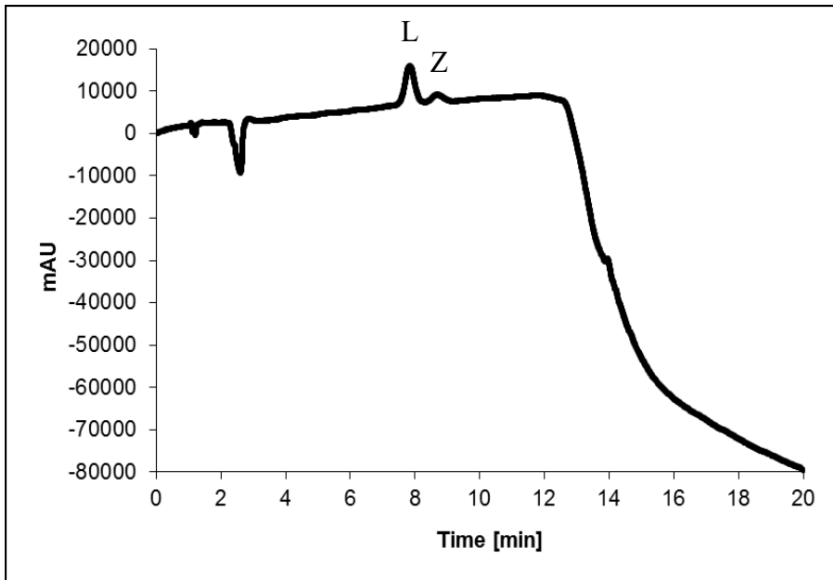


Fig. 2. Chromatogram of methanol extract of lutein and zeaxanthin from knotweed (*Reynoutria spp.*) (L – lutein, Z – zeaxanthin).

In knotweed the mean contain of lutein and zeaxanthin was 800 $\mu\text{g/g}$ f.w. and 300 $\mu\text{g/g}$ f.w.. The extraction yield was 16.7 $\mu\text{g/g}$ f.w. of lutein and 5.6 $\mu\text{g/g}$ f.w. of zeaxanthin. In comparison with other extraction techniques our method has lower extraction yield but costs for extraction are much lower. Designed method is simple, effective, fast and mainly economic. The most beneficial aspect of this method is that we found use for invasive plant knotweed.

4 Conclusion

All over the world millions of Czech crowns or dollars are spent for liquidation of invasive plant called knotweed (*Reynoutria spp.*) [5]. This plant is liquidated without any usage. For this purpose we wanted to find some usage for this plant. Due to our research we found out the knotweed contains high concentration of two very important carotenoids (lutein and zeaxanthin). Therefore we proposed the simple method for selective extraction of lutein and zeaxanthin. Designed method has the following advantages: it is simple, fast, effective and cheap.

5 Acknowledgment

This work was supported by Project CZ.1.05/2.1.00/03.0100 (IET) financed by the Structural Funds of the Europe Union, by Project LO1208 of the National Feasibility Programme I of the Czech Republic, by project of the Moravian-Silesian Region Reg. No. 02508/2013/RRC., by student grant identification number sgs06/PřF/2014, and by the project SV SP2014/194 “Biomedical engineering systems X”.

6 References

1. ARNAL, E., M. MIRANDA, J. BARCIA, F. BOSH-MORELL a F. J. ROMERO. Lutein and docosahexaenoic acid prevent cortex lipid peroxidation in streptozotocin-induced diabetic rat cerebral cortex. *Neurosci.* 2010, 166, 271-278.
2. AUSICH, R. L. a D. J. SANDERS. Process for the formation, isolation and purification of comestible xanthophyll crystals from plants. US Patent 5. 1997, 648, 564.
3. BONE, R. A., J. T. LAUDRUM a S. L. TARSIS. Preliminary identification of the human macular pigment. *Vis Res.* 1985, č. 25, s. 1531-1535.
4. LAKSHMINARAYANA, R., G. ARUNA, R. K. SANGEETHA, N. BHASKAR, S. DIVAKAR a V. BASKARAN. Possible degradation/biotransformation of lutein in vitro and in vivo: isolation and structural elucidation of lutein metabolites by HPLC and LC-MS (APCI). *Free Radic Biol Med.* 2008, 45, 982-993.
5. Project „Záchrana lužních stanovišť v povodí Morávky“. In: Projekt Záchrana lužních stanovišť v povodí Morávky [online]. (c) 2007 ZLSPM, [cit. 2013-12-01]. Dostupné z: <http://www.life-moravka.cz/>.
6. VILKHU, K., R. MAWSON, L. SIMONS a D BATES. Applications and opportunities for ultrasound assisted extraction in the food industry – A review. *Innovative food science & emerging technologies.* 2008, 9, 161 – 169.
7. YOU, X., Z. XU, W. PRINYAWIWATKUL, M. J. KING. Improving extraction of lutein from egg yolk using an ultrasound- assisted solvent method. *Journal of food science.* 2006, 71/4, 239 – 241.

Application of fuel cell technologies

Jiří Kazárik and Bohumil Horák

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jiri.kazarik, bohumil.horak}@vsb.cz

Abstract. The fuel cell is increasingly used as a stationary unit for the production of electricity. To his greater demand but also helps the automotive industry, where the fuel cell is increasingly starting to use as a primary or secondary source of electrical power for electric vehicles. In this expansion of the fuel cell occurs requirements the fuel cell, and especially his fuel consumption. As part of the Shell company organized a competition for the most fuel-efficient vehicle powered by a fuel cell as well as the combustion engine. This article describes how to upgrade a vehicle VŠB - TU Ostrava for the Shell Eco Marathon, and especially the fuel cell as a power source for the latest edition.

Keywords: fuel cell, hydrogen technologies, hydrogen

1 Introduction

A fuel cell is a transducer in which the chemical energy released during the oxide-reduction reactions and transformed in electrical energy, allows direct conversion of chemical energy bound in the fuel into electrical energy without the need for heat or mechanical transition (transformation) intermediate. This electricity can be used for example to power electric motors that can propel the vehicle. A fuel cell can imagine as electrolysis completely upside down. The one we bring hydrogen electrode (or hydrocarbon fuel), the second electrode oxygen or oxidizing agent, is taking place between the electrodes of a suitable water and electrolytes in the presence of a catalyst here is the chemical combining oxygen and hydrogen in the water, while the electrodes of an electrical voltage. The energy released by chemical reaction between oxygen and hydrogen. This reaction usually arises as heat, which can also be consumed. Moreover, practically there are no harmful emissions, only water vapor (using hydrocarbons also carbon dioxide).

A fuel cell has two electrodes, cathode and anode. The term electrode called the cathode, which takes place the reductive reaction (increase in negative charge of the chemical elements and compounds) and called the anode electrode deadline, which is an oxidation reaction (increase of positive charge of the chemical elements and com-

pounds). The reaction in the fuel cell cathode is the positively charged anode and the negative. In the backlash, electrolysis, is electrically negative cathode and anode electrically positive. Electrons flowing spontaneously, from the negative electrical pole to electrically positive pole.

The fuel cell supplying fuel and oxidizing gas directly to the anode and cathode. The physical structure of the fuel cell is such that the gases flowing through the channels on both sides of the electrolyte. The electrolyte is the basis for allocation of fuel cells for different types various different types of electrolytes lead ions. The principle is illustrated in Fig. 1.

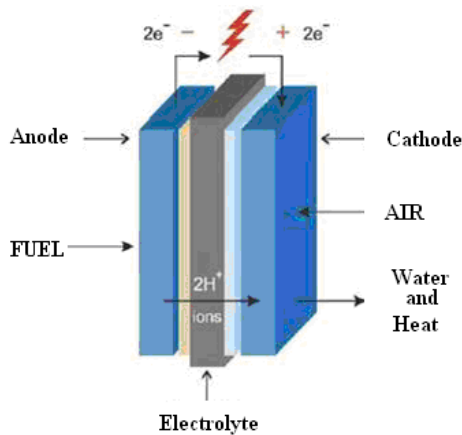


Fig. 1. General illustration features a hydrogen fuel cell

2 Fuel cell

The new version of the vehicle HydrogenIX for Shell Eco Marathon in the category prototype uses a new fuel cell companies Horizon. The fuel cell from the previous fuel cell performance and his different alternative involving all the component. The fuel cell needs to be able to operate an external power source, that source is used only for starting the fuel cell. After starting the fuel cell can be an external power supply to disconnect power and control electronics is already implemented in the fuel cell.

Fuel cell is made up of plate-like cells with air channels to allow the flow of air across the membrane. The membrane facilitates the flow of Hydrogen creating the release of electrons. Electrically conductive separator plates between each pair of cells enable the flow of electrons. The stack aspect is that they are all placed on top of each other and held together by epoxy endplates.

Numbers of cells	48
Rated power	1000W
Reactants	Hydrogen and air
Max stack temperature	65°C
Low voltage shut down	24V
Over current shut down	42A



Fig. 2. Fuel cell Horizon

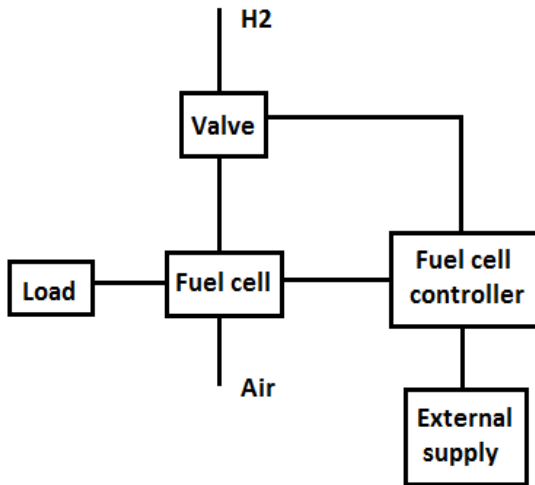


Fig. 3. – Block diagram

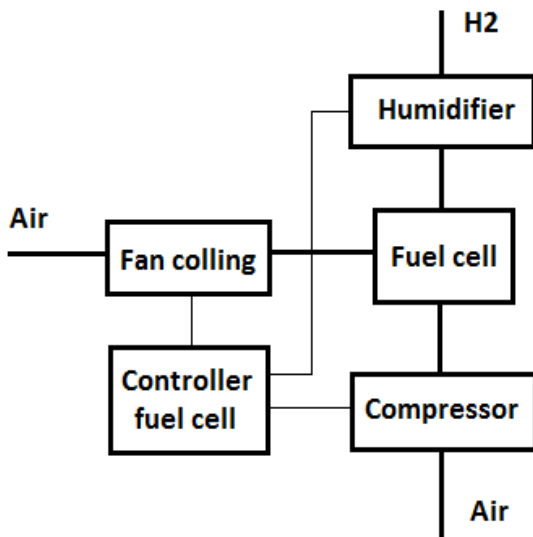


Fig. 4. – Block diagram Fuel cell

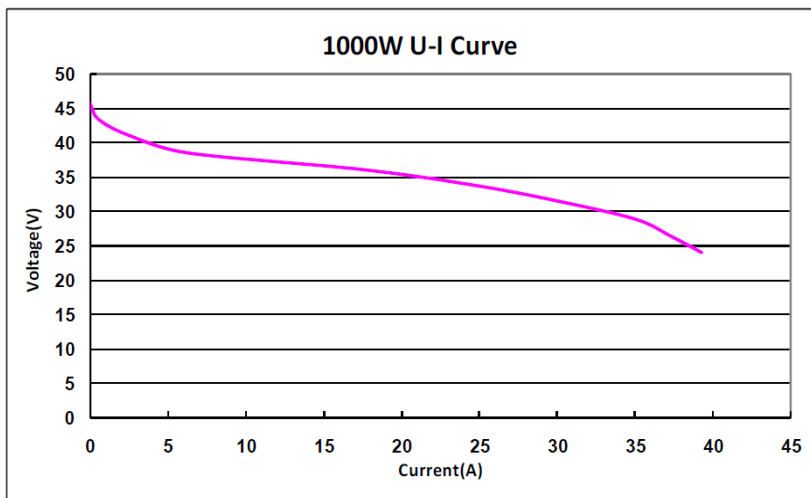


Fig. 5. U - I Curve

3 Practical verification in the car

The practical test and verify theories based on measurement and experimentation on the stationary unit in the laboratory is a laboratory prototype HydrogenIX. The prototype is powered by an electric motor and a generator of electricity is low-temperature PEM fuel cell.

HydrogenIX prototype also serves as a mobile laboratory, which allows measurements and running tests to drive changes in structure and changes in the structure of the software control algorithm fuel cell and also for the practical test and verify theories based on measurement and experimentation on the stationary unit in the laboratory located in the laboratory role.

4 Conclusion

The fuel cell was not tested in the laboratory role in measuring the volt-ampere characteristics and performance of all components and control electronics provided by the manufacturer. After mallow the fuel cell is connected via electronics developed which will be manually and automatically controlled by the individual components of the fuel cell. All measurements have been carried out in a hydrogen laboratory VŠB - TU Ostrava. With proper operation of the fuel cell with developed electronics, will be the fuel cell is connected to the inverter and the motor, and then moved into the vehicle HydrogenIX and optimized for racing Shell ECO Marathon.

5 Acknowledgement

This work was supported by project SP2014/188 “Control of technological systems with OAZE providing an independent sustainable development of complex system 2014” within the student’s grant competition (VŠB - TU Ostrava)

References

1. Horák B., Koziolek J., Kopřiva M., Papoušek M., Slanina Z.: Studie *pohonu mobilních pohonu s palivovým článkem*, Studie ČEA, Ostrava 2005, 203 s
2. *FuelCellStore.com* [online]. 2014 [cit. 2014-07-25]. Available z WWW: <http://www.fuelcellstore.com/>
3. *Horizonfuelcell.com* [online]. 2013 [cit. 2014-07-25]. Available z WWW: <http://www.horizonfuelcell.com/#lh-series-stacks/c52t>

Universal embedded platform for low-cost industrial image processing analysis

Michal Kelnar, Ondřej Horáček, and Miloslav Holba

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{michal.kelnar, ondrej.horacek, miloslav.holba}@vsb.cz

Abstract. Image processing is an increasingly common tool for a quality control of products in many industrial areas. Use of conventional approaches based on computers to implement image analysis in industrial quality control of products is sometimes inconvenient for many reasons. A more preferable solution then uses an embedded device with a built-in camera or camera chip and a connection to an industrial bus.

This article describes the design of an embedded device based on the universal development platform Raspberry Pi which cooperates with MODBUS - TCP. Control Web 6.1 from Czech Moravian Instruments is used for visualization of the results and other variables of the image processing application on PC.

Keywords: Raspberry Pi, MODBUS-TCP, Raspicam, Control Web

1 Introduction

Image analysis is becoming increasingly popular in quality management. This tool is used for eliminating subjectivity of visual evaluation of product quality by persons. At the same time it is used to increase the efficiency of quality control. The use of industrial cameras such as Basler cameras and image processing on an ordinary computer or industrial computer or on conventional graphics card is nowadays very common. This solution enables rapid development of quality control chain using standardized components for the application. Furthermore, in combination with the software applications developed for example on LabVIEW, the development of a complete inspection station is very fast, but the price of such equipment is rather high.

Another variant of image analysis applications in industrial quality control is the use of an embedded device that is equipped with a camera chip and the necessary peripherals. The hardware portion of the device may then be made of the available development kit, such as i.MX, Raspberry or other, usually MCUs based on the ARM core. Such a device is mainly characterized by very low production cost, but on the other hand, more complicated development of software applications. It is the complicated software application development for embedded devices that in many cases

exceeds the cost of development of a prototype in a conventional way. In case of a significantly larger number of devices the embedded device is definitely more convenient.

This article deals with the development of a universal embedded device for use of image analysis in industrial quality control. The hardware platform is based on the ultra-low cost development platform Raspberry Pi, to which the camera chip and optics are available, and the use of visualization in PC to connect to this device.

Visualization program is based on the system Control Web 6.1. This is inexpensive software for a rapid development of industrial applications, developed by Czech Moravian Instruments. Data transfer between the visualization and the device is handled by the industrial protocol MODBUS - TCP. The communication protocol is implemented at the application layer. This open protocol was originally developed for programmable logic controllers (PLC), but its simplicity and ease of implementation makes it a perfect tool for moving data between the visualization and the platform Raspberry Pi.

2 Hardware concept

The hardware unit consists of the development kit Raspberry Pi, a camera module Raspberry - RPI Board Camera, DC/DC 5V power supply voltage for Raspberry, lighting and power switch trigger input circuit.

The base of Raspberry Pi is a microcontroller BROADCOM BCM2835 (ARM11) clocked at 700MHz, with 32b wide bus. The working memory version of Raspberry Pi B is 512MB. With the price of \$ 35 it is a very cheap and powerful microcomputer. RPI Board Camera is equipped with a 5MP camera chip Omnivision 5647 with maximal resolution of 2592x1944 pixels. The supported video formats are 1080p 30fps, 720p 60fps and for 60/90fps recording 640x480p. Connection of the camera module with Raspberry is done through standard interfaces CSI on 15 pin Ribbon Cable. (1)

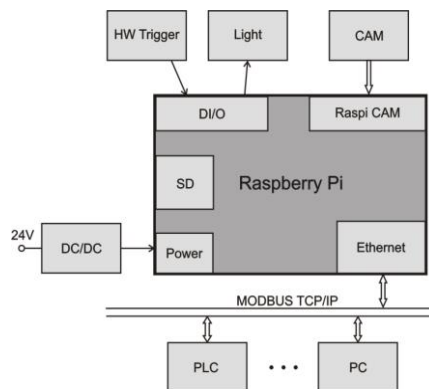


Fig. 1. Embedded HW idea schema

3 Application structure

Raspberry Pi is equipped with the operating system Raspbian, which is based on Debian optimized for hardware Raspberry Pi. (2)

The software application uses several software modules to provide all necessary functionalities. These modules are mainly:

- libmodbus – library for receiving and transmitting data according to the Modbus protocol. This library supports serial communication (RTU) and TCP communication (Ethernet). (3)
- libraspicam – driver for the camera module.
- Boost - a set of cross-platform libraries to support the creation of portable applications and increase the efficiency of application development.
- vsftpd - very fast, stable and secure FTP server.(4)
- SSH – for development, testing and application configuration on the device.

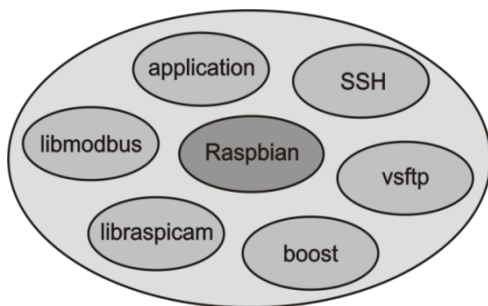


Fig. 2. SW parts

Design and implementation of the applications are based on the structure of the standard requirements of general application that implements image analysis. Basically, beside the image itself the input parameters are entered into the image analysis algorithms and the output results are provided. Very often it is also necessary to store the changed parameters. Similarly, the results of the changes are often stored. Another output of image analysis algorithms is errors. Errors mean wrong configuration, its unavailability or other errors arising from the image processing algorithms. Of course, a record of errors is provided, too. The whole application is developed in C++ under Visual Studio 2012 and VisualGDB.

A record of all instances of classes inherited from abstract class ILoggable is processed in a separate thread. Processing logging in this thread is not dependent on a specific class, because it generally operates with the base class ILoggable.

Operation communication via Modbus / TCP is implemented in another - communication thread. This thread handles all sessions on port 502. Data are synchronized with the communication by exclusive access.

By connecting the client to the port the data synchronization starts. First of all, the current application data are read and converted into a table of registers, accessible

through library libmodbus. In case of a request for change of data from the client the data are subsequently parsed from the registers back to the application data objects.

4 Modbus TCP/IP

MODBUS is a standard communication protocol used in industrial buses. It was developed by Modicon (now Schneider Electric) already in 1979. It is an open protocol at the application layer ISO / OSI model, originally designed for programmable logic controllers (PLC - programmable logic controllers), which allows the transfer of data over different buses and networks.

The report on the application level MODBUS defines MBAP header (MODBUS Application Protocol header). (5)

Table 1. MODBUS Protocol

MBAP	FUNCTION CODE	DATA
------	---------------	------

It consists of four parts.

Table 2. MODBUS Application Protocol Header

Transaction ID	Protocol ID	Length	Slave ID
----------------	-------------	--------	----------

The function code indicates what operation a server should perform. It is a one byte number. There are three groups of function codes defined: public function codes, user-defined function codes and reserved function codes.

The public function codes are clearly defined and uniqueness is guaranteed. They are public and approved by Modbus-Ida.org.

The user defined codes allow to implement the undefined specifications. Uniqueness is not guaranteed and for different devices is different, but after negotiation they can be assigned to the public. The reserved function codes are currently used by some companies and firms and are not available for public use.

ID Transaction is set by the client. It is a unique double-byte number that identifies the client's request. They may not come in the same order as they were sent. The server response therefore has the same ID and the client is able to recognize which demand the response belongs to.

ID Protocol is a Double-byte number adjusted by the client. For MODBUS TCP / IP it is always 0.

The length in two bytes number defines how many bytes will follow.

ID of the client device is a single-byte address used to identify the client by server. Some devices require this, some do not. (6)

5 MODBUS and Control Web 6.1

Control Web is software for developing visualization and control technology. It allows you to inexpensively implement various industrial applications. Control Web is designed to run on 32 or 64 bit Windows operating systems on standard PCs.

Drivers are used for communication with input-output devices. They communicate with the system via channels. Channels represent the data variables to be read or written depending on the type of channel.

The driver configuration is stored in two files, a parametric and mapping file. Both are generated automatically, a parametric file must be set in the inspector. It is run directly from the Control Web. Driver for MODBUS TCP / IP is not in the standard set of drivers supplied together with Control web, it has to be bought separately and installed. The configuration proceeds on in the inspector of the drivers. The main configuration is in the Channels section.

Table 3. MODBUS control web configuration

<i>IP ADDRESS</i>	<i>FIRST CHAN.</i>	<i>LAST CHAN.</i>	<i>DATA FIELD</i>	<i>OFFSET</i>	<i>BIDIRECT</i>	<i>CHANNEL DIRECTION</i>
1@192.168.8.10	100	1099	4X	1	No	input
1@192.168.8.10	1100	2099	4X	1001	Yes	bidirectional
1@192.168.8.10	2100	2599	4X	2001	No	input
1@192.168.8.10	2600	3099	4X	2501	yes	bidirectional

IP address of the server is written in the form 1@192.168.8.10. The first number indicates the client ID, which is transmitted in the message. If the MODBUS run on a port different from the standard definition 502, the IP address is written in the form 1@192.168.8.10: 503, where the number after the colon indicates the port number.

The data area 4X indicates holding registers. The first and the last channels indicate the channel range which the driver will write into. The driver transmits information about the state of communication through specific channels that are mapped to the range of 1 to 99. If a communication error occurs, an exception is generated and the drivers of each channel can be determined by a variety of data, which specify an error or an event. Channel No. 1 distinguishes the case of an error while reading or writing, or if TCP / IP connection to the station has been established. Group of channels No.10 to 18 ads data error while reading resp. 20 to 28 at enrolment.

The offset refers to the first register, from which to read or write the first channel.

There are three directions of the channel input (read only), output (write only) and bidirectional (read and write).

In order to communicate with Raspberry-PI, Control Web 6.1 application was created, enabling imaging and data storage. Application is made as an application library that will be repeatedly inserted into a project, depending on the number of devices.

6 Conclusion

Integration of image analysis to embedded devices in a non-standard use of quality management is becoming an increasingly common tool for eliminating the subjectivity of visual assessment of product quality carried out by persons, while increasing the efficiency of quality control. The basic idea of creating a default development platform is based upon the use of the low-cost hardware platform Raspberry Pi. Together with an optimized operating system for the hardware platform Raspbian and a created development platform it represents a relatively powerful and extensible foundation for further development of the applications processing visual information.

For the device control and visualization of the analyzed data an application using the programming system Control Web 6.1 was created. The application was developed as an application library, which will be inserted into a new visualization and will fully operate a particular device. Depending on the number of devices in the industrial application it can be inserted in the project library many times. This provides a modular tool for numerous variations of application and significantly reduces the time required to create complete applications. Communication between the visualization and the device is via Modbus TCP / IP, which was implanted to the platform Raspberry at the application layer.

Acknowledgements

This work was supported by contract research HS45032, Innovation of perforating belt material with parameters detection based on the principles of image signal processing, VSB-TU Ostrava. Special thanks to our supervisors, Jan Žídek, Jiří Koziorek and Petr Bilík.

References

1. Raspberry Pi [online]. [cit. 2014-07-29]. Available from: <http://www.raspberrypi.org/>
2. Raspbian [online]. [cit. 2014-07-29]. Available from: <http://www.raspbian.org/>
3. A Modbus library for Linux, Mac OS X, FreeBSD, QNX and Win32 [online]. [cit. 2014-07-29]. Available from: <http://libmodbus.org/documentation/>
4. Probably the most secure and fastest FTP server for UNIX-like system [online]. [cit. 2014-07-29]. Available from: <https://security.appspot.com/vsftpd.html/>
5. RONEŠOVÁ, Andrea. ZÁPADOČESKÁ UNIVERZITA. Protocol brief MODBUS [online]. [cit. 2014-07-29]. Available from: home.zcu.cz/~ronesova/bastl/files/modbus.pdf/
6. MODICON. Modbus Protocol Reference Guide. June 1996. [online]. [cit. 2014-07-29]. Available from: http://web.eecs.umich.edu/~modbus/documents/PI_MBUS_300.pdf

Interactive Rehabilitation System for Balance Exercises

Jana Krohova and Maros Smondrk

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jana.krohova, maros.smondrk}@vsb.cz

Abstract. This paper describes the design and implementation of an interactive rehabilitation system for balance exercises. The balance exercises are used to restore the function of the locomotors system after previous injury or disease. The quality of rehabilitation and its progression is usually evaluated only subjectively by physiotherapist. The proposed rehabilitation system is used for objective evaluation of the quality of rehabilitation.

Keywords: accelerometers, Bluetooth 4.0, visual feedback, medical application, telemetry

1 Introduction

The balance exercises are nowadays used in many rehabilitation centres. The benefit of this exercise has been demonstrated in a number of significant research studies [1], [2]. Balance exercises are used to restore functional states after the previous musculo-skeletal injury or disease of, for example, the knee (functional training along with the balance exercise counselling significantly improves the condition of people with a plastic of the knee joint, especially the walking speed, standing on the one leg and synostosis [3]), the ankle, the hip, the spine and the shoulder. These exercises are also used for the rehabilitation of hyperkinetic and postural disorders in children. [1], [2]

There are many types of rehabilitation tools for balance exercises, for example: large inflatable balls, small soft balls, wobble boards, the Posturomed platform, balance steps, balance cylindrical segments, stabilizer plates, etc [4], [5]. In our work we used the wobble board.

2 Problem Definition

In general, the rehabilitation centres use the wobble board as a rehabilitation balance tool to practise basic exercises under the supervision of the physiotherapist. The evaluation of this exercise is only subjective and depends on the experience of the physiotherapist. The monitoring and evaluation of the patient's progress and balance exercise is very difficult to realize under these terms and conditions. The system for interactive rehabilitation balance exercises on the wobble board has been designed and tested. The main aim of this work was to design and implement a rehabilitation tool

which will be able to evaluate objectively the progress of rehabilitation and display both qualitative and quantitative information about the performed exercise.

3 System Description

The proposed rehabilitation system for balance exercises consists of the wireless telemetric wobble board and a computer (Fig. 1).

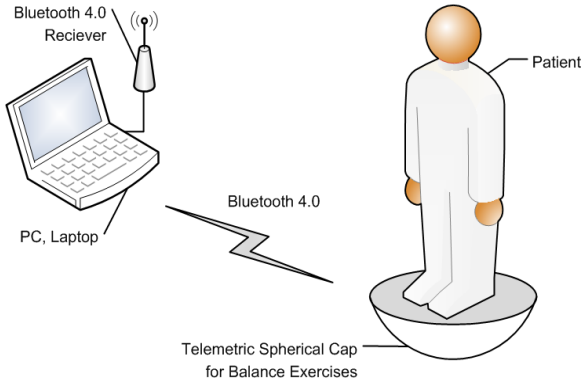


Fig. 1. System features.

The rehabilitation system for balance exercises and the computer communicate through the Bluetooth 4.0 protocol. The Bluetooth 4.0 wireless communication is a very important part of this proposed device. This communication protocol was chosen on the basis of the low power and low latency capabilities and for the easy retrieval of data from the telemetric wobble board during the rehabilitation exercises without limiting the movements of the patient. [6]

The most important part of the whole system is the wooden wobble board with an integrated telemetry module for measuring angular deflection. The measured angular deflection gives objective, qualitative and quantitative information about the performance of the rehabilitation exercise and the patient's progress.

3.1 Hardware

The telemetric wobble board consists of two parts: the wobble board and the measuring part with its own power supply and switch. The measuring part is located in the exact geometric centre of the wobble board.

The commercially available Bluetooth cB-OLP425 platform was used as the measurement part. This platform was selected due to its low energy consumption and the new wireless communication technology Bluetooth 4.0. It is based on the CC2540, which is a system on chip solution from Texas Instruments. It combines a 2.4 GHz transceiver, a micro-controller, a 256 kB of in-system programmable flash memory, a 8 kB of RAM and peripherals. In addition to the peripherals integrated in the CC2540

system on chip solution, the cB-OLP425 includes a temperature sensor, a red and a green LED and the 3-axis digital accelerometer LIS3DH from ST Microelectronics. The communication between the CC2540 and peripheral sensors is implemented using I²C bus. [7], [8], [9]

The LIS3DH has dynamic user selectable full scales of ± 2 g/ ± 4 g/ ± 8 g/ ± 16 g, it is capable of measuring accelerations with output data rates from 1 Hz to 5 kHz and it provides an embedded high-pass filtering with cut-off frequency 0.2 Hz which delete the DC component of the measured acceleration [9], [10]. The sampling frequency was set to 10 Hz at ± 2 g in high resolution mode with a precision of 12 mg/digit. These values were empirically determined and consulted with physiotherapists from the University Hospital of Ostrava. The output value of acceleration is converted to the angular deflection (Fig. 2) for each axis according to the following formulas [11]:

$$\alpha = \tan^{-1}(g_x/\sqrt{g_y^2 + g_z^2}) \quad (1)$$

$$\beta = \tan^{-1}(g_y/\sqrt{g_x^2 + g_z^2}) \quad (2)$$

$$\gamma = \tan^{-1}(\sqrt{g_x^2 + g_y^2}/g_z) \quad (3)$$

where g_z , g_y , g_x are the Cartesian components of acceleration g , which are projections of acceleration g onto axes x_0 , y_0 , and z_0 , respectively. α , β , γ are angles included between acceleration g and axes x_0 , y_0 , and z_0 . [11]

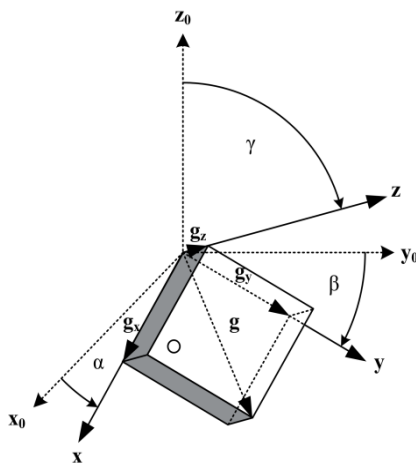


Fig. 2. The Cartesian components of acceleration g and angles between acceleration g and axes x_0 , y_0 , and z_0 . [11], [12]

4 Software

The developed software consists of two separate parts. The first part represents the firmware of the measuring segment for the acquisition of accelerometric data and

ensures wireless communication with the USB dongle plugged into the computer. The second part is the rehabilitation software used by physiotherapists and patients. Its main function is to provide real-time graphical representation of the patient angular displacement or the wobble board movement during the rehabilitation process. The designed rehabilitation software was created in the development environment MATLAB. It has three main functions: the mutual connection establishment with the measuring part, the acquisition of measured data and their graphical visualisation and saving the measured data into the database.

5 Preliminary testing of the proposed system

The proposed rehabilitation system is tested at the Therapeutic Rehabilitation Clinic of the University Hospital of Ostrava. In order to demonstrate the potential of our proposed system we made preliminary measuring. Three volunteers at the age of 25 to 29 participated in the experiment. The first volunteer was healthy. The second one had the plastic surgical procedure of the anterior cruciate ligament one month ago and the plastic surgical procedure of the meniscus of the left knee one year ago. The third volunteer had the plastic surgical procedure of the anterior cruciate ligament and meniscus of the left knee two years ago. He also suffered from the chondromalacia of the cartilage and muscle imbalance of the left leg. The volunteers practised in a relaxed sitting position for each leg. Their main task was to achieve the maximum possible deflection in each axis and stay in the white tolerance area. The tolerance was set to 5° . The experiment was performed three times for each leg and volunteer.

For an objective evaluation of the performed exercise we compute the error, which is the ratio between the number of samples which were inside the tolerance area and the number of samples which were outside. The healthy volunteer achieved an average error equal to 0.43 % by the left leg (LL) and 0.57 % by the right leg (RL), respectively. The second volunteer reached an average error equal to 3.46 % by the LL and 1.53 % by the RL. The last one reached an average error equal to 2.21 % by the LL and 0.98 % by the RL. From the above measurements it can be concluded that the post-surgical volunteers achieved a higher error by the LL than by the RL. It is important to note that the error is affected by the diagnosis, physical training, etc. In Fig. 3 the comparison of the measured deflection in the x -axis of the first and third volunteer is shown.

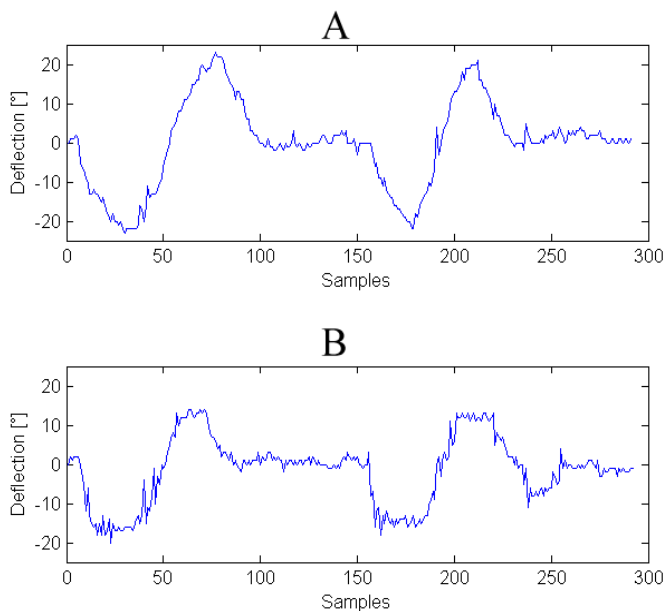


Fig. 3. The measured deflection in the x -axis of the first volunteer (A) and third volunteer (B). The horizontal axis - samples; the vertical axis - deflection in angular degrees.

Another significant feature can be seen there. Volunteer no. 3 suffers from the imbalance of the LL, among other things. In general, the imbalance causes worse muscle coordination and impairs the articulated mobility. This can be easily seen in the Fig. 3B as small oscillations. The imbalance can be treated by long-term rehabilitation which can be subsequently evaluated by the proposed system.

6 Conclusion

The proposed interactive system for the balance exercise on the wobble board allows the wireless data transfer to the computer and a real-time visualisation of the patient's position during the rehabilitation and allows the patient to perform rehabilitation at home on his own. The implementation of hardware and software in the rehabilitation process makes the rehabilitation process more attractive for the patients, mainly for children and more objective for the physiotherapists. The measured data can be saved for further processing in order to evaluate the patient's rehabilitation progress. The long-term measurement of the angular displacement provides the quantitative and qualitative information about the performance of the balance exercise.

7 Acknowledgment

The work and the contributions were supported by the project SV SP 2012/2512 „Biomedical engineering systems XI“ and TACR TA01010632 “SCADA system for the control and measurement of process in real time”, also supported by the project MSM6198910027 “Consuming Computer Simulation and Optimization”. This paper has been also elaborated in the framework of the project Opportunity for young researchers, reg. no. CZ.1.07/2.3.00/30.0016, supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic.

8 References

1. D. S. Kaesler, R. B. Mellifont, P. Swete Kelly and D. R. Taaffe, “A novel balance exercise program for postural stability in older adults: A pilot study”, *Journal of Bodywork and Movement Therapies*, 11(1), p.37-43, 2007.
2. C. Liebenson, “Better balance exercises”, *Journal of Bodywork and Movement Therapies*, 9(2), pp.148-149, 2005.
3. S. R. Piva, et al., “A balance exercise program appears to improve function for patients with total knee arthroplasty: a randomized clinical trial”, *Physical Therapy*, 90(6), p.880-894, 2010.
4. R. Jebavy, T. Zumr, *Strengthening with the balance tools*. 1st edition, Praha: Grada, 2009.
5. K. Kriegelová, M. Černý, “Software tool for direction deflection recognition based on patient's postural dysfunction,” *Electronics and Information Engineering (ICEIE)*, 2010 International Conference On , vol.1, no., pp.V1-295,V1-299, 1-3 Aug. 2010
6. Bluetooth, S. I. G., 2010. Bluetooth Core Specification v4.0. 20 January 2013, [Online]. Available: <https://www.bluetooth.org/Technical/Specifications/adopted.htm>
7. Connectblue, 2012. cB-OLP425 cB-OLS425 cB-OLS426 Electrical Mechanical Data Sheet. 18 December 2012, [Online]. Available: <http://support.connectblue.com/display/PRODBTSPA/cB-OLP425+cB-OLS425+cB-OLS426+Electrical+Mechanical+Data+Sheet>
8. Connectblue, 2012. cB-OLP425 Development Kit Getting started. 18 December 2012, [Online]. Available: <http://support.connectblue.com/display/PRODBTPLA/cB-OLP425+Development+Kit+Getting+started>
9. STMicroelectronics, 2010. LIS3DH - MEMS digital output motion sensor ultra low-power high performance 3-axes “nano” accelerometer. 18 December 2012, [Online]. Available: <http://www.st.com/web/en/resource/technical/document/datasheet/CD00274221.pdf>
10. STMicroelectronics, 2010. Application note LIS3DH - MEMS digital output motion sensor ultra low-power high performance 3-axes “nano” accelerometer. 18 December 2012, [Online]. Available: http://www.st.com/web/en/resource/technical/document/application_note/CD00290365.pdf
11. Ch. J. Fisher, *Using an Accelerometer for Inclination Sensing*. 18 December 2012, [Online]. Available: http://www.analog.com/static/imported-files/application_notes/AN-1057.pdf
12. S. Luczak, W. Oleksiuk, M. Bodnick, “Sensing Tilt with MEMS Accelerometers”, *Sensors Journal*, IEEE , vol.6, no.6, pp.1669-1675, 2006.

WiFi Localization of Exploration Vehicle Based on Multilateration

Ales Kurecka

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

Abstract. This work deals with the extension of the localization system in an existing mobile exploration vehicle with global indoor localization. The presented solution uses existing WiFi infrastructure. Resulting localization module evaluates approximate position of the vehicle based on measured signal strength (RSSI) from visible WiFi Access Points and their location. The Path-Lost model of signal propagation and the multilateration technique is used to achieve this goal.

Keywords: WiFi, RSSI, localization, multilateration, Gaussian filter

1 Introduction

Localization is one of the most fundamental problems associated with the control of autonomous mobile robots. These robots are usually equipped with a number of different sensors that can be used to localize. One of the basic ways of tracking robot position is odometry. That include incremental or inertial sensors. For accurate localization is possible to use rangefinders such as laser sensors. Methods based on these sensors allow the relative localization referenced to eg. the starting position of the robot. These techniques are known as position tracking [1].

In many applications it is necessary to know or determine the absolute position. Some localization methods such as Monte Carlo localization (MCL), are able to determine the absolute position of the management of multiple hypotheses, which after a few iterations outweigh the only one. Complications can occur due to the environment symmetry (eg. several identical floors) and the absolute positioning can converge very slowly. As a useful complement these localization systems is the use of wireless networks [2],[3].

In outdoor areas can take advantage of the ubiquitous GPS. The GPS works on frequencies in the band 1176.45 MHz (L5) to 1841.40 MHz (L4) (so in the UHF band) and is limited to line of sight. In buildings or heavily wooded areas is GPS usually useless as a result of lack of signal.

For indoor environments the WiFi networks seems as a suitable solution. These are already very widespread and in order to convey a quality connection APs

installation is dense. It can be expected that this situation will continue to improve. Due to the fact that the planned deployment (i.e. expansion of the existing continuous positioning system with global localization) is primarily to exclude hypotheses and outweigh the actual position, the demands on the precision of the localization of low.

This work uses localization with WiFi.

2 New solution

Two popular methods for indoor localization utilized WiFi networks is used: K-Nearest neighbour algorithm based RF fingerprinting and multilateration methods. For deployment fingerprinting methods are required signal intensity maps of the planned deployment, that must be known apriori. That can be obtained with simulations or measurements. Multilateration method is based on the position estimation based on distances to visible APs [4].

Fingerprinting can achieve much greater accuracy than multilateration, but it is also more difficult to input data. To deploy multilateration only knowledge of the location of network nodes is needed.

For the actual location is especially important to recalculate the value of the RSSI on the distance from the transmitter. Because the signal is affected by losses, needs to be filtered before processing. Filtration is carried out with Gaussian filter.

2.1 The empirical model

For barrier-free environment, where we expect a smooth sloping curve, can be used to predict losses spread basic empirical model [5],[6].

$$L(d) = L_1(d_1) + 10 \cdot \gamma \cdot \log \frac{d}{d_1}, \quad d \geq d_1, \quad (1)$$

Where $L(d)$ is a path lost (dB), γ is a propagation exponent (-), d is a distance between TX and RX antenna (m), d_1 is a reference distance (m) and L_1 is a reference attenuation on d_1 (dB).

By modifying the basic empirical model can get Lost-Path model, which further expands upon the relationship of the antenna gain, loss and slow the spread of losses:

$$r = t - l_0 - 10 \cdot \gamma \cdot \log \frac{d}{d_1} - S, \quad d \geq d_1, \quad (2)$$

Where r is a received signal power (dBm), t is a transmit signal power (dBm), l_0 is a path lost on d_1 (dBm), γ is a propagation exponent (-) and S is a log-normal shadow fading (dB).

Lost-path model is the basis for this work.

2.2 Multilateration

Multilateration is a mathematical method, which based on distances from known reference nodes and evaluate position of unknown node [fig. 1].

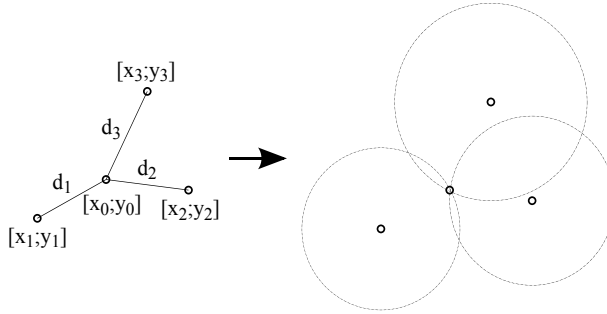


Fig. 1. Principle of multilateration

A special case of multilateration is trilateration. If we know the position of three nodes and their distances to the unknown node, we can calculate its position. The situation can be described analytically by a system of equations (analytical writing of circles).

After the generalization of the system for n-nodes and modify to the matrix form, the relationship can be obtained, from which we can directly calculate the coordinates of the unknown node x [[7]]:

$$\mathbf{H} \cdot \bar{x} = \bar{b} \tag{3}$$

where:

$$\mathbf{H} = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \\ \dots & \dots \\ x_n - x_1 & y_n - y_1 \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \quad \bar{b} = \frac{1}{2} \cdot \begin{bmatrix} (x_2^2 + y_2^2 - d_2^2)(x_1^2 + y_1^2 - d_1^2) \\ (x_3^2 + y_3^2 - d_3^2)(x_1^2 + y_1^2 - d_1^2) \\ \dots \\ (x_n^2 + y_n^2 - d_n^2)(x_1^2 + y_1^2 - d_1^2) \end{bmatrix} \tag{4}$$

In practice, the error bound distances. This error causes the system of equations is incompatible and $x \notin R(\mathbf{H})$. The circles may not intersect or intersect at multiple points and therefore this simple calculation can not be used. This problem is known as noise away and can be addressed by using some approximation methods. In this work we used the method of least squares. The general formula for calculating multilateration is the following:

$$\bar{x} = (\mathbf{H}^T \cdot \mathbf{H})^{-1} \mathbf{H}^T \cdot \bar{b} \tag{5}$$

2.3 Gaussian filter

The measured RSSI value - signal loss from each AP is very volatile and time variant. These changes are called fast signal losses and come from various sources. Error due to rapid losses can be very large, the order of units to tens of *dBm*. In order to evaluate the RSSI at all, it is necessary to filter out those losses. One of the methods suitable for the filtration of these signals is Gaussian filter [7] and for discrete areas has the form

$$(f \times g)(k) = \sum_{i=0}^k f(i) \cdot g(k-i), \quad g = \frac{1}{\sigma \cdot \sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \quad (6)$$

where σ is the standard deviation and μ is the mean value (Gaussian function maximum).

For practical application of Gaussian filter is used only a limited number of samples of Gaussian functions. Thus the sampled Gaussian function makes a vector, called the kernel. To apply a filter, must be kept the history of the filtered values and history depth must be equal (or greater) than number of kernel elements. For each sample, history is updated, multiplied by the result obtained by the kernel and summed. It is important to choose the correct value of σ , which affects the quality of filtering, but also introduces delay and inertia. When $\sigma \rightarrow \infty$ Gaussian filter is passed into moving average.

3 Implementation

The implemented localization module is written in C++. The entire process of locating shows the chart [fig. 2].

The WiFi module periodically scans visible APs. Measured RSSI and MAC address pairs are periodically read from the WiFi module. The list of pairs is forwarded to the thread that the stream expands by information from the map. This information is a coordinates, antenna gain, path loss and propagation exponent. The RSSI is filtered by Gaussian filter. The result of this step is a list of AP items, which have position, filtered RSSI value and other necessary data to calculate multilateration.

Multilateration first sorts the items according to signal strength and then remove items with a very weak signal. These APs are typically far and the signal is strongly affected by the fraud and bring it calculation error, which is too large. In addition, they can be after the Fresnel zone. Matrix H and B are created from the remaining items. The calculation is then performed using the multilateration formula listed above.

4 Testing

4.1 Evaluation of Gaussian filter contribution

Test of Gaussian filter influence is based on data from measurements with HP iPAQ hx4700. AP was used as the ZyXEL P660HW-T3 device. The [fig. 3a)]

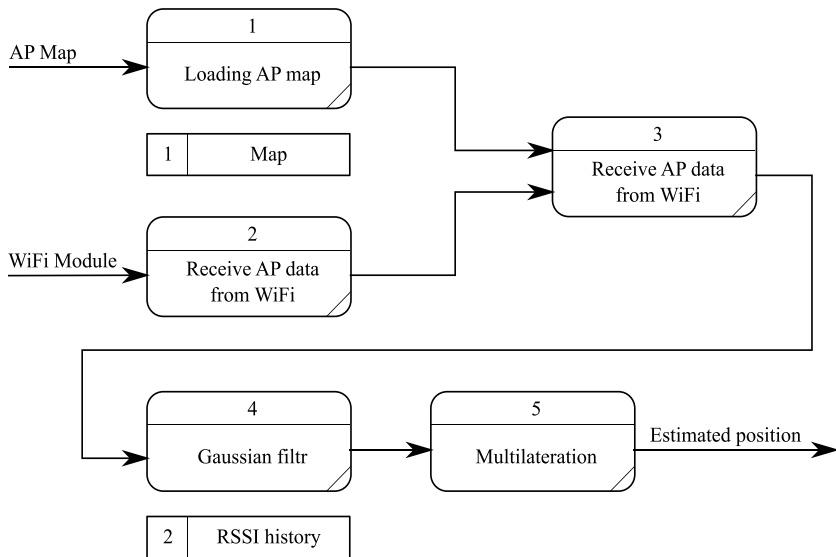


Fig. 2. Location data flow diagram

shows the unfiltered samples and filtered samples with Gaussian filter with different sizes and sigma parameter of kernel.

The kernel size of 20 is seen that the suppression of leakage is not fast enough, while at 200 the other hand, system response is too slow. As a pretty good compromise seems to be size of kernel as one hundredth.

4.2 Localization accuracy

To verify the accuracy, the scheme has been formulated with 4 transmitter (AP). The situation demonstrates [fig. 3b)]. The black dashed line is the actual trajectory. The blue trajectory is localized and gray points indicate the location, which represent locations without filter.

The figure shows that the localized path quite successfully approximates the actual trajectory. It is also evident influence of Gaussian filter. Location without Gaussian filter is practically useless, and burdened with very large deviations. Moreover, localized position abruptly changes with each scan. Figure also demonstrates the influence of distance on the accuracy of localization. If the AP is closer, more accurate results can be expected.

It should be noted that these are ideal conditions. Not calculated with multipath propagations or attenuation caused by obstacles.

5 Conclusion

This paper presents a global localization algorithm used WiFi networks based on multilateration. Multilateration accuracy is affected by the quality of the signal

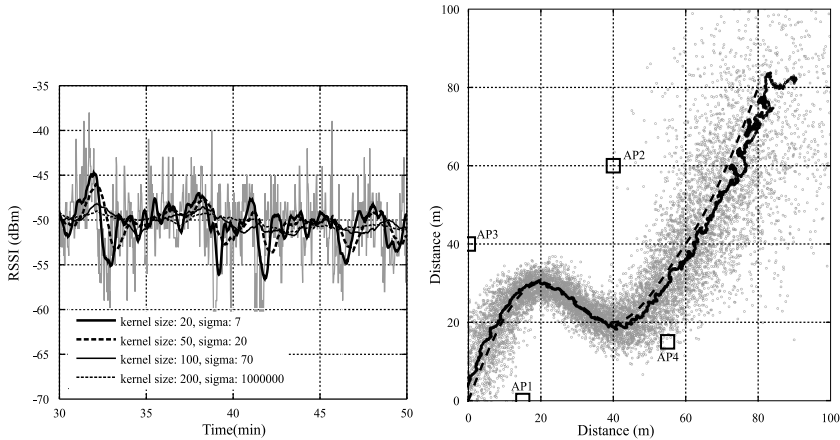


Fig. 3. a) Application of Gaussian filter to the measured data, b) Localization with 4 APs

propagation model. For this work, the basic empirical model is used. Expected accuracy is in meters, which is sufficient.

References

1. Y. Liu, Z. Yang, *Location, Localization, and Localizability: Location-awareness; Technology for Wireless Networks* (2010) Springer, 162, ISBN 1441973710, 9781441973719
2. S. Thrun, W. Burgard, D. Fox, *Probabilistic Robotics* (2005) Mit Press, 647, ISBN 0262201623, 9780262201629
3. F. Dellaert, D. Fox, W. Burgard, S. Thrun, *Monte Carlo localization for mobile robots, Robotics and Automation* (1999). 1999 IEEE International Conference on, vol. 2, no., pp.1322,1328 vol.2, 1999, doi: 10.1109/ROBOT.1999.772544
4. Alkasi, U.; Al Shayokh, M.; Partal, H.P., *An experimental comparison study on indoor localization: RF fingerprinting and Multilateration methods, Electronics, Computer and Computation (ICECCO), 2013 International Conference on*, vol., no., pp.255,259, 7-9 Nov. 2013
5. Y. Mostofi, A. Gonzalez-Ruiz, A. Gaffarkhah, D. Li, *Characterization and modeling of wireless channels for networked robotic and control systems - A comprehensive overview* (2009) 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, art. no. 5354765, pp. 4849-4854.
6. S. Varzandian, H. Zakeri, S. Ozgol, *Locating WiFi access points in indoor environments using non-monotonic signal propagation model; Control Conference (ASCC), 2013 9th Asian*, doi: 10.1109/ASCC.2013.6606157
7. Y. Zhang, L. Bao., S.-H. Yang, M. Welling, D. Wu., *Localization algorithms for wireless sensor retrieval; Computer Journal*, 2010, 53 (10), pp. 1594-1605.

The Bioelectrical Impedance Spectroscopy and Electrical Impedance Tomography Measurement Device

Marketa Michalikova

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

Abstract. This paper presents the design and construction of the device intended for measuring bioelectrical impedance spectroscopy with extension for electrical impedance tomography measurement. The device is based on The Freescale Tower system with S08MM Freescale micro-controller. Method used to measure bioimpedance is Magnitude Ratio and Phase Difference Detection (MRPDD) method. The measurement is loaded with the systematic error, therefore the calibration is needed, particularly the Quadratic Interpolation Self-calibration algorithm was used for calculation of the measured load. The relative error of impedance measurement after calibration ranges from -4.52 % to 5.98 % with mean value of -0.02 %.

Keywords: Bioelectrical impedance spectroscopy, Electrical impedance tomography, Four-electrode configuration, MRPDD method

1 Introduction

Nowadays, the bioelectrical impedance analysis (BIA) is used in a wide range of diagnostic methods, for example to estimate the body composition, especially the body fluid volume and percentage of body fat as used in [1], also to measure various cardiac parameters with the bioelectrical impedance plethysmography method [2],[3], in vivo blood characterization e.g. hematocrit, level of glucose and hydration [4] and other methods like bioimpedance spectroscopy for tissue characterization or electrical impedance tomography [3]. BIA can be also used to predict the risk in hemorrhagic dengue fever [5].

Some appliances use only single frequency method (mostly 50 kHz) to determine total body water (TBW) and fat-free mass (FFM), but this method cannot give parameters characterizing intracellular water (ICW) and extracellular water (ECW). Therefore the multifrequency BIA is used to obtain these parameters of the body composition. Bioelectrical impedance spectroscopy (BIS) uses measurement in the whole given frequency range, so the tissue characterization can be made [6].

In this paper the design of the device providing measurement of both bioelectrical impedance spectroscopy and electrical impedance tomography in four-electrode configuration with frequency range from 1 kHz to 1 MHz is described. The Magnitude Ratio and Phase Difference Detection method (MRPDD) is used for measurement of the bioelectrical impedance as described in [7].

2 Methods and the device description

Magnitude Ratio and Phase Difference Detection method (as described in [7]) is based on comparison of two signals - one measured from unknown impedance Z_x , the second one is measured from reference resistor R_s connected in series with unknown impedance. Current sine-wave excitation signal of constant amplitude $100 \mu A_{PP}$ is applied by two electrodes to Z_x and appropriate voltage drop is measured with two measurement electrodes. The device is based on Freescale Tower modular system. The module with S08MM microcontroller controls two other module boards. The first board contains measurement circuit based on MRPDD method, on the second module board there are four multiplexers enabling multiplexing of sixteen electrodes. The electrodes are connected to the second module board via a connector. The device combines two methods of bioelectrical measurement - bioelectrical impedance spectroscopy and electrical impedance tomography. The device is captured in figure 1.

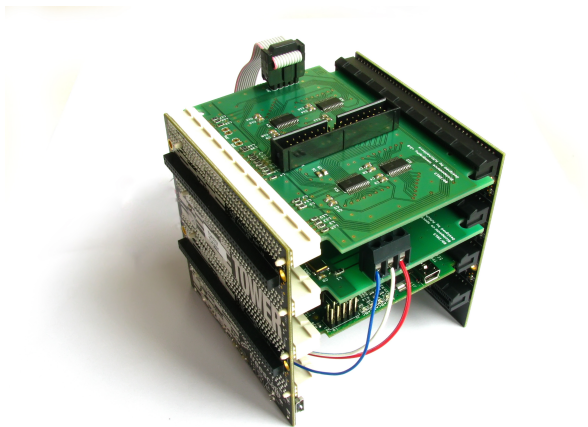


Fig. 1. Tower-based device for bioimpedance measurement

2.1 Measurement module board

This module consists of two main parts - signal generator and signal detector. Block diagram is depicted in fig. 2. Signal generator provides voltage sine-wave

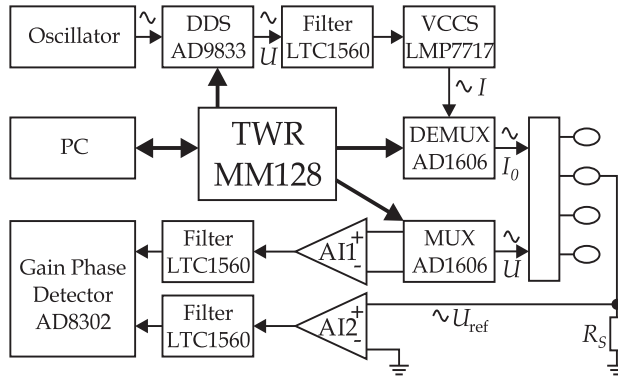


Fig. 2. Block diagram of device

signal of required frequency, which is transferred to current using voltage controlled current source (VCCS). The current sine-wave is then applied via electrodes to measured impedance Z_x . The signal detection part is intended to detect and process measured voltage sine-wave. First the signal is measured by instrumental amplifier (IA). There are two IAs. One measures voltage signal from unknown impedance Z_x and the other voltage drop across the reference resistor R_S . Both signals are then processed the same way. Then signals are compared in gain-phase detector (GPD), which compares amplitudes of signals (in other words degree of signal attenuation caused by passing through the impedance load) and phase shift between signals. There are two DC signals on the output of GPD, which correspond to amplitude ratio and phase difference of two input signals. Then unknown impedance Z_x can be calculated according to equation 1.:

$$Z = \frac{|A_1|}{|A_2|} \cdot R_S \cdot |K| e^{j\Phi} \quad (1)$$

where A_1 and A_2 represent gain of IA1 and IA2, K is the amplitude ratio and Φ corresponds to phase difference of the signals. Assuming that IAs are identical and have the same gain the member with gain ratio equals 1 so it does not influence calculation of unknown impedance and can be left out from equation.

3 Results

In order to test the device, simple RC circuit, which replaces skin-tissue interface, was constructed and the amplitude characteristic across wide frequency

range from 1 kHz to 1 MHz with 1 kHz frequency step was measured. Measured amplitude frequency characteristic before calibration is depicted in figure 3. A corresponding ideal frequency characteristics of given RC circuit was also calculated for comparison, see figure 4.

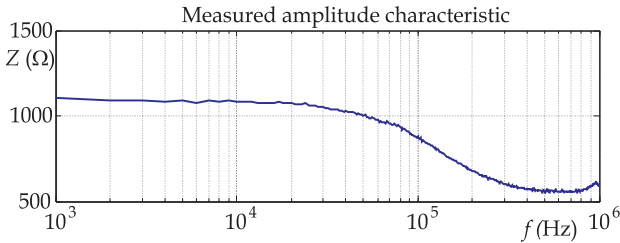


Fig. 3. Amplitude and Phase Characteristics of Measured RC circuit

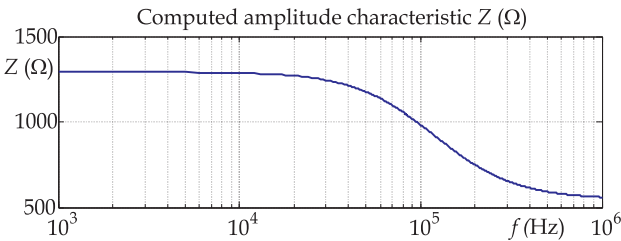


Fig. 4. Calculated Amplitude and Phase Characteristics of RC circuit

While measuring impedance the systematic error caused by stray capacitances between signal leads on input of instrumental amplifiers occurs, so the calibration is required. I used the Quadratic Interpolation Self-calibration algorithm for nonlinear measuring systems proposed in [8]. There are needed three reference quantities of impedance, so I constructed three different RC circuits with the same cut-off frequency of 100 kHz (as used in [7]) and used them as the three reference points which determine the correlation between the real value of impedance and the measured value. The calibration is carried out according the equation (2) [7]:

$$\begin{aligned}
 Z_x = & Z_{R0} \frac{(Z_{meas} - Z_{mR1})(Z_{meas} - Z_{mR2})}{(Z_{mR0} - Z_{mR1})(Z_{mR0} - Z_{mR2})} \\
 & + Z_{R1} \frac{(Z_{meas} - Z_{mR2})(Z_{meas} - Z_{mR0})}{(Z_{mR1} - Z_{mR2})(Z_{mR1} - Z_{mR0})} \\
 & + Z_{R2} \frac{(Z_{meas} - Z_{mR0})(Z_{meas} - Z_{mR1})}{(Z_{mR2} - Z_{mR0})(Z_{mR2} - Z_{mR1})}
 \end{aligned} \tag{2}$$

Calibrated amplitude characteristic is depicted together with expected amplitude characteristic in figure 5. Then I calculated relative error of Z_X after calibration, see figure 6.

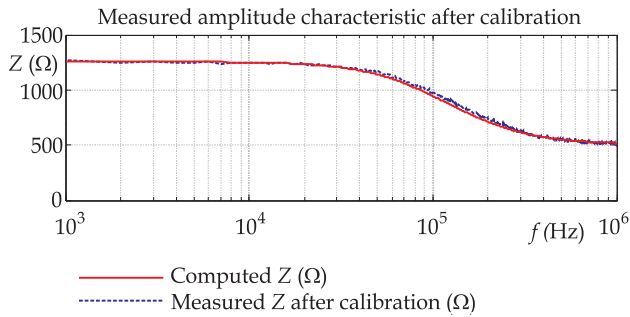


Fig. 5. Amplitude and Phase Characteristic After Calibration

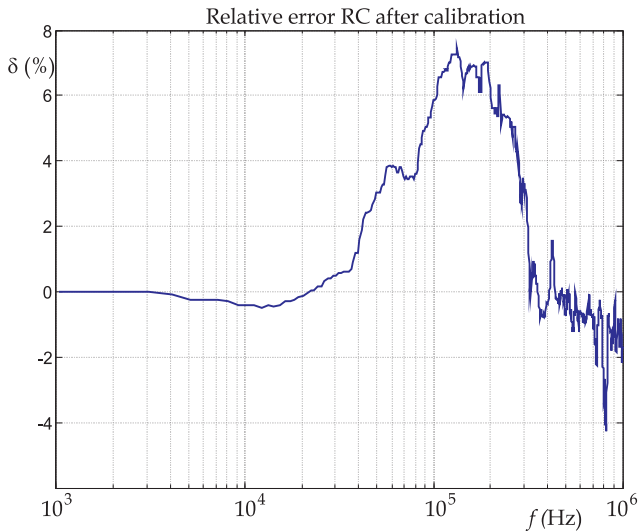


Fig. 6. Relative Error of Measurement After Calibration

4 Conclusion and future work

The constructed device is able to measure impedance in frequency range from 1 kHz to 1 MHz. The impedance measurement inaccuracy is caused by a systematic and a random error. The systematic error was eliminated thanks to

calibration. The calibration was applied on absolute value of the impedance, calibration on the phase shift has not been done yet, since both input signals of GPD are highly disturbed with noise, so the GPD cannot give the right results on its output and the phase shift cannot be calculated correctly. The relative error after calibration occurs from -4.52% to 5.98% with mean value of -0.02% . The most significant relative errors occur on high frequencies. In the future work it will be crucial to eliminate the noise, which disturbs signals on GPD input. Therefore the 5th order low-pass elliptic filters with 1 MHz cut-off frequency were used to filter these signals for GPD input. This approach has not brought expected results, since the filtration was not successful enough and the noise still remained superposed on signals. The noise is produced mostly by Tower elevators, which proves the Tower usage approach as a not good solution for the bioimpedance measurements and will not be used in future realization of the device. The phase calibration improvement is a subject of ongoing work.

References

1. M. Kishore, V. Saphthagirivasan, V. Manohar, and M. Anburajan, "Total body composition measurement by using bioelectric impedance analysis and its impact on obesity," vol. 1, 2010, pp. V196–V199.
2. C. Corciova, M. Turnea, D. Matei, and D. Andritoi, "Evaluation of cardiac parameters using electrical impedance plethysmography." *Revista medico-chirurgical a Societii de Medici i Naturaliti din Iai*, vol. 116, no. 3, pp. 927–932, 2012, c.
3. A. Nowakowski, T. Palko, and J. Wtorek, "Advances in electrical impedance methods in medical diagnostics," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 53, no. 3, pp. 231–243, 2005.
4. T. Dai and A. Adler, "In vivo blood characterization from bioimpedance spectroscopy of blood pooling," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 11, pp. 3831–3838, 2009.
5. H. Abdul Rahim, E. Shair, and M. Taib, "A non-invasive approach to predict risk in dengue hemorrhagic fever (dhf) using bioelectrical impedance analysis," 2009, pp. 44–47.
6. U. Kyle, I. Bosaeus, A. De Lorenzo, P. Deurenberg, M. Elia, J. Gmez, B. Heitmann, L. Kent-Smith, J.-C. Melchior, M. Pirlich, H. Scharfetter, A. Schols, and C. Pichard, "Bioelectrical impedance analysis - part i: Review of principles and methods," *Clinical Nutrition*, vol. 23, no. 5, pp. 1226–1243, 2004.
7. Y. Yang, J. Wang, G. Yu, F. Niu, and P. He, "Design and preliminary evaluation of a portable device for the measurement of bioimpedance spectroscopy," *Physiological Measurement*, vol. 27, no. 12, 2006.
8. J.-G. Liu, U. Fruehauf, and A. Schoenecker, "On the application of special self-calibration algorithm to improve impedance measurement by standard measuring systems," vol. 2, 1999, pp. 1017–1022.

Methods verification of image recognition in metallurgy industry

Martin Mikolajek

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

Abstract. This paper presents further possibilities of image alphanumeric recognition in metallurgy industry. In normal situations, the images are scanned and digitized by standard optical camera. But in some situations it is not possible to use the optical camera, for example due to contamination dust of metallurgical materials. Therefore input numeric information can be bad readable. It is necessary to search for new methods of image recognition for metallurgical use. Research focuses on analyzing the usefulness of modern sensor technology for special cases in industrial automation for the steel industry where produced materials are signed by using stamped numbers. Therefore this paper research is specialized for various methods that can be used for reading alphanumeric marks on stamped surfaces materials recognizing instead of the standard industrial cameras usage. Especially are using the optical and laser scanners for surface differences recognition etc.

Keywords: Recognition in metallurgy industry, Laser scanner, Industrial automation, Measurement methods, Stamped characters.

1 Introduction

The paper includes a summary of various special sensor types for the alphanumeric marking recognition in metallurgy industry, which should give data about the type and other parameters of the tested product. Identification of the product is a topic research interest, which is necessary in today's manufacturing processes. The practical part describes the use of a special sensor technology in specific problematic case of alphanumeric marking recognition. There are presented tested methods and possibility of reading embossed digits in metal which are effects by dust or a considerably high temperature. Specifically, the results are applicable for cases where is necessary to identify the figures or marks in metal products produced in industrial automation. The innovation of embossed alphanumeric marks recognition sensor methods is very needful, because during the reading marking numbers on materials in process for metallurgy industry can be included dust faults. [1]

2 Information about problematic reading stamped characters

This section provides more information about the selection of the sensor technology type, in order to achieve the expected optimal results. Marked data in the metal material also contain information about the type and quality of the metal slab, which is recorded to database information system. The marking process realizes marking of alphanumeric characters in the material form of the embossed depth of 0.5 - 1 mm. Size of embossed alphanumeric characters is representing by 15 x 10 mm surface. Width of embossed alphanumeric figures is 0.5 to 1 mm size. The hole diameter is about 1 mm in case of using the stamping holes. [1]

Embossed alphanumeric characters may be contaminated by dust or other impurity; therefore we tried to testing another method for recognition instead of camera techniques. This problem can be resolved for example by using scanner method for material surfaces detection [2]. Another problem may be high temperature of measured product what depends on the location of the sensing device in the industrial technology. The measured metal product can be measured in temperature up to about 800 ° C. [3]

There is provides more information about the selection of the sensor technology type, in order to achieve the expected optimal results. Marked data in the metal material also contain information about the type and quality of the metal slab, which is recorded to database information system. The marking process realizes marking of alphanumeric characters in the material form of the embossed depth of 0.5 - 1 mm. Size of embossed alphanumeric characters is representing by 15 x 10 mm surface. Width of embossed alphanumeric figures is 0.5 to 1 mm size. The hole diameter is about 1 mm in case of using the stamping holes.

Embossed alphanumeric characters may be contaminated by dust or other impurity; therefore we tried to testing another method for recognition instead of camera techniques. This problem can be resolved for example by using scanner method for material surfaces detection. [3]

Another problem may be high temperature of measured product what depends on the location of the sensing device in the industrial technology. The measured metal product can be measured in temperature up to about 800 ° C.

Request for alphanumeric marking recognition immediately after embossing heating process

This process type is usable to ensure proper readability of embossing alphanumeric data. Thus it is executed for ensuring that the embossing is done correctly for other automatic identification of product. Incorrect elements in alphanumeric marking are usually carried out immediately after the thermal processing of metallurgical materials. The advantage is the fact that immediately after punching alphanumeric marking are recognized, which can guarantee the reliability of reading accuracy characters. There could be very useful thermal camera for identification of embossing alphanumeric markers. There is not guaranteed enough precision with surface optical sensors usage because of high temperature of marked product. Next implementation of using sensor can be alphanumeric marking recognition for cold material state especially for information to automated system.

This process type is loaded with deterioration due and oxidation progress and dusty contamination of the applied material. Thus dirty material is not possible to identify with the standard industry camera. Therefore, there is another possibility for identification by a scanner for measuring surfaces of materials.

The verification of useful sensor type was focused to ensure proper data capturing, such as a scanning resolution, range, sensitivity, operating temperature, mechanical strength, etc. The example of embossing alphanumeric marker characters is shown in Fig. 1.



Fig. 1. Example of embossed alphanumeric marking characters.

3 Suitable sensing devices for reading embossed characters

There is presented a selection of the available types of sensing devices for measuring surface. It discusses the available scanners that are suitable for measuring the metal surface so that it is possible to analyse the embossed alphanumeric characters. The possible marking numbers measuring can be performed using accurate laser scanners allow full scanning the surface of a desired object with great precision. Selecting the type of measuring device must be a compromise between price and value of the measuring device.

First scanner type description is based on method of optical principle. This type of scanners is designed primarily for use in the laboratory and clean environment. Scanned slabs are exposed in the environment for which it is not intended, it would be necessary to use sensor element with protection solution. The required resolution for the measurement of embossed characters is 0.5 - 1 mm, what is suitable only for scanners with EVA series Artec, Artec Artec Eva lite or Spider. These types of scanners are generally designed for fast 3D digitization of image objects with a wide range of technology fields for medicine, engineering, archaeology , etc. Other types of scanners are not suitable because of lower resolution for embossed alphanumeric characters.

These optical 3D scanners so unlike the laser need no reference marks are highly portable, lightweight. Scanners allow capturing a large field of view. Scanning occurs in real time and the user sees on the computer monitor that part of the model has already been scanned. [5]

Second type of scanner works on method of laser light principle. The available laser scanners are suitable for use in many industrial applications. This is not a 3D scanner type but it is only line scanner type. There is necessary to realize the moving of the material under the scanner assign Y coordinates by using trigger or CMM counters.

These types of scanners must operate in a relatively small measuring distances of tens millimetres. Their advantage is the high precision measurement of the surface on the order of micrometres. Evaluation of the distance is performed using a triangulation method. There is changed the position of incidence of the light beam to the sensor depend on varying distance between the sensor and the surface. [5]

Third method is infra-camera type description based on high temperature detection. The purpose of the Infra-camera is a non-contact surface temperature measurement element. Choosing the type of camera is always a compromise between practical value thermo camera and its cost, which depends on elementary parameters: working temperature, dpi resolution, sampling frequency, temperature sensitivity, the quality of the camera optics, etc. The resolution affects the price of the camera. The most common resolution thermal camera is around 160 x 120 pixels or 382 x 288 pixels. These are the lowest price, which is usable.

4 Practical measurement reading quality characters

The first method of measurement verification was performed with 3D optical scanner EVA series Artec [6]. This type of scanner was used for test measuring at laboratory. There was necessary placed around measured metal surface non-flat paper for connectivity of optical methodology algorithms. There is implemented special software application for PC, which gives computed result image shown in Fig. 2.

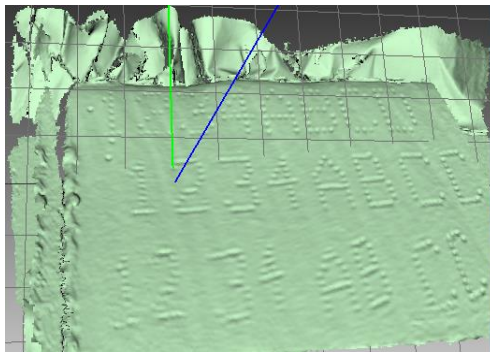


Fig. 2. Verification result of scanner sensor method based on optical principle

Next method of measurement verification was performed verification of scanner sensor method based on laser light principle. The verification was performed with Microepsilon scanCONTROL 2750-100 scanner sensor. This type of scanner was used for

test measuring at laboratory for 2D measurement and continuously movement of intensity surface differences measurement. Compare to previous method, there is necessary to movement of scanner because of only line measurement by sensor. Measurement speed is set to 250 frames per second. The measured results are presented graphically in the following figure Fig. 3. There is not assured the ideal constant velocity motion of sensor above the measured product, reading characters can be not clearly readable. The other laboratory measurement of Microepsilon scanCONTROL 2750-100 scanner sensor was performed with linear unit for moving the sensor over the steel plate shown in Fig. 3. The surface movement speed was 9 mm/s and a profile frequency of 160 per second. The maximal possible movement speed to scan over the surface is about 60 mm per second. If this isn't enough either there is possible to reduce the measured points per scan from 1280 pixel resolution to 640 pixel resolution. In increases possibility of movement speed to 120 mm per second, but it has to be proved if the dots in the metallurgy material will still be detected.

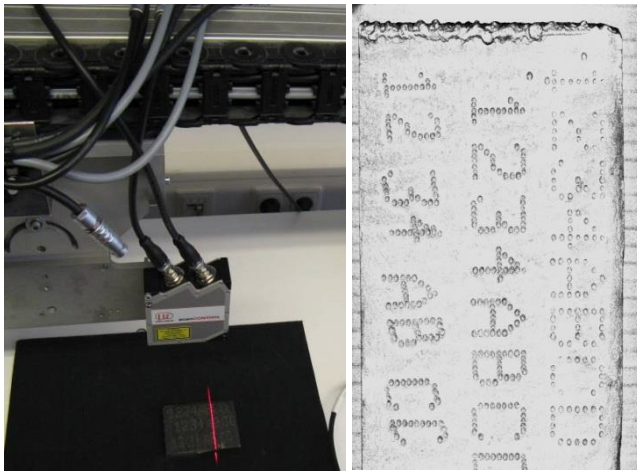


Fig. 3. Measurement with constant movement and verification result of scanner sensor method based on laser light principle with constant movement [1]

The last type of measurement verification was performed with infra-camera Micro - Epsilon TIM 200. This type of infra-camera has resolution 160x120 pixels with possibility of 1500 Celsius degree measuring. Compare to previous methods, this measurement is usable for high temperature surfaces especially immediately after embossing of marking characters, because there is for short time different temperature in the embossed places. There is problem with different temperature field on the others places not only where are embossed characters.



Fig. 4. Verification result of thermo-camera type based on high temperature detection

5 Conclusion

The purpose of the research was to realize verification of the possible special sensor technology methods in specific problematic case of embossed alphanumeric marking recognition for metallurgy industry. These method types are verified in laboratory as well as in industry implementation. The measured metal product can be measured in cold temperature or in high temperature up to about 800 Celsius degree.

There are presenting processes and results for scanner type based on method of optical principle, scanner type based on method of laser light principle, infra - camera type description based on high temperature detection. The paper analyzed possibility and limits of usage these special sensor methods for precisely image recognition of embossed alphanumeric characters in metallurgy material.

Acknowledgements

This work was supported by project SP2014/156, "Microprocessor based systems for control and measurement applications." of Student Grant System, VSB-TU Ostrava.

References

1. Mikolajek M.; Machacek Z.; Koziorek J. Modern Sensor Technology for Alphanumeric Recognition in Metallurgy Industry. *Elektronika Ir Elektrotechnika*. 2014, Vol.20, No.5. pp. 3-7. ISSN: 1392-1215.
2. Hajovsky R.; Pies M. Complex Measuring System for Longtime Monitoring and Visualization of Temperature and Toxic Gases Concentration. *Elektronika Ir Elektrotechnika*. 2012, Vol. 122, Issue 6, pp. 129-132. ISSN 1392-1215.
3. Machacek Z.; Hajovsky R. Modeling of Temperature Field Distribution in Mine Dumps with Spread Prediction.; *Elektronika Ir Elektrotechnika*. 2013, Vol.19, No.7. p.53-56. ISSN: 1392-1215
4. <http://www.micro-epsilon.cz/laser-scanner/index.html> [download 20.12.2013].
5. <http://www.artec3d.com/hardware/artec-eva/specifications/> [download 20.12.2013].

Usage of Genetics Algorithms in Task of System Identification

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](mailto:jana.nowakova@vsb.cz), [miroslav.pokorny](mailto:miroslav.pokorny@vsb.cz)}@vsb.cz

Abstract. System identification is one of the necessary tasks in controller design and its adaptation. Many identification methods are known, and new ones are still being developed in order to find a better solution for huge scale of cases. In the paper identification of system of 2nd order systems using genetic algorithms is demonstrated. In presented case genetic algorithms are used for finding parameters of difference equation of the controlled system and it substitutes classic, conventional optimization methods. Proposed method can be used for continuous identification or it can be activated in defined time points on stored data. And on the other hand, presented task is also a case of a specific usage of genetic algorithms and it can serve as a proof of efficiency of this non-conventional optimization method (simulated in the Matlab&Simulink software environment).

Key words: Identification, system, genetic algorithms, optimization.

1 Introduction

Identification of controlled system is a known and necessary task in controller design and its adaptation [1], [2]. According to the parameters of identified system, controller is adapted but the method of controller adaptation is another problem, which is not considered in paper. Presented proposed method can be used both for continuous and discrete systems.

Genetic algorithms (GA) are non-conventional optimization methods [3]. In our case GA are used for finding of coefficients of difference equation of the controlled system according to the fitness function.

2 Difference Equation of the Identified System

The first task is to define general transfer function of the identified system of the second order. For our proposal the transfer function of the identified systems is chosen as follows

$$G_S(s) = \frac{1}{c(s+a)(s+b)}, \quad (1)$$

where a , b and c are real numbers - parameters of identified system [4]. The proposed method can be also used for systems with another shapes and orders of transfer function.

2.1 Discretization of the Transfers Function of the System

The next step is to discretize the defined transfer function (1) using \mathcal{Z} -transform considering zero-order hold [5]

$$G_S(z) = \mathcal{Z} \left\{ \frac{1}{c(s+a)(s+b)} \cdot \frac{1-e^{-sT}}{s} \right\} = \frac{z-1}{z} \cdot \frac{1}{c} \mathcal{Z} \left\{ \frac{1}{s(s+a)(s+b)} \right\}, \tag{2}$$

where T , is optional sampling period.

After decomposition to the partial-fractions

$$G_S(z) = \frac{z-1}{z} \cdot \frac{1}{c} \mathcal{Z} \left\{ \frac{A}{s} + \frac{B}{s+a} + \frac{C}{s+b} \right\}, \tag{3}$$

where

$$\begin{aligned} A &= \frac{1}{ab}, \\ B &= \frac{1}{a(a-b)}, \\ C &= \frac{-1}{b(a-b)}, \end{aligned} \tag{4}$$

discrete transfer function of the identified system (1) could be written as

$$\begin{aligned} G_S(z) &= \frac{z^2(A+B+C)}{c(z-e^{-aT})(z-e^{-bT})} + \\ &+ \frac{z(-Ae^{-aT} - Ae^{-bT} - Be^{-bT} - B - Ce^{-aT} - C)}{c(z-e^{-aT})(z-e^{-bT})} + \\ &+ \frac{Ae^{-aT}e^{-bT} + Be^{-bT} + Ce^{-aT}}{c(z-e^{-aT})(z-e^{-bT})}. \end{aligned} \tag{5}$$

2.2 Parameters of Difference Equation of the System

As the transfer function is ratio of \mathcal{Z} -transform of the output signal to the \mathcal{Z} -transform of the input signal [4]

$$G_S(z) = \frac{Y(z)}{U(z)}, \tag{6}$$

difference equation of the system can be expressed as

$$Y(z) [c(z - e^{-aT})(z - e^{-bT})] = U(z) [z^2(A + B + C) + z(-Ae^{-aT} - Ae^{-bT} - Be^{-bT} - B - Ce^{-aT} - C) + Ae^{-aT}e^{-bT} + Be^{-bT} + Ce^{-aT}] \quad (7)$$

and according to this it could be written as

$$cy(k) + (-ce^{-aT} - ce^{-bT})y(k-1) + ce^{-aT}e^{-bT}y(k-2) = (-Ae^{-aT} - Ae^{-bT} - Be^{-bT} - B - Ce^{-aT} - C)u(k-1) + (Ae^{-aT}e^{-bT} + Be^{-bT} + Ce^{-aT})u(k-2), \quad (8)$$

note that

$$A + B + C = 0. \quad (9)$$

For bigger clarity which leads to a simple relation

$$Dy(k) + Ey(k-1) + Fy(k-2) = Gu(k-1) + Hu(k-2). \quad (10)$$

The parameters D , E , F , G and H are parameters of the difference equation to be identified using GA.

3 Genetic Algorithm Synthesis

The aim of the work is to identify the parameters of the difference equation of the system using GA. As it was mentioned before the classical optimization method for identification of coefficients of difference equation was substituted by use of genetic algorithm.

For the identification of the coefficients of difference equation the minimization of fitness function [6] can be expressed in the form of

$$J = \frac{1}{N} \sum_{j=1}^N [y^0(x_j) - y^*(x_j)]^2 \rightarrow 0, \quad (11)$$

where N is size of the set of input-output. The minimization of fitness function J (11) by embedded function of genetic algorithm in Optimtool in Matlab environment is used. The parameters of GA are selected as

- population type - double vector,
- scaling function - rank,
- population size - 20,
- mutation function - constraint dependent,
- crossover function - scattered,
- selection - stochastic uniform,
- migration - forward,
- stop criterion - 300 generations.

4 Numerical Examples

For verification of the proposed method two systems are chosen. It includes simulation of both of them while input and output signals are stored in memory.

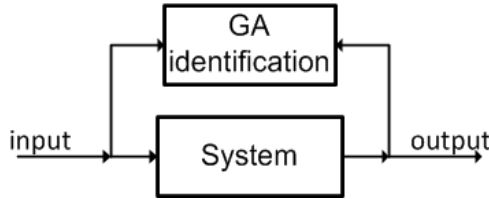


Fig. 1. Identification block with the identified system [7]

4.1 The First Verification

Considering the system with transfer function

$$G_{S1}(s) = \frac{1}{8s^2 + 48s + 64}, \tag{12}$$

using \mathcal{Z} -transform, where $T = 0.1$ sec, the difference equation (10) is determined

$$-1.4891y(k) + 0.5488y(k-1) + 5.1341 \cdot 10^{-4}y(k-2) = 4.2035 \cdot 10^{-4}u(k-1) + u(k-2). \tag{13}$$

The first step is to generate set of input and output data using Matlab&Simulink with a consequent verification by means of GA performed at this data set. GA is used with the parameters defined in Chapter 3. After 300 generations the fitness function J_1

$$J_1 = 1.397 \cdot 10^{-19}. \tag{14}$$

Difference equation is then determined as

$$40.5098y(k) - 14.9305y(k-1) - 0.0140y(k-2) = -0.0114u(k-1) - 27.2051u(k-2), \tag{15}$$

which is practically the same as (13) multiplied by -27.2051 (identified coefficient H, which does not have any effect, it can be cancelled out). After inverse \mathcal{Z} -transform with the same $T = 0.1$ sec, the continuous transfer function works out as

$$G_{S1ident}(s) = \frac{0.125}{s^2 + 6s + 8} = \frac{1}{8s^2 + 48s + 64}, \tag{16}$$

after identification.

4.2 The Second Verification

For the second verification, the system with transfer function with parameter $a = 1$, $b = 0.5$ and $c = 0.8$ is created, transfer function then is written as

$$G_{S2}(s) = \frac{1}{0.8s^2 + 1.2s + 0.4}, \quad (17)$$

and again using \mathcal{Z} -transform, where $T = 1$ sec, the difference equation (10) is determined

$$-0.9744y(k) + 0.2231y(k-1) + 0.3870y(k-2) = 0.2348u(k-1) + u(k-2). \quad (18)$$

The process of the verification is also the same as in the previous case, the set of input and output signal is determined and the verification is done on the stored data. After 300 generations the fitness function J_2 works out as

$$J_2 = 1.1376 \cdot 10^{-11}. \quad (19)$$

Difference equation is again determined with only small differences (equation should be again cancelled out by parameter $H = 1.0744$) compared to the calculated difference equation

$$-1.0469y(k) + 0.2397y(k-1) + 0.4158y(k-2) = 0.2522u(k-1) + 1.0744u(k-2). \quad (20)$$

After reverse \mathcal{Z} -transform with the same $T = 1$ sec, the continuous transfer function $G_{S2ident}(s)$ where $a = 1$, $b = 0.5$ and $c = 0.8$ are identified, which corresponds to the specified system.

5 Conclusion

It was presented specific usage of GA in identification of of 2nd order systems with transfer function in the shape (1). For optimization by GA the fitness function is defined as the root mean square difference between the model and system outputs. It is possible to use proposed idea also for another types of system, the only necessary task is to define the difference equation. For using of GA, the calculation time could be considered as a disadvantage, but it is not a problem for such systems as systems of 2nd order.

Subscribed solution is being one of the part of algorithm of continuous adaptation of controller using soft-computing approach [8].

Acknowledgements This work has been supported by Project SP2014/156, "Microprocessor based systems for control and measurement applications", of the Student Grant System, VŠB - Technical University of Ostrava.

References

1. Friedland, B.: Advanced control system design. Prentice-Hall, USA (1996)
2. Keesman, Karel J.: System Identification. Springer-Verlag, London (2011)
3. Goldberg ,D., Sastry, K.: Genetic Algorithms. The design of Innovation. Springer-Verlag, NY (2007)
4. Levine, W. S.: The Control Handbook. Jaico Publishing House, Mumbai (1999)
5. El Attar, R.: Lecture Notes on Z-Transform. Lulu.com, USA (2006)
6. Ajith, A., Solar, J.R., Koppen, M.: Soft Computing Systems Design, Management and Application. IOS Press, the Netherlands (2002)
7. Garnier, H., Wang, L.: Identification of Continuous-time Models from Sampled Data. Springer-Verlag, London (2008)
8. Nowaková, J. - Pokorný, M.: On PID Controller Design Using Knowledge Based Fuzzy System. Advances in Electrical and Electronic Engineering, **10(1)**, 18-27 (2012)

User identification by biometric methods

Pavlina Nudzikova

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

Abstract. The concept of user identification by biometric methods is described. The paper summarizes techniques that are routinely used in biometric identification, gives the reader basic awareness of the background of biometrics and explains the meaning of basic terms used in this area.

Keywords: Biometrics, recognition, identification, gait recognition, finger recognition

1 Introduction

Biometrics is an automatic method serving to verify the identity of a person based on recognition of the individual's unique biological characteristics. It is based on the belief that the biological characteristics in question are unique in human population and constant in time. Anatomic or physiological properties are used for identification purposes. To be usable in biometrics, the characteristics must be unique, invariable, practically measurable and amenable to subsequent processing for evaluation by comparison with the same characteristics of other individuals. Biometrics is used to identify a person (identification) or to verify a person's identity (verification).

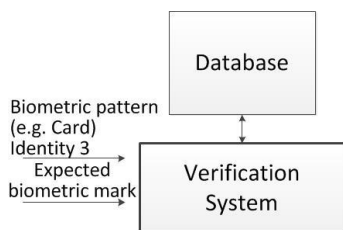


Fig. 1. Principle of verification.

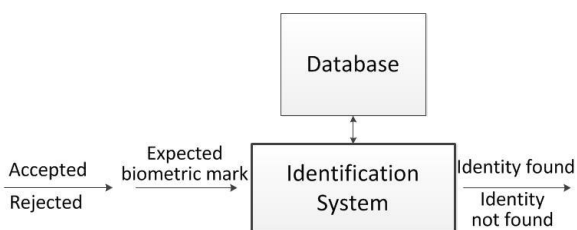


Fig. 2. Principle of identification.

Although biometrics has been in the focus recently particularly in relation to computer security, recognition by using biological characteristics has been used for centuries or, actually, millennia. Palm prints as sort of author's signature have been found in archaeological caves. Such prints are as much as 30,000 years old. Humans are generally able to tell one individual apart from another based on the individual's phys-

iological features, typically the face and voice. However, it was not until the late 1960s that biometrics of humans started to be automated owing to the development of computer techniques.

Biometrics of humans is superior to other means such as passwords or chip cards in that the characteristics cannot be lost or forgotten. Also, biometric means of authentication are simple to use: it suffices to put a finger or palm of one's hand on a device or to look into a camera. As another asset, it is very difficult – actually nearly impossible – to falsify the data that are required for identification. Also, biometric characteristics cannot be transferred to any other individual.

The use of biometrics has also its drawbacks, though. While you can always tell if a digital password is valid or not, this is never that unambiguous in biometrics. A certain fraction of erroneous identifications, although very very small, always exists. Biometrics is not, and never will be, absolutely reliable. This is so because one will never be able to provide a sample which is absolutely identical with that stored in the template.

Biometric systems are used in a number of industrial areas, among them automotive systems. Biometrics is used in active security systems, for driver and passenger authorization, in comfort systems, systems monitoring the driver for any signs of fatigue and systems to detect situations where the driver fails to pay full attention to driving. For use of biometrics in vehicles see ref. [1]

2 Basic methods of the biometrics of humans

A biometric system is basically a recognition system which acquires biometric data from the individual logging into it. The system extracts a set of characteristic features from the information obtained and compares it to data stored in the database. Subsequently the system responds in some way based on the result of the comparison.

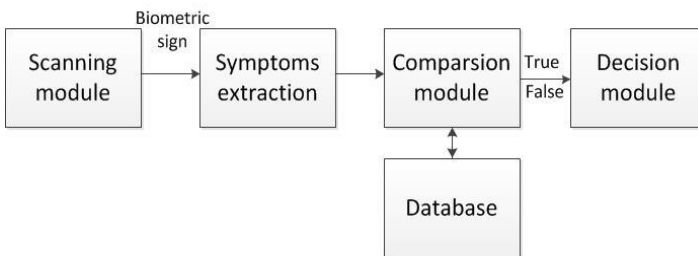


Fig. 3. Biometric system modules.

3 Biometrics of the eye

Identification of humans through their retina and/or iris is among established biometric approaches [7], [8], based on the fact that those parts of human eye are unique in any individual. Identification methods based on those properties are highly reliable.

3.1 Recognition through the iris

This is a method that enables reliable identification of a person based on the unique reaction of the person's iris to incident light. The iris has a random appearance, with a physical structure which is extremely rich in data patterns that differ from one person to another, including monozygotic twins. The process of user identification through iris recognition is shown in Fig. 4 below. The steps and methods are described in refs. [2], [3]. Methods for iris location in the face are described in ref. [6]. An optimized Daugman's algorithm for iris location is described in the paper by Hebaishy [14]. A comparison between Daugman's algorithm and the Hough transform is presented in ref. [16]. A newer iris recognition method through mutual information estimation is introduced in ref. [13].

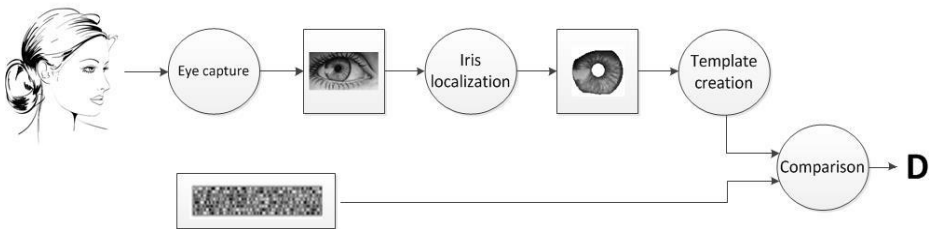


Fig. 4. User identification by iris recognition [2].

3.2 Recognition through the retina

Recognition through the retina is a biometric method that compares individuals based on scanning and comparison of their retinal patterns. A special optical camera is required to obtain the image of the retinal blood vessels. The retina is scanned by a low-intensity infrared beam which follows the circular path of the retina passing through the pupil. The retina is more or less transparent to light at the wavelengths used. The image of the retina is actually formed only by the blood vessel network behind the retina, and it is that image that is used for the recognition of individuals. The scanner beam runs around a certain area of the retina and captures a circular picture of the retina where the blood vessel structures are darker than the surrounding tissue. The methods to acquire and process the image are described in refs. [7], [8].

4 Identification through the fingerprint

This method is based scanning the person's fingerprint and its comparison with the database. Every person has their own unique fingerprints. The fingerprint pattern is even different in monozygotic twins. The pattern consists of papillary lines [2], [10] which are present on the finger pads. Algorithms that are currently applied to the identification of persons based on their fingerprints are described in ref. [11]. Various contact sensors and contactless sensors are used for fingerprint scanning [7], [8], [10], [12].

5 Biometrics of the face

Identification of persons through face recognition is the oldest method. For millennia, people have been recognizing their family, friends and other contacts intuitively. Recognition of other persons based on their facial features has been the most natural and established method during the history of mankind. We perform this "operation" routinely during the day with a high reliability. It is natural, then, that efforts to achieve automatic recognition based on human face characteristics are among prominent ones in biometrics. Face identification systems have been described in refs. [2], [8], [7].

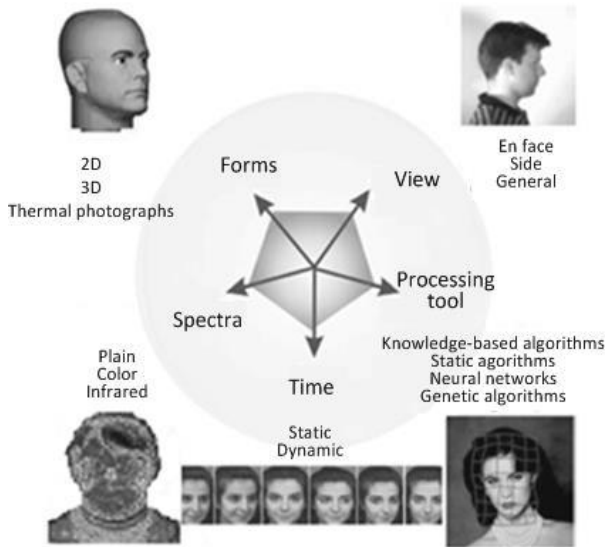


Fig. 5. The different approaches to machine processing of the human face. [8]

6 Gait biometrics

Bipedal locomotion, or walking dynamics, is a next parameter usable to identify individuals. Locomotion is defined as the act or power of moving from place to place. Specifically in humans, locomotion is the motion of an individual in the gravitation field by the individual's own power, using the limbs or other anatomic parts of human body. Recognition of humans based on their gait is one of the newly emerging areas of biometric applications. For the principle of the method refer to [4], [5], [8]. Horák and Richter [17] achieved image segmentation through motion detection and by using the mean displacement method for clustering of points. The segmentation methods are also described in ref. [18].

7 Other biometric methods

Biometry of the bloodstream is a next biometric method [7], [19]. The shape of the blood vessels is measured when scanning the hand (either the back or the palm of the hand). Additional methods include the use of the DNA, voice features, signature dynamics, keyboard typing dynamics, computer mouse moving dynamics, shape of the auricle, and others [2], [7], [8], [9], [15].

8 Comparison of the biometric methods

The basic biometric methods and their properties and compared in Table 1 [2]. The table demonstrates that the most reliable techniques include the fingerprint method, which is also most beneficial from the financial aspect and is user friendly.

Table 1. List of basic biometric methods and some of their properties.

Biometrics	Precision	Cost	Variability in time	User friendliness	Total
Fingerprint	●●●	●○○	●○○	●●○	●●●
Hand geometry	●●○	●●○	●●○	●●○	●●○
Face recognition	●●○	●●○	●●○	●○○	●●○
Iris recognition	●●●	●●●	●○○	●○○	●●●
Retina recognition	●●●	●●●	●●○	●●●	●●○
Nail bed	●●●	●●○	●●○	●●○	●●○
DNA	●●●	●●●	●○○	●●●	●●○
Voice recognition	●○○	●○○	●●●	●○○	●●○
Signature dynamics	●○○	●○○	●●○	●○○	●●○
Key pressing dynamics	●●○	●○○	●○○	●○○	●●○
Level	Low ●○○		Medium ●●○	High ●●●	

9 Conclusion

Identity verification based on an individual's biometric properties is a rapidly developing field of science and technology, with a constantly increasing potential, number of methods and their combinations. The use of biometrics in the automotive industry is a current topic of interest. The methods are only at the onset of their development, and it will certainly be very interesting to engage oneself in this field of application for the car and truck industry. Various application areas are open to use, from inexpensive and simple ones for basic verification purposes to very sophisticated ones, it is only up to you to choose the option that suits you best.

Based on the knowledge acquired I decided to use the gait recognition method, which will pre-identify the user at the electromobile charging stations, followed by a more reliable method, such as the fingerprint scanning method where the image template is compared to a smaller selection of images stored in the database. A combination of two different methods will be used for the identification. The benefits of this approach include rapid user identification as well as enhanced identification precision.

10 Acknowledgment

This work is also supported by project SP2014/188, named "Control of technological systems with OAZE providing an independent sustainable development of complex systems" of Student Grant Agency (VSB - Technical University of Ostrava).

11 References

1. Plíhal, J. Využití biometrie ve vozidle (Use of biometrics in vehicles). Proceedings of the international professional conference Improving the Safety of Vehicle Operation in Armed Forces (Zvýšení bezpečnosti provozu vozidel ozbrojených sil), 2011, pp 93-98, ISBN 978-80-904625-2-6.
2. Nůdziková, P., Slanina, Z., Vala, D., Drábek, P., Electromobility I – User identification, ISBN 978-80-248-3531-0.
3. Daugman, J. How iris recognition works. IEEE Trans. CSVT, vol. 14, Issue 1, 2004, pp 21-30. doi: 10.1109/TCSVT.2003.818350.
4. Nixon, M. S., Carter, J. N., Automatic recognition by gait. Proceedings of the IEEE, vol. 94, Issue 11, 2006, pp 2013-2024. doi: 10.1109/JPROC.2006.886018.
5. Choudhury, S. D., Tjahjadi, T. Gait recognition based on shape and motion analysis of silhouette contours. Computer Vision and Image Understanding, vol. 117, Issue 12, December 2013, pp 1770-1785. doi: 10.1016/j.cviu.2013.08.003.
6. Jaskovský, P. Person identification using iris recognition. Master's thesis, Czech Technical University, Prague, 2013.
7. Drahanský, M., Orság, F. and coworkers, Biometrie, Brno 2011, 1st edition, ISBN 978-80-254-8979-6.
8. Rak, R., Matyáš, V., Říha, Z. and coworkers, Biometrie a identita člověka, Prague: Grada Publishing, a.s., 2008, ISBN 978-80-247-2365-5.
9. Biometric signature. Available from:
10. Kazik, M. Fingerprint processing. Master's thesis, VUT Brno, 2011.
11. Kovac, P. Project of biometrical identification system for small organization. Master's thesis, UTB Zlin, 2009.
12. Technology – Biometrics. [online]. Available from:
13. Dobeš, M. Image processing: New methods of localization and recognition human's iris. Dissertation, VSB-Technical University of Ostrava, 2011
14. Hebaishy, M. A. POSTER: Optimized Daugman's algorithm for iris, National Authority for Remote Sensing and Space Science, Egypt, Cairo
15. DNA. [online]. Available from:
16. Mehta, H. R., Mehta, R. G., Compare the techniques for iris localization, Journal of Information, Knowledge and Research in Electronics and Communication Engineering, Vol. 02, Issue 02, November 2012 to October 2013, pp. 666-670, ISSN 0975 – 6779.
17. Horák, K., Richter, M., Segmentace obrazu pro identifikaci osob pomocí bipedální lokomoce. (Image segmentation to identify persons using bipedal locomotion), Proceedings of the international conference Technical Computing Prague 2009, Prague: HUMUSOFT, 2009, pp 36, ISBN 978-80-7080-733-0.
18. Kryzaneck, J. Person identification by means of bipedal locomotion. Master's thesis, VUT Brno, 2010.
19. Biometrics - Systems. [online]. Available from: <http://www.biometricke-systemy.cz/>

Remote control of electric car system

Vít Otevřel and Bohumil Horák

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{vit.otevrel, bohumil.horak}@vsb.cz

Abstract. This article deals with remote control of electric power generating unit. There was a PC linked with LAN controller, which controlled the system, used for the verification of the conception. Important part of this research was preparation of the software for generator's control.

Key words: Remote control, generator, Ethernet, PowerLine, WiFi¹, programming, programming language C#, .NET Framework.

1 Introduction

Remote control of varied facilities and appliances is currently quite wide spread. There are several possibilities how to control a facility, in our situation a generator. The control can occur via some of well-known buses for houses or buildings control, for example, KNX² etc., or via power lines – the so-called PowerLine. The inclusion of the facility control a generator into the certified KNX facilities is lengthy and expensive because of the need of certification. For this reason, we decided to follow a different route – the combination of communication via Ethernet (and possibly also WiFi) and the communication via PowerLine. The following chapters describe the experiment, the components and the control application.

2 Experiment

The system consists of several components. The main part is the alternating current generator, which was controlled. Fig. 1 illustrates the block scheme of the connection. The PowerLine shows the power feeding used for the communication, but also the circuit into which electric power is delivered from the generator after its switching on. The control unit is connected directly to the generator and controls the running of the generator. Sensors from this unit are connected to the generator – the fuel checking in the tank, the measuring of the output

¹ Marking of several standards IEEE 802.11 describing the wireless communication within computer networks.

² The World standard for the home and building control.

voltage and current from the generator etc. The blue dashed line divides a part in the block scheme which is deal within this article. The generator control unit is connected with the LAN³ controller which allow the remote control. LAN module is then connected with the Ethernet/PowerLine converter serving for the simple transmission within a building. The converter Ethernet/PowerLine is located also in the room with the PC, connected by an Ethernet cable. The block scheme indicates also the chance to control the generator with a smart phone or tablet. That might be simply organised by WiFi AP/router connection instead of the computer. Established wireless network allows connection with a phone having the control application installed.

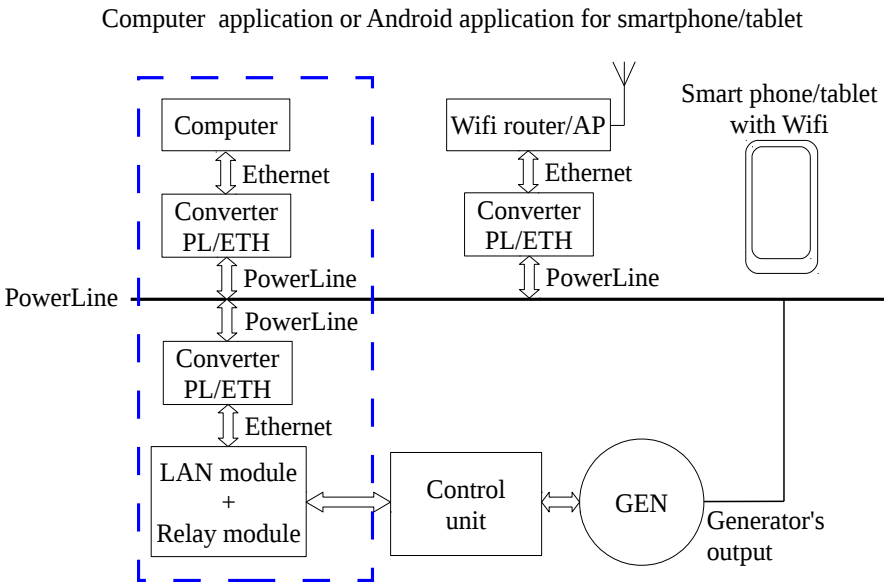


Fig. 1. Block scheme of the generator connection

2.1 LAN module

The LAN module is universal unit on the market accessible. This module includes the firmware with the embedded web server and easy web management – the control peripherals, the module settings, scheduler etc. The module has got varied peripherals integrated. The most important ones are as follows: [4]

³ Local Area Network.

- Embedded web server with web UI⁴.
- Scheduler for the relay outputs.
- Prepared inputs for thermometers – PT1000 or DS18B20.
- 1 relay output with an extended board for up to 5 relay outputs.
- PWM⁵ output with the possibility to change the frequency and the duty.
- 4 digital inputs.

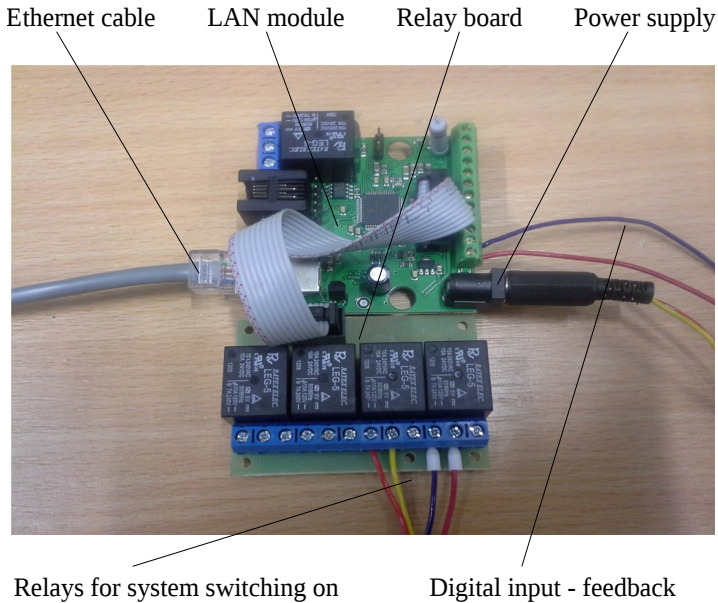


Fig. 2. LAN module connection

As seen on Fig. 2 it was used the additional board to extend the number of relay outputs. There was one digital input used for the feedback from the generator control unit used for the control application. It was checking if the generator was switched on/off. Also, there were two relay outputs utilised for the generator switching on/off. [4]

3 Control application

The control application has been written for .NET Framework of the version 2.0 with the programming language C#. The application was tested within the operating system Windows having the installed environment of .NET Framework.

⁴ User Interface.

⁵ Pulse-width modulation.

The application was also tested in the operating system Linux with the Mono platform installed.

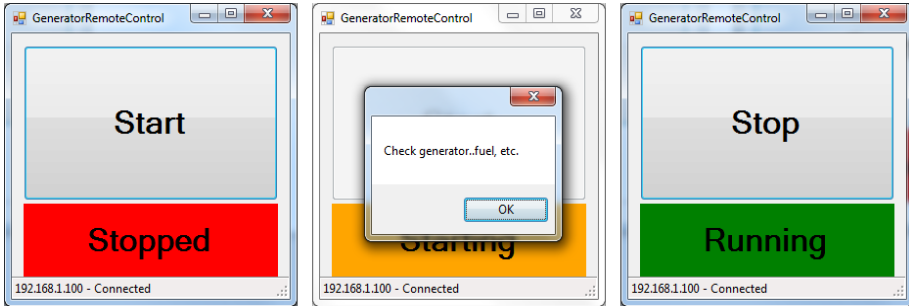


Fig. 3. The user's interface of the control application

The user's application has got a simple UI with the only push button, a single text box and a status bar. The push button switches the remote generator on/off. The text box indicates the generator's status – stopped/starting/running. The status bar displays the information about IP address of the LAN controller on which communication is transmitted and the information if the application is connected or not. The application configuration could be done only by editing XML⁶ file in which the IP address of the LAN module is saved.

3.1 Function

The application accesses the XML file with the configuration after it was switched on. Then, it reads the IP address of the LAN module. If there is no XML file, a new XML file is created with the default configuration. The set up of the application into the initial set up follows and the thread updating the generator's status is switched on.

The thread function is illustrated in Fig. 4. The thread periodically updates the statuses of inputs/outputs from LAN module and reflects their statuses in UI. The first step downloads the *st0.xml* file from the LAN module where the statuses of all LAN module peripherals are saved. If the download is successful, the connection with the LAN module will start. Consequently, the start of the generator is resolved. That must take place within 5 seconds and this is signalled by the digital input *DIO* of the LAN module. If there is no feedback, the user is informed and the generator will stop. The thread shuts "gets asleep" for 500 ms at the end of each cycle.

The method for generator switching is invoked by the click on push button. The HTTP⁷ request is send to LAN module by GET method. The request is

⁶ eXtensible Markup Language

⁷ HyperText Transfer Protocol

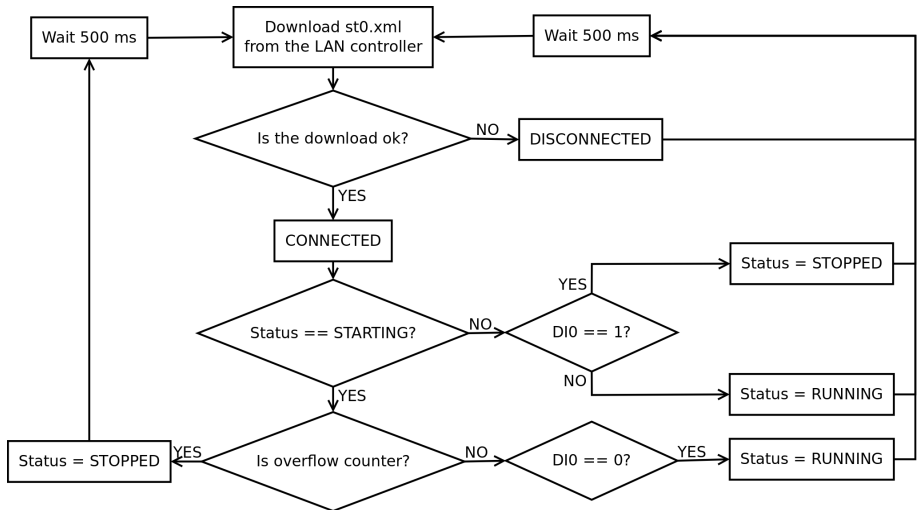


Fig. 4. Function of the thread

prepared by LAN module and switch relay – the generator is running/stopped. On the next time the thread wait on feedback and updates the generator status. [5] [6]

4 Acknowledgement

This work was supported by the project SP2014/188 "Control of technological systems with OAZE providing an independent sustainable development of complex systems 2014" within the Student Grant Agency (VŠB-TU Ostrava).

5 Conclusion

The function of remote control of the electric car system was implemented within the experiment with the universal module – the LAN controller. This module created a "bridge" between the generator's control unit and the user's software running in a computer. The future development of the task could consider the creation of an application for smart phones which would allow controls of the generator via WiFi. That would be much more comfortable nowadays when compared with controls of the system via the computer. Another possibility of extension could be the creation of an own control module equipped with the Ethernet socket or WiFi interface instead on one module.

References

1. IEEE-SA -IEEE Get 802 Program - 802.3: Ethernet. [online]. [cit. 2014-07-28]. Available from: <http://standards.ieee.org/about/get/802/802.3.html>

2. IEEE Standard for Ethernet. In: IEEE Std 802.3-2012 [online]. 2012 [cit. 2014-07-28]. Available from: http://standards.ieee.org/getieee802/download/802.3-2012_section1.pdf
3. Power-line communication. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2014-07-28]. Available from: http://en.wikipedia.org/wiki/Power-line_communication
4. LAN Controller v2.0. In: *Manual* [online]. [cit. 2014-07-28]. Available from: http://files.i4Wifi.cz/inc/_doc/attach/StoItem/2988/LAN_MODULE_manual_EN_v2.pdf
5. *C#: programujeme profesionálně*. 1. vyd. Brno: Computer Press, 2003, xxx, 1130 s. ISBN 80-251-0085-5.
6. Microsoft Developer Network. *System.Net Namespace ()* [online]. [cit. 2014-07-28]. Available from: <http://msdn.microsoft.com/en-us/library/System.Net%28v=vs.110%29.aspx>
7. Microsoft Developer Network. *System.Threading Namespace ()* [online]. [cit. 2014-07-28]. Available from: <http://msdn.microsoft.com/en-us/library/system.threading%28v=vs.110%29.aspx>

Pulse transit time: promising method for continuous non-invasive blood pressure monitoring

Lukas Peter

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

Abstract. Monitoring of changes in blood pressure is one of the most commonly used medical methods because it can be measured non-invasively and it is a very good indicator about the status of the cardiovascular system, and it helps medical doctors to adjust the ideal treatment. Currently is non-invasive blood pressure monitored mainly at discrete time intervals and it is used an inflatable cuff placed around the arm, which is uncomfortable for the patient. It could be very useful to know continuous profile of pressure values during long time period, especially during standard daily activities. It could be used for homecare monitoring and prevention of premature death. Nowadays there are a few technical solutions for continuous non-invasive measurement of blood pressure. This paper is focused on the description of methods for continuous blood pressure measurement, especially on PTT (Pulse transit time) method, which is based on time delay measuring in pulse wave propagation during vascular system. This method could be future of blood pressure measuring.

Keywords: Non-invasive Blood pressure measurement, Auscultatory method, Oscillometric method, Tonometry, Volume clamp method, Pulse transit time method

1 Introduction

On each beat of the heart, the blood is ejected from the ventricular to the aorta. Each ejection produces a force on blood, which creates a pressure wave travelling from the heart to the peripherals vessels. This pressure wave travels along the arterial tree and its flow depends on arterial properties as elasticity, stiffness or thickness of artery wall. The dependence of blood pressure with arterial properties makes it a very good indicator of the status of the cardiovascular system. The monitoring of the cardiovascular system is therefore very important.

The long time continuous monitoring can be very useful to medical doctor for following up the reactions of cardiovascular system in presence of some drugs or medical treatment. Continuous monitoring of BP is useful in many situations, such as in cardiovascular diseases (hypertension), or to prevent falling accidents related to orthostatic hypotension (low pressure).

2 Continuous non-invasive blood pressure measurement methods

When we consider noninvasive and continuous measurement methods, we mean methods not needing skin penetration and the possibility to measure blood pressure values on long time periods.

2.1 Auscultatory and oscillometric method

The most known and the most widely used noninvasive methods for blood pressure measurements are auscultatory and oscillometric methods. These two methods are commonly used during standard examination at the office of medical doctors. Both methods use a cuff placed around the arm, which is inflated at a pressure higher than the systolic pressure, then slowly deflated.

The auscultatory method is based on listening to Korotkoff's sounds under the cuff, which appear when the cuff pressure equal to the SP and disappear when the remaining pressure equal to DP.

The oscillometric method is based on measuring of the pressure oscillations which appear when the pressure in the cuff equals to the SP, are maximum at the mean pressure, and disappear at the DP. These methods cannot be used for long time monitoring because medical doctor's supervision is necessary. Blood pressure can be measured only at time intervals, with more than 3 minutes time delay between successive measurements.

2.2 Tonometry method

The tonometry method is more adapted for continuous monitoring of the blood pressure waveforms. This method is based on applying a controlled force orthogonally to the wall of a superficial artery against a bone.

A force sensor measures the pressure at contact, thus this action on superficial artery produces a local occlusion. It doesn't need a cuff but it needs some supporting device which creates pushing force. The applied force must be small in order not to completely close the artery, as in this case BP is not measured and there is a risk of ischemia.

Tonometry is highly sensitive to motion so it need of a continuous precise positioning of the sensor. So the ideal situation for measurement with tonometry is in the rest position. Device is not comfortable for patient during daily activities.

2.3 Volume clamp method

The first continual noninvasive method for blood pressure monitoring, was introduced and patented by Czech medical doctor Jan Peňáz in 1969. This method is still partly occlusive because it uses a small cuff around the finger to maintain constant the flow of blood under this cuff during each heartbeat. The device uses an inflatable

finger cuff with a built-in photoplethysmography (PPG) sensor and a pressure controller unit placed on wrist of arm.

PPG does not access directly to blood pressure values but can measure blood volume changes in artery. Unfortunately, these volume changes cannot be transformed into pressure values because of the non-linearity of the elastic components of the arterial wall. PPG measures pulse wave it means changes in volume of blood under cuff. Based on these changes is changing pressure in cuff. If blood volume under a cuff is constant, blood pressure value can be evaluated as pressure value in a cuff. Therefore, the continuously changing pressure, which is applied from the outside of finger, corresponds to the intra-arterial pressure and thus it is an instantaneous, continuous measure for arterial blood pressure.

Volume clamp method can be used for long time period monitoring of blood pressure waves. Although in 1985, a study about tissue hypoxia revealed an impairing effect on blood circulation in the finger tip on the distal side of the finger cuff.

2.4 Pulse wave velocity method

One possible way to access BP without using a cuff is to measure the velocity of propagation of pulse waves along the arterial tree. Actually, the velocity of pulse pressure waves propagating along the arterial tree depends on the value of blood pressure. Indirect measurement of blood pressure was performed by continuously measuring pulse wave velocity. The best site for measuring this velocity is located in the aorta because relationship between PWV and BP is only exploitable in central elastic arteries. But it is difficult to perform non-invasively. It is easier to measure arrival of the pulse wave from the first arteries, which begin at the aorta and are more accessible, the common carotids and the common femoral arteries.

The measurement of PWV is the gold standard for investigating arterial stiffness, but it also gives indirect measurements of blood pressure by continuously measuring pulse wave velocity.

The *in vivo* determination of aortic PWV is difficult because it is necessary to detect the time of arrival of a pressure pulse at both the ascending aorta and the iliac bifurcation and to precisely measure the travelling distance of the pulses. The aorta is not easily accessible, neither by optical nor by mechanical means. One possibility is to measure arrival of a pressure pulse at two substitute arterial sites which are close to aorta. The best is measured arrival of pulse wave from the first arteries, which start from the aorta and are accessible, the common carotids and the common femoral arteries.

The PWV measurement method needs to know precisely the distance between sites where signals are collected. Each technique provides different and inconsistent recommendations on how to derive this distance, which is measured manually. These devices require the presence of a skilled operator to manually localize the carotid and femoral arteries. For this reason this measurement is operator dependent and prone to errors.

2.5 Pulse transit time method

Practically the PTT is derived from the time delay between the R-wave of ECG and the arrival of pulse wave on finger, which is measured from PPG. The R-wave occurs at the beginning of the heart systole when the blood ejected from heart to aorta creates the pulse pressure wave which further travels to the periphery. Therefore, the R-wave is a good reference for the beginning of the pressure cycle. Additionally, it is easy to detect.

The PTT method is based on the evaluation of the PW velocity from the travel time between two known locations. Because there is a relationship between PWV and BP, there is the same relationship between time of travel and BP.

It was described that elasticity of an artery is related to velocity of the volume pulses propagating through it. Each contraction of the left ventricle ejects the blood from the heart into the arterial system, which affects velocity of the blood and produces a pressure wave which travels along the elastic arteries. It is this pressure wave, not blood flow, which is felt as the pulse beat. The main factors that determine the speed of propagation of the pulse wave, thus affect PTT value, are the elasticity coefficient, the thickness of the arterial wall, the end-diastolic diameter of the vessel lumen and blood density.

Recently, several methods for cuff-less blood pressure estimation using PTT value were described in the literature. Experiments were conducted on different groups with independent devices used to acquire ECG and PPG and BP simultaneously. The most frequently used method for statistical data analysis is linear regression.

Method based on pulse transit time was tested in clinical conditions.

3 Materials and methods

3.1 Methods description

Clinical test of methodology of blood pressure measurement based on PTT was made in Hospital Podlesi a.s., Trinec. Measuring was agreed by ethical committee. It was measured 15 patients after surgery. Each patient was measured approximately 5mn. It was measured ECG, PPG from finger and ear lobe and invasive blood pressure by catheter. Signals were acquired by developed device.

3.2 Device description

In this study was used measuring circuit for ECG, PPG and blood pressure measuring simultaneously. It was used developed analog device which consisted from six channels. One of them was analog part for one lead ECG measuring, one was used for blood pressure measurement, based on resistor bridge measuring, and four of them was used for transmit PPG measurement. Device was powered by 9V battery which was necessary for safety of patient.

Signal was acquired by measuring card NI6009. For first signal analysis and for saving of data was prepared visualization software in LabView2012.

3.3 Signal analysis

For signal analysis was used MatlabR2013b. Signal analysis is consisted from three parts:

- Detection of R-wave from ECG signal
- Segmentation PPG and BP signal based on time of R-wave
- Searching of significant points from PPG and BP in each segment

ECG signal was filtered by moving average filter with very small width of window, just for smoothing of noise which could affect detection of R-wave. Because of bigger amplitude of ECG signal and because of no so many moving artifacts as during PPG and blood pressure measuring, it wasn't necessary used another filtering method on ECG signal.

Because of very small amplitude of PPG and blood pressure signal, these signals were affected by noise and also by offset voltage. It was necessary used high pass 0.5Hz and low pass 15Hz filter. Values for each filter were chosen because of properties of signal.

Detection of significant points from PPG and blood pressure signal is based on first derivation of signal. For evaluation and testing relationship between blood pressure and PTT it had to look for peak, foot and steepest part of PPG and blood pressure wave.

Currently we have complex of data from each patient where are time interval for R-wave, peak, foot and steepest part from PPG and blood pressure wave and everything is ready for statistic analysis.

4 Discussion

From the doctors' point of view, the better method for continuous noninvasive BP measurement should offer high accuracy and low risk to patient. From the patient point of view, the idealistic method for continuous noninvasive BP measurement should be comfortable, painless, non-occlusive, needing no supervision and with no risk during long time monitoring.

There is still unsatisfied needs for a cuff less, non-invasive method for the continuous monitoring of blood pressure during anesthesia and in critical-care. Among all candidate methods, pulse transit time (PTT) offers the possibility of fulfilling the requirements

PTT method can bring benefice because it indirectly evaluates the BP from existing standard biological signals, it does not need medical supervision, it doesn't use uncomfortable cuff and it can offer long time measurement with no risk to the patient.

PTT is measured as the time interval between R wave from ECG and the time of arrival of pulse wave into the periphery during same cardiac cycle. The time markers on PPG waves differ in a lot of experiments. Some use foot or peak of PPG wave, other

use some point between foot and peak. The peak of second derivation of PPG wave is also used as it is the sharpest change in the signal.

Methods for obtaining PTT from ECG neglect the electromechanical delay between electrical and mechanical work of heart. During systolic phase there exists a time delay between R waves and the opening of aortic valve, called the pre ejection period (PEP). It means that pressure wave doesn't start with R wave but after a short time delay. The PTT time, measured as time delay between R wave and PPG wave, therefore includes the time delay between electrical and mechanical work of heart.

5 Conclusion

In summary, commercial possibilities currently exist for non-invasive monitoring of blood pressure. The devices mostly in used clinically, are based on volume clamp method. But these methods are uncomfortable for the patient for the long term monitoring during normal daily activities. Alternatively, the blood pressure can be evaluated from other bio signals such as ECG or PPG, which can be more conveniently measured from body surface, with no limitation and no disturbance to the patient.

Several authors described the relationship between blood pressure value and PTT. Many authors described different possibilities for evaluation of blood pressure based on PTT. PTT value was compared to other methods, such as invasive blood pressure, volume clamp method, and standard oscillometric method. The results showed high correlation between BP and PTT.

6 Publications

1. Peter, Lukas, and Martin Cerny. "Pulse Transmit Time Laboratory Measurement Solution." *Programmable Devices and Embedded Systems*. Vol. 12. No. 1. 2013.
2. Peter, Lukas, and Martin Cerny. "Hardware for precise in vivo pulse transmit time CNIBP tests." *Applied Electronics (AE), 2013 International Conference on*. IEEE, 2013.
3. Peter, Lukas, Noury, Norbert, and Cerny Martin. "A Review of methods for non-invasive and continuous blood pressure monitoring: Pulse Transit Time method is promising?" (in proceeding IRBM)

The color analysis of transfusion bag

Petra Rajmanová

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`petra.rajmanova.st@vsb.cz`

Abstract. Visual detection of hemolysis is arbitrary and therefore mostly unreliable since it may over and underestimate the actual prevalence of hemolyzed serum specimens (i.e., trained observers are unable to accurately rank the degree of interference in serum). Hemolysis is still one of the biggest challenges to the laboratory specialists. Currently, hemolysis in blood bag is recognized by a visual estimation. Nurses compare the color of blood plasma with reference samples and according to their own opinion they evaluate whether blood plasma is damaged or not. Probably the best sensor for automated analysis is to use image sensors – camera or video camera. Software that is designed in MATLAB programming environment is able to evaluate color of blood plasma according to the reference samples. Determination of color is done in three color spaces, namely, RGB, Lab, and xyY.

Keywords: blood plasma, hemolysis, color recognition.

1 Introduction

Donated blood is process that is indispensable for medicine because donated blood cannot be artificially manufactured or substituted. Blood may be obtained in two ways. The first possibility is to take of full blood with components. The second possibility is a targeted sampling of individual blood components. All the blood is processed into erythrocytes, plasma, and leukocytes. In plasmapheresis the donors took off only blood plasma and other blood components are returned to their body. Hemolysis may occur in a red blood cell unit during blood collection, transportation, preservation and or different stages of handling in the blood bank [1, 2]. Visual detection of hemolysis is possible usually by observing the color of the blood plasma and comparing with table with reference samples.

2 Blood plasma processing

About 55% of blood is blood plasma, a fluid that is the blood's liquid medium, which by itself is straw-yellow in color. The blood plasma volume totals of 2.7–3.0 liters (2.8–3.2 quarts) in an average human. It is essentially an aqueous solution containing 92% water, 8% blood plasma proteins, and trace amounts of other materials.

There are two ways of taking of blood plasma. Early transfusions used whole blood, but modern medical practice commonly uses only components of the blood, such as red blood cells, white blood cells, plasma, clotting factors, and platelets. The blood is taken into plastic bag, see Fig.1. Donated blood is usually subjected to processing after it is collected, to make it suitable for use in specific patient populations. Collected blood is then separated into blood components by centrifugation: red blood cells, plasma, platelets, albumin protein, clotting factor concentrates, cryoprecipitate, fibrinogen concentrate, and immunoglobulins (antibodies). Red cells, plasma and platelets can also be donated individually via a more complex process called apheresis. Blood plasma can be used in hospital or it can be shipped to industrial factory for processing.



Fig. 1. Transfusion plastic bag

3 Damage to the blood plasma

Hemolysis is the breakage of the red blood cell's membrane, causing the release of the hemoglobin and other internal components into the surrounding fluid. Hemolysis is visually detected by showing a pink to red tinge in serum or plasma. Hemolysis can occur from two sources. The first in-vivo hemolysis may be due to pathological conditions, such as autoimmune hemolytic anemia or transfusion reaction. And second in-vitro hemolysis may be due to improper specimen collection, specimen processing, or specimen transport [3,4].

4 Visual inspection

It has long been known that visual assessment of the degree of actual concentration of hemolysis is mostly unreliable. Visual inspection by laboratory personnel is highly unreliable, depends on the sample type and may overestimate the actual prevalence of hemolyzed serum specimens while underestimate it in plasma samples. Visual detec-

tion is based on comparison with photographs of samples containing various concentrations of hemoglobin. The maximum amount of hemoglobin in the blood plasma could be 1g/L. If there is more than 1g/L, the blood plasma is damaged. Visual inspection is based on color analysis.

5 Color analysis

Color detection is done by optical methods that may be spectrophotometric, objective and subjective methods. Spectrophotometric method works by absorption and emission of radiation. The measurement uses a spectrophotometer. Objective methods detect colors such as photographic apparatus or camera and subjective methods used to analyze color vision, glasses or microscope. To detect hemolysis of the blood product is best to use the camera or camcorder.

Color is associated with the properties of the object, of the light and of the human perception mechanisms. Color is a psycho - physical phenomenon and this is the reason why it is impossible to use the human eye considered for accurate meter of color.

6 Automated processing

Automated processing could be a new way of analysis of blood plasma. For this method could be used optical sensors. Color can be recorded by camera or video camera. Information is summarized on the image. This picture is processed by a special program. Besides recognition of color, there is possible to get information about barcode and check integrity of the bag. The measure chain of automated analysis consists of light source, blood plasma, recorder and processing on computer.

Light is a physical point of view the source of electromagnetic radiation and it has a very big effect on the perception of color. Light sources can be natural such as the sun, fire, and electricity in gases or artificial, for they are considered light bulbs, fluorescent lamps or LEDs. When shooting is best to use natural light source, the picture then resembles reality more than using artificial sources. Blood plasma has to be checked before its freezing because after freezing, blood plasma is covered by a white frost. White frost has its influence on recognition. Recording medium can be camera or video camera. Processing on computer is last step in recognition [5, 6].

6.1 Designed automated processing

The algorithm for automated recognition was designed in MATLAB. Pictures have been taken by camera in Blood Centre of the Faculty Hospital in Ostrava. Hemolysis has been recognized in three color spaces. After that, it has been compared which model is the best. And the best of them will be used for next processing. The color spaces are RGB, Lab, xyY. And reference colors come from the table that is used by visual inspection, see Fig. 2.

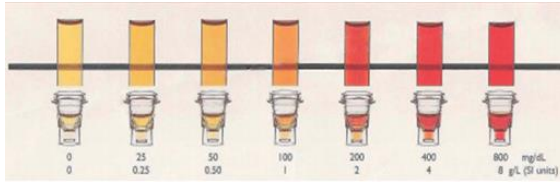


Fig. 2. Color scale for assessing degrees of hemolysis, expressed in g/l [mg/dl] hemoglobin.

RGB color space.

The RGB model is composed of R = red color, G = green color and B = blue color. The model is based on knowledge reference colors. These references colors signify the level of hemolysis. These samples have been taken from the table that is used for visual analyzing, see Fig.2. The algorithm evaluated five blood plasmas wrongly. It detected hemolysis for clear blood plasma. Mistakes were in the first half of testing in RGB color space.

LAB model.

The second model is the Lab model. L is brightness component “a, b” and represents colors of components. L component is unnecessary. The algorithm does not use it for recognition. The algorithm uses only component “a, b”. The reference colors come from tab; see Fig. 2. The values are transferred from RGB to XYZ and from XYZ to Lab by transformational equation:

$$\begin{aligned}
 X &= 0.430574 \cdot R + 0.341550 \cdot G + 0.178325 \cdot B \\
 Y &= 0.222015 \cdot R + 0.706655 \cdot G + 0.071330 \cdot B \\
 Z &= 0.020183 \cdot R + 0.129553 \cdot G + 0.939180 \cdot B
 \end{aligned}$$

$$\begin{aligned}
 L &= 116 \cdot \left(\frac{Y}{Y_n}\right)^{1/3} - 16 \\
 a &= 500 \cdot \left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Y}{Y_n}\right)^{1/3} \right] \\
 b &= 500 \cdot \left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Z}{Z_n}\right)^{1/3} \right]
 \end{aligned}$$

There is focus calculated for every individual reference sample with coordinates “x, y”. Blood plasma is classified to the nearest sample. The algorithm evaluated two blood plasmas wrongly.

xyY model.

Last model is xyY model. This model includes the extent of human seeing. Y is brightness and colors components are “x, y”. Transformational equation transfers values from RGB to XYZ and from XYZ to xyY. Algorithm does not use Y component.

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

$$Y = Y$$

The model xyY is the worst way for recognition of hemolysis in blood plasma. The number of wrong results is eleven and software detected two hemolytic blood plasmas as clear blood plasma. The algorithm can recognize color of blood plasma on knowledge reference colors. These colors come from the table; see Fig. 2. The plan uses real color of blood plasma and does reference colors from these samples. When blood plasma was on the LCD pad, the barcode was in the middle. Hemolysis was detected from whole bag. The plan is recognize hemolysis just from peripheral parts without middle part, where the barcode is.

This method can be enlarged on other functions. One of the functions can be checking barcodes. Nurses do this checking. They are checking integrity of bag too. This can be done by algorithm too. The nurses could have more time for another more important work.

7 Results

The aim of this work was to design a method for automatic analysis of blood plasma. This method is designed in close cooperation with Blood Centre of Faculty Hospital Ostrava. Color of blood plasma was analyzed in three color spaces RGB, Lab, xyY. Pictures have been taken by camera and light sources, such us daily light, LED light and LCD display light. The worst results were in xyY color space, because this color space is based on human seeing. This can probably be the reason for the worst recognition. Majority of mistakes were caused by bad light source. Software was designed in MATLAB. It has many functions. Software can recognize hemolysis of blood plasma in three color space. Future plan is to work out the software in detail, to make new reference samples from real plasma and to use light conditions at high quality.

Acknowledgment

This work is also supported by project SP2014/188, named "Control of technological systems with OAZE providing an independent sustainable development of complex systems" of Student Grant Agency (VSB - Technical University of Ostrava).

References

1. Goumassi, E., Dijkstra – Tiekstra, M.J., Hoentjen, D. and de Wildt, J. – Eggen, J. Haemolysis of red blood cells during processing and storage. *Transfusion*, vol 52, Mar. 2012, pp-489-492, doi:10.1111/j.1537-2995.2011.03298.
2. Lippi, G. and Franchini, M. Advancements in laboratory diagnostics: an invaluable tool for assessing quality of blood transfusions. *Blood Transfusion*, vol. 58, Dec. 2012, pp. 1-2, doi:10.2450/2012.0226-12.
3. Hess, J.R., Sparrow, R.L., van der Meer, P.F., Acker, J.P., Cardigan, R. A., and Devine, D.V. Red blood cell haemolysis during blood bank storage: using national quality management data to answer basic scientific questions. *Transfusion*, vol. 49, Dec. 2009, pp. 2599-2603, doi:10.1111/j.1537-2995.2009.02275.
4. Janatpour, K.A., Paglieroni, T.G., Crocker, V.L., Du Bois, D.J., and Holland, P.V. Visual assessment of haemolysis in red blood cell units and segments can be deceptive. *Transfusion*, vol. 58, Jul. 2004, pp. 984-989.
5. Machacek, Z., Slaby, R., Hercik, R., and Koziorek, J. Advanced System for Consumption Meters with Recognition of Video Camera Signal. *Elektronika Issue*, vol. 18, no. 10, 2012, pp. 57-60, ISSN 1392-1215.
6. Slaby, R., Hercik, R., and Machacek, Z. Compression methods for image processing implementation into the low capacity device. In *Technical Gazette*, vol. 20, no. 6, 2013, pp. 1087-1090, ISSN 1330-3651.
7. Fraser, B. Murphy, C., and Bunting, F. *Real World Color Management*. 2nd ed. Peachpit Press, 2005, ISBN 0 – 321 – 26722 – 2.

Measurements for remotely controlled power generator

Jakub Rawicki¹, Vít Otevřel², and Bohumil Horák²

¹Faculty of Energy and Environmental Engineering, Power Engineering,

ul.Kon:

Poland

²Department of Energy and Environmental Engineering, FEECS,

VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{vit.otevrel, bohumil.horak}@vsb.cz

Abstract. Nowadays distributed generation gained relevant meaning throughout the world, especially in places where electricity supplies are threatened because of wide range of reasons. The problem concerns countries where energy consumption grew rapidly in past years (existing electrical grid system not sufficiently developed) as well as areas which are difficult to be supplied from different causes. Among them distance from main energy sources, specific weather conditions or landform can be highlighted [1]. The way of managing the process of delivering electrical energy to local grids is crucial not only from economical point of view but also for our safety as receivers. It is what this project is about.

Keywords: generator, distributed generation, measurement, microcontroller, converter, solar inverter, energy sources

1 Introduction

In this case the 5,5 kW petrol power generator is the main energy source. The AC from it must be transformed to DC by AC/DC converter. Then current should be delivered to solar inverter which is able to adjust its parameters to the grid (this kind of device is commonly used to transfer power from solar panels). Below general schematic is presented:



Fig. 1. General schematic of the system

The point is to enable potential owner to operate the whole system by Ethernet using power line. It concerns being able to turn on the generator as well as being fully informed about current parameters such as:

- temperature,
- angular velocity of internal rotor,
- value of current from generator and amount of power delivering to the grid.

This paper will be focused on measurement part of the issue and transferring data into computer using ADC converters built in microcontroller.

2 Measuring and controlling the power

The solar inverter adopted to this project enables monitoring of current power value sending to the grid[6]. What is more, there is possibility to control this parameter by software called Sunny Data Control [4] along with parameters from DC side. The problem is to send data into PC. Nevertheless the solar inverter provides owners with communication interface based on RS 485 protocol. To make signal understandable for computer we need to convert it for example to USB. There are at least two solutions:

- connecting solar inverter directly to computer using RS485/USB wire,
- connecting solar inverter with RS485/Ethernet converter.

The disadvantage of first method is that owner is limited by the length of the cable. Although this is the simplest way of connection. The second option is much more convenient for user but it requires access to the Ethernet and is more expensive because of price of the converter. An another flaw is necessity of using additional power supplier for the converter.

The experiment of discharging the car battery was undertaken, to check if everything works correctly. Despite listed disadvantages the second possibility was chosen as more comfortable.



Fig. 2. The general view of the connection



Fig. 3. The RS485/ethernet converter

The part of data, concerning amount of power delivering to the grid, received from the experiment is presented on the graph below. The measurement was executed every 60 seconds.

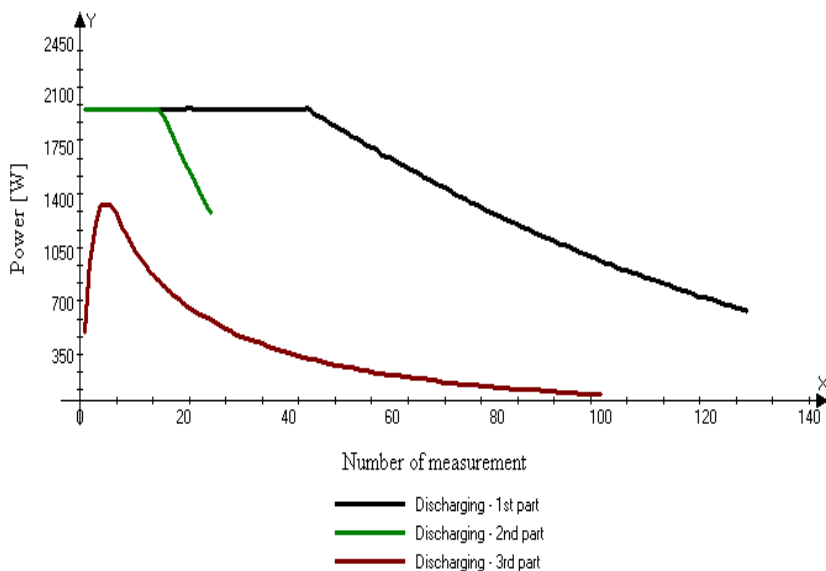


Fig. 4. The characteristics of discharging car battery

The breaking point in each characteristic is the point when the amount of energy in the car battery was too small to send set value of power to the grid, which in this case was 2kW. Because voltage of DC side was set to certain value (at start 320V) it was current which was getting smaller. To make solar inverter sending 2kW of power again, it was necessary to decrease voltage. In third stage of discharging the level of 2kW was unachievable. The peak of the last characteristic is about 1,4 kW.

To equip a future user with software already fulfilling all requirements, probably some functions from YASDI [5] (yet another SMA data implementation) will be used. It will liquidate a necessity of additionally downloading Sunny Data Control.

3 Measuring AC from generator

The value of current, to be accurate its root mean square, can exceed 10 A. Because of that a conventional way of measuring current is restricted. The alternative solution of this problem has been figured out.

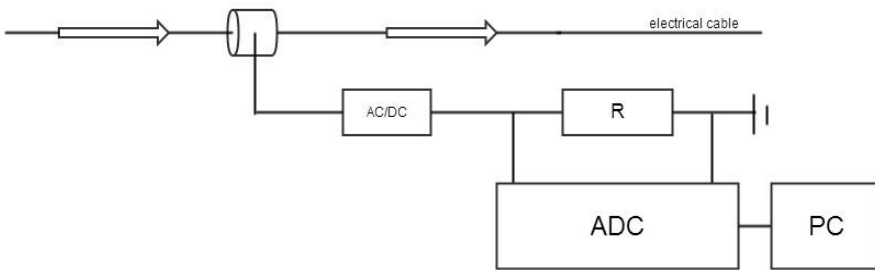


Fig. 5. Measuring AC in non-invasive way

In this method the non-invasive current sensor will be used like a main measurement part. Then transferring data to PC requires a few additional elements. First and foremost there is a need of conversion the current into the voltage. It can be achieved by AC/DC converter and resistor (for devices which are about to be applied, resistance will be 60Ω).

Next stage is to deliver information about voltage, which is kind of analog signal, to computer which requires digital number. To do it ADC converter is necessary. In case of this project embedded Pi will be used. That device has 3 ADC converters built in. Its main part is STM 32F103 microcontroller[2].

4 Temperature and angular velocity

Temperature will be measured by K type thermocouple. It will give an information about current temperature of exhaust fumes. This parameter is essential for controlling the whole process because it shows important changes in way of engine work. To receive accurate temperature it is inevitable to use additional temperature sensor (for instance one wire digital temperature sensor such as DS18B20 [3]), which will deliver

temperature of the generator surrounding (thermocouple provides user only with difference of temperatures between its two endings).

The voltage from thermocouple will be too small taking quantization of microcontroller into consideration. Assuming maximal possible temperature (about 200°C) its value will be approximately $8000\mu\text{V} = 0,008\text{V}$. The quantization number is:

$$2^{12} = 4096$$

The range of voltage for microcontroller is from 0 to 3,3V. After dividing this value by quantization number the range of every state will be a score:

$$\frac{3,3\text{V}}{4096} \approx 0,0008\text{V}$$

Another calculation shows :

$$\frac{0,008}{0,0008} = 10$$

that there would be only 10 states of measuring temperature. It is clearly not enough. That is why additional transistor for amplifying the voltage from thermocouple is necessary.

In the measurement of angular velocity of internal rotor barrier sensors as well as optic sensor can be used. Thus, for this project targets typical bike counter turned out to be sufficient.

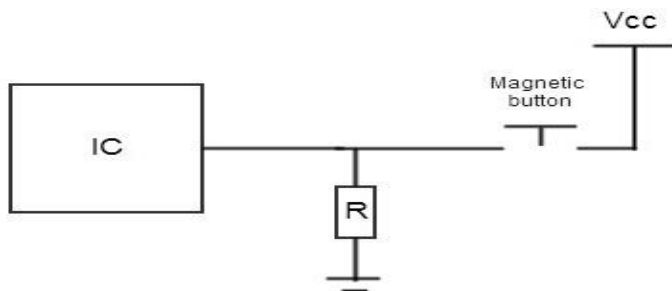


Fig. 6. Schematic of circuit with bike counter as a magnetic button.

Every time when there is a contact between sensor and magnet mounted on the rotor, the circuit is being closed. The voltage triggers high state at input pin of microcontrol-

ler and can be adjusted by well chosen power supply. There is only need to count the number of impulses received by microcontroller.

5 Conclusion

All solutions suggested here make handling of generator much more convenient for the users. In spite of the fact that some parts of this project can be treated like a kind of curiosity (such as automatic turning the generator on) the whole issue is undoubtedly significant. The process of controlling electricity production in self-sufficient energy systems has an influence on both finances and safety. That matter becomes even more relevant when there are more than one energy sources (wind turbines, solar panels etc.). It concerns also heating systems like heating pumps or solar thermal collectors. A possibility of taking some actions from the distance spares the time and restricts the necessity of manual interventions.

Acknowledgement

This work is supported by project SP2014/188, named "Control of technological systems with OAZE providing an independent sustainable development of complex systems" of Student Grant Agency (VSB - Technical University of Ostrava).

References

1. Chmielniak T., *Technologie energetyczne*, Wydawnictwo Naukowo Techniczne, 2013
2. Paprocki A., *Mikrokontrolery STM 32 w praktyce*, Wydawnictwo BTC, Legionowo 2011
3. *DS18B20 Programmable Resolution 1-Wire Digital Thermometer*. 2008. [31.07.2014]. Available from: <http://datasheets.maximintegrated.com/en/ds/DS18B20.pdf>
4. *Sunny Data Control*. [31.07.2014]. Available from: http://www.sma-america.com/en_US/products/software/sunny-data-control.html
5. *Implementation of the SMA Data Protocol*. [31.07.2014]. Available from: <http://www.sma-iberica.com/fileadmin/fm-dam/documents/YASDI/YASDI-10NE1106.pdf>
6. *Sunny Boy 3000TL/4000TL/5000TL*. 2014. [31.07.2014]. Available from: <http://www.solarcentury.com/downloads/SB5000TL-DEN102030.pdf>

Remote Control of Power Line Socket

Junwen Shen¹, Vít Otevřel², and Bohumil Horák²

¹Electrical Engineering and the Automatization Specialty,
Shanghai University
99 Shangda Road, BaoShan District, Shanghai
shenjw92@gmail.com

²Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{vit.otevrel, bohumil.horak}@vsb.cz

Abstract. In this paper, it introduces the way to control the power socket wirelessly. Through the message sent from the computer, microcontroller controls the sockets' on and off. Then devices get energy and run. People can run the whole power system in the office or even at home by controlling the socket.

Keywords: Socket, Remote control, Network, Microcontroller

1 Introduction

As the developing of science and technology, the power socket also need some improvement on the command field. For the convenient of controlling the whole power system, the paper shows how to long-ranged control the sockets.

1.1 Similar solution

There is some existed solutions. For example, LAN controller (Fig. 1) can serves as a web server ,but it has long been lacking in the market network solution, only in ISP network.



Fig. 1. LAN controller

On the figure 2 is shown control program for computer. This program sent a request to the LAN module and LAN controller switching the relay.

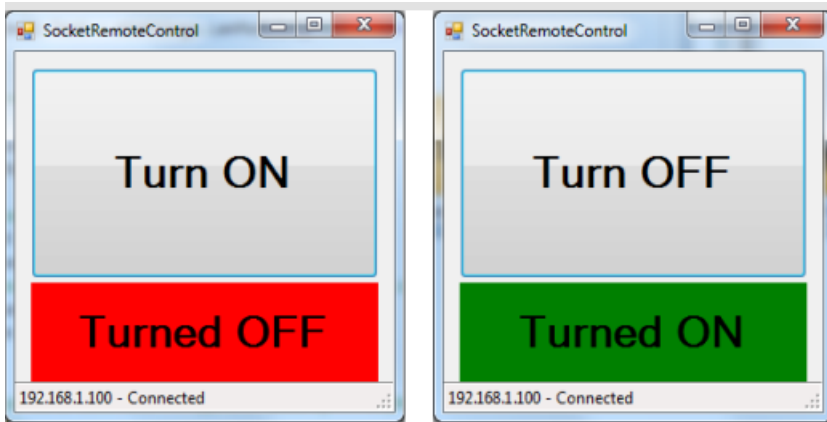


Fig. 2. Screenshot of LAN controller application

1.2 Present solution

The present solution allowed many computers to control the system in the same time. And it can also control the sockets for long distance. The designed codes help people to control devices automatically. It makes the power system easier to be controlled without people.

2 Experiment

The solution need raspberry pi, wires, remote controller and sockets. The remote controller is FHT-7901-T transmitter. It includes 1 remote and 3 power plug. It could run in 433.05 - 434.79MHz. There are a couple of buttons to control the same outlet. One is for run, another is for stop. The chip on the FHT-7901-T transmitter is PT2272. It works as an information sender. And the chip on the socket runs as a receiver. The solution remove the buttons and uses microcontroller to control the socket. Firstly, computer sends a message to the microcontroller through network. Then, the chip on the raspberry pi receives the message and gives order to the remote controller. The remote controller runs as a switch. It sends a radio to control the sockets on and off. Block scheme of the experiment is shown on Fig. 3.

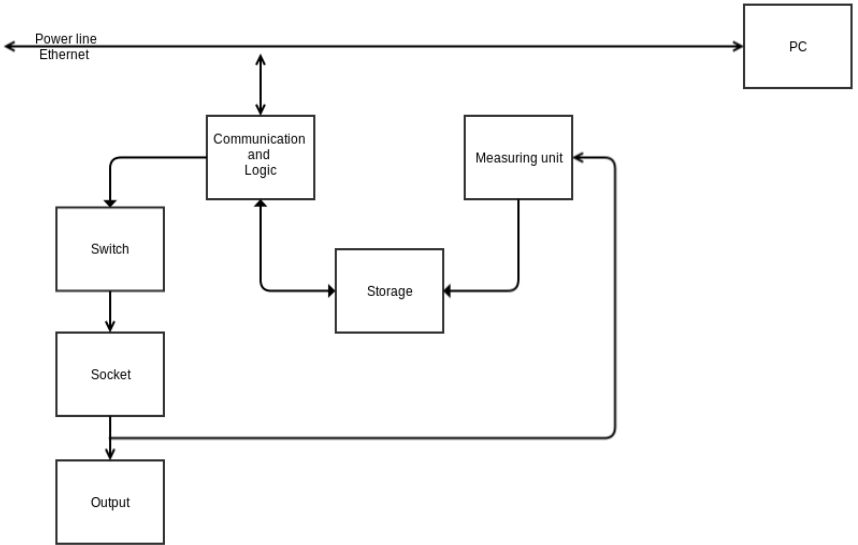


Fig. 3. Block scheme of the system

3 Control application

Remove the button from the board. Use wires to link the positive side and ground. Connect the raspberry pi with remote controller by pins. And configure display and keyboard to the pi.

At last, when the codes run, the board can run the ‘on’ button for B outlet. It means that the microcontroller could replace the button and control the socket.

The microcontroller on raspberry pi can control the voltage high or low. So link wires to the relay. It can also control the outlet directly, because it is operated by voltage.

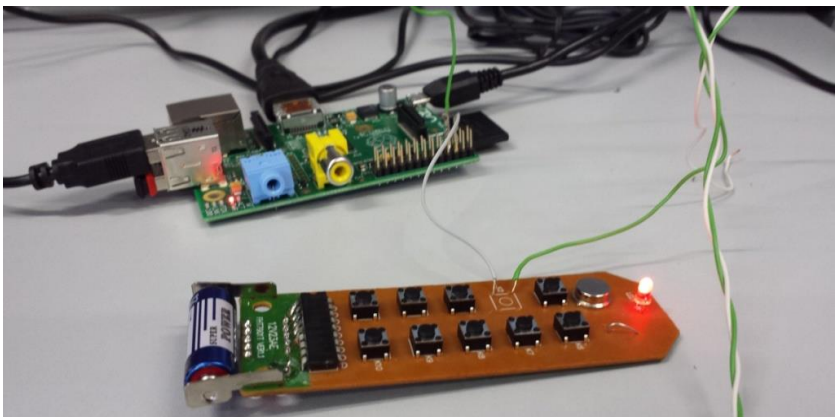


Fig. 4. Raspberypi and controller for sockets

For the solution, it need an SD card that is 4GB or greater in size. Then put it into computer. Download the NOOBS zip file which is necessary for the raspberry pi. Extract the zip. Copy the extracted files onto the SD card that you just formatted so that this file is at the root directory of the SD card. Next, insert the SD card into the Pi and connect the power supply. Finally, run the system and insert the name and password. Write codes through the python GPIO. Function of the python program shows on the Fig.5.

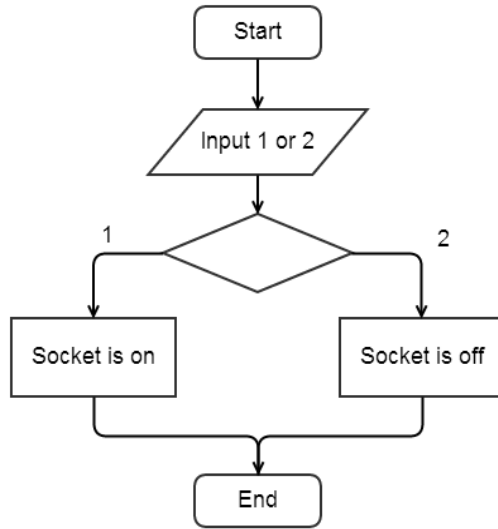


Fig. 5. Diagram of the function

First note the IP address of the Raspberry Pi. Open the python. Write codes ‘ifconfig’ in python. On the second line, it shows "inet addr:69.164.xxx.xxx" on the screen. It is the IP address of your Raspberry Pi.

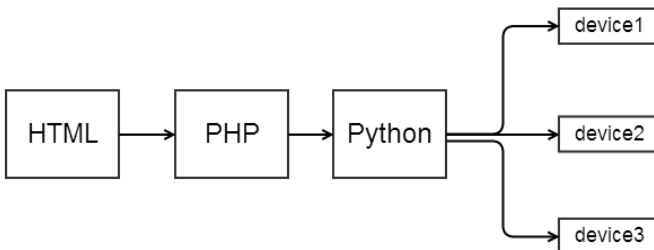


Fig. 6. Diagram of the web application

HTML is used to build the website. Type the Notepad with html to create the heading and buttons. The title is “Virtual Switch” and two buttons are ‘On’ and ‘Off’. The website has the same role of the button. If was clicked on the ‘ON’ button, the socket is on. If click ‘OFF’ button, the socket is off. Python make this function come true. So it needs GPIO codes as below to control several sockets. PHP works as a bridge. It links HTML and Python together. Through the button on the website, it can run the codes typed by python, and then redirect to the first page. The website runs as following.



Fig. 7. Web UI for control

All computers in the local wifi environment can control sockets through controlling the raspberry pi wirelessly.

4 Acknowledgement

This work was supported by the project SP2014/188 ”Control of technological systems with OAZE providing an independent sustainable development of complex systems 2014” within the Student Grant Agency (VSB-TU Ostrava).

5 Conclusion

The solution uses the network to control the microcontroller. Through the orders from microcontroller, the socket is run or stop. It is good to control the devices which work in day and night. People only need to work inside the room and control the socket outside. It still has some space to improve. The power between the raspberry pi is small (3.3V). And it doesn't have PWM pin. It needs to connect with Embedded Pi, which is better on controlling devices.

Reference

1. Embedded Pi User Manual. [On-line]. [2013-05-16]. Available from: http://www.coocox.org/Embedded_Pi/Embedded_Pi_User_Manual.pdf

2. *How to Make a Raspberry Pi Web Server*. [On-line]. [2014-08-02]. Available from:
<http://www.wikihow.com/Make-a-Raspberry-Pi-Web-Server>
3. *LAN Controller V2.0*. [On-line]. Available from:
http://files.i4wifi.cz/inc/_doc/attach/StoItem/2988/LAN_MODULE_manual_EN_v2.pdf
4. *Setting up your Raspberry Pi*. [On-line]. Available from:
<http://www.raspberrypi.org/help/quick-start-guide/>
5. *循序渐进学Python之函数入门 – The gradual learning Python function entry (1)*. [On-line]. [2008-09-01]. Available from:
<http://developer.51cto.com/art/200809/88052.htm>

Distance Determination of Objects of Cluster of n-dimensional Structure

Roman Slabý, Jana Nowaková, and Martin Hasal

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{roman.slaby, jana.nowakova, martin.hasal}@vsb.cz

Abstract. In the paper, the way how to find distance of the objects of the cluster of n-dimensional structures is described. As a cluster it can be considered any objects, where object can be e.g. elements of substance or any points, which create coherent cluster. When structure is defected it is not possible, in all cases, to determine the distance between objects clearly and that is why the distance can be determined only on the base of statistical methods. The next step will be design of the method, which will use basic characteristics of exploratory analysis for distance determination of objects of cluster. Method is appropriate and useful for cluster of n-dimensional structures created by cells with object situated in their tops.

Key words: Cluster, matrix structure, distance determination.

1 Introduction

Cluster analysis divides data into groups (clusters) that are meaningful, useful, or both. It is a data reduction tool that creates subgroups that are more manageable than individual data. Like factor analysis, it examines the full complement of inter-relationships between variables. In some cases, however, cluster analysis is only a useful starting point for other purposes, such as data summarization. Whether for understanding or utility, cluster analysis has long played an important role in a wide variety of fields: psychology and other social sciences, biology, statistics, pattern recognition, information retrieval, machine learning, and data mining. [1]

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them.

As a cluster of n-dimensional structure it is considered cluster with defined location of objects. The smallest part of n-dimensional structure is called cell. Repeated cells fill the space of matrix. The basic cell contains only objects, which are situated in its tops as it is depicted in Figure (1). [2], [3]

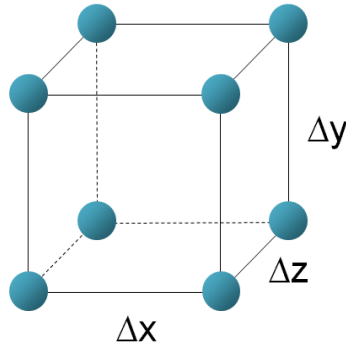


Fig. 1. Basic (unit) cell

Different shapes of cells are known according to the structure of cluster

1. 2D structures - flat structures, where distances of objects located on axis x and y can be any and
 - the axis are perpendicular to each other or
 - not,
2. 3D structures,
 - triclinic - the structure is described by three vectors of unequal length and none of the three vectors is orthogonal to another,
 - monoclinic - the structure is described by three vectors of unequal lengths and they form a rectangular prism with a parallelogram as its base. Hence two pairs of vectors are perpendicular, while the third pair makes an angle other than 90° ,
 - orthorhombic - the structure is described by three vectors of unequal length and all three vectors intersect at 90° angles,
 - tetragonal - the structure is described by three vectors, two of them create a square base and the third one is any, but all three vectors intersect at 90° angles,
 - trigonal - the structure is described by three vectors of equal length, but the angles of intersection of the vectors are not equal to 90° ,
 - hexagonal - see Figure (2),
 - cubic - the cell is in the shape of a cube, it is one of the most common and simplest shapes.

Undisturbed basic cell, sometimes called as ideal cell, has regular shape and contain objects in every its top. But in many cases, because in real life nothing is ideal, cell can be disturbed - it means that some object can be deflected, some object can be missed or other foreign object can be in addition. The disturbances of cells cause damages in structure of cluster and problems in finding of distance between objects on defined axes, or in other way said, assessment of roughness of cluster structure. [4], [5], [6]

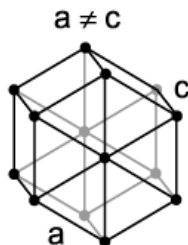


Fig. 2. Hexagonal structure

Roughness of cluster structure is defined as a distance between objects on the axes. The smaller the distance is, the finer the cluster structure is.

For finding of the distance it is needed to know the location of objects, which is determined conclusively by coordinates.

2 Determination of Distance between Objects

The distance of cluster objects of n-dimensional structures can be determined in many ways, but the simplest way is to find adjoining object on axis with the smallest distance from a reference object $r = [x_r, y_r, z_r, \dots, n_r]$. This can be done using simple algorithm with low requirement for calculation and calculation time for determining of roughness of cluster structure. According to the shape of the cell, it is necessary to calculate the distances $\Delta x, \Delta y, \Delta z \dots \Delta n$ for all axes for selected object $o = [x_o, y_o, z_o, \dots, n_o]$

$$\begin{aligned} \Delta x &= |x_o - x_r|, \\ \Delta y &= |y_o - y_r|, \\ \Delta z &= |z_o - z_r|, \\ &\vdots, \\ \Delta n &= |n_o - n_r|. \end{aligned}$$

(1)

Described way can be realized only for structures, which are created of undisturbed basic cells, it means the cells without any disturbances. [5], [8], [9]

In case when the cell is damaged, it is not so easy to determine the distance between objects conclusively. When at least one object of grid is deflected from axis, the calculation is not right, because for any other objects located on the same edge, the smallest distance of the objects is defined as a difference of object location and location of the deflected object in the edge. The same problem is caused by presence of foreign object. According to the outlined problems, if it is the cluster of points in space, it is necessary not only to find the nearest object

on the same edge of the cell and assess the distance Δx , Δy and Δz (in case of 3D space).

It is needed for every object to determine the distance to the nearest object in the axis. Doing this, the sample to every axis is get and it contains data represented the distances. The way of assembling of the sample for cluster of n-dimensional structures with flat shape is defined in Figure (3).

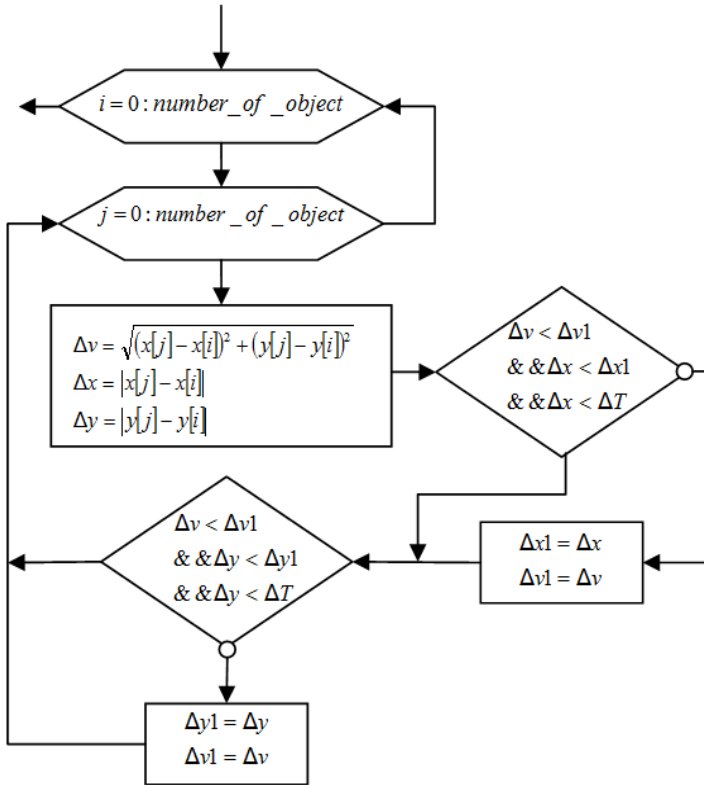


Fig. 3. Flowchart for setting of the sample

Proposed method is based on going through of all objects of the cluster and specification of distance Δv of selected object and reference object in the plain and withal distances Δx , Δy and Δz between objects in all axis

$$\Delta v = \sqrt{(x_o - x_r)^2 + (y_o - y_r)^2 + (z_o - z_r)^2}. \tag{2}$$

It allows us to find distance between objects for every axis of axis cross. The deflection of any object or presence of foreign object is needed to eliminate, if the smallest distance should be found.

The elimination of errors in grid structure is realized using thresholding level ΔT . The distances between objects in axis, which are smaller than the thresholding level ΔT are not considered as right values, because probably, they are caused by error in matrix structure. The thresholding level ΔT could acquire for every axis different value, which could eliminate the effect of errors in structure in particular axis. Cluster structure with disturbances in plain XY (for easy graphical representation) is shown in Figure (4). The structure of the depicted cluster is disturbed due to deflection of some objects from the tops of the cell and also due to presence of foreign objects depicted in red colour.

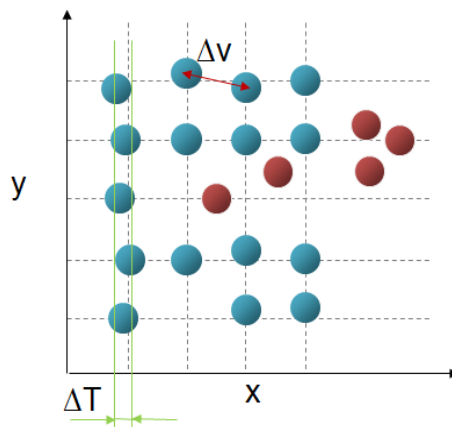


Fig. 4. Cluster with disturbed matrix structure

3 Conclusion

It was presented method for easy clustering of object in n-dimensional space. The method was implemented and the next step is to defined method, how the thresholding level ΔT is find and assess the distance of the objects. The proposed method will use the basic characteristics of exploratory analysis for distance determination of objects of cluster. Method is appropriate and useful for cluster of n-dimensional structures created by cells with object situated in their tops.

Acknowledgements This work has been supported by Project SP2014/156, "Microprocessor based systems for control and measurement applications", of the Student Grant System, VŠB - Technical University of Ostrava.

References

1. Aldenderfer, M. S. and Blashfield, R. K.: Cluster Analysis. SAGE Publications (1984)
2. Brito, P., Bertrand, P., Cucumel, G. and De Carvalho, F.: Selected Contributions in Data Analysis and Classification. Springer Verlag (2007)
3. DuToit, S. H. C., Steyn, A. G. W. and Stumpf, R. H.: Graphical Exploratory Data Analysis. Springer London, Limited, London (2011)
4. Everitt, B. S., Landau, S. and Leese, M.: Cluster Analysis. Wiley (2001)
5. Frederick H. and Dearing B. E.: Exploratory Data Analysis. SAGE Publications (1979)
6. Mirkin, B. G.: Clustering for Data Mining: A Data Recovery Approach. Taylor & Francis (2005)
7. Mirkin, B. G.: Mathematical Classification and Clustering. Kluwer Academic Publishers (1996)
8. Romesburg, C.: Cluster Analysis for Researchers. Lulu.com (2004)
9. Walter, S. and Deloudi, S.: Crystallography of Quasicrystals: Concepts, Methods and Structures. Springer (2009)

Exercise-based Games for Balance Rehabilitation

Maros Smondrk and Jana Krohova

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{maros.smondrek, jana.krohova}@vsb.cz

Abstract. This paper deals with the development of the exercise-based games for balance rehabilitation. The balance rehabilitation as a part of physical therapy is used for the functional diseases of the musculoskeletal system. The proposed rehabilitation software is used for capturing the patient movements on the wobble board and should lead to objective evaluation of the quality of rehabilitation and its progression. The proposed exercise-based games were designed in cooperation with the Therapeutic Rehabilitation Clinic of the University Hospital of Ostrava.

Keywords: exercise-based games, balance rehabilitation, wobble board

1 Introduction

In general, sedentary lifestyle and lack of the regular exercise usually leads to formation of functional diseases of the musculoskeletal system. These factors are significantly involved in the formation of various muscle imbalances, incorrect posture, incorrect movement patterns, increased risk of accident conditions and risk of falls in elderly, [1]. The balance exercise is commonly used to treat aforementioned problems. The benefits of these exercises have been reported in a number of significant research studies. [2], [3]

The main principle of balance exercises is to balance on the rehabilitation tool. The rehabilitation tool could be in form of balance pads, balance boards and balls. The most commonly used balance tool is a wobble board. Its spherical shape allows a patient to balance in the range of 0 to 30 degrees in each direction. Balance exercises could be summarized as general motion patterns such as motion in a specific direction (e.g. ahead, back, left, right) or maintain equilibrium posture. Standard rehabilitation program consists of few exercises that are performed several times per week.

The main task of the physiotherapist is visual control of patient's motion in order to ensure that the patient performs the exercise correctly. The physiotherapist controls the patient posture and involved muscle groups, as well. The patient's and the physiotherapist's active participation is essential because of the overall effectiveness of the therapy. The quality of the rehabilitation and patient's progress is usually evaluated only subjectively by physiotherapist.

2 Problem Definition

The idea of objective assessment of important parameters such the quality of rehabilitation and patient's progress has reasonable importance and it was discussed in many research studies [4] - [6]. These studies mainly deal with using visual feedback in balance exercise and motion sensors integrated into balance tool in order to measure patient's motion during exercise. Measured movement of the rehabilitation tool or patient's center of mass is visualized in software as a virtual point or is represented as a control input for motion of an object (ball) in 3D virtual game. Most of the aforementioned authors used open source games which could be changed only by programmers, but not by physiotherapist. Moreover, the usability of the rehabilitation games has been demonstrated on the healthy participants. From this point of view, the design of the games or rehabilitation exercises should be more addressed to the patients with different types of motion ability. The most appropriate approach is to implement rehabilitation exercises which are common in rehabilitation practice. An equally important part of software should be a user selectable difficulty levels with respect to patient's diagnosis. For instance many injuries of lower extremities do not allow a patient to maintain equilibrium posture on the balance tool, even tracking specific motion pattern. The measured data management is another important part of the rehabilitation software. The measured data should be stored for post-processing analysis in order to evaluate quality of the performed exercise. This value should be represented in meaningful way i.e. as a score. This score would have great effect on the patient motivation in rehabilitation. To overcome the aforementioned constraints we have developed rehabilitation software for balance exercise, which is discussed in more detail in the next section.

3 Balance Rehabilitation Software

The rehabilitation software was created in the development environment MATLAB R2011b. The proposed rehabilitation software was designed and tested in cooperation with the Therapeutic Rehabilitation Clinic of the University Hospital of Ostrava. The main window can be divided into two parts. Part A entitled "Registration of a new patient" and Part B entitled "Selection of registered patients" (Fig. 1).

The part A is intended to register new patient to the patient database. The user has to enter the basic patient information such as name, personal identity number (ID), height, weight, diagnosis, additional diagnosis, affected limb, and should enter notes, if it is needed.

The part B of the program is designed for search and selection of patients who are already registered. The user can choose from two search criteria: search by patient's first name or patient's personal identity number. Results matching the search criteria are dynamically displayed in the section "Records".

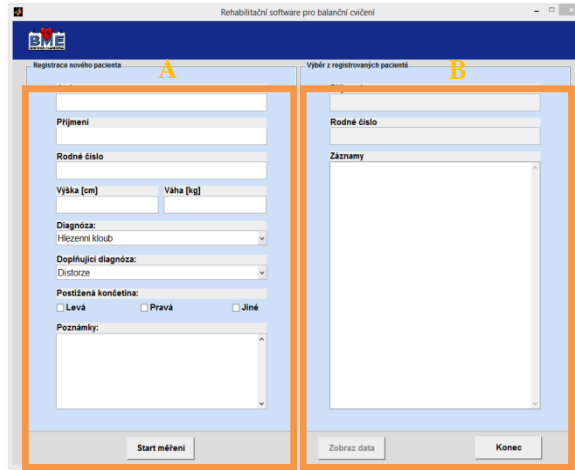


Fig. 1. The main application window.

The main window also contains three buttons "Start Measurement", "Display Data", and "Close", by which the user controls the program.

Each button has its own application window and function. After clicking the "Start Measurement" button, the measurement window appears (Fig. 2).

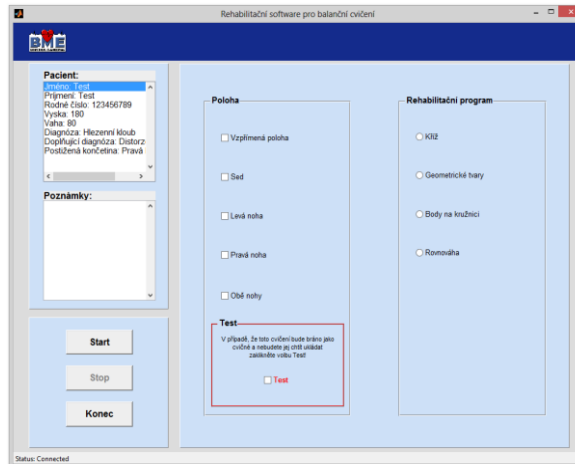


Fig. 2. The window for measurement.

The measurement window consists of four basic parts: basic information about the patient, notes, and parameters of the rehabilitation program. The basic information about the patient summarizes the data entered during the initial registration of the patient into the database. The parameters of the rehabilitation program are the most important part of this window. In this part the user chooses a rehabilitation program and parameters of the rehabilitation. A particular rehabilitation task could be performed in a sitting or standing position. These two cases are mutually exclusive. Sub-

sequently, a user has to specify the leg on which the task is performed. The patient can practice the left, right or both legs. These three cases are also mutually exclusive. In the part “Rehabilitation Program” user can choose a particular exercise from four available ones: “Cross”, “Geometric Shapes”, ”Points on the Circle”, and “Stability”.

In the rehabilitation program “Cross”, a user can choose from four different tolerance bands in the range from 5 to 20 degrees with step 5 and adjust the size of maximal deflection in the range from 10° to 30° . In this rehabilitation program the patient has to move in the white area and try to achieve the maximum possible deflection in horizontal and vertical axis (Fig. 3). After clicking the “Start” button the chosen rehabilitation program with its adjusted parameters is displayed. The current patient movement is represented by a two-dimensional representation, the black point is the current patient position and grey points are representing recent samples. In Fig. 3 you can also see that in the left side of the window there is place for notes, which is very useful for the physiotherapist who can enter some important notes about the exercise performance (e.g. falls, problems, etc.).

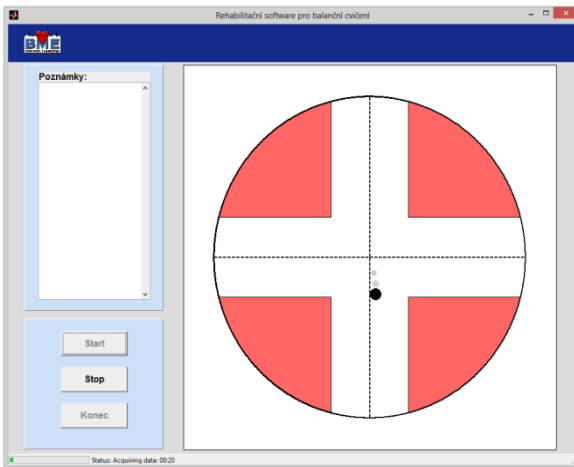


Fig. 3. The rehabilitation program “Cross”.

In the rehabilitation program “Geometric Shapes”, a user can choose one from three different geometric shapes: circle, square and triangle. After the shape and the size of the angular deflection in degrees (in the range from 10° to 30°) is selected and the “Start” button is clicked the selected geometric object of the chosen size is displayed (Fig. 4A). The main task is that the patient has to describe the shape of the selected object (by his own movement) on the wobble board. Green area represents the tolerance band, in which the patient should move and the black square in the middle represents the selected angular deflection, which should the patient achieve during the rehabilitation.

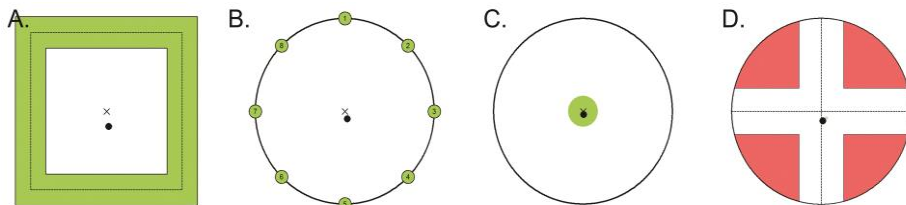


Fig. 4. The rehabilitation programs.

The third rehabilitation program is represented by points located on the circle’s circumference (Fig. 4B). During this exercise, the patient is trying to reach a particular point according to the physiotherapist’s instruction. The number of points could be selected in the range from 4 to 8. The number of points and motion sequence is intended by physiotherapist with respect to patient’s diagnosis.

The last rehabilitation programme is called “Stability”. In this exercise, the patient is trying to maintain equilibrium posture on the wobble board (Fig. 4C). This task has to be performed within tolerance area. The tolerance area is represented by green region. The black circle represents the maximum achievable angular deflection.

After clicking the "Display Data", the patient data management window appears (Fig. 5). This window is designed to view measured patient data. The score represents patient’s success achieved in particular exercise. It is simple computed as a ratio between number of samples which were inside the tolerance area and the number of samples which were outside. The recorded samples, which represent the patient’s movement performed during the exercise, are visualized with respect to the selected rehabilitation program.

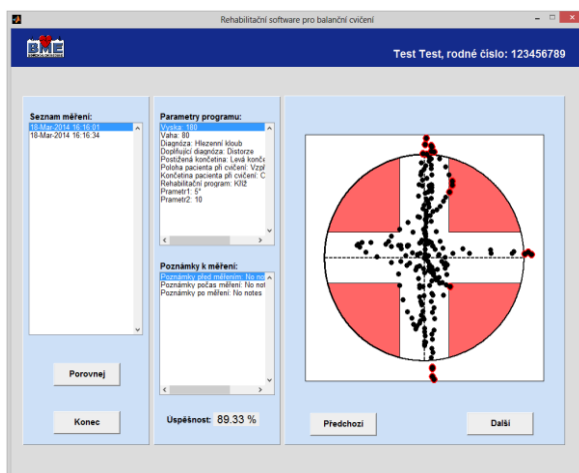


Fig. 5. The patient data management window.

4 Conclusion

The proposed rehabilitation software for balance exercise on the wobble board allows a physiotherapist to design specific rehabilitation program with respect to the patient's diagnosis. The rehabilitation programs encompass the motion patterns which are commonly used in rehabilitation practice. The measured data are stored and processed in order to compute patient's score. The score simply represents patient's ability to perform selected rehabilitation exercise. The patient data management part of proposed software allows a physiotherapist to view patient's progress. The implementation of hardware and software in the rehabilitation process makes the rehabilitation more attractive for patients and more objective for the physiotherapists. The future work will be the extension of the rehabilitation programs to deliver more focused therapy for different types of injuries as well as the development of signal processing techniques for complex evaluation of the patient's progress.

5 Acknowledgement

The work and the contributions were supported by the project SV SP2014/194 "Bio-medical engineering systems X" and the project Opportunity for young researchers, reg. no. CZ.1.07/2.3.00/30.0016 supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic.

6 References

1. Dvorak, R. Fundamentals of kinesiotherapy. 3th ed. Olomouc: Palacky University in Olomouc, 2007.
2. Kaesler, D. S., Mellifont, R. B., Swete Kelly, P., and Taaffe, D. R. A novel balance exercise program for postural stability in older adults: A pilot study. *Journal of Bodywork and Movement Therapies*, 2007, 11(1), p.37-43.
3. Liebenson, C. Better balance exercises. *Journal of Bodywork and Movement Therapies*, 2005, 9(2), pp.148-149.
4. Karime, A., Al Osman, H., Alja'am, J. M., Gueaieb, W., El-Saddik, A. Tele-Wobble: A Telerehabilitation Wobble Board for Lower Extremity Therapy. *IEEE Transactions on Instrumentation and Measurement*, July 2012, vol.61, no.7, pp.1816-1824, doi: 10.1109/TIM.2012.2192338.
5. Fitzgerald, D., Trakarnratanakul, N., Dunne, L., Smyth, B., Caulfield, B. Development and user evaluation of a virtual rehabilitation system for wobble board balance training. *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology"*, August 2008, art. no. 4650134, pp. 4194-4198.
6. Nilsson, N.C., Serafin, S., Nordahl, R. Poster: The Fwobble: Continuous audio-haptic feedback for balance control, 2012 IEEE Symposium on 3D User Interfaces (3DUI), 4-5 March 2012, pp.153 – 154, doi: 10.1109/3DUI.2012.6184207

Fast sequence anomaly detection

Andrey V. Sukhanov¹, Sergey M. Kovalev¹, and Jan Otýpka²

¹Department of Automatics, Telemech. and Comm. on Railway Transport, RSTU,
Rostovskogo Strelkovogo Polka Narodnogo Opolcheniya sq., 2,
Rostov-on-Don, 344038, Russia
drewnia@rambler.ru, ksm@rfniias.ru

²Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.otypka@vsb.cz

Abstract. Anomaly detection is an important problem of finding instances in data that do not conform to expected behavior. In general, instances are related to each other. That is why sequence anomaly detection is the most important area of anomaly detection problems, which is widely used in variety of application domains such as intrusion detection in computer networks, health care, economics, etc. This paper presents a new approach dedicated to sequence anomaly detection problem. Proposed technique is based on Least-Squares temporal - difference learning (LSTD). There is suggested the update of LSTD which use of current data instance pre-history in test phase. It is shown that such technique has high detection rate in computational experiments.

Keywords: Anomaly detection, sequence anomaly, LSTD, anomaly prediction

1 Introduction

Anomaly detection dedicated to the problems that have been popularly researched within diverse research areas and application domains [1]. Anomaly detection applications are medicine, economics, cyber-security, space craft fault detection, etc.

One of the open problems in anomaly detection is the modeling and prediction of complex sequential anomalies, which consist of a series of temporally related data instances. In such case classic techniques cannot be used for efficient anomaly detection because they exclude sequence aspect. Due to this fact anomaly detection in sequence data is actively developed last years.

Anomaly detection in sequence data is the one of important anomaly detection sub-problems [2]. Its main aim is sequences that do not to conform to normal class. These nonconforming patterns are also named as anomaly patterns.

Existing anomaly sequences detection approaches are divided into 3 large groups [3]:

1. Supervised anomaly detection. Approaches of this group assume the availability of a training data set which has labeled instances for normal as well as anomaly class [4].

2. Semi-supervised anomaly detection. Approaches that operate in semi-supervised mode assume the training data which consists only of elements belongs to one class (normal or anomaly) [5].

3. Unsupervised anomaly detection. Unsupervised approaches don't require any training data. The techniques in this category make the implicit assumption that normal instances are far more frequent than anomalies (or vice versa) in the test data. If this assumption is not true then such techniques suffer from high false alarm rate [6].

This paper presents semi-supervised anomaly patterns detection approach. Our technique is based on Least-Squares temporal-difference learning [7, 8]. We improve this technique in test phase with including information about current instance pre-history.

2 Background and related works

One of the main problems, which are decided by sequence anomaly detection techniques, is intrusion detection in computer networks. The purpose of intrusion detection is finding cyber-attacks or non-permitted deviations of the characteristic properties [7, 8]. Classic techniques dedicated to this problem extract information of known attacks and compare them with observed data. However, in real cases intrusion detection systems must detect an anomaly as fast as possible before this anomaly arrives completely to input of classifier. So fast anomaly detection has been widely researched in this area last years.

The main aim of fast anomaly detection (or anomaly prediction) is the affinity recognition of test pattern to aim objective (or anomaly) class at first states of this pattern. Such recognition is achieved due to apriority information about statistics of anomaly appearance, which is extract from labeled train set.

Recognition rate is based on probability of target pattern occurring in the next steps of classification. Moreover, as more states sequence of pattern are recognized as higher accuracy of its affinity to objective (non-objective) class. In this case we must add the "affinity threshold" ε of test pattern, i.e. a minimum value of occurring veracity when we can say that test pattern is identical to target pattern.

3 Update of Least-Squares temporal-difference learning (LSTD)

Markov decision process is popular model of stochastic process description. Classic Markov modeling assumes property according to which probability transition of next state depends on current state and doesn't depend on the sequence of previous states.

Real systems don't satisfy to Markov property. That is why Markov decision model was changed by Least-Squares temporal-difference learning (also named as LSTD) [7]. LSTD learning algorithm is next:

Algorithm 1:

Given:

- a state space $S = \{s_i\}$;
- a set of feature vectors $\Phi = \{\varphi_{s_i}\}$, where $\varphi_{s_1} = (1,0,0, \dots, 0)$, $\varphi_{s_2} = (0,2,0, \dots, 0)$, etc.;
- a coefficient $\lambda \in [0,1]$;
- a set of train patterns $PAT = \{pat_k\}$.

Set $P = 0, B = 0$.For each train pattern $pat_i = (x_{i_1}, x_{i_2}, \dots, x_{i(n_i)})$ do {Set $t = 1$;Set $z_{i_t} = \varphi_{x_{i_t}}$;While $x_{i_1} \leq x_{i(n_i)}$ do {

$$P = P + z_{i_t} \left(\varphi_{x_{i_t}} - \varphi_{x_{i_{t+1}}} \right)^T.$$

 $B = B + z_{i_t} r(x_{i_t})$. $z_{i_{t+1}} = \lambda z_{i_t} + \varphi_{x_{i_{t+1}}}$. $t = t + 1$.

}

}

Set $W = P^{-1}B$.

Coefficient λ plays the main role in Algorithm 1 since its value means the dependence degree from all pre-histories of each state s in the train data. Matrix P with dimension $|S| \times |S|$ is transition matrix. When λ converges to 0, information, which extracts from transition matrix P , converges to information extracting from Markov transition matrix:

$$P = (T - N)^{-1}, \quad (1)$$

where T – transition number matrix (T_{ij} is the number of transitions from state s_i to state s_j);

N – matrix, the main diagonal of which is the number of times each state has been visited.

Vector B with dimension $|S| \times 1$ is reward vector. It is depended from reward function $R: x \rightarrow r(x)$, which is defined as:

$$r(x) = \begin{cases} 1, & \text{if } x = x_{i(n_i)} \text{ and pattern}_i \text{ is target;} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

When λ converges to 0, B converges to vector $(0, \dots, 1, \dots, 0)$.

The result of LSTD learning is weight coefficients w_i for each state from state space.

Classic anomaly detection based on LSTD learning for testing sequence $TEST = \{x_t\} (t = 1, \dots, m)$ is performed by the next formula:

$$V_t = \sum_{t=1}^m \varphi_{x_t} W, \quad (3)$$

where V_t – veracity of target pattern finish state forthcoming from current state.

As we can see such detection classify only all sequence and don't let to evaluate each test state.

Target pattern recognition from test state sequence in our experiments is free from this disadvantage and performed by:

$$V_t = \alpha V_{t-1} + \varphi_{x_t} W, \quad (4)$$

where α – depending coefficient of current pre-history.

Detection based on (4) allows considering not only pre-histories of states from train data but also the current pre-history from test data set. Classic LSTD based approaches don't have this opportunity.

Affinity threshold can be calculated by analyzing of train data set via formula (4). In this case we can say that testing sequence tends to anomaly if forthcoming veracity of some consecutive states will be higher than affinity threshold ε .

4 Computational experiments

Updates of LSTD learning were tested by set of experiments. Known benchmark patterns Coffee [9] were used in one of such experiments. Coffee consists of 28 implementations. Each of these implementations was quantized and resampled. In result we got 28 patterns with $|S| = 37$ and length 28. One of patterns is illustrated in fig. 1.

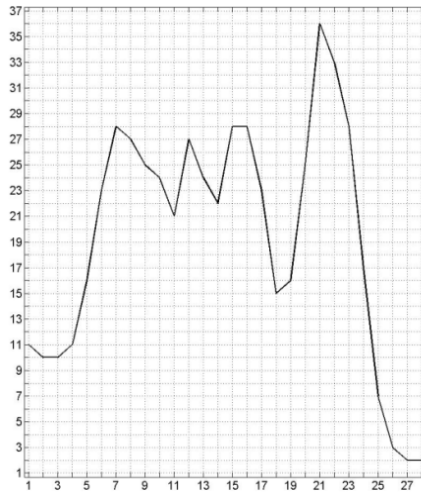


Fig. 1. Coffee implementation

One Coffee implementation was used as sequence, after that we considered pattern as target. Train data was made of 250 target patterns. We made set of heuristic tests on the train data to get optimal coefficients of λ , α and ε . These tests were concluded in finding of the best prediction error from formula (4). When we were finding ε , we

took 5 sequence elements with $V_t > \varepsilon$. Scheme of training is illustrated in fig. 2. Fig. 3 depicts the evaluation of λ and fig. 4 depicts the evaluation of α .

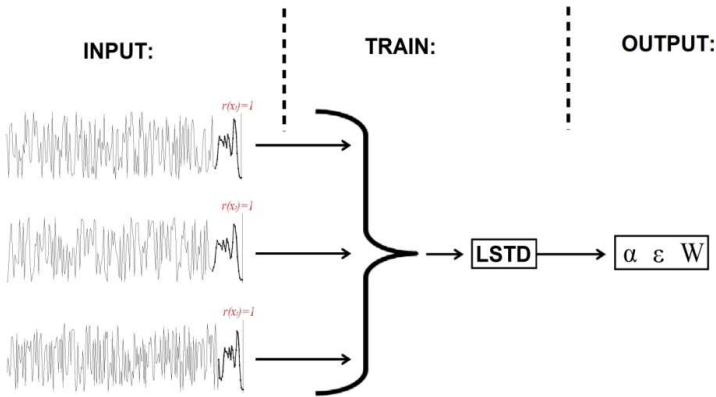


Fig. 2. Training scheme

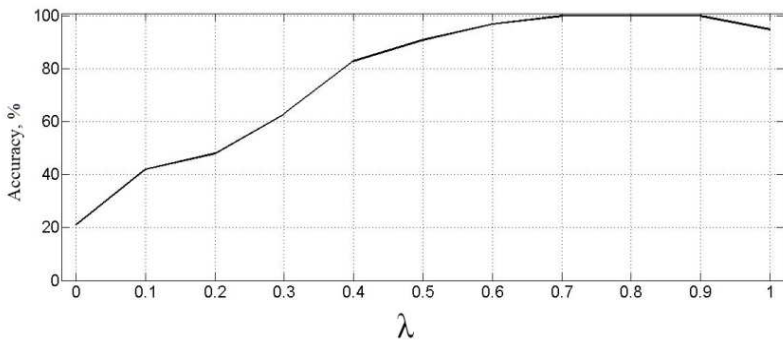


Fig. 3. Dependence of classification accuracy from λ with best α

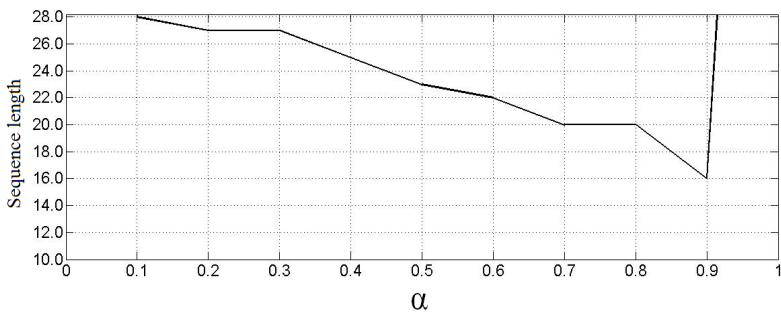


Fig. 4. Dependence of sequence length for best classification accuracy from α with best λ

As we can see in fig. 4, the prediction of target pattern was at 16 state of sequence with best values of coefficients.

Further we make some difficulties in the test phase to show advantages of our approach. We take fragments of other Coffee implementations and shuffle them. Then we add 500 target patterns to this sequence and test it with best values of coefficients (fig. 5).

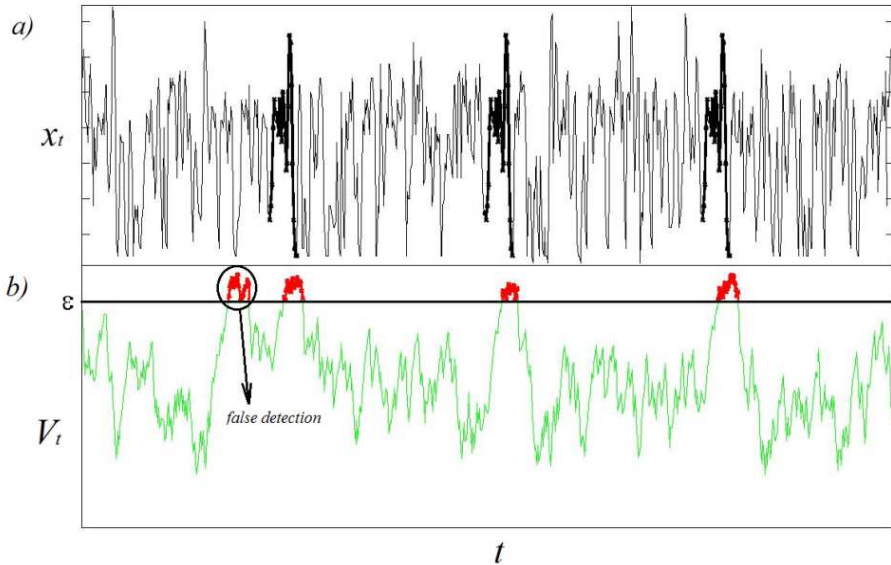


Fig. 5. Fragment of test data set (a) and its classification (b)

As we can see we had some wrong target pattern detection in this case. Accuracy of test classification was 93%. Nevertheless we can find all of patterns Coffee and predict them at 20 - 25 states.

5 Conclusions

This paper presents a new approach for decision of fast detection problem, which is dedicated to applications of symbol sequences classification. The new update for Least-Squares temporal-difference learning have been proposed in the work, which is concluded in adding of information about current state pre-history in test phase. It has been shown in computational results with known patterns Coffee that our technique allows to get excellent results in sequence anomaly detection. So, presented technique can be used in prediction systems for such applications as intrusion detection, healthcare, economic, etc.

Acknowledgment

This paper has been supported by project of SGS №. SP2014/45 and project of Technology Agency of Czech Republic №. TA02030946.

References

1. CHANDOLA, V.: *Anomaly Detection: A Survey* / Varun Chandola, Arindam Banerjee and Vipin Kumar // *ACM Computing Surveys*. Vol. 41(3). Article 15, July 2009. – pp. 1–7.
2. KOVALEV, S. M.: *Operativnoye detektirovaniye temporal'nykh patternov v sekventsial'nykh dannyykh (Operative temporal pattern detection in sequence data)* / S. M. Kovalev, A. V. Muravskiy // *Izvestiya YUFU. Technicheskiye nauki. Tematicheskiy vipusk «Intellektual'nye SAPR» (Proceedings of the SFU. Engineering science. Special Issue "Intelligent CAD")*. — Taganrog: Publ. TTI YUFU, 2012, No 7 (132), pp. 92–98.
3. CHANDOLA, V.: *Anomaly Detection for Syambolic Sequences and Time Series Data. PhD Thesis, Computer Science Department, University of Minnesota* — 2009. — 173p.
4. ARTEMYEV, I. S., DOLGYI, A. I., SUKHANOV, A. V., KHATLAMGIYAN, A. E.: *Neyroimmunnaya model' klassifikatsii v zadachakh identifikatsii na transporte (Neuroimmune model of classification in transport identification problems)* // *Integrirrovannyye modeli i myagkiye vychisleniya v iskusstvennom intellekte. Sbornik nauchnykh trudov VII-y Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii (Kolomna, 20-22 maya 2013 g.). V 3-kh tomakh. T.3. (Integrated models and soft computing in artificial intelligence. Collection of Scientific Papers VII-th International Scientific and Technical Conference (Kolomna, May 20-22, 2013). In 3 volumes. V.3.)* – Publ.: Fizmatlit, 2013. pp. 980 – 987.
5. DASUPTA, D. A.: *comparison of negative and positive selection algorithms in novel pattern detection* / D. Dasgupta, F. Nino // *In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Vol. 1. Nashville, TN, 2000.* — pp. 125–130.
6. KOVALEV, S. M., SUKHANOV, A. V.: *Anomaly detection based on Markov chain model with production rules* // *Program products and systems.* – 2014. – № 3 – pp. 40 – 44.
7. BOYAN, J. A.: *Technical update: least-squares temporal difference learning* // *Machine Learning* 49. – 2002. – pp. 233–246.
8. XU, X., WANG, X. N.: *Adaptive network intrusion detection method based on PCA and support vector machines, ADMA 2005, Lecture Notes in Artificial Intelligence, LNAI 3584.* – 2005. – P. 696–703.
9. KEOGH, E.: *The UCR time series classification/clustering homepage* / E. Keogh, X. Xi, L. Wei, C.A. Ratanamahatana // http://www.cs.ucr.edu/~eamonn/time_series_data/.
10. SUKHANOV, A. V.: *Stokhasticheskaya Markovskaya model' poiska anomalii v temporal'nykh dannyykh (Markov model for anomaly detection in the temporal data)* // *Trudy Kongressa po intellektual'nym sistemam i informatsionnym tekhnologiyam «IS&IT'13» Nauchnoye izdaniye v 4-kh tomakh. (Proceedings of the Intelligent Systems and Information Technology «IS&IT'13» Congress. A scholarly edition in 4 vol.)*. – Publ.: Fizmatlit, 2013. – V.1. pp. 177 – 181.
11. SUKHANOV, A. V., KOVALEV, S. M.: *Nechetkiye markovskie algoritmy (Fuzzy Markov algorithms)* // *«Trudy RGUPS» ("Proceedings of RSTU")*. – 2013. – №4 (25) . – pp. 104 - 108.

Advanced method for control of magnetization of AMR sensing element

David Vala

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

Abstract. Electromagnetic compatibility of electronic devices is the area which significance grows nowadays. As the consequence of this growth there is a continuous process of making more strict standards focused on electromagnetic radiation of electronic devices. Sensors technology begins to be a part of these issues due sensors bandwidth increasing and approaching to frequency of radio communication band. Nowadays microcontrollers and similar digital circuits are integrated into sensors devices and it brings new sources of electromagnetic radiation in modern smart sensors. The aim of this work is to show possibility of fusion of technologies from telecommunication and sensors devices to reduced electromagnetic radiation of magnetic sensors, especially advanced of clock pulses generation in the area of weak magnetic field measurement. This contribution is aim on Anisotropic Magnetoresistors (AMR) from wide range of magnetic field sensors and the problematic of its demagnetization drive control.

Keywords: Anisotropic magnetoresistors, AMR, magnetic sensor, EMC, weak magnetic field

1 Introduction

Anisotropic Magneto-Resistive (AMR) sensors are fabricated with Permalloy (NiFe) thin films that create changes in resistivity with respect to external magnetic fields. These film materials are similar to magnetic recording tapes in that strong magnetic fields can disrupt the magnetic domains of the film particles from a smooth factory orientation to arbitrary directions. Accuracy and resolution of these sensors will suffer until the film magnetic domains are “reset” to recreate a uniform direction. This application note shall detail the set and reset functions for AMR sensors including the reasons to perform this function, characteristics of set/reset components, and example circuits showing the present state of the art in set/reset pulse generation.

AMR sensors are designed in Wheatstone bridge configurations, with four magneto-resistive elements that remain identical electrically when no external magnetic

fields are applied. With linear-mode AMR sensors, the externally applied magnetic fields are to be limited in strength so as to not disrupt the factory set magnetic domains of the permalloy thin-film elements. Figure 1 shows three examples of magnetic orientation of the film domain structure.

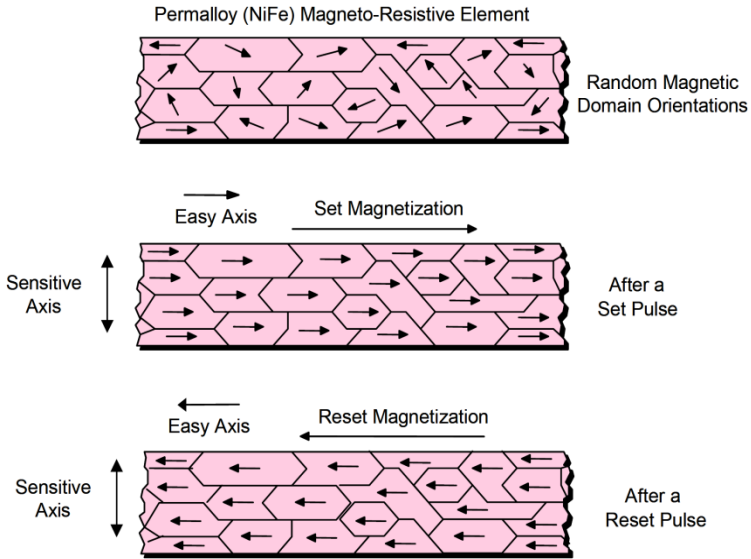


Fig. 1. Domain orientation in permalloy magneto-resistive element [8]

The alignment of the magnetic domains is up along the “easy” axis of the material. The “sensitive” axis is perpendicular to the easy axis direction and serves as the driving function of the magneto-resistive characteristic curve shown in Figure 2. When an external magnetic field has a vector component in the sensitive axis, the field rotates the magnetic moment creating a change in resistance. If the external magnetic fields are in the operational field range, the magnetic moments will return to their set or reset orientations after the field is removed. To put some numbers on the magnetic fields involved, typically a “set” or “reset” field requires about 4 mT or more to be applied to completely orient the magnetic domains.

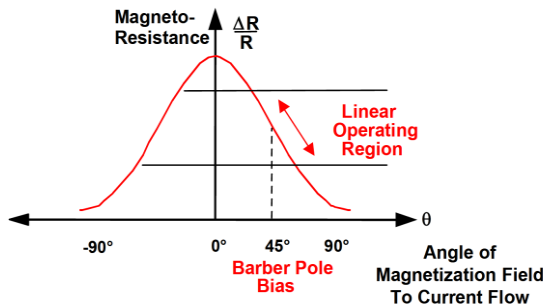


Fig. 2. Characteristic curve of AMR sensor [7]

2 Why bi-directional magnetization

The reasons to perform a set or reset on an AMR sensor are:

1. recover from a strong external magnetic field that likely has re-magnetized the sensor
2. optimize the magnetic domains for most sensitive performance
3. flip the domains for extraction of bridge offset under changing temperature conditions

Strong external magnetic fields that exceed a 1 to 2 mT “disturbing field” limit, can come from a variety of sources. The most common types of strong field sources come from permanent magnets such as speaker magnets, nearby high-current conductors such as welding cables and power feeder cables, and by magnetic coils in electronic equipment such as CRT monitors and power transformers. Magnets exhibit pole face strengths up to unit of Tesla. These high intensity magnetic field sources do not permanently damage the sensor elements, but the magnetic domains will be realigned to the exposed fields rather than the required easy axis directions. The result of this re-magnetization of the sensor elements will be erroneous measurements and indications of “stuck” sensor outputs. Using the set and reset pulses will magnetically “restore” the sensor.

AMR sensors are also ferromagnetic devices with a crystalline structure. This same thin film structure that makes the sensor sensitive to external magnetic fields also has the downside that changing magnetic field directions and thermal energy over time will increase the self-noise of the sensor elements. This noise, while very small, does impair the accurate measurement of sub-micro Tesla field strengths or changes in field strength in nano Tesla increments. By employing frequent set and reset fields on the sensor, the alignment of the magnetic domains in each permalloy element drops the self-noise to its lowest possible level.

As the sensor element temperature changes, either due to self-heating or external environments, each element’s resistance will change in proportion to the temperature. In the Wheatstone bridge configuration with the elements configured as a sensor, the bridge offset voltage and bridge sensitivity will drift with temperature. In compassing applications, the sensitivity drift of multiple sensor bridges are ignored due to the proportional method of deriving heading, but the bridge offset voltages must be updated and corrected for best accuracy as the temperature changes. One way to eliminate the bridge offset voltage is to make stable magnetic field measurements of the bridge output voltage in between each set and reset field application. Since the external field components of the bridge output voltage will flip polarity, the set and reset bridge output voltages can be subtracted and the result divided by two to calculate the bridge offset.

3 Magnetization process

The above description explained that providing pulses of electrical current creates the needed magnetic fields to realign the magnetic domains of the sensor resistive elements. Also the rationale for performing these set and reset pulses has been justified. The following paragraphs shall show when and how to apply these pulsed currents, and circuits to implement them. The temperature coefficient math shows the importance of choosing appropriate strap voltages to create the correct strap currents and that both resistances and currents are variables to be dealt with. The direction of the strap currents also determines what is considered a “set” or a “reset” pulse. Set pulses are defined as pulsed currents that enter the positive pin of the set/reset strap. Likewise negative pulsed currents are considered reset pulses. Figure 3 shows a simplistic schematic of a set/reset circuit.

These set and reset pulses are shown in Figure 3 as dampened exponential pulse waveforms because the most popular method of generating these relatively high current, short duration pulses is via a capacitive “charge and dump” type of circuit. Most electronics, especially in consumer battery powered devices, do not have the capability to supply these high current pulses from their existing power supply sources. Thus “Vsr” is actually a charged up capacitor that is suddenly switched across the set/reset strap. The value of this capacitor is usually a couple hundred nano-Farads to a few micro-Farads depending on the strap resistance to be driven. The decay of the exponential waveform will mostly be governed by a time constant (τ or Tau) that is the capacitance in farads multiplied by the resistance, and is measured in seconds.

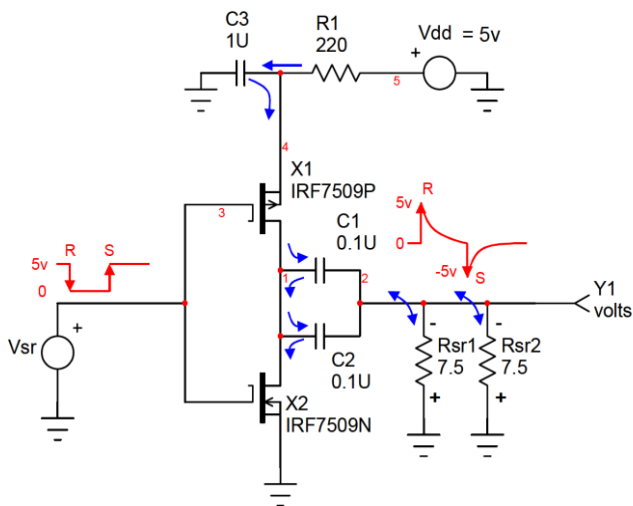


Fig. 3. Simplified schematic of set/reset pulse [8]

4 Measurement and reduction of set reset pulse radiation

Each current pulse can product wide spectrum of electromagnetic radiation. This radiation can disturb another devices or measurement itself especially in multisensory application. For this reason I'm aimed to reduce this source of radiation because classic technique as electromagnetic shielding of sensor is in direct conflict with function of the sensor.



Fig. 4. Measurement equipment and workplace for measurement of electromagnetic radiation

As is shown on figure 5 by the fusion of techniques from telecommunication technologies whit sensor techniques is possible to reduce the radiation of AMR magnetic sensor at least of 3 dB in worst case but generally is the radiation reduction significantly higher. Another advantages designed techniques detail described in [14] and [15] is lower possibilities of detection of AMR magnetometer when is switched on.



Fig. 5. Measurement of spectrum of magnetization circuit of AMR sensor without and with reduction techniques.

5 Conclusion

Detailed description of techniques for reducing of electromagnetic radiation of AMR magnetoresistive sensor is possible to find in internal documents [14] and [15]. This technique is also detailed described in application of the invention on Czech Industrial property office under number E193142 from 30 of January 2014 which is not public document in days when this contribution is written.

Acknowledgement

This work was supported by the project SP2014/188 “*Řízení technologických soustav s OAZ pro komplexní energetické systémy 2014*“ within the student’s grant competition (VŠB-TU Ostrava) and This work is supported by project SP2013/135, named “Control of technological systems with OAZE providing an independent sustainable development of complex systems” of Student Grant Agency (VSB - Technical University of Ostrava)

Reference

1. Ripka, P., Magnetic sensor and magnetometers, Artech House, Norwood, MA, USA, 2001 ISBN 1-58053-057-5
2. Fuka, J., Havelka, B.: Elektřina a magnetismus, vydání 3, SPN, Praha, 1979
3. Stork, T.: Electronic Compass Design using KMZ51 and KMZ52, Philips Semiconductors Systems Laboratory Hamburg, Germany, 2000
4. Data sheet: KMZ52 Magnetic Field Sensor, Philips Semiconductors
5. Hauser, H., Stangl, G., Fallmann, W., Chabicovsky, R., Riedling, K.: Magnetoresistive Sensors, Preparations, Properties and Applications of Thin Ferromagnetic Films, June 2000, Vienna, Austria
6. Handling Sensor Bridge Offset AN212, Honeywell, 2002
7. Set/Reset Pulse Circuits for Magnetic Sensors AN-201, Honeywell
8. Reference Design: Low Cost Compass AN214, Honeywell, 2002
9. Horák, B., Vala, D.: Smart Digital Magnetometer Application in Navigation Subsystem of Mobile Robot, VŠB – TU Ostrava
10. Vala, D.: Mini Flux-gate Sensor, INPF-LMN/LEG Grenoble, 2002
11. Caruso, M., J., Bratland, T., Smith, C., H., Schneider, R.: A New Perspective on Magnetic Field Sensing, Honeywell, SSEC Nonvolatile Electronics Inc.
12. Vala, D., Návrh a realizace systému autonomního měření polohy bezpilotního letového prostředku Kavka, Diplomová práce, kat.455, VŠB-TU Ostrava, 2003
13. Boashash, B., Time Frequency Signal Analysis and Processing, Elsevier, Oxford, 2003 ISBN 0-08-044335-4
14. Vala, D. Způsob zvýšení utajení zařízení pro měření změny slabých magnetických polí před vyhledáváním radioelektronickými prostředky. Interní dokument VŠB-TU Ostrava, ev. č. 040/02-09-2011, Ostrava, 2011
15. Vala, D., Buzení demagnetizace AMR magnetorezistoru rozprostřeným spektrem, Interní dokument VŠB-TU Ostrava, ev. č. 039/02-09-2011, Ostrava, 2011

Artificial Neural Network Based Derivation of Vectorcardiographic Leads

Michal Vozda, Martin Cerny, and Jindrich Cernohorsky

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

Abstract. Vectorcardiography (VCG) represents of electromotive forces generated during cardiac activity in te form of loops in the orthogonal three-dimensional space. Although there are ways how to measure VCG, direct measurement is not common in clinical practice. In this paper we present nonlinear transformation method for deriving VCG leads from conventional 12-lead electrocardiogram (ECG) based on Artificial Neural Network (ANN). ANN based derived VCG achieved significantly lower Mean Squared Error (MSE) and higher correlation coefficient than commonly used linear transformation based inverse Dower method.

1 Introduction

According the World Health Organisation (WHO), died 17.3 million people from Cardio-Vascular Diseases (CVDs) in 2008. It represents 30% of all global deaths. Of these deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million were due to myocardial infarction (MI). WHO estimates that by 2030 more than 23 million people will die annually from CVDs.

Standard tool for diagnosis of heart diseases is electrocardiography (ECG), whose analysis is usually based on empirical base. Different representation of the electrical activity of the heart provides vectorcardiography which is on the other side more suitable for automatic computerized analysis. VCG was long considered as a diagnostic method in many fields with higher specificity and sensitivity compared to conventional electrocardiography. Unfortunately it was gradually replaced with 12-lead electrocardiography in common clinical practice. However, the diagnostic sensitivity of vectorcardiography still predominates in specific cases. The highest potential is offered by computational vectorcardiography in the diagnostics of ischemic heart disease, right ventricular hypertrophy, MI localization, and other diseases.

Because VCG is not commonly directly measured in clinical practice, there were published linear transformations between ECG and VCG leads [1]. Instead of direct measurement it is possible to compute VCG leads from conventional 12-lead ECG which is usually used in clinical practice.

Published transformations are based on transformation matrices \mathbf{D} whose coefficients are possible determine by linear regression between directly measured

ECG and VCG leads. The most used method is inverse Dower transformation [3].

$$\begin{aligned} [V1V2V3V4V5V6I II]^T &= \mathbf{D} [XYZ]^T, \\ \mathbf{E} &= \mathbf{D}\mathbf{V}, \end{aligned} \quad (1)$$

where \mathbf{E} represents ECG leads (V1 . . . V6, I, II) and \mathbf{V} are VCG leads (X,Y,Z). \mathbf{D} is transformation matrix.

The aim of our work is detection and prediction heart diseases which can endanger suddenly and unexpectedly people everyday [4],[5],[6]. We use automatical analysis of features obtained from directly measured or derived VCG. Accuracy of derivation of VCG is important for next work. In this paper we present new technique for derivation VCG based on the ANN.

2 Material and Methods

2.1 Study Population

Tested records are chosen from a PTB diagnostic database that has been recorded using healthy volunteers and patients with different heart diseases at the Department of Cardiology of the University Clinic Benjamin Franklin in Berlin, Germany. The database contains 549 records from 286 subjects. Each subject is represented by one to five records. Each record includes 15 simultaneously measured signals: the conventional 12 leads together with the 3 Frank VCG. Each signal is digitized at 1000 samples per second, with 16 bit resolution over a range of ± 16.384 mV [2].

Only the first records for healthy volunteers were taken into account. The records were band-passed by a FIR filter with linear phase response in the band from 0.25 to 150 Hz (-3dB). From the filtered records, representative beats excluding PVCs and artefacts were chosen. The beginning of the beat is defined as the distance from an R wave: $T_R - 0.4T_{min(RR)}$ and the end of the beat is defined as $T_R + 0.6T_{min(RR)}$. Where T_R is the length of presence of an R wave and $T_{min(RR)}$ is the minimal pulse period for the record. Individual representative beats were averaged in each VCG lead.

2.2 Design of Neural Network Architecture

To synthesize VCG leads from the 12-lead ECG we use a multilayer feed-forward ANN trained by means of the supervised back-propagation algorithm. The architecture includes input layer, hidden and output layers. Each layer includes neurons with specific activation function. ANN in Fig. 1 with 2 hidden layers and 4 neurons in each hidden layer achieved the best accuracy of transformation. With increasing number of layers and/or neurons increased the training time of ANN with insignificant impact on the accuracy of the transformation. The input

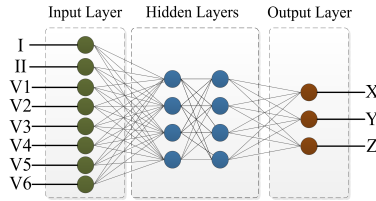


Fig. 1. Architecture of ANN.

layer contains 8 neurons. Input and hidden layers contain neurons with hyperbolic tangent sigmoid activation function. The output layer contains 3 neurons with linear activation function.

For derivation of VCG we use the committee of ANNs for higher robustness of the transformation method. Committee machine method is popular in processing with machine learning algorithms. This method uses a divide and conquer strategy in which the responses of multiple ANNs are averaged into a single response. The committee of 15 ANNs was tested. Each ANN has 8 inputs represented by 8 independent ECG leads. Six of them are leads V1...V6 and the last two leads are some of the 15 combinations of limb leads pairs (I; II; I; III; I; aVL; . . . ; aVR; aVF). Each ANN has different combination of independent ECG leads what is shown in Fig. 2.

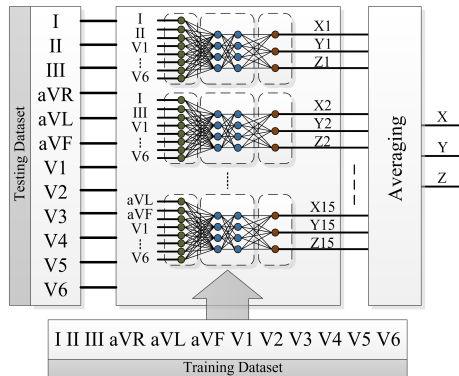


Fig. 2. Architecture of committee of ANN.

2.3 Verification and Validation

ANN based transformation was compared with commonly used inverse Dower transformation method which is realised according to equation (1). The both methods were tested under the same conditions, with the same testing datasets and with 10-folds cross validation. This approach is usually used for testing the

performance of supervised learning-based methods. Records were randomly divided into 10 groups, 9 of them (255 records) served for training the ANNs as the training dataset. One remaining group (28 records) served for testing of all methods as the testing dataset. In the next step other 9 groups were chosen and 1 remaining group was again used as the testing dataset. We continued in this way 10 times. This is the most efficient way how to test the supervised learning-based methods with maximum usage of the training data. Performance of all methods was evaluated by the two most often used parameters: Pearson correlation coefficient and MSE which were computed between derived and directly measured VCG in each lead for all methods. Pearson correlation coefficient indicates the degree of similarity between two signals and is independent from the differences in their amplitudes. The MSE is a measure of differences in amplitudes between two signals. Each method was tested for all VCG leads. MSEs and correlations between leads computed by the tested methods and directly measured Frank leads were evaluated.

3 Results

One sample t-test was used for testing the normality of MSE and correlation coefficients for all methods and in all leads. Whereas that data are not from a normal distribution, the nonparametric Mann-Whitney U-test was used for testing the differences between the methods. The null hypothesis was tested that the data obtained from two compared methods are samples with equal medians. Rejecting the null hypothesis we prove statistical difference between the two methods at the 5% significance level ($\alpha = 0.05$).

The statistical tests prove that the differences between the ANN based and inverse Dower method are significant. For ANN based method is the MSE significantly lower than for inverse Dower transformation for all leads see Fig. 3. The correlation coefficient is significantly higher for X and Z lead. For lead Y are differences insignificant.

4 Discussion

In this study, new ANN based method was tested and compared under the same conditions with commonly used method. Differences between methods are statistically significant and ANN based method is more accurate than inverse Dower method. Our results suggest that nonlinear regression method could improve the accuracy of transformation which is validated by MSE and correlation coefficient.

Evaluation of a new method based on MSE and correlation is often used but it cannot replace the experience of a skilled cardiologist who can evaluate differences in diagnostic information which is the most important parameter. For future work we would like to test nonlinear regression methods on larger datasets measured in cooperation with cardiologists at different workplaces.

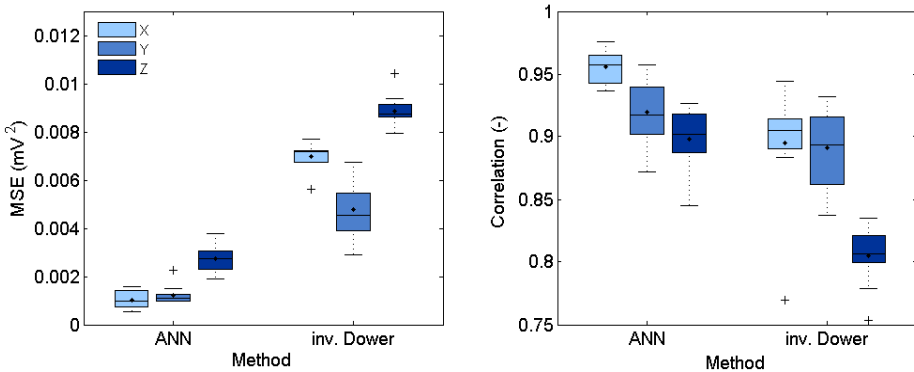


Fig. 3. Box-Whisker diagram of the MSE and correlation for individual methods in leads X, Y and Z.

5 Conclusion

The 12-lead vectorcardiography is based on transformations of the 12-lead ECG to VCG leads. Each transformation introduces certain error of amplitude and shape of the derived VCG leads. In this article we proposed a new method based on ANNs that provides significantly higher accuracy of transformation when compared with conventional commonly used method. Method which is based on ANNs can improve the state of the art in transformations between lead systems.

References

1. Rubel P, Fayn J. Quantitative assessment of eight different methods for synthesizing frank VCGs from simultaneously recorded standard ECG leads. *Journal of Electrocardiology*. 1991, vol. 24, s. 197-202.
2. Bousseljot R, Kreisler D, Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB ber das Internet. *Biomedizinische Technik, Band 40, Ergänzungsband 1 (1995) S 317*.
3. Edenbrandt L, Pahlm O. Vectorcardiogram synthesized from a 12-lead ECG: Superiority of the inverse Dower matrix. *Journal of Electrocardiology*. 1988, vol. 21, issue 4, s. 361-367. DOI: 10.1016/0022-0736(88)90113-6
4. Vozda M, Peterek T, Cerny M. Novel Method for Deriving Vectorcardiographic Leads Based on Artificial Neural Networks. In: *Proceedings of the 41st International Congress on Electrocardiology*. Slovakia, Bratislava: Slovak Academy of Sciences, 2014. s. 175-178. ISBN 978-80-969-6727-8. AR 2013/2014
5. Vozda M, Cerny M, Jurek F. Fitting parameters of the vectorcardiographic model by Particle swarm optimization method. In: *Yberc 2014: Proceedings of the 6th International Young Biomedical Engineers and Researchers Conference, Bratislava July 2014*, ISBN 9788097169701. AR 2013/2014
6. Vozda M, Hrvolova B, Krohova J, Smondrk M, Penhaker M, Computer-Based Vectorcardiograph for Research Purposes, In *17th International Conference ELECTRONICS '2013, Palanga, Lithuania, June 17-19, 2013 AR 2012/2013*

Remote-controlled generator

Boris Vukojević¹, Jiří Kazárik², and Bohumil Horák²

¹Josip Juraj Strossmayer University of Osijek,
Faculty of Electrical Engineering, Communication and Informatics,
Kneza Trpimira 2B, 31000 Osijek, Croatia
`boris.vukojevic@etfos.hr`

²Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{`jiri.kazarik, bohumil.horak`}@vsb.cz

Abstract. Remotely starting secondary power supply is useful for many applications. In this paper, system for remote operations of a standard fuel generator is described. Topology of the system is shown and explained. Major parts, communication and control, are described in more details. For some parts of the system, solutions are proposed. In the end, safety and security are addressed.

Keywords: fuel, generator, automation, microcontroller, control, remote, Ethernet

1 Introduction

One of the problems of solar energy is lack of control over produced power. Produced power depends on time of the day and weather. When power consumptions exceed power production, difference needs to be supplied from other sources. Monitoring of power consumption and production can be done by computer, and if it's needed, computer can start power production from second power source. Fuel generator is a good choice for this job. This paper addresses the problem of operating fuel generator using computer.

2 Generator remote control

Generator needs to be controlled over power line. Existing technology enables Ethernet communication over power line [1, 2]. With this, problem is simplified to controlling generator over Ethernet. Remote control system has to be able to:

- start/stop the generator on demand without human interference
- regulate power production
- measure and record: produced power, generator temperature and generator rpm

- send measuring data on request, over power line

System can be divided at two main parts: communication unit and generator operating device. In this setup, communication unit is used for communicating over Ethernet, providing user interface for human operator and sending measurement data. Generator operating device can start and stop the generator and performs measurements.

Generator operating device is a machine-to-machine interface. It is consisted of two types of devices: controller and action device. Controller translates high level orders to low level instructions that can be performed by action devices. The generator controller also collects measurements and passes them to communication unit. Action device perform a simple task such as: pushing, pulling, twisting or rotating. In this paper it is assumed that generator has three elements that need to be operated: choke, start key and starter. For each element there is one action device. Topology of described system is given in Fig. 1.

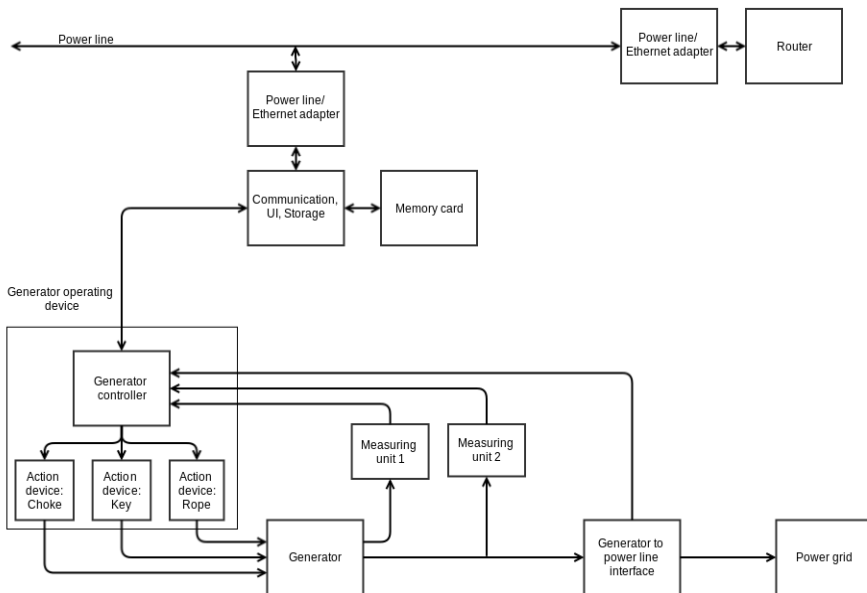


Fig. 1. Topology of a system for automating generator operations.

3 Generator controller

Generator controller has two operating modes: collecting data and executing generator operations. After controller has been powered on it runs initializing procedures. At this time, all devices and interfaces are configured and prepared for operation. After startup, controller enters his usual state: collecting and storing data. When order is received, interrupt is triggered. Interrupt procedure is started and this

procedure executes the order. After interrupt procedure, controller continues with collecting and storing measurements. Described work flow is shown on Fig. 2.

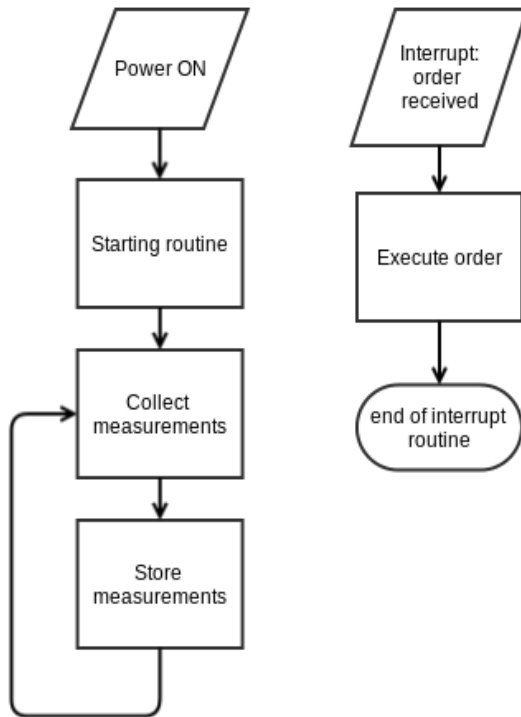


Fig. 2. Controller operation with interrupts.

Orders are:

- stop generator
- start generator

For every order, controller has to perform several operations in right sequence. Operation can be: reading measurement to determine the state of generator, using action device, sending information and so on. Sequences for listed orders are:

Stop routine:

```

if(generator is not working):
    respond(generator is off)
    exit routine
turn key to off
if(generator_is_not_working):
    respond(generator_is_off)
  
```

```
else
    respond(error)
```

Start routine:

```
if(generator_is_working):
    respond(generator is on)
    exit routine
turn key to star
if(generator is working):
    turn key to on
    respond(generator is on)
else
    turn choke to on
    for 3 times:
        pull rope
        if(generator is working):
            exit loop
turn choke to off
if(generator_is_working):
    respond(generator is on)
else
    respond(error)
```

4 Building the prototype

Communication unit is built on Raspberry Pi [3]. Raspberry Pi is one board computer that can run Linux operating system. With operating system, Raspberry Pi has built in support for communications over Ethernet. Same device runs web server. Web server is used to host page with user interface for human operator.

Data storage is solved using SQL database hosted on Raspberry Pi. Database makes handling data easier. Data can be added, received and transfer using any programming language with support for SQL database. In this setup, data can be added using low level language that's used by microcontroller and presentation of data can be done using higher level or web language, suitable for web page user interface. Database is stored on memory card, which makes transferring whole database trivial.

Generator controller operations are performed by microcontroller. Because of requirement of communication with Raspberry Pi, Embedded Pi [4] was chosen. Embedded Pi is a development platform for Raspberry Pi, Arduino and 32-bit embedded ARM. It is based on the STMicroelectronics STM32F103RB MCU[5]. It is used in standalone mode as a Cortex-M3 board connected with Raspberry Pi.

Microcontroller reads, process and stores measurements made by measuring devices. State of generator is also determined from measurements. Some states of

generator are: `generator_is_on`, `generator_is_off`, `generator_temperature_high` and so on. Change of state can trigger procedures and in that way system regulate itself.

Second task of microcontroller is to control action devices, and thru them, operate the generator. Action devices are controlled with PWM or other digital signals. Action devices have yet to be made.

4.1 Safety and security

Safety of this system needs to be address. Basic safety mechanism is a kill-switch, button or a switch that stops all action. Kill-switch is operated by human. System is able to stop itself if some conditions are met. For example, if generator is overheating, system can stop it. There are two levels of safety shutdown: generator shutdown and whole system shutdown.

Error can appear in every part of the system. Errors have to be detected, logged and processed. Error detections can be performed by action devices and measuring units.

Security problems for this system are the same as with every Ethernet network over power line. Using standard security measures for such networks is recommended [6].

5 Conclusion

System for automated generator operation is described. Topology of this system is given and explained. Communication and control are described in more details. Controller workflow is shown. Solutions for some parts of system are proposed. Realization of whole system is work in progress. After system is built and tested, integrations and optimizations can take place. Whole system can be built on single, custom made board.

6 Acknowledgement

This work was supported by project SP2014/188 "Control of technological systems with OAZE providing an independent sustainable development of complex system 2014" within the student's grant competition (VŠB - TU Ostrava).

References

1. Broadridge, R. (1989). "Power line modems and networks". Second IEE National Conference on Telecommunications. London UK. pp. 294–296.
2. Dostert, K (1997). "Telecommunications over the Power Distribution Grid-Possibilities and Limitations". Proc 1997 Internat. Symp. on Power Line Comms and its Applications: 1–9.

3. Datasheet Raspberry – Pi [online]. 2014 [cit. 2014-07-31]. Available from: www.element14.com/community/servlet/JiveServlet/downloadBody/65470-102-1-287848/Raspberry-Pi%20Technical%20Data%20Sheet.pdf
4. Datasheet Embedded Pi board [online]. 2014 [cit. 2014-07-31]. Available from: <http://www.farnell.com/datasheets/1705504.pdf>
5. Datasheet STM32F103 [online]. 2014 [cit. 2014-07-31]. Available from: <http://www.st.com/web/en/resource/technical/document/datasheet/CD00161566.pdf>
6. Newman, Richard; Gavette, Sherman; Yonge, Larry; and Anderson, Ross. Protecting Domestic Power-line Communications. Retrieved 2008-01-06.

Dynamic Network Analysis of 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. The study of the dynamic directed network is relatively new and interesting subject in the research of the network analysis. The purposes of this paper are represented and analyze DBLP as a directed co-author network with dynamic quality.

Keywords: Co-author Network, Stationarity, Context, Directed Network

1 Introduction

The research of this paper is focused to study and find dynamic quality of co-authors in dynamic directed network of DBLP. Co-author network of computer science bibliography (DBLP)¹ represents an example of dynamic complex network which can be analysed by various methods from the point of view focused on network evolution [1, 2]. It is also possible to investigate DBLP as a heterogeneous bibliographic network which contains multiple types of objects, such as authors, venues, topics and papers, as well as multiple types of edges denoting different relations among these objects.

In this paper, a directed co-author network is constructed using context. Context is created by extracted terms which author used in titles of articles. Different sets of terms used by different authors then give the orientation and the weight in a new evaluation of relations in the co-author graph using the context. Therefore, a dynamics of evaluated directed network is determined not for the whole network or co-author communities, but for individual authors and their co-authors. At final we proposed a modified method for determination of some dynamic centrality for a directed dynamic network.

2 Directed Co-Author Network with Context

DBLP (Digital Bibliography Library Project) is a computer science bibliography database hosted at University of Trier, in Germany. It was started at the end

^{*} This paper is based on article named ‘Dynamic Centrality for Directed Co-Author Network with Context’. Therefore I would like to thank co-authors, namely Pavla Dráždilová, Jan Martinovič, Kateřina Slaninová for cooperation with this article and agreement to its publication.

¹ Computer science bibliography (DBLP) website: <http://dblp.uni-trier.de/>

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.

From DBLP dataset we create graph of co-authors. We use terms for evaluation of the relation between co-authors. We extend standard evaluation of the relation, which is based on the number of the join publications or articles, by a factor that represent context between author and terms selected from the term set. Term set is understood as a collection of all keywords, which are extracted from titles of articles.

In paper [3] for evaluation the relation between co-authors in undirected graph used *ContextScore*. We enriched this evaluation for directed graph and used *ContextScoreD* and calculated it for selected time periods as *ContextScoreP*. A detailed description of new *ContextScoreD* and *ContextScoreP* was published in [4].

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.

We focus to analyse the behaviour of neighbourhood (exactly adjacent vertices) of selected author extracted from the network during a time period. The proposed approach which use dynamic metrics is inspired by work of Palla et al. [5].

Palla et al. [5] evaluate communities in the network using *AutoCorrelation* function. However, we are interested in dynamics of individual nodes in the network and their neighbourhood rather than dynamics of different communities in the network. Therefore, we have defined *Stationarity* for $N(A_i, t_v)$, a set of the all neighbour nodes.

Provided that we consider for each moment an unitary relation weight $w(A_i, A_j, t_v) = 1$ then original *AutoCorrelation* can be modified to Eq. 1:

$$C(A_i, t_v, t) = \frac{|N(A_i, t_v) \cap N(A_i, t_v + t)|}{|N(A_i, t_v) \cup N(A_i, t_v + t)|} = \quad (1)$$

$$= \frac{\sum_{A_j \in (N(A_i, t_v) \cup N(A_i, t_v + t))} w(A_i, A_j, t_v)w(A_i, A_j, t_v + t)}{\sum_{A_j \in (N(A_i, t_v) \cup N(A_i, t_v + t))} (\max(w(A_i, A_j, t_v), w(A_i, A_j, t_v + t)))^2}$$

Then consider, that the time axis is equidistantly divided into the years, for example $t_0 = 2000, t_1 = 2001, \dots, t_{max} = 2014$ and t is 1 year.

The *Stationarity* of neighbourhood of author A_i is defined as the average *AutoCorrelation* between subsequent states:

$$\zeta(A_i) = \frac{\sum_{t_v=t_0}^{t_{max}-1} C(A_i, t_v, t)}{t_{max} - 1 - t_0}, \tag{2}$$

where t_0 denotes the begin of the observation, t_{max} is the end of the observation and t is a step. Thus, $(1 - \zeta)$ represents the average ratio of members changed in the period [5].

We extend our approach for the directed network with context which is created from terms. We look on the *Stationarity* of neighbourhood from directed point of view. The directed edges evaluated by *ContextScoreD*(A_i, A_j) describe the influence power of author A_i into author A_j . *AutoCorrelation* is defined by a number of neighbours of the selected node. Due to this reason, the original definition would be $C_{in} = C_{out}$. However, this approach is not sufficient. Therefore, we have decided to eliminate a specific amount of edges by specific parameter *diff_{ij}*. A detailed description of *diff_{ij}* was published in [4].

We have left the edges in both directions, if the authors influence each other by the nearly same power. If one of the authors influences the other *bound* more, the stronger edge has been left during the reduction.

Then we definite the *AutoCorrelation* in directed way by the Eq.3 The *AutoCorrelation* function $C_{in/out}(A_i, t_v, t)$ is used to quantify the relative overlap of directed weighted edges between two neighbourhoods $N_{in/out}(A_i, t_v) = \{A_j; ContextScore(A_{j/i}, A_{i/j}, t_v) > 0\}$ of the same author A_i at t time steps apart:

$$C_{in/out}(A_i, t_v, t) = \frac{|N_{in/out}(A_i, t_v) \cap N_{in/out}(A_i, t_v + t)|}{|N_{in/out}(A_i, t_v) \cup N_{in/out}(A_i, t_v + t)|} \quad i = 1, \dots, |A|. \tag{3}$$

The *StationarityD* of neighbourhood of author A_i in directed graph is defined as the average *AutoCorrelation* between subsequent states:

$$\zeta_{in/out}(A_i) = \frac{\sum_{t_v=t_0}^{t_{max}-1} C_{in/out}(A_i, t_v, t)}{t_{max} - 1 - t_0}, \tag{4}$$

The more increases $C_{in}(A_i, t_0, t)$, the more dynamically is A_i influenced by its neighbourhood. If $\zeta_{in}(A_i) = 1$ then the influence of neighbourhood into A_i is more static in time. If $\zeta_{in}(A_i) < 1$ then the influence of neighbourhood is more dynamic.

The more increases $C_{out}(t)$, the more dynamic is the influence of A_i into its neighbourhood. If $\zeta_{out}(A_i) = 1$ then the influence of A_i into its neighbourhood is more static. If $\zeta_{out}(A_i) < 1$ then the influence of A_i into its neighbourhood is more dynamic.

4 Experiments

To demonstrate experiments, we have generated weighted directed co-author graphs for each year from 1980 to 2014 in the experiments. The first phase of the experiments was focused on directed edges weighted by context $ContextScoreP(A_i, A_j, t_0, t_{max})$, see [4]. Based on the consideration about the graph reduction and the reduction of less important edges due to obtaining a real image of the author’s neighbourhood, we have defined $diff_{ij} = 0,01$ and have removed the edges.

We have concentrated on several selected authors during last years in the experiments.

We can see the values of $Stationarity_{in}$ and $Stationarity_{out}$ for the author A1 in Tab. 1 and Tab. 2. The values are absolutely identical, which in our evaluation means that the dynamics of the neighbourhood that influences the author A1 and the dynamics of the neighbourhood that the author A1 influences is the same. Since the values are small, we are talking about a relatively dynamic neighbourhood of the author A1. Very similar situation is for the author A2 with the difference that his/her neighbourhood is more stable then the neighbourhood of the author A1. Observing the author A3, see Tab. 2, we can find the gradual increase of the values of $Stationarity_{out}$. This can be interpreted as a possible stabilisation of the neighbourhood, to which has the author A3 influence. It can be possible to predict its better stabilisation in the future. Considering the neighbourhood, which has the influence to the author A3, we can see in Tab. 1 that $Stationarity_{in}$ stays nearly on the same value during the analysed time period. That means that the co-author community of the author A3 that influences him/her is permanently dynamic and do not stabilises.

Author	$\zeta_{in}2005 - 2010$	$\zeta_{in}2006 - 2011$	$\zeta_{in}2007 - 2012$	$\zeta_{in}2008 - 2013$
A1	0.17334	0.17123	0.16423	0.16728
A2	0.22614	0.23612	0.22833	0.21554
A3	0.24186	0.27344	0.27344	0.29395
A4	0	0	0.08	0.18
A5	0	0	0	0.033
A6	0.22051	0.21232	0.180505	0.20777

Table 1. $Stationarity_{in}$ for selected authors and selected time period

Author	$\zeta_{out}2005 - 2010$	$\zeta_{out}2006 - 2011$	$\zeta_{out}2007 - 2012$	$\zeta_{out}2008 - 2013$
A1	0.17334	0.17123	0.16423	0.16728
A2	0.21446	0.22445	0.22375	0.21554
A3	0.22282	0.24504	0.23852	0.24236
A4	0	0	0.08	0.18
A5	0	0	0	0.075
A6	0.19422	0.19068	0.16129	0.19462

Table 2. $Stationarity_{out}$ for selected authors and selected time period

5 Conclusion

In the paper, we proposed a modified method for determination of $Stationarity$ in a directed network. As the edge evaluation by $ContextScoreD$ means the knowledge scope, which one author can provide the other author, the $Stationarity_{out}$ during the time corresponds with the influence power, which one author could have to the other co-authors. Contrary to the previous statement, $Stationarity_{in}$ during the time corresponds with influence power from the other co-authors to the given author. Presented experiments show that the stability measure of the selected authors is low and the set of co-authors change in time. Moreover, the influence power of the author to his/her neighbourhood differs from the influence power from his/her neighbourhood to the author. We intent to focus on other types of weighted directed networks, in which is important to determine $Stationarity$ of its members in the future.

6 List of publications

Dynamic Centrality for Directed Co-Author Network with Context

A. Babskova, P. Dráždilová, J. Martinovič, K. Slaninová, In proceedings of the CISIM2014, 2014

References

1. A.-L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 3311(3):590-614, 2002.
2. I. Belykh, M. di Bernardo, J. Kurths, and M. Porfiri. Evolving dynamical networks. *Physica D: Nonlinear Phenomena*, 267(Complete):1-6 2014.
3. A. Babskova, P. Dráždilová, J. Martinovič, K. Slaninová. Topic extension using the network extracted from DBLP, *EMSS2013*, 2013
4. A. Babskova, P. Dráždilová, J. Martinovič, K. Slaninová. Dynamic Centrality for Directed Co-Author Network with Context, In proceedings of the *CISIM2014*, 2014
5. G. Palla, Albert-lászló Barabási and Tamás Vicsek, and B. Hungary, Guy I. Quantifying social group evolution. *Nature*, (46):664667, 2007.

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 the 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 $q_{l_1} \leq T.attr_1 \leq q_{h_1}$` . Multidimensional range query is often implemented by multidimensional data structures, e.g. n -dimensional B-tree [5], R-tree [6] or the R*-tree [3] and it can be processed for the following SQL statement: `SELECT * FROM T WHERE $q_{l_1} \leq T.attr_1 \leq q_{h_1}$ AND ... AND $q_{l_n} \leq T.attr_n \leq q_{h_n}$` . 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.

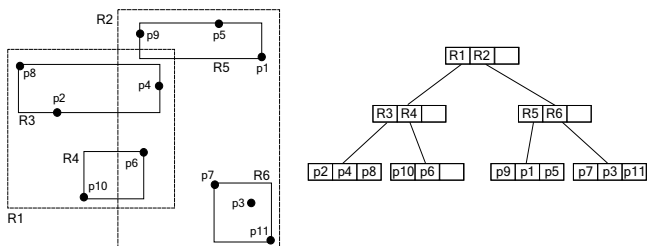


Fig. 1. Planar representation and general structure of the R-tree

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 the 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.

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.

2.1 Range Query Processing Using CPU

We utilize three common approaches to traverse the r-tree. In the case of **depth-first search (DFS)** 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.

Second approach is based on **breadth-first search (BFS)** algorithm. In this case algorithm traverses the r-tree level by level. We utilize two buffers. One for current level and another for relevant child nodes. At the beginning first buffer contains only root node. We search all nodes in first buffer and if there are relevant child nodes matched by the query rectangle they are placed in the second buffer. We use same methods `IsIntersected` for inner nodes and `IsInRectangle` for leaf nodes. When first buffer is completely searched we switch buffers. The first buffer contains relevant nodes from previous search and second one is empty. The algorithm continues until it reaches leaf level.

We also propose combined algorithm called **depth-breadth first search (DBFS)**. This combines both previous approaches. In this case we need buffer for each level of the r-tree. We start searching the r-tree by DFS manner. When an item is relevant instead of searching we place it in buffer for next level. If the buffer for next level is full we search it in order to clear buffer and continue with current level. When all nodes in current level buffer were searched we continue by searching next level.

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 same approaches as for the CPU. Only search part is executed on the GPU. The difference between CPU algorithm is that we need to transfer all nodes to GPU's memory first. After all leaf nodes are transfered to GPU's memory we invoke GPU algorithm. The GPU algorithm cannot access CPU's memory so we cannot add corresponding items directly into result set. Instead we return only vector of informations which tuples should be added to result. The result set is then filled on CPU based on returned vector.

3 Experimental Results

In our experiments, we compare the performance of range query processing using CPU and GPU algorithms on various data collections and devices. We tests all three approaches (BFS, DFS an DBFS).

Due to lack of space we only present result for POKER¹ data 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 $10\times$ 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.

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 Tables 2, 3 and 4 we show performance of CPU and GPU algorithms². In the case of GPU we consider only situation where all data are already stored on the GPU. For all search methods we can see GPU's performance is up to 40% better in the case of low selectivity. For higher selectivity the performance of CPU is better or similar. The reason of this is because the work load for such situations is low and we cannot fully utilize GPU resources. We also compare the performance between different search techniques. From results it is obvious breadth first search has slightly better performance than depth first search or their combination depth-breadth first search. In case of BFS we use large buffer to contain all nodes to be searched. But in most situations the buffer capacity will not be sufficient. We handle buffer overflow with combined algorithm DBFS. The results show the DBFS performance is about 10% worse over BFS.

As we described in the previous chapter. GPU's algorithm cannot fill result set directly. In Table 5 we depicted performance of both algorithms without adding to the result set. The performance of single GPU algorithm is up $2.1\times$ better.

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

² The experiments were executed on NVIDIA Geforce GTX 550 Ti with 1GB of DDR5 (Memory speed 4,104 Mhz), 4 SM, 48 cores/SM; Intel Core i5-P2450 3.2Ghz, 6.0 MB L2 cache; 8GB of DDR3; Windows 7 SP1.

Table 2. R-tree Range Query Results for DFS

Result Size	DFS			
	CPU	GPU	# IN	# LN
1	0.0001	0.0002	53.2	10.8
283.2	0.0033	0.0034	1722.8	1046.4
22194.8	0.0248	0.0222	10757.2	8328.8
425567.3	0.1121	0.0791	32264.5	27101.5

Table 3. R-tree Range Query Results for BFS

Result Size	BFS			
	CPU	GPU	# IN	# LN
1	0.0001	0.0006	21.7	10.8
283.2	0.0033	0.0041	338.7	1046.4
22194.8	0.0267	0.022	1214.7	8328.8
425567.3	0.1142	0.0748	2582	27101.5

4 Conclusion

In this article, we compared CPU and GPU variants of range query algorithms in the R-tree. Based on our experiments we identified some bottlenecks of processing on GPU.

We summarize the results. (1) The results of GPU shows the algorithm can be faster for low selectivity. (2) Performance of GPU algorithm is significantly reduced by data transfer (3) In the case of insufficient work load the GPU can be very ineffective.

As a result to an integration into a DBMS kernel we must solve mainly these issues. In current state of the art we already make some precautions

- We included GPU node cache in order to reduce data transfers to GPU.
- We designed GPU load balancing algorithm to fully utilize search on GPU.

In our future work we need to implement an query processor which will be able to plan range query execution for best performance using all available devices.

Table 4. R-tree Range Query Results for DBFS

Result Size	DBFS			
	CPU	GPU	# IN	# LN
1	0.0001	0.0008	21.7	10.8
283.2	0.0032	0.0033	338.7	1046.4
22194.8	0.0259	0.0182	1214.7	8328.8
425567.3	0.1093	0.0674	2582	27101.5

Table 5. Comparison of Range Query Algorithms Without Adding to Result

Result Size	CPU Algorithm [ms]	GPU Algorithm [ms]	Speedup
1	0.0001	0.0002	0.4x
283.2	0.0025	0.0017	1.5x
22 194.8	0.0189	0.0106	1.8x
425 567.3	0.0671	0.0322	2.1x

5 Projects and Publications

Pavel Bednář is a member of team solving the grants SP2014/187 – Analýza spolehlivosti sítí a optimalizace využití OZE, SP2014/211 – Škálovatelnost databázových systémů and FRVŠ 230/2014 - Inovace kontroly projektů a portálu pro výuku databázových předmětů. 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.

Stains detection on nanoparts images

Petr Berek, Jan Platos, and Vaclav Snasel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{petr.berek, jan.platos, vaclav.snasel}@vsb.cz

Abstract. In this paper we describe how we used fuzzy k-means method for stains identifying and counting. We used well known method for image segmentation to identify stains. We describe how we created some stain patterns and how we used them to stains bordering and counting by hierarchical clustering.

Keywords: image analysis, image processing, image segmentation, fuzzy, clustering, fuzzy clustering, fuzzy c-means, nanoparts

1 Introduction

Image analysis is very important in a lot of tasks - i.e. in medical images we can detect cancer[6], in production we can detect some defects, etc...[1] Image analysis or image processing contains a lot of methods which we can use to detect some features (as shapes, colors, textures, ...).[9] We can divide image analysis into several phases. First phase is about image capturing. It is very important phase because of noisy and demands on equipment.[9] In case of nanomaterials we need accurate electron microscopes. If there is noise on medical images there may be some feature false detected.

In next phase we need divide image into some classes by some features. This phase is called as segmenantation. It is very important in automatic image analysis because entities of interest are extracted for next processing as description and recognition.[9] There are a lot of methods how to image can be divided into classes. Fuzzy k-means (c-means) is well known method for fuzzy clustering and image segmentation.[2][3][4][5][6] The aim of clustering is divided data into clusters so that entities in clusters are as similar as possible. In fuzzy clustering each entities is in each cluster but it is associated with each cluster by set of membership level.[7]

Fuzzy k-means algorithm is based on iteratively minimizing cost function 1 where N is set of data, C is set of cluters, u_{ij} is membership data x_j to c_i and v_i is i -th cluster center. Membership functions are updated by equation 2 and cluster center are updated by 3. Parameter m is fuzziness parameter. In most of

application it is set to 2.[8]

$$J = \sum_{j=1}^N \sum_{i=1}^C u_{ij}^m \|x_j - v_i\|^2 \quad (1)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^{2/(m-1)}} \quad (2)$$

$$v_i = \frac{\sum_{j=1}^N u_{ij}^m x_j}{\sum_{j=1}^N u_{ij}^m} \quad (3)$$

2 Experiment

2.1 Collection

Data collection was created by two images of nanoparts. These images were in gray scale. Because we didn't know which area is created by stains and which didn't we created collection of ten images for each origin image. Images in this collection were in black and white only, didn't in gray scale. These images were created by c-means method known as fuzzy k-means. This method is known in area of computer graphics. We used non-overlap blocks created by 3x3 pixels from original image. These blocks were clustered by fuzzy k-means into 10 clusters. We gave color to each cluster so that cluster contained darker blocks assigned color was warmer and vice versa.

We used both algorithms, k-means and c-means, with different block size - 2x2, 3x3, 4x4 and 8x8 pixels. When we used blocks 8x8 pixels for clustering, results were very poor. So we could use 2x2 or 4x4 pixels so we decided to use average - 3x3 pixels. We tried on different clusters count - 3, 5 and 10 clusters. If we use more cluster result is more precise. So we used 10 clusters after that. Both algorithms were set to same finish condition parameter. This parameter is distance between two last iterations. We set it to 0.01. This distance was measured by euclidean distance. Fuzziness parameter m was set to 2.

2.2 Pattern of stains

We created some patterns because of search of stains. These patterns had 3x3 pixels. There were some shapes which can identify stains. We generated all shapes by combination of black pixels -from one black pixel and 8 white pixels to 9 black pixels. For our work we assumed only pattern with more then five black pixels can be interested. From these patterns of interested we selected manually about 30 patterns which can identify stains. There are some instances of patterns on Figure 1.

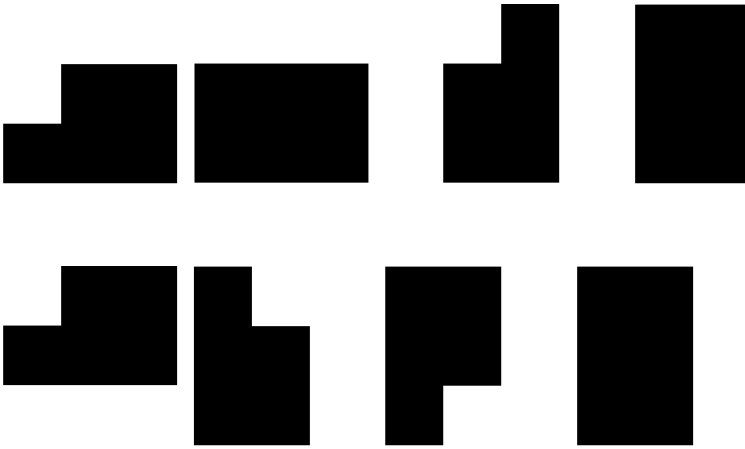


Fig. 1: Instances of stain patterns

2.3 Search of stains

In order to identify stains we created overlapping blocks 3x3 pixels from black and white image created by c-means by sliding window about 1 pixel in x and y axis. Each block was compared with each pattern. If block was same as pattern then it was marked as part of a stain. Blocks which was not marked was removed. We assumed if blocks are overlapping then they must created stain.

We looked for stains in two phases. At first we did not use pattern with all black pixels. Result of this was method for stains bordering. At second phase we used all patterns. This method created similar image as c-means method. The goal of this method was identify what is stain. We edited this method so that we assigned a number to each block. If two blocks are overlapped they are created by same stain. We changed block number to same. It is like hierarchical clustering. But we don't know count of clusters. We clustered blocks by count of overlap pixels.

3 Results

There are some information about input images in table 1. We tried k-means to identify similar blocks at first. Then we compared these results with c-means algorithm. You can see it on Figure 2 and 3. As we wrote above darker points have warmer color and vice versa. But image created by c-means algorithm can be created with better representation of color. You can see c-means algorithm identifies stains better than k-means. Darkest areas - areas with warmest color - are smaller on image created by c-means then the darkest areas on image created by k-means.

On Figure 4 is shown how to bordering works. At first there is image created by c-means. This image is created by drawing of clusters which contain blocks

with the first, second and third darkest stains. Then we applied comparison of overlap blocks and our patterns and created stains border. Finally, we redrew borders to input image.

In table 2 and 3 is count of stains on images. We mention only two images created by c-means for each input image. These images was created by the first and the second darkest color by c-means clusters. X in table means count of overlap blocks. The first row is for first input image so that we created new image only from the first darkest cluster. Second row is for image created by the first and the second darkest cluster. Third row is for second input image etc...

Image	Count of non overlapping blocks	Count of overlapping blocks	Resize of image
4_17	87552	782595	1024x768
4_18	157896	1413846	1376x1032

Table 1: Information about input images

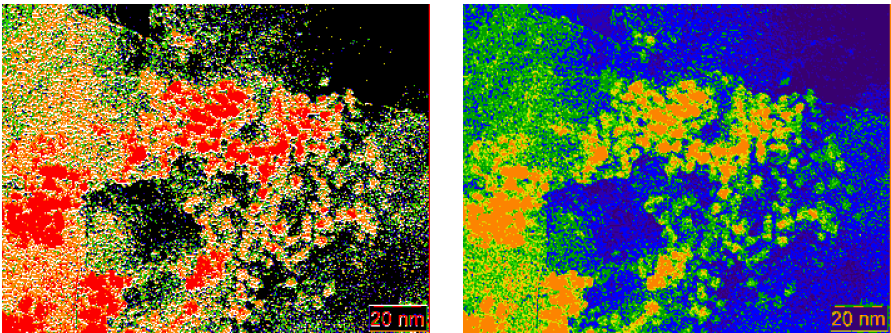


Fig. 2: Comparison of images created by k-means and c-means

Image	$X \geq 1000$	$1000 > X \geq 500$	$500 > X \geq 250$
1_1	4	14	11
1_2	14	5	26
2_1	18	14	19
2_2	29	24	43

Table 2: Count of stains on images

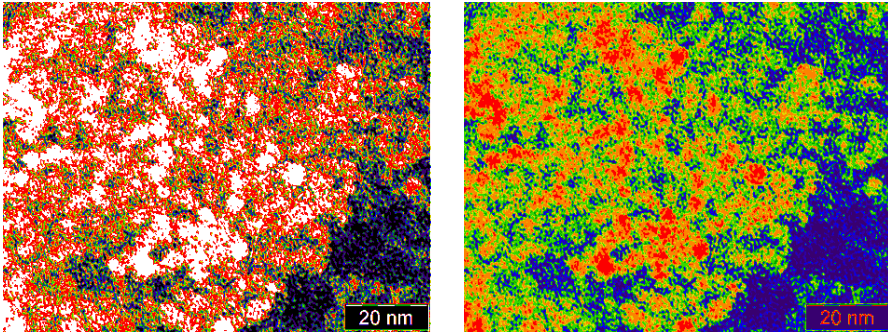


Fig. 3: Comparison of images created by k-means and c-means

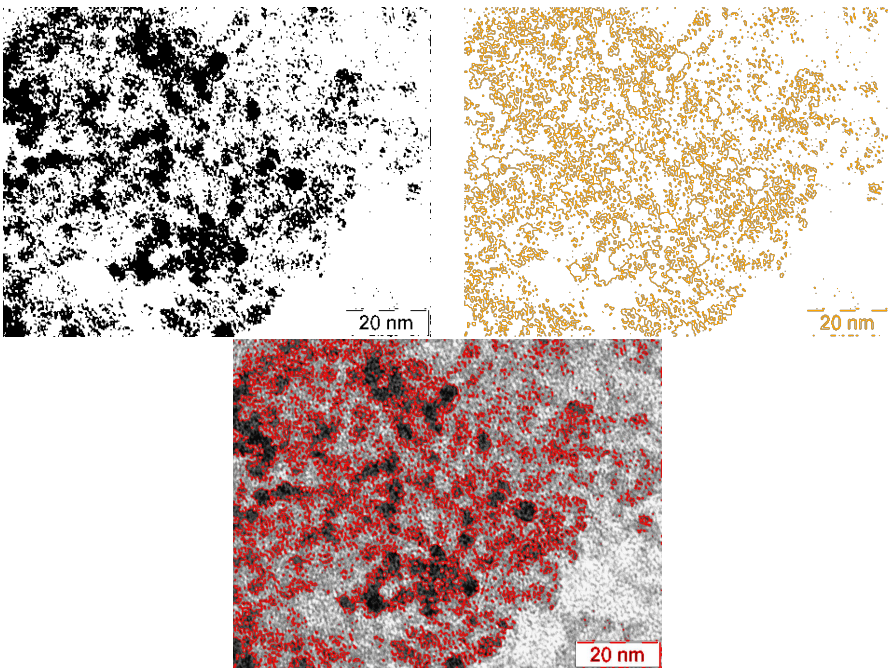


Fig. 4: Stains bordering

Image	$250 > X \geq 125$	$125 > X \geq 62$	$62 > X \geq 31$
1_1	16	15	35
1_2	22	45	65
2_1	38	67	153
2_2	87	205	285

Table 3: Count of stains on images

4 Conclusion

We described our method for stains counting on images. We used standard method used for image segmentation - fuzzy k-means - to identify which pixels could create stains. Then we created some stains patterns and we showed how to use them to stains bordering and counting. We carried out this method on two images only because we don't have larger dataset. But we want to improve this method and try it on larger dataset.

References

1. E R Davies, IMAGE PROCESSING FOR THE FOOD INDUSTRY, World Scientific (2000), ISBN 981-02-4022-8
2. NIKHIL R. PAL and SANKAR K. PAL, A REVIEW ON IMAGE SEGMENTATION TECHNIQUES (1993)
3. Bezdek J, Hall L, Clarke L. Review of MR image segmentation using pattern recognition. *Med Phys* 1993;20:1033â48.
4. Lyer NS, Kandel A, Schneider M. Feature-based fuzzy classification for interpretation of mammograms. *Fuzzy Sets Syst* 2002;114: 271â80.
5. Yang MS, Hu YJ, Lin KCR, Lin CCL. Segmentation techniques for tissue differentiation in MRI of Ophthalmology using fuzzy clustering algorithms. *Magn Reson Imaging* 2002;20:173â9.
6. Tobias Chriiian CahoonA, Melanie A. Suttorf, James e. Bezdek, BREAST CANCER DETECTION USING IMAGE PROCESSING TECHNIQUES, *Fuzzy Systems*, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference; 973 - 976 vol.2
7. Bezdek, James C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. ISBN 0-306-40671-3
8. Chuanga KS., Tzenga HL., Chena S., Wua J., Chenc TJ., *Fuzzy c-means clustering with spatial information for image segmentation*, *Computerized Medical Imaging and Graphics*, 2006, 9â15
9. Tinku Acharya, Ajoy K. Ray, *Image Processing - Principles and Applications*, 2005, ISBN:0471719986

The stock market's solution synthesized by Analytical programming

Marian Bielík and Tomas Vantuch

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{marian.bielik.st, tomas.vantuch.st}@vsb.cz

Abstract. Analytical programming is a method that generates the elementary functions often very complex functionalities that can be used in the symbolic regression. Symbolic regression is data substitution with appropriate mathematical formulas that best fits a given dataset, in this case stock markets data.

1 Introduction

Symbolic regression [1] represents a process during which measured data sets are fitted, thereby a corresponding mathematical formula is obtained in an analytical way. Algorithms that solve symbolic regression are Genetic programming, Grammatical evolution, Analytical programming. AP is a new method of symbolic regression and is described and used in this paper. AP is a grammar free algorithmic structure. It can use any Evolutionary Algorithm, in this case Genetic algorithm (GA) and Particle swarm optimization (PSO), described later. AP is used with markets data and implementation, results are described in this paper together with suggestions for future work.

1.1 Analytical programming

Analytic programming (AP) is a new method of symbolic regression. The basic principles were developed and published in [2, 3]. It is based on different principles than genetic programming or grammatical evolution, these principles are described in [2]. AP can use any genetic algorithm. This paper uses genetic algorithm (GA) and particle swarm optimization (PSO) described later. AP is based on the set of functions, operators and terminals. All these mathematical objects form a functional set with different number of arguments called a general functional set (GFS), that AP uses to synthesize appropriate solutions. This paper uses these subsets:

The AP individuals consist of non-numeric expressions (operators, functions) that are represented in the evolutionary process by the integer indexes. This index is then used as a pointer to a set of terms that AP uses to synthesize a target program to evaluate the objective function.

- GFS(all)= {+, -, *, /, Sin, Cos, Tan, Cot, 1, X_t , X_{t-1} , Y_t , Y_{t-1} , Z_{t-1} , Y_{t-2} }
- GFS(2 arguments)= {+, -, *, / }
- GFS(1 arguments)= {Sin, Cos, Tan, Cot}
- GFS(0 arguments)= {1, X_t , X_{t-1} , Y_t , Y_{t-1} , Z_{t-1} , Y_{t-2} }

Table 1. Subsets of the grammar's symbols

AP is therefore a series of functional mapping. Though it is necessary to comply with the rules, to ensure that every individual will represent a unique function. Example of this mapping is shown in the [obrazok 1]

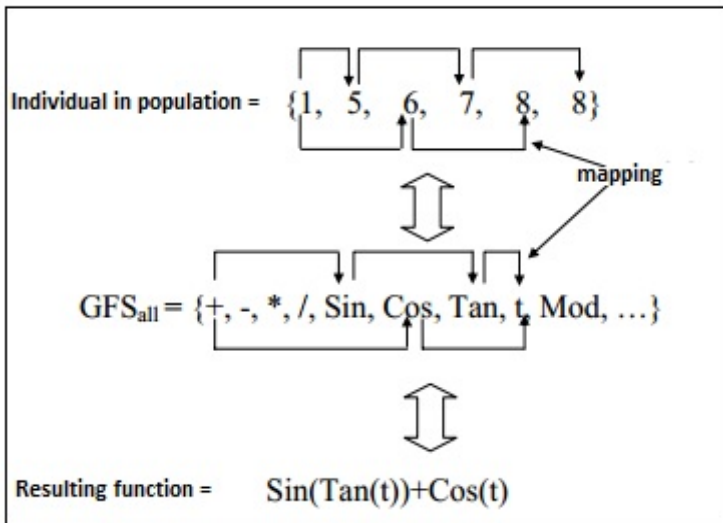


Fig. 1. Substitution algorithm of analytic programming method

1.2 Particle swarm optimization

The PSO is the evolution algorithm based on swarm intelligence [7]. The main part of this approach is to look for the answer in n-dimensional hyperspace. In this area there are spread all individuals, called particles, to come thru it, communicate to each other and evaluate themselves to get the best position.

Each particle is defined by its vector determining its position, its velocity to proceed the movement and the memory of the latest best position. The best position could be taken from all particles in the space or from the smaller amount of them. In case of creating local extremes instead of global ones, we call, the PSO is using the item of neighborhood [8]. In this case the particles are not

affected by the whole group, but only by their adjusted neighbors, that were chosen randomly to get more opportunities in finding the solution.

The PSO is using the objective function to get the fitness evaluation of the particle in each iteration. The objective function is the same as in the GA approach, to fit the historical data of the predicted time set.

1.3 Genetic evolution

The GA is an algorithm inspired by the evolution theory described by Darwin [5]. There is a defined set of individuals forming the population and each of this individual is simply described only by its vector of defined length. The first population is created from the seed, the random number, obviously taken from some random number generator.

There is a fitness function for candidates evaluation and according to the evolutions theory, the best one has to be kept and used for creating the new generation. The new generation is made by crossover of the best individuals from the previous generation. This approach has to keep the best properties, that are founded in the generation. In the end of crossover, there can be optionally used the added mutation, which stands for the change of the one random bit in breded individual. This brings to the system some added randomness.

After the defined amount of generations, there is a hypothesis that the best candidate from the last population will brings the closes answer to the given problem. The problem was adjusted the same as in the case of the PSO algorithm.

2 Experiment Design

Experiments in this paper are focused to obtain polynomial that will represents stock markets data. These data are Microsoft daily data from years 2010, 2011 and 2012. Variables for polynomials are markets daily open, close prices and volumes.

Analytical programming is implemented in Java and connected firstly to PSO and then to GA. PSO is supported by Java library JSwarm and GA was obtained from the ECJ framework [9]. Created polynomials by AP has to be evaluated and for this purpose is used connection to Matlab, where in loop we can compute the sum of all differences between compared time sets and time sets made by polynomials.

Open and close prices are highly correlated and AP can this issue reflect by generating bad results when it will return $close_t = open_t$ as enough good polynomial. To avoid this not wanted results, there is implemented checking of the length of the polynomials and all polynomials of length one are penalized as individuals with zero fitness value.

3 Conclusions

At first we have to say that the quality of results does not depend on analytic programming but on chosen evolution algorithms and their setting. This becomes

	GA	PSO
individuals / particles	150	40
generations / iterations	100	100
genotype length	1000	1000
mutation probability	100%	
max velocity		50

Table 2. Adjustment of the optimization algorithms

the weakest part of our concept. As it was mentioned in previous chapter, the fitness was not only to minimize the differentiation between polynomials time set and the fitted time set, but it was adjusted to find good enough polynomial. There was issues like, the polynomial consists only of one item, the polynomial generates time set that contains equals values to open or close prices, etc. All of these candidates was penalized by zero fitness value. The other issue appears to us the crossover in case of GA and the orientated movement in case of PSO. The one-point crossover [6] does not guarantee that the new individual will make better polynomial or even valid one. Here you can see the progress of fitness values over 100 generation.

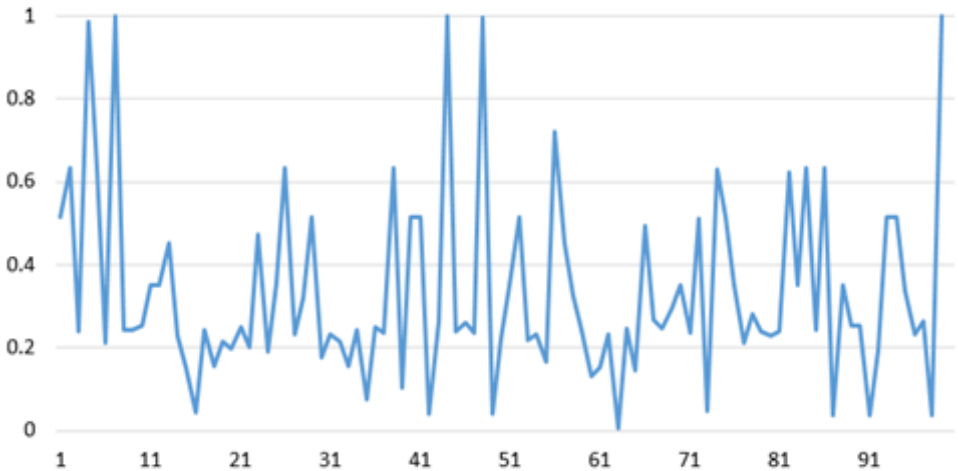


Fig. 2. GA's fitness progress

The progress is unstable and its almost randomly option to get better individual in next generation from the previous.

The combination of the Analytical programming and GA created polynomial that consists of three parts with closely fitting to close values.

$$close_t = open_{t-1}/(volume_{t-1}-close_{t-1})-close_{t-1}-sin(open_t/volume_{t-1}) \quad (1)$$

As you can see, the parts that contains division by volume are very close to zero, because of high values of volume time set and this makes the result almost equal to close value moved one iteration back. Unfortunately this result is insufficient. The progress of resulted polynomial according to fitted time set you can see in chart below.

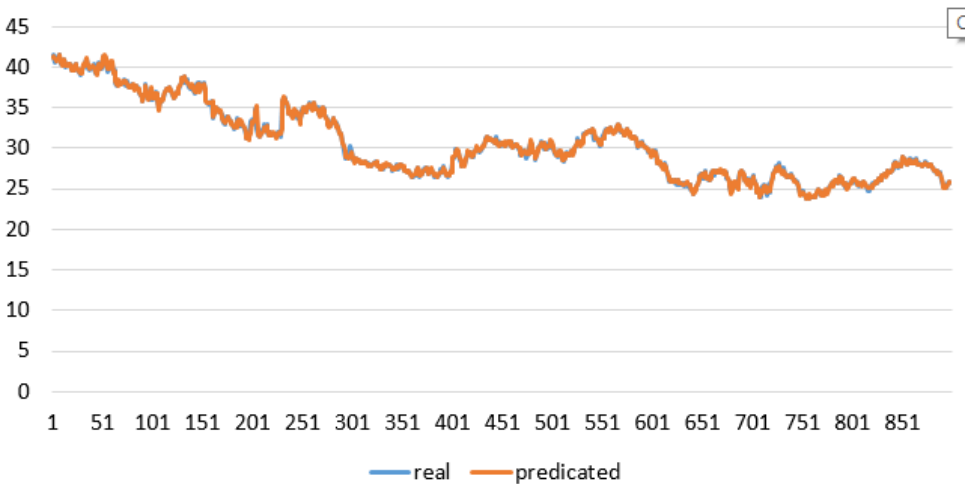


Fig. 3. Progress of polynomial’s time set according to fitting data

Results of combination of the AP and the PSO are similar as before. It created polynomial that generates data very close to original markets data. The movement of PSOs particles was very early affected by best position of founded polynomial that reflects only values of the open prices time set. Adding more and more iterations does not improve the result.

$$close_t = tan(open_t/open_{t-1}/open_{t-1}) - open_{t-1} \quad (2)$$

3.1 Future work

Because the results of our experiments was insufficient, there we have to introduce our next steps in our work.

First of all, it has to be implemented the crossover that works based on the symbolic regression parts [4]. Crossover like this will split the whole individuals vector into blocks, each of them will represent the part of the polynomial and

the crossing of them will guarantee that all new candidates will keep the best properties from previous generation and will be valid.

The next step will to extend the amount of added constants. This we can proceed by finding correlated time series thru market data. In this task we can use some kind of self-organizing maps [10] or other algorithms. The more constants we have, the more combination has our evolution algorithm.

And the last task that seems to be really necessary is to do some preprocessing of input data to prevent result like this. The preprocessing can be for example differentiation of the time set.

ACKNOWLEDGMENT

The following grants are acknowledged for the financial support provided for this research: Grant Agency of the Czech Republic - GACR P103/13/08195S, is partially supported by Grant of SGS No. SP2014/159, VB - Technical University of Ostrava, Czech Republic, 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.

References

1. Ivan Zelinka, Zuzana Oplatkov, Milos Seda, Pavel Osmera, Frantisek Vcelar, 2008, Evolucni vypočetni techniky principy a aplikace.
2. Evolutionary Computation - Principles and application (2008), Ivan Zelinka, Donald Davendra, Roman Senkerik, Roman Jasek and Zuzana Oplatkova, chapter 9
3. Michael O'Neil, Conor Ryan, 2003, Grammatical Evolution
4. Nguyen Quang Uy, Nguyen Xuan Hoai, Michael O'Neill, R. I. McKay, Edgar Galvnlpez, 2010, Semantically-based crossover in genetic programming: application to real-valued symbolic regression
5. John R. Koza, 1992, Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, Massachusetts.
6. Riccardo Poli and W.B. Langdon, Genetic Programming with One-Point Crossover
7. J. Kennedy and R. C. Eberhart, Particle swarm optimization, in Proc. IEEE Int. Conf. Neural Networks, Perth, Australia, Nov. 1995, pp. 1942-1948.
8. J. Kennedy, The particle swarm: Social adaptation of knowledge, in Proc. 1997 Int. Conf. Evolutionary Computation, Indianapolis, IN, Apr. 1997, pp. 3033-3038.
9. David R. White, Software review: the ECJ toolkit, March 2012
10. Kohonen, T., 1990, The self-organizing map

Flexible Neural Tree in Pseudo-Periodic Time-Series Forecasting

Tomas Burianek

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`tomas.burianek.st1@vsb.cz`

Abstract. The time-series forecasting is a significant area of Machine Learning field. Neural Networks are heavily used models for forecasting. This paper presents type of Neural Network under the name Flexible Neural Tree and its performance in Pseudo-Periodic Time-Series forecasting.

Keywords: Flexible Neural Tree, Genetic Programming, Particle Swarm Optimization, Time-Series Forecasting

1 Introduction

The *Flexible Neural Tree* (FNT) has been used in many time-series forecasting problems in past. This model has irregular flexible tree architecture and it is optimized during learning procedure by one of the available tree optimization algorithms. Tree parameters including weights of connections between nodes and parameters of activation function in functional nodes may be optimized by global continuous optimization techniques. This paper presents performance of FNT on simulated Pseudo-Periodic time-series and it is organized as follows: in section 2 theory of FNT is presented with Genetic Programming and Particle Swarm Optimization algorithms used for topology and parameters optimization respectively, in section 3 performed experiments are described with results discussion and finally in section 4 conclusion of FNT performance is presented.

2 Flexible Neural Tree

Flexible Neural Tree (FNT) is a kind of *Neural Network* (NN) introduced in [4, 5]. It is multilayer feed-forward NN where the topology is represented as a tree containing functional nodes with activation functions and leaf nodes as inputs to the network. Output of the network is the root functional node in the tree. This tree topology is constructed in an automatic way during evolution process. This allows to find the optimal tree structure design and also selects appropriate inputs to the network. Also connection weights between nodes and activation function parameters are optimized during evolution process. Advantage of FNT

model is that for topology and tree parameters different optimization techniques may be used. Topology construction may be done by *Probabilistic Incremental Program Evolution* (PIPE) [11, 4], *Genetic Programming* (GP) [2, 1] or *Ant Programming* (AP) [3]. Tree parameters (connection weights and activation function parameter) may be optimized by global continuous optimization techniques. Heuristic methods such as *Genetic Algorithm* (GA) [9], *Differential Evolution* (DE) [14, 10] and *Particle Swarm Optimization* (PSO) [6, 13] are good for this task.

2.1 Tree construction

A new tree is constructed from pre-defined instruction set $\mathcal{S} = \mathcal{F} \cup \mathcal{T}$. Terminal set $\mathcal{T} = \{t_1, t_2, \dots, t_{N_t}\}$ consists of input nodes to the neural tree. Functional set $\mathcal{F} = \{+_2, +_3, \dots, +_{N_f}\}$ consists of functional node operators where $+_i$ has i inputs. This functional node has a Gaussian parametric activation function. Then output of functional node is:

$$out = f(a, b, net) = e^{-\left(\frac{net-a}{b}\right)^2}. \quad (1)$$

Parameters a and b are adjustable. Parameter net is a total input charge of functional node $+_i$:

$$net = \sum_{j=1}^i w_j x_j, \quad (2)$$

where w_j is connection weight between functional node and input x_j to this node. Input x_j to the functional node could be output from another functional node $+_i \in \mathcal{F}$ or input terminal node $t_k \in \mathcal{T}$. Output from network is computed by depth first method for tree traversing from the root functional node to terminal nodes.

2.2 Objective function

Accuracy evaluation of a neural tree during optimization is obtained using objective function. Supervised learning is processed on training set $\{(x^p, y_{tar}^p)\}$ where $p = 1, \dots, P$. The goal is to predict target output pattern y_{tar}^p for given input pattern x^p . Inaccuracy measure of obtained prediction y_{pre}^p is computed by *Mean Square Error* (MSE) defined in expression:

$$MSE = \frac{1}{P} \sum_{p=1}^P (y_{pre}^p - y_{tar}^p)^2. \quad (3)$$

2.3 Topology optimization

Optimization of neural tree topology is performed by *Genetic Programming* (GP). Theory of GP was introduced by Koza [8]. The algorithm starts with initial

population of random created computer programs here presented as the trees. Then iteration process is performed where in each step a new generation of trees is created from previous generation. First, fittest individuals are selected using *tournament selection* according to minimal MSE described in equation 3. Then a new trees are created by genetic operators namely: *reproduction*, *crossover* and *mutation*. The reproduction operation only copies one individual tree to a new generation. The crossover operation creates two new trees from two old trees by swapping their random picked sub-trees. The mutation is processed on one tree by random changing of its structure by changing leaf nodes to another leaf nodes or replacing one leaf node by new generated sub-tree or changing one functional node to leaf node.

2.4 Parameter optimization

There are two kinds of parameters to optimize. Activation function parameters a and b from each functional node and weight w connections between each node of a tree. Optimization is performed by well known evolutionary algorithm *Particle Swarm Optimization* (PSO). The PSO algorithm introduced by Kennedy and Eberhart in 1995 [6] is inspired in social behaviors in swarm population. Swarm consists of set of particles. Each particle i has its position p_i and velocity v_i in multidimensional space of defined by optimized problem. Each particles stores its best (local) position b_i and also best (global) position b_g in a swarm is stored. Particles flies in multidimensional space and its movement is influenced by local and global best positions. In each evolution step a new velocity for every particle i on every dimension j is obtained by equation:

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 r_{1j}(b_{ij}(t) - p_{ij}(t)) + c_2 r_{2j}(b_{gj}(t) - p_{ij}(t)) , \quad (4)$$

where ω is inertia weight introduced in [12], c_1 and c_2 are constants that regulate influence of local and global positions on particle. Variables r_{1j} and r_{2j} are uniformly selected random numbers. New position of particle is computed from previous position and a new velocity using following equation:

$$p_{ij}(t+1) = p_{ij}(t) + v_{ij}(t+1) . \quad (5)$$

After position update of every particle in a swarm, particles are evaluated by objective function and a new local b_i and global b_g optimal solutions are selected according to the best evaluations.

Tree parameters a , b and w are encoded as a vector:

$$(a_1, \dots, a_{N_f}, b_1, \dots, b_{N_f}, w_1, \dots, w_{N_{con}}), \quad (6)$$

where N_f is number of functional nodes and N_{con} is number of connections and it represents position of a one particle with size $2N_f N_{con}$. Output from PSO is then best evaluated particle and it represents best parameters setting. As objective function the MSE from equation 3 is used.

2.5 FNT general learning algorithm

First, initial population of trees is randomly created. Then topology optimization is processed by Genetic Programming algorithm for given number of iterations. After optimization of topology, parameters of every tree in population are optimized by Particle Swarm Optimization, each for given number of iterations. Tree with best objective function evaluation is stored. Then topology optimization followed by parameters optimization is repeated until number of iterations of whole learning algorithm is reached. Solution is the best evaluated tree of the population.

3 Experiments and Results

Experiments were performed on Pseudo-Periodic Synthetic Time-Series Data Set. This data were obtained from UCI Machine Learning Repository [7]. The data appears highly periodic, but never repeats itself. It was originally designed for testing indexing schemes in time series databases. There are 10 data columns, each has 100000 samples. Each of this data column was generated by independent invocations of the function:

$$y = \sum_{i=3}^7 \sin(2\pi(2^{2+i} + \text{rand}(2^i))t), \quad (7)$$

where $0 \leq t \leq 1$. The goal for FNT was to predict next sample value from window of previous samples. For experiments were picked first two data columns. Experiments were performed for windows of 10 sample values to the past to predict next 1st and 5th sample value to the future. Data sets were split to training set with 70% of data and to testing set of remaining data. Final results of FNT prediction on testing data set are shown in figure 1, comparing 1st and 5th sample value prediction. It is obvious, that 1st sample value prediction has better results than for 5th sample value prediction. Comparison of FNT performance on different data sets and different predictions are shown in table 1 with final MSE for each result. Best result was obtained by predicting 1st data sample on *Pseudo-Periodic Data Set 1*.

Prediction	Pseudo-Periodic Data Set 1	Pseudo-Periodic Data Set 2
1st data sample	2.8319×10^{-5}	4.9124×10^{-5}
5th data sample	7.5873×10^{-5}	8.5625×10^{-5}

Table 1. Table of errors showing MSE for two data sets with 1st and 5th data sample prediction with black solid line as a prediction and dashed red line as a target.

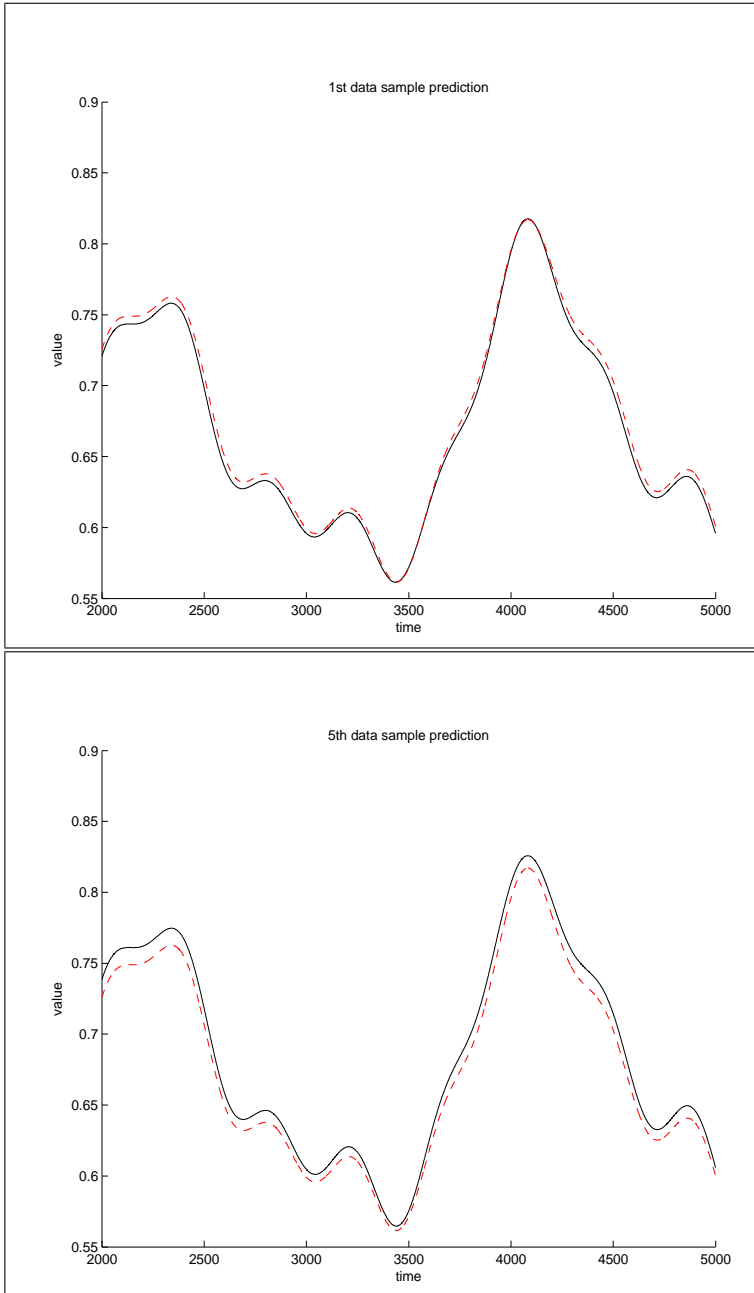


Fig. 1. Results for *Pseudo-Periodic Data Set 1* comparing 1st and 5th data sample prediction.

4 Conclusion

The *Flexible Neural Tree* is a specific type of Neural Network with meta-heuristic techniques for finding good topology and settings of parameters in a tree. FNT model is presented as a very powerful tool for time-series prediction. Its abilities were proven for two benchmark Pseudo-Periodic Time-Series data sets with good prediction results for both data sets. Best result was obtained for the first Pseudo-Periodic Data Set with MSE equal to 2.8319×10^{-5} .

References

1. Chen, Y., Abraham, A., Yang, B.: Feature selection and classification using flexible neural tree. *Neurocomputing* 70(1-3), 305–313 (2006)
2. Chen, Y., Yang, B., Abraham, A.: Flexible neural trees ensemble for stock index modeling. *Neurocomputing* 70(4-6), 697–703 (2007)
3. Chen, Y., Yang, B., Dong, J.: Evolving flexible neural networks using ant programming and pso algorithm. In: Yin, F.L., Wang, J., Guo, C. (eds.) *Advances in Neural Networks* ISSN 2004, *Lecture Notes in Computer Science*, vol. 3173, pp. 211–216. Springer Berlin Heidelberg (2004)
4. Chen, Y., Yang, B., Dong, J.: Nonlinear System Modelling Via Optimal Design of Neural Trees. *International Journal of Neural Systems* 14(02), 125–137 (2004)
5. Chen, Y., Yang, B., Dong, J., Abraham, A.: Time-series forecasting using flexible neural tree model. *Inf. Sci.* 174(3-4), 219–235 (Aug 2005)
6. Kennedy, J., Eberhart, R.: Particle Swarm Optimization. In: *Neural Networks, 1995. Proceedings., IEEE International Conference on.* vol. 4, pp. 1942–1948 (1995)
7. Keogh, E.J., Pazzani, M.J.: Uci machine learning repository - pseudo periodic synthetic time series data set (1999), <http://archive.ics.uci.edu/ml/datasets/Pseudo+Periodic+Synthetic+Time+Series>
8. Koza, J.R., Bennett III, F.H., Stiffelman, O.: Genetic Programming as a Darwinian Invention Machine. In: Poli, R., Nordin, P., Langdon, W., Fogarty, T. (eds.) *Genetic Programming, Lecture Notes in Computer Science*, vol. 1598, pp. 93–108. Springer Berlin Heidelberg (1999)
9. Mitchell, M.: *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, USA (1998)
10. Price, K., Storn, R.M., Lampinen, J.A.: *Differential Evolution: A Practical Approach to Global Optimization (Natural Computing Series)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005)
11. Salustowicz, R., Schmidhuber, J.: Probabilistic incremental program evolution. *Evolutionary Computation* 5(2), 123–141 (1997)
12. Shi, Y., Eberhart, R.: A modified particle swarm optimizer. In: *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on.* pp. 69–73 (May 1998)
13. Shi, Y., Eberhart, R.: Parameter Selection in Particle Swarm Optimization. In: Porto, V., Saravanan, N., Waagen, D., Eiben, A. (eds.) *Evolutionary Programming VII, Lecture Notes in Computer Science*, vol. 1447, pp. 591–600. Springer Berlin Heidelberg (1998)
14. Storn, R., Price, K.: Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization* 11(4), 341–359 (1997)

Human Activity Recognition: Classifier Performance Evaluation on Multiple Datasets

Pavel Dohnálek

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

Abstract. Human activity recognition is an active research area with new datasets and new methods of solving the problem emerging every year. In this paper, we focus on evaluating the performance of both classic and less commonly known classifiers with application to three distinct human activity recognition datasets freely available in the UCI Machine Learning Repository. During the research, we placed considerable limitations on how to approach the problem. We decided to test the classifiers on raw, unprocessed data received directly from the sensors and attempt to classify it in every single time-point, thus ignoring potentially beneficial properties of the provided time-series. This approach is beneficial as it alleviates the problem of classifiers having to be fast enough to process data coming from the sensors in real-time. The results show that even under these heavy restrictions, it is possible to achieve classification accuracy of up to 98.16%. Implicitly, the results also suggest which of the three sensor configurations is the most suitable for this particular setting of the human activity recognition problem.

Keywords: Human activity recognition, pattern matching, classification, comparison

1 Introduction

Human activity recognition (HAR) is one of the more recent research topics that recently gained on popularity and focus of both academic and commercial researchers. Since human activity monitoring has a broad range of applications like homecare systems, prisoner monitoring, physical therapy and rehabilitation, public security, military uses and others, the motivation to create a reliable human activity recognition system is considerable. More often than not, time complexity of a classification algorithm is a limiting issue. While developing new methods and optimizing the existing ones is certainly the correct way of approaching the problem, it may not always be the most feasible as in some signal processing problems it may prove difficult to classify the data as fast as it is acquired. This is usually necessary because the signal is then processed as a time-series and as such needs to be as continuous as possible. In this paper,

we attempt to alleviate this problem and focus on the possibility of recognizing human activities from single time points (measurements) given only raw, unprocessed data from the sensors. Should this approach prove possible and even reliable, the classifiers would no longer place limits on the operating frequencies of the sensors, simply discarding measurements when the classifier is not ready to process them without the risk of losing vital information.

2 Datasets

The datasets used in this paper have been acquired from the UCI Machine Learning Repository where they are freely available to download along with their description and related research papers. The goal of this paper is to learn if human activities can be reliably recognized from single time points rather than time series of a signal generated by the sensor systems without any prior pre-processing. These conditions were met by three HAR datasets available in the UCI repository: Physical Activity Monitoring for Aging People 2, OPPORTUNITY Activity Recognition Data Set and Localization Data for Person Activity Data Set, shortened in the following text as the PAMAP2, Opportunity and Localization datasets, respectively.

3 Orthogonal Matching Pursuit

Well described in [1], OMP is an iterative sparse approximation algorithm that reduces data into a given number of sparse coefficients and thus can be considered a dimensionality reduction method. Given an overcomplete dictionary of observations (measurements), for each observation to be classified, OMP picks a number of the best fitting observations from the dictionary and uses them to compute the sparse coefficients. Those are then checked against the dictionary itself for similarity and classified.

The dictionary can be represented as an $m \times n$ real-valued matrix A , where m is the length of an observation and n is the number of observations in the dictionary (training observations). The iterative nature of the algorithm allows for sparse coefficient number to be chosen in advance. It stands to reason to limit the number of sparse coefficients s such that $s \leq m$, although the number can be truly limited only by the number of training observations, n .

Originally, the classifier proposed in [2] requires $n = t \times c$, where t is the number of training observations for a given class and c is the number of classes. The classification algorithm requires the training set to contain the same number of training observations for each class. It is also necessary to keep the observations of a given class grouped together. Therefore, the training matrix has the form of $A = [\mathbf{a}_{11}, \mathbf{a}_{21}, \dots, \mathbf{a}_{t1}, \mathbf{a}_{12}, \dots, \mathbf{a}_{tc}]$, where $\mathbf{a}_{ij}, i = 1..t, j = 1..c$ is the i th training observation of class j and length m .

The proposed modification changes the meaning of t and the resulting number of observations in the training set. Here,

$$n = \|\mathbf{t}\|_1 = \sum_{x=1}^c \mathbf{t}_x, \quad (1)$$

where \mathbf{t} is a c -dimensional vector consisting of the numbers of training observations for a given class. By this, the limitation imposed on the number of training observations in the original classification approach is lifted, yielding a training matrix in the form of $A = [\mathbf{a}_{11}, \mathbf{a}_{21}, \dots, \mathbf{a}_{t_11}, \mathbf{a}_{12}, \dots, \mathbf{a}_{t_c c}]$, where \mathbf{a}_{ij} , $i = 1..t_x$, $x = 1..c$, $j = 1..c$ is the i th training observation of class j and length m .

The sparse coefficients are obtained by finding the sparse solution to the equation

$$\mathbf{y} = A\mathbf{s}, \quad (2)$$

where $\mathbf{y} \in \mathbb{R}^m$ is the query vector, $A \in \mathbb{R}^{m \times n}$ is the training matrix and $\mathbf{s} \in \mathbb{R}^n$ is the sparse coefficient vector. The stopping criterion in the implementation is reaching the sparse coefficient vector with the desired number of non-zero values.

3.1 Classification

To classify the query signal vector, a strategy of computing the residual value from the difference between the query vector and its sparse representation converted into the vector space of the training matrix vectors is employed. This is performed for each class resulting in c residuals. The classification is then based on the minimum residual. Formally, the classification problem can be stated as follows:

$$\arg \min r_k(\mathbf{y}) = \|\mathbf{y} - A\mathbf{s}_k\|_2. \quad (3)$$

Here, \mathbf{s}_k is an n -dimensional vector with non-zero elements located only on the indices corresponding to the k th class in the training matrix, hence the need for the training samples of a given class to be grouped together in the matrix. The algorithm could be described by the following steps:

- Set the iteration variable i to 1
- Replace all sparse coefficients with indices not corresponding to class i with zeros
- Multiply the training matrix with the modified vector \mathbf{s}
- Compute the ℓ^2 -norm of the resulting vector
- Increase i by 1 and repeat for all classes
- Output the class whose ℓ^2 -norm is the lowest

Computing the residuals is generally not computationally expensive and can be performed in real time, depending on the size of the training matrix. Only very large training matrices can slow the process down significantly.

Aside from the modification above, we also used a tensor adaptation that we developed and described in [3]. This adaptation restructures the data into a

Table 1. Averaged recognition accuracy for each of the classifier per dataset size (Localization dataset)

Classifier	Training set size					TPM (ms)
	10%	20%	30%	40%	50%	
10-NN	70.36	71.89	72.72	73.26	73.55	6.699
5-NN	70.28	71.53	72.22	72.64	72.93	6.059
3-NN	69.58	70.64	71.23	71.58	71.90	5.784
RF	67.01	68.57	69.35	69.92	70.30	8.276×10^{-3}
1-NN	67.23	68.08	68.54	68.82	69.03	5.901
CART	64.28	66.02	66.75	67.35	67.76	8.883×10^{-4}
OMP _{mod}	64.17	65.10	65.62	65.95	66.21	9.597
OMP	63.17	64.22	64.82	65.17	65.43	6.495
OMP _{ten}	55.61	56.40	56.75	56.97	57.27	8.427
QDA	46.50	46.39	46.37	46.45	46.45	5.547×10^{-4}
LDA	37.12	36.96	36.95	36.90	36.99	5.125×10^{-4}
NCC	28.92	28.55	28.76	28.72	28.77	2.829×10^{-4}

tensor and implements elements of the ensemble paradigm into its classification procedure. Experiments confirming its ability to generally perform better than the original OMP version in a HAR problem was included in [3]. Other classifiers used in this paper were k-Nearest Neighbors (k-NN), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Nearest Centroid Classifier (NCC), Random Forest (RF) and Classification and Regression Tree (CART).

4 Experiments

The following section describes the workflow used to evaluate the performance of the classifiers as well as the evaluation process and its results.

4.1 Experimental settings

The execution of some of the algorithms can be customized through execution parameters which, for these experiments, were set according to the best empirical speed/accuracy ratio. While the goal was to make the results as comparable as possible despite using datasets with greatly dissimilar properties, the Localization dataset contained so few parameters that the evaluation required a different setting. The common settings in terms of customizable algorithms for all dataset were as follows: as all algorithms were implemented in the latest version of MATLAB, the default MATLAB settings were used for CART. k-NN was experimented upon with four settings of the k parameter: 1, 3, 5 and 10. As k-NN is a classic and very well known classifier, we decided to include these four most successful settings as opposed to just picking one that performed the best. LDA, QDA and NCC were not customizable.

Table 2. Averaged recognition accuracy for each of the classifier per dataset size (PAMAP2 dataset)

Classifier	Training set size					TPM (ms)
	10%	20%	30%	40%	50%	
1-NN	95.49	96.78	97.41	97.91	98.16	13.737
OMP _{mod}	95.80	96.89	97.46	97.89	98.12	200.040
RF	96.28	97.24	97.61	97.86	98.05	1.320
3-NN	94.94	96.21	96.86	97.33	97.65	13.748
5-NN	94.55	95.84	96.49	96.93	97.25	13.212
10-NN	93.98	95.34	95.98	96.44	96.75	13.631
OMP _{ten}	93.15	94.47	95.12	95.53	95.91	123.903
OMP	91.15	92.75	93.66	94.29	94.67	80.547
CART	91.52	93.00	93.66	94.16	94.58	7.228×10^{-4}
QDA	69.34	68.87	68.75	68.70	68.80	2.332×10^{-3}
LDA	64.32	64.40	64.52	64.46	64.47	2.407×10^{-3}
NCC	54.76	54.55	54.57	54.60	54.61	1.378×10^{-3}

4.2 Results

The averaged classification accuracies given as percentual success rates are shown in Tables 1 through 3 where the classifiers are sorted according to their best results (regardless of the training set size) in the descending order. For all methods except for LDA, QDA and NCC, larger training set resulted in more or less improved accuracy. LDA, QDA and NCC, on the other hand, suggest that in HAR problems, the training set size does not matter to them. This does not come as a surprise as no matter the training set size, data distribution (for LDA and QDA) and the centroid location (for NCC) remain the same.

For the Localization dataset (Table 1), the results are generally discouragingly low, topping at 73.55% for k-NN with $k = 10$. This suggests that while the dataset can provide some accuracy, the data collection setting using only four localization tags is not suitable for the task as specified in this paper.

The PAMAP2 dataset (Table 2) has shown promising results before, and the results were confirmed in these experiments as well. With the maximum accuracy at 98.16% by 1-NN followed very closely with the difference of only 0.04% by our proposed method, the dataset appears to present a highly reasonable compromise between accuracy and price of the data capturing solution.

Finally, the Opportunity dataset (Table 3) shows promising and satisfactory results while its setting seems to be the most suitable for our modification of the OMP based classifier. The classifier surpassed all other classifiers with the lead of 0.78%, 1.45% and 7.75% against RF, 1-NN and CART, respectively. The approximate times required to classify one measurement for each classifier are listed in each table in the TPM column. The values are in milliseconds.

Table 3. Averaged recognition accuracy for each of the classifier per dataset size (Opportunity dataset)

Classifier	Training set size					TPM (ms)
	10%	20%	30%	40%	50%	
OMP _{mod}	89.98	92.51	93.95	94.96	95.71	604.396
RF	91.53	92.85	93.73	94.41	94.93	2.235
1-NN	88.10	90.66	92.27	93.44	94.29	36.047
3-NN	86.83	89.04	90.49	91.59	92.51	37.488
5-NN	86.13	88.10	89.43	90.44	91.25	36.592
10-NN	84.73	86.37	87.44	88.30	89.03	41.847
CART	83.63	85.45	86.65	87.37	87.96	9.775×10^{-4}
OMP _{ten}	78.33	81.45	83.36	84.68	85.58	82.948
OMP	78.60	81.36	83.05	84.28	85.13	55.745
QDA	78.36	78.45	78.40	78.37	78.41	7.826
LDA	68.45	68.48	68.51	68.46	68.42	7.891
NCC	66.11	66.11	66.12	66.10	66.09	5.301

5 Conclusion

This paper evaluated several classification techniques and presented their success rates in human activity recognition from single time points and without any prior preprocessing. Given the sensor technology that was used to create the PAMAP2 dataset, it was shown that basic activities (walking, lying and sitting) can be recognized reliably and very precisely. Two more datasets with different sensor settings were compared with lesser accuracy results. While for the Opportunity dataset the methods performed only slightly worse, the Localization dataset has proven unsuitable for the defined task. The approach to the HAR problem proposed in this paper suggests that using the right sensor setting, problems with the follow-up classifier's time complexity can be greatly alleviated. The proposed modification of the OMP based classifier has shown excellent results for two of the datasets, but the modification's increased performance comes at the price of significantly higher time complexity.

References

1. Thomas Blumensath and Mike E. Davies. On the difference between orthogonal matching pursuit and orthogonal least squares, 2007.
2. John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:210–227, 2009.
3. Pavel Dohnálek, Petr Gajdoš, and Tomáš Peterek. Tensor modification of orthogonal matching pursuit based classifier in human activity recognition. In *Nostradamus 2013: Prediction, Modeling and Analysis of Complex Systems*, volume 210 of *Advances in Intelligent Systems and Computing*, pages 497–505. Springer International Publishing, 2013.

Distance and Energy-based Descriptors for Object Detection

Radovan Fusek and Eduard Sojka

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

Abstract. Many feature-based object detectors have shown that the use of gradient image information can be a very efficient way to describe the appearance of objects. Especially, the gradient sizes, directions and histograms are commonly used. Nevertheless, many works proved that the features based on the edge information require large training sets due to their high dimensionality. This was the motivation to find alternatively ways how to encode image information with the least possible redundant information. In this paper, we propose novelty methods that are used to effectively encode the image information for object detection.

1 Introduction

In the feature-based detectors based on the sliding window technique, the selection of relevant features that are able to reliably describe the objects of interest is a key point. In this area, the Histograms of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Haar features are considered as the state-of-the-art methods.

The Haar-like features which are similar to Haar basis function were proposed by Papageorgiou and Poggio [9]. In [11], the Haar-like features was popularized by Viola and Jones. The authors developed the framework consists of the image representation called the integral image combined with the rectangular Haar-like features, and AdaBoost algorithm [4]. The extension of the Haar feature set has been presented by Lienhart et al.[7]. In [12], the authors proposed the front-view car and bus detection method based on the AdaBoost and Haar-like features.

In [2], the authors proposed the method in that the Histograms of Oriented Gradients (HOG) are used to encode the appearance of the object. Many methods and applications that are based were presented in recent years. The PHOG descriptors that use the pyramid image representation were presented in [1]. The classical HOG-based detector was used for detecting upper bodies for automated upper body pose estimation in [3]. In this work, the authors used the upper bodies HOG-based detector to obtain a weak model of person and reduce the search space for body parts. Similarly, the HOG-based head-shoulder detection algorithm was used in the method that is focused on estimating the number of people in surveillance scenes [6].

Ojala et al.[8] proposed the Local Binary Patterns (LBP) in that the local image structures (e.g. lines, edges, spots, and flat areas) can be efficiently encoded by comparing every pixel with its neighboring pixels. The important properties of LBP are resistance to lighting changes and low computational complexity. Duo to their properties, LBP were used in many recognition tasks, especially for facial image analysis. In [5], LBP were used for solving the face detection problem in low-resolution images. Multi-block Local Binary Patterns (MB-LBP) for face detection were proposed in [13].

The mentioned features are based on the fact that the appearance of the objects is described by the image edge information (intensity differences). In general, the features based on the edge information (e.g. length, magnitude, orientation, localization) require large training sets due to their high dimensionality. Additionally, in the cases that the edges are very thin, it is obvious that the edges information is difficult to hit (by the samples). Therefore, the proposed methods are based on the distance function (or energy function) in that the information about its changes is not so important.

2 Hierarchical Energy-transfer Features

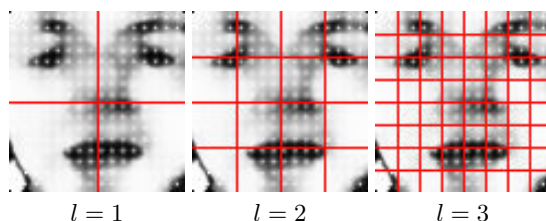


Fig. 1. The different hierarchical levels of the cells.

In this paper, we proposed the interesting method for object description. The method is based on the fact that the appearance of the objects can be described using the function of temperature distribution. The distribution of temperature is investigated after the temperature transfer. The vector of features that is composed of this distribution is then used as an input for the SVM classifier. The temperature distribution reflects the presence of objects and their parts and the appearance of object of interest can be described by the distribution (with a relatively small amount of descriptors), which is the main idea of the method we propose.

Once the temperature transfer inside the image is obtained, the function of temperature distribution inside the image is investigated (the example of temperature distribution is shown in Fig. 1). For this purpose, the image is iteratively divided into the finer spatial cells; i.e. we recursively divide the image into the

cells of varying size (Fig. 1). In general, the image at hierarchical level l has 4^l cells. Inside each cell, the distribution is investigated. We use the mean cell temperatures as the values in the feature vector. For the additional information and for the precise description of the temperature distribution, we use the histogram of the temperature distribution that is also determined inside the cells.

3 An Improvement of Energy-Transfer Features Using DCT

In this work, we proposed an improvement of previously mentioned method; instead of the mean temperature inside each cell, we use the Discrete Cosine Transform (DCT) coefficients to encode the function of temperature distribution. After DCT, the DC coefficients represent the average temperatures of the regions and the AC coefficients represent temperature changes across the regions. It is obvious that the information obtained after DCT is more descriptive and can be used to effectively encode the function of temperature distribution. Finally, the PCA (Principal Component Analysis) is used to create the feature vector with the relatively small dimensionality.

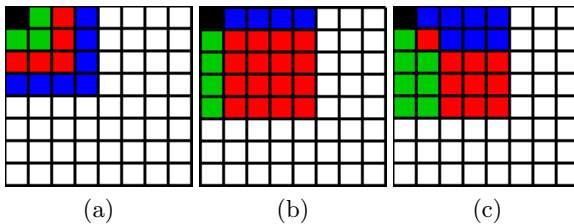


Fig. 2. The three different options of AC patterns. The areas are depicted by three different colors in that the averages of coefficients are calculated.

We also experimented with the different sizes of the blocks and cells. We observed that the best results were obtained using 16×16 blocks and 8×8 cells; inside the cells the DCT coefficients are computed and composed to the final feature vector. In the case of cells with 8×8 pixels (similarly in JPEG compression), each cell consists of 1 DC coefficient and 15 AC coefficients after DCT. The coefficients that are located in the upper left corner contain the most of information (low frequencies). On the other hand, the bottom right coefficients represent higher frequencies that can be discarded. Therefore, instead of the encoding the whole set of the coefficients, we encode the upper left coefficients only.

To encode the upper left coefficients, we create three patterns of these coefficients (Fig. 2) for our experiments (similarly in [10]). In these patterns, the three AC regions are created. These regions represent the different frequencies

and different information can be encoded using different patterns. To reduce the quantity of the coefficients, the mean of coefficients is calculated inside these regions. It means that each 8×8 cell is represented by four values; 1 DC coefficient + 3 averages of AC coefficients. The final feature vector is composed from these values.

4 Distance-based Descriptors

In this paper, we proposed detection method that is based on the distance function. The proposed method is based on the fact that the properties of the image (especially the properties of the objects) can effectively be described by the distance function. In essence, we divide the image inside the sliding window into the blocks and cells (similarly as in HOG), but instead of the histograms of gradients we encode the values of distance function inside each cell. The goal is to obtain more meaningful values for recognition than the classical state-of-the-art method. This leads to the reasonable dimensionality of the feature vector; furthermore, the values of distance function can be easily obtained by sampling. The feature vector that contains the distance function values is then used as an input for the SVM classifier. In this particular case, we use the geodesic distance, nevertheless, it is important to note that any appropriately distance function can be used in the proposed detection framework (e.g. resistance distance, diffusion distance). The visualizations of geodesic distance values inside the cells of different sizes are shown in Fig. 3. Based on the cell sizes, information with various levels of details is obtained. To compress the information contained in the distance function in to a reasonable number of values, we use four values from each cell only.

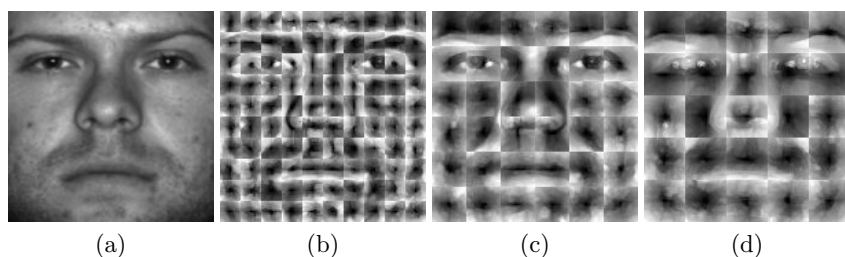


Fig. 3. The visualization of the distance function values inside each cell. The example of face image (a). The sizes of cell 15×15 (b), 25×25 (c), 35×35 (d).

5 Conclusion

In the paper, we presented efficient ways how the image information can be encoded into the feature vector, which can be used in sliding-window-based tech-

niques of recognition. The methods are based on the fact that the appearance of the objects can be described using the function of temperature distribution. We also proposed the improvement of encoding the temperature (energy) distribution that is useful for object description. The improvement is based on the fact that the important image information can be described using DCT coefficients. Finally, we proposed the detection method that is based on the geodesic distance function. We leave experiments with various types of distances for future work.

Acknowledgments

This work was supported by the SGS in VSB Technical University of Ostrava, Czech Republic, under the grant No. SP2014/170.

Author's Publications

The parts of results were presented in:

- Fusek, R., Sojka, E.: Distance-based Descriptors and their Application in the Task of Object Detection. 36th German Conference on Pattern Recognition (GCPR 2014), In Lecture Notes in Computer Science, to appear 2014
- Fusek, R., Sojka, E.: Gradient-DCT (G-DCT) Descriptors. In Proceedings of 4th International Conference on Image Processing Theory, Tools and Applications, IPTA 2014, to appear 2014
- Fusek, R., Sojka, E., Mozdren, K., Surkala, M.: An Improvement of Energy-Transfer Features Using DCT for Face Detection. In Elmoataz, Abderrahim and Lezoray, Olivier and Nouboud, Fathallah and Mammass, Driss (eds.) Image and Signal Processing, Lecture Notes in Computer Science, vol. 8509, pp. 511-519. Springer International Publishing (2014)
- Fusek, R., Sojka, E., Mozdren, K., Surkala, M.: Energy Based Descriptors and their Application for Car Detection. In VISAPP 2014 - Proceedings of the International Conference on Computer Vision Theory and Applications, to appear 2014
- Fusek, R., Sojka, E., Mozdren, K., Surkala, M.: Hierarchical Energy-transfer Features. In: Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods, pp. 695-702. SciTePress, (2014)

Author's Awards

"An Improvement of Energy-Transfer Features Using DCT for Face Detection" has been selected among the best papers of the ICISP 2014 conference; the paper was invited to submit for a special issue of the Springer journal "Signal, Image and Video Processing" on "Advances in Low-Level Image representations for processing and analysis".

References

1. Anna Bosch, Andrew Zisserman, and Xavier Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on Image and video retrieval, CIVR '07*, pages 401–408, New York, NY, USA, 2007. ACM.
2. 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.
3. V. Ferrari, M. Marin-Jimenez, and A. Zisserman. Progressive search space reduction for human pose estimation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.
4. Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proceedings of the Second European Conference on Computational Learning Theory, EuroCOLT '95*, pages 23–37, London, UK, UK, 1995. Springer-Verlag.
5. A. Hadid, M. Pietikainen, and T. Ahonen. A discriminative feature space for detecting and recognizing faces. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II-797–II-804 Vol.2, 2004.
6. Min Li, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In *ICPR*, pages 1–4. IEEE, 2008.
7. R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I-900–I-903 vol.1, 2002.
8. Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, January 1996.
9. Constantine Papageorgiou and Tomaso Poggio. A trainable system for object detection. *Int. J. Comput. Vision*, 38(1):15–33, June 2000.
10. Tienwei Tsai, Yo-Ping Huang, and Te-Wei Chiang. Image retrieval based on dominant texture features. In *Industrial Electronics, 2006 IEEE International Symposium on*, volume 1, pages 441–446, July 2006.
11. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I-511 – I-518 vol.1, 2001.
12. Chunpeng Wu, Lijuan Duan, Jun Miao, Faming Fang, and Xuebin Wang. Detection of front-view vehicle with occlusions using adaboost. In *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*, pages 1–4, 2009.
13. Lun Zhang, Rufeng Chu, Shiming Xiang, Shengcai Liao, and Stan Z. Li. Face detection based on multi-block lbp representation. In *Proceedings of the 2007 international conference on Advances in Biometrics, ICB'07*, pages 11–18, Berlin, Heidelberg, 2007. Springer-Verlag.

Eye and Iris Detection in Images with Active Near-infrared Illumination

Michael Holuša and Radovan Fusek

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

Abstract. In this paper, we present a method for eye and iris detection in grayscale images with active near-infrared illumination. The eye detector is based on well-known supervised machine learning algorithms. The extracted eye images are passed to the iris detection algorithm. This algorithm uses the Hough circle transform with some image preprocessing methods utilizing the geodesic distance and Canny edge detector. Performance of our method is evaluated on own dataset containing images with and without IR illumination. The results are presented.

1 Introduction

Detection and recognition of certain parts of human face are important in many areas, such as medicine, personal identification, cryptography or safety. The eye is one of the most convenient human part used for these purposes. For example, in a safety area, according to the position of iris, it can be estimated where the person is watching or can be measured frequency of eye blinking. This information may be used in vehicles to recognize fatigue of the driver or if the driver is focusing on driving. Iris localization is currently solved by many different approaches. In [1], the AdaBoost method in combination with a color segmentation is used. The authors in [2] use local brightness change and contrast to detect iris and pupil. In [3], the authors use the modified Hough Circle Transform (HCT) [4] to detect iris.

In our work, we focus on detection from grayscale images with active near-infrared illumination. This illumination allows us to detect eyes and irises in dark conditions, which would be impossible with normal images without the illumination. Our algorithm consists of two major parts - eye detection in images containing a human face, and iris detection in the extracted eye images. For the eye detection, we use a supervised machine learning algorithm. In the iris detection, we use brightness information for image preprocessing. Iris detection itself is realized by HCT. After that, the detected circle is postprocessed to check if our algorithm really detects iris or it is a false detection.

The paper is organized as follows. In Section 2, we briefly introduce the eye detection algorithm, Section 3 contains our iris detection method based on HCT. The experimental results are presented in Section 4, Section 5 is a conclusion.

2 Eye Detection

For eye localization, we use the sliding window detection framework. In general, in the area of object detection, the methods based on the sliding window became very popular in recent years. The idea behind this approach is based on the fact that a rectangular window scans the input image at different sizes and scales. After the scanning process, many sub-images (windows) are obtained. Inside each sub-image, the appropriate image features are computed and composed to the final feature vector. This vector is then used as an input for the trainable classifier (e.g. support vector machines, AdaBoost, neural networks). After the classification process, each sub-image is marked as background or object of interest. Goal is to effectively encode image information inside the sliding window. In this area, three types of features are considered as the state-of-the-art over the last few years; Local Binary Patterns (LBP) [5, 6], Haar features [7, 8], and Histogram of Oriented Gradient (HOG) [9].

We experimented with each type of these features and we decided to use the HOG based eye detector combined with the Support Vector Machine (SVM) classifier due to the faster training phase (faster than LBP and Haar features) and satisfactory detection results. Similarly, the eye detection based on the HOG features were presented in [10, 11].

3 Iris Detection

In the situation when the eye detector returns the sub-image containing the eye, our goal is to detect iris in this sub-image. An example of the detected eye is shown in Fig.1(a). Our first step is a transformation of the image to the representation in which we would be able to differ the brightness of iris area and the rest of image around the eye. As the transformation we use the geodesic distance [12, 13]. We find a point inside the pupil and compute the geodesic distance to the all other image points. Let x_p be a pupil point. Then x_p is the point with minimum brightness lying on one of three shifted horizontal lines in the center of image, because color of the human pupil is darker than colors of other eye parts. Since the eye detector returns sub-images where the eye is centered, one (or more) of these lines should pass through the pupil.

After this transformation, the area of pupil has the lowest brightness. Due to the closeness of iris and pupil, the iris brightness is also low and other areas have higher brightness (Fig.1(b)). This property leads to the image thresholding, which removes image information that are not important for us, such as areas with higher brightness located around the iris. Since human eyes have different colors and images may be obtained in different lighting conditions, it is difficult to define the threshold globally. Therefore, we define this threshold T as the first quantile of cumulative image histogram. In other words, we assume that the iris brightness is located in 20 percents of the histogram. This value is estimated experimentally on the basis of ratio of the iris area and the rest of eye image. The result of threshold function is in Fig.1(c).

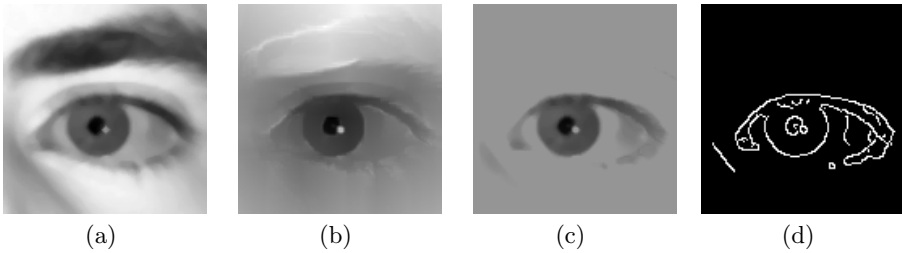


Fig. 1. The process of iris detection. (a) Eye image detected by the method described in Sec.2, (b) Geodesic distance applied on the input image, (c) Thresholded image, (d) Binary image after Canny edge detector.

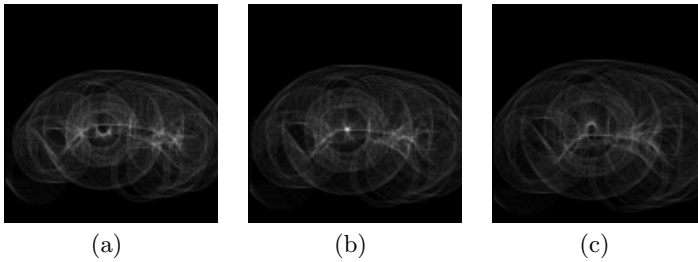


Fig. 2. The accumulator visualization of Hough circle transform with various radiuses. The white spot in (b) signalizes a center of circle with the current radius.

After that, we apply the Canny edge detector on the thresholded image to find candidates on the iris border (see Fig.1(d)). Since iris has a circular shape, we use Hough Circle Transform (HCT) to find the border. The goal of HCT is to find three parameters to define a circle (center point x_{center}, y_{center} and radius r). It leads to the complexity $O(N^3)$ in the number of non-zero image pixels in the binary image after the edge detection. Therefore, the previously applied threshold function reduces the iris center candidate points for sake of efficiency.

The visualization of HCT accumulators with three different radiuses are shown in Fig.2. We choose the point with the highest value from all the computed accumulator arrays as the center of possible iris circle. To avoid false detections, we set two thresholds that help us to decide whether HCT correctly found the iris circle. The first threshold is an accumulator value in the possible iris center (the white spot in Fig.2(b)). This value represents an amount of circles that intersect in this point, i.e. how many points create the circle of current radius in the image. If the number is low, there can be only few points that create the circle (for example semicircle) and the detection is possibly false. The threshold is set to πr , which means that at least half of the circle has to be visible. The second threshold is a mean of brightness in the inner area of the detected circle. Since the iris brightness should be lower than the brightness of surrounding area, we compute the mean of brightness in the circle and compare it with the

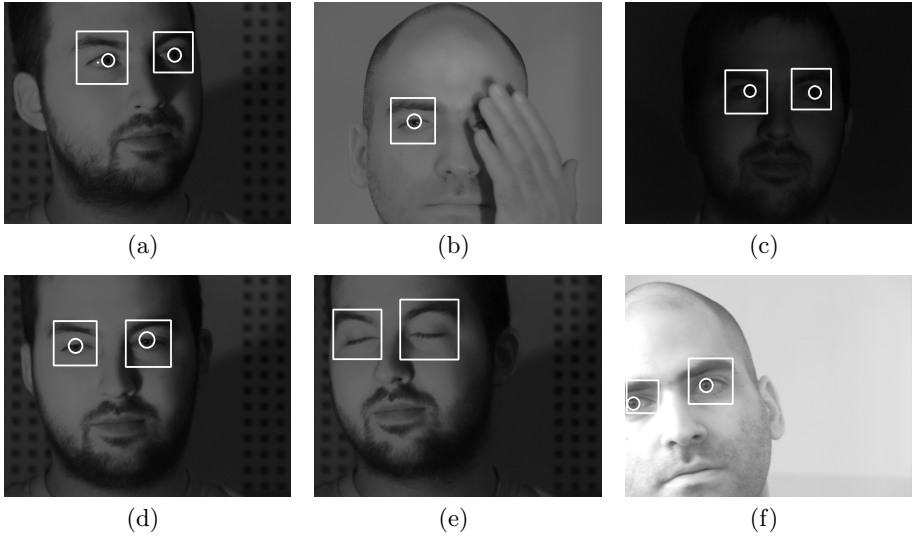


Fig. 3. Examples of correct eye and iris detections. (a) Detection of irises located on the side of sclera. (b) Detection of only one eye while the second eye is covered. (c) Detection of eyes and irises in very dark environment. (d) Detection of irises in partially closed eyes. (e) Detection of closed eyes. (f) Detection of eye located near the image border.

threshold T that was described before. If the threshold is higher than T , the iris is detected incorrectly. If the circle detected by HCT passes these two tests, it is labeled as the iris.

4 Experiments

In this section, the experimental results of our method are validated. For the training phase of the eye detector (based on HOG and SVM), we used approximately 12000 positive images and 25000 negative images. For the evaluation process, we prepared image sequences of different people in two different lighting conditions (light conditions, dark conditions with IR illumination). The images are recorded from the camera placed in front of a person and contain faces with various head and eyes positions. The goal is detect open and also closed eyes in every image. If the eye is open and at least half of its iris is visible, the iris should be detected as well. The method has been tested on 2236 images that contain 4243 detectable eyes and 3753 irises (the rest of eyes are closed or irises are not visible enough). The examples of successful detections are shown in Fig.3. The examples of images where iris is not detected are in Fig.4, the reasons of mis-detections are mentioned below this image. The effectiveness of the presented algorithm is in Table 1. We measured 3 values - Sensitivity, Precision and F1

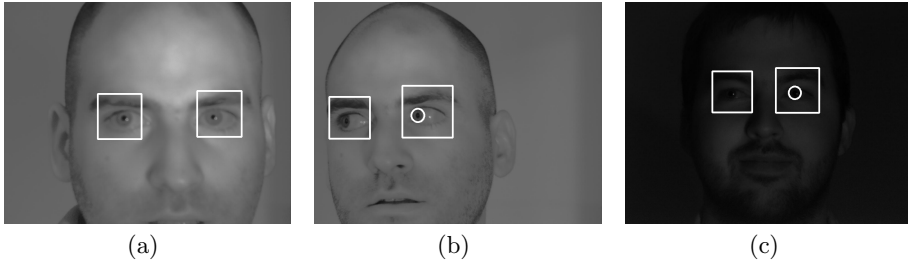


Fig. 4. Examples of missed iris detection. (a) Face is not correctly focused, the image is blurred and the gradient between iris and sclera is not correctly detected. (b) Due to the eye position, iris shape is distorted and has not circular shape, so that the Hough transform do not detect the circle. (c) Very dark environment and direction of IR illumination caused that the colors of iris and sclera are similar and the gradient is not detected.

score. The values are computed as

$$S = \frac{TP}{TP + FN} \quad P = \frac{TP}{TP + FP} \quad F1 = \frac{2TP}{2TP + FP + FN}, \quad (1)$$

where TP (True Positive) is a number of correctly labeled objects, FP (False Positive) is a number of incorrectly labeled objects, and FN (False Negative) is a number of objects that were not labeled. The average time of detection is 28 ms per eyes localization and 53 ms per iris detection (Intel Core i5 2.30GHz).

	Sensitivity	Precision	F1 Score
Eye Detection	90.42 %	99.81 %	94.89 %
Iris Detection	85.36 %	99.44 %	91.87 %

Table 1. The effectiveness of our algorithm.

5 Conclusion

We have introduced the algorithm for iris detection in grayscale images with infra-red illumination. Our method is based on Histogram of Oriented Gradients and Support Vector Machine for eye detection in images. Iris detection is based on Hough circle transform that is applied on image preprocessed by geodesic distance transform and Canny edge detector. The algorithm has been tested on more than two thousand face images and performs promising results. It detects open and closed eyes with F1 score close to 95 %, irises are detected with almost 92 % of F1 score. In the case of precision, we achieved false positive object detection less than 1 %. The running time is approximately 130 ms for one face image, which makes the method usable in real-time applications with only a short

delay. The main future work is an improvement of iris detection in problematic images whose examples are shown in Fig. 4. The another work may be also focused on speed improvement of our method.

References

1. Jeong, D.S., Hwang, J.W., Kang, B.J., Park, K.R., Won, C.S., Park, D.K., Kim, J.: A new iris segmentation method for non-ideal iris images. *Image Vision Comput.* **28** (2010) 254–260
2. Cho, D.H., Park, K.R., Rhee, D.W.: Real-time iris localization for iris recognition in cellular phone. *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, International Conference on Self-Assembling Wireless Networks, International Workshop on* **0** (2005) 254–259
3. Dobe, M., Martinek, J., Skoupil, D., Dobeov, Z., Pospil, J.: Human eye localization using the modified hough transform. *Optik - International Journal for Light and Electron Optics* **117** (2006) 468 – 473
4. Kimme, C., Ballard, D., Sklansky, J.: Finding circles by an array of accumulators. *Commun. ACM* **18** (1975) 120–122
5. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* **29** (1996) 51–59
6. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In Pajdla, T., Matas, J., eds.: *Computer Vision - ECCV 2004*. Volume 3021 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg (2004) 469–481
7. Papageorgiou, C., Poggio, T.: A trainable system for object detection. *Int. J. Comput. Vision* **38** (2000) 15–33
8. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Volume 1. (2001) I–511 – I–518 vol.1
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Volume 1. (2005) 886 –893 vol. 1
10. Monzo, D., Albiol, A., Sastre, J., Albiol, A.: Precise eye localization using hog descriptors. *Machine Vision and Applications* **22** (2011) 471–480
11. Chen, S., Liu, C.: Precise eye detection using discriminating hog features. In Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W., eds.: *Computer Analysis of Images and Patterns*. Volume 6854 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg (2011) 443–450
12. Toivanen, P.J.: New geodesic distance transforms for gray-scale images. *Pattern Recogn. Lett.* **17** (1996) 437–450
13. Criminisi, A., Sharp, T., Rother, C., P'erez, P.: Geodesic image and video editing. *ACM Trans. Graph.* **29** (2010) 134:1–134:15

LIDAR Function Transformation: A New Algorithm for Registration of LIDAR Data via SOMA Optimization

Branislav Hóly

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

Abstract. In this paper, a new algorithm for registration of LIDAR points is introduced. The algorithm consists of two steps – construction of LIDAR functions for a fast correspondence search, and a minimization of the error function which describes error of alignment between two range images. The minimization is done via SOMA (Self-Organizing Migrating Algorithm) for a robust search of a global minimum. The algorithm has been tested and is able to align two LIDAR images with high accuracy and with short computation time even if an initial guess is unknown.

Keywords: registration, scan matching, LIDAR, SOMA

1 Introduction

Registration of LIDAR data is an important problem in many applications like mapping, robot navigation, or autonomous car driving. These robots and cars are equipped with a laser scanner to scan a surrounding area. The laser usually emits one or more rays per degree. Reflections of the rays from the closest obstacles provide information about the surrounding area.

The main goal of the registration is to find a rigid transformation between two range images to align them with the best fitness. Since the LIDAR images are scanned from different poses, they only partially overlap. Also, correspondences between points are unknown which makes this problem difficult. The correct rigid transformation can be used to refine the new pose of the robot for further usage (e.g. in algorithms like SLAM).

This paper is organized as follows. The next section contains brief descriptions of previous registration methods and Sec. 3 describes the new algorithm. Results of experiments are shown in Sec. 4 and the conclusion is presented in the final section.

2 Previous Work

Iterative Closest Points (ICP) [1, 10] is mostly used in registration of two point sets. The algorithm minimizes an error function iteratively with SVD until con-

verged. The error function is defined as (b_i and m_i are corresponding points):

$$E(T) = \sum_{i=0}^N \|T \cdot b_i - m_i\|^2 \quad (1)$$

There are other variants of ICP like TrICP [3], LM-ICP [6], Generalized-ICP [8] with a probabilistic framework, or Adaptive Iterative Closest Key-point [5] for RGB-D data.

Normal Distribution Transform (NDT) [2] is a different kind of a registration method for 2D space. In this registration method, space is divided into cells and for every occupied cell, a normal distribution of probability is computed from points inside the cell. The error function (minimized by Newton's algorithm) is computed as a negative of sum of corresponding point (from the first scan) and normal distribution of probability (from the second scan). A variant of NDT for 3D space is described in [7].

A similar description of space like ours is used in PCM (Polar Coordinates Matching) [4]. It uses sets of points in polar coordinate system for a fast correspondence search.

3 New Algorithm

As mentioned above, the ICP-like algorithms are minimizing the function of sum of square distances between two corresponding points (Eq. (1)). The correspondence is mostly found by searching for the closest point in second set. Even an optimized search with use of k -d trees has asymptotic complexity $\mathcal{O}(\log(n))$, which makes computation of the error function expensive with asymptotic complexity $\mathcal{O}(n \log(n))$. For this reason, different computation of an error function with asymptotic complexity $\mathcal{O}(n)$ is used. The computation is done only over overlapping parts for a better resistance to partially overlapping range images.

Unlike standard algorithms, the new algorithm does not use points to describe the surrounding area because they are not suitable for a fast correspondence search. Better description of the surrounding area is given by the LIDAR function defined as:

$$area = \int_0^{2\pi} lf(\theta) d\theta \quad (2)$$

where the LIDAR function $lf(\theta)$ returns only one distance for θ (Fig. 1(a)).

This representation of the area is not suitable for any set of points in general but is very useful for LIDAR points. Sampling of the LIDAR function results in a set of points in polar coordinate system which is used in PCM as mentioned earlier.

Error of alignment of two range images (described by lf_a and lf_b) can be computed by the following error function:

$$E(T) = \int_0^{2\pi} (lf_a(\theta) - \hat{lf}_b(\theta))^2 d\theta \quad (3)$$

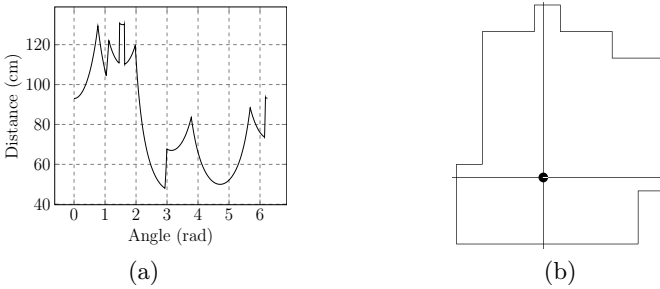


Fig. 1. LIDAR function (a) of area (b).

where $\hat{lf}(\theta)$ is a transformed LIDAR function according to T . The correct rigid transformation T is found by minimizing the error function E with SOMA algorithm as shown in Sec. 3.2. The LIDAR function for discrete space is defined as:

$$area = \sum_{i=0}^N lf(i) \tag{4}$$

where N is a resolution of the LIDAR function and $lf(i)$ is a sample of the LIDAR function in angle $\theta = \frac{2\pi i}{N}$. The error function is then defined as follows:

$$E(T) = \frac{N}{p^2} \sum_{i=0}^N c(lf_a(i), \hat{lf}_b(i)) \left(lf_a(i) - \hat{lf}_b(i) \right)^2 \tag{5}$$

where N has to be same for both LIDAR functions, function c is defined as:

$$c(x, y) = \begin{cases} 1 & \text{if } x > 0 \wedge y > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

and p is a number of samples where both functions are defined which is:

$$p = \sum_{i=0}^N c(lf_a(i), \hat{lf}_b(i)) \tag{7}$$

Coefficient $\frac{N}{p}$ ensures higher error for transformations with a low number of correspondences.

3.1 Construction of LIDAR Functions

Construction of LIDAR functions and transformed LIDAR functions is similar and requires to have points A_j in the same order as they were scanned by a laser rangefinder. For every point A_j , the value of the LIDAR function will be computed as:

$$lf(i_j) = \sqrt{x_j^2 + y_j^2} \tag{8}$$

where index i_j is defined as:

$$i_j = \text{integer} \left(\frac{\text{atan2}(y_j, x_j)}{2\pi} N \right) \quad (9)$$

where function atan2 returns angle in $\langle 0; 2\pi \rangle$ and N is a resolution of the LIDAR function. The value of $\text{lf}(i)$ is computed as an average of consecutive points with the same i . Non-consecutive points with the same i can appear in the transformed LIDAR function and will be stored at index i until the final step.

For consecutive points with different i (within some distance threshold), new values will be inserted (computed with linear interpolation) to have occupied every cell in the LIDAR function.

For transformed LIDAR functions, some points can not be visible from pose described by T . The point A_j with neighbouring points A_{j-1} and A_{j+1} is invisible if $i_{j-1} > i_j > i_{j+1}$.

In the final step, only one, minimal, and visible value for every i is chosen. For those cells which have only one invisible value, value 0 is used.

3.2 Optimization with SOMA

BIO inspired algorithm SOMA (Self-Organizing Migrating Algorithm) [9] is able to find minima of a function, mostly a global minimum, even if an initial guess is wrong or unknown.

To compute the best rigid transformation to align two range images, minimization of the error function (5) is needed. The input of the error function is a transformation matrix T defined by three parameters φ , t_x , t_y . The error function can be rewritten for SOMA algorithm as:

$$E(\varphi, t_x, t_y) = \frac{N}{p^2} \sum_{i=0}^N c(\text{lf}_a(i), \hat{\text{lf}}_b(i)) \left(\text{lf}_a(i) - \hat{\text{lf}}_b(i) \right)^2 \quad (10)$$

where $\hat{\text{lf}}$ is a transformed LIDAR function according to:

$$T = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) & t_x \\ \sin(\varphi) & \cos(\varphi) & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

Setting of SOMA parameters depends on required speed and final fitness. The only exception for our case is parameter $D = 3$. Experiments showed that optimal parameters for error function (10) for fast convergence to the global minimum are following: Mass $\in \langle 2; 3 \rangle$, Step $\in \langle 0.11; 0.31 \rangle$, PRT $\in \langle 0.5; 0.8 \rangle$, NP $\in \langle 10; 25 \rangle$, Migrations $\in \langle 100; 1000 \rangle$ and AcceptedError $\in \langle 0.0001; 0.00001 \rangle$.

Initial individuals can be randomly distributed with a uniform distribution of probability in the case if a guess of pose (for example from an odometer) is unknown. With odometry data, we can distribute the individuals with the

normal distribution of probability with mean vector $\boldsymbol{\mu} = (\varphi_o, t_{x_o}, t_{y_o})$ where $\varphi_o, t_{x_o}, t_{y_o}$ are extracted from the odometer and covariance matrix $\boldsymbol{\Sigma}$ describes motion error of the robot or the vehicle.

4 Results

The algorithm has been tested with different laser scans of the corridor at our university. Each scan contains about 1,000 points. The same scans have been tested with ICP and NDT for comparison (with and without initial guess).

Input LIDAR images are shown in Fig. 2 and initial guess is $\varphi_o = 0, t_{x_o} = 1m$ and $t_{y_o} = 0$ because the robot moved 1 meter towards.

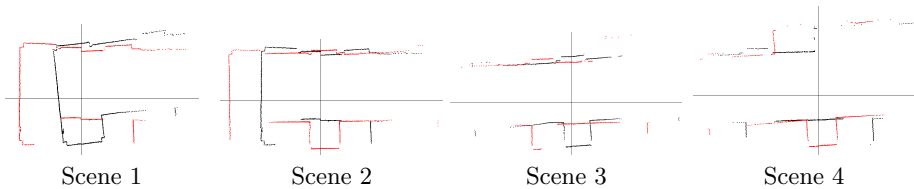


Fig. 2. Input scenes without the initial guess.

All tested algorithms were able to align these frames correctly if the guess of new pose was known. Different situation happened when the initial guess was unknown. In this case, traditional algorithms stuck in local minima and were not able to align the same scenes correctly. ICP was able to align only scene 1 and scene 2, but not as good as with known guess and NDT failed in all scenes (Fig. 3). Only the new LFT algorithm was able to align all scenes correctly. Results are shown in Fig. 4.

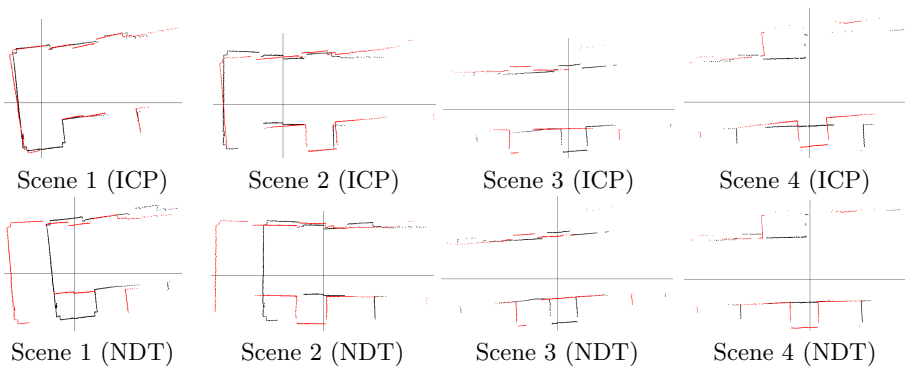


Fig. 3. Results of ICP and NDT with unknown guess.

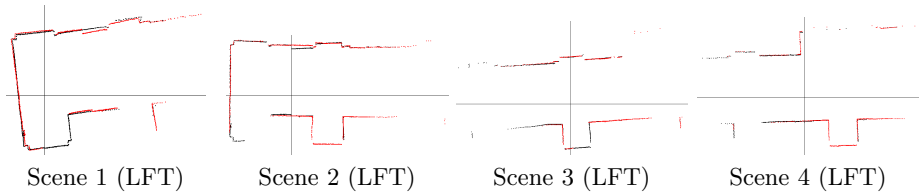


Fig. 4. Results of LFT with unknown guess.

5 Conclusion and Future Work

A new algorithm for registration of LIDAR data is presented. The algorithm consists of two parts, creating LIDAR functions and minimization which is done by BIO inspired algorithm SOMA. Thanks to this optimization algorithm, the new algorithm is able to find a global minimum of the error function to align two scans with the best fitness. Experiments proved that this is possible even if the initial guess is not presented. Future work will focus on the part of creating the LIDAR function which takes most of the computation time. Variant for 3D data is considered as well.

References

1. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), 239–256 (1992)
2. Biber, P., Straßer, W.: The normal distributions transform: A new approach to laser scan matching. In: *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on.* vol. 3, pp. 2743–2748. IEEE (2003)
3. Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P.: The trimmed iterative closest point algorithm. In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on.* vol. 3, pp. 545–548. IEEE (2002)
4. Diosi, A., Kleeman, L.: Laser scan matching in polar coordinates with application to slam. In: *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on.* pp. 3317–3322. IEEE (2005)
5. Ekekrantz, J., Pronobis, A., Folkesson, J., Jensfelt, P.: Adaptive iterative closest keypoint. In: *Mobile Robots (ECMR), 2013 European Conference on.* pp. 80–87. IEEE (2013)
6. Fitzgibbon, A.W.: Robust registration of 2d and 3d point sets. *Image and Vision Computing* 21(13), 1145–1153 (2003)
7. Magnusson, M., Lilienthal, A., Duckett, T.: Scan registration for autonomous mining vehicles using 3d-ndt. *Journal of Field Robotics* 24(10), 803–827 (2007)
8. Segal, A., Haehnel, D., Thrun, S.: Generalized-icp. In: *Robotics: Science and Systems.* vol. 2 (2009)
9. Zelinka, I.: Soma self-organizing migrating algorithm. In: *New optimization techniques in engineering,* pp. 167–217. Springer (2004)
10. Zhang, Z.: On local matching of free-form curves. In: *BMVC92,* pp. 347–356. Springer (1992)

Fast Parallel Artificial Bee Colony on CUDA

Jan Janoušek

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

Abstract. Artificial bee colony is a metaheuristic optimization algorithm based on the behaviour of honey bee swarm. These bees work largely independently of other bees, making the algorithm suitable for parallel implementation. Within this paper, we introduce the algorithm itself and its subsequent parallelization utilizing the CUDA platform. The runtime speedup is demonstrated on several commonly used test functions for optimization.

Keywords: parallel algorithm, artificial bee colony, CUDA

1 Introduction

Artificial bee colony is a relatively young algorithm, first published by Dervis Karaboga in 2005 [1]. The algorithm is based on the behaviour of a bee colony when searching for food. There are three types of bees: employed, onlooker and scout bees. A food source represents a possible solution to the problem. An employed bee visits a source in their memory and finds a source in its surroundings/neighbourhood. It then determines the amount of food in this source, and if it is better than the current source, it initiates a waggle dance whose length corresponds to the quality of the source. The quality of the source is represented by a fitness function. Onlookers then choose a source based on the length of a waggle dance and subsequently find a new source in the surroundings of these sources and assess its quality. If its quality is better than the quality of the current food source, onlookers share this information with other bees. If a bee cannot find a better food source for a certain amount of time, it becomes a scout which searches for a new food source regardless of existing sources. The existence of scouts ensures that ABC cannot remain stuck in a local extreme.

During the existence of ABC, several variants and spin-offs of this algorithm have been developed. These include for instance Interactive Artificial Bee Colony (IABC) [2], Upgraded Artificial Bee Colony (UABC) [3], Discrete Artificial Bee Colony (DABC) [4] or Fast Mutation Artificial Bee Colony (FMABC) [5].

In the past years, several parallel implementations were published. One example is an implementation from 2011 [6] which was built as a multi-thread Java application. This improved the runtime on a problem with dimension 10 by a factor of 3. Another implementation from 2013 [7] which was based on Open MPI exhibited a runtime improvement for a problem dimension of 100 by a factor of

2. A paper published in 2013 [8] presented a parallel implementation on CUDA which focused on the optimization of very small dimensions of problems. This implementation improved the runtime of the computation, which took place only over shared memory, by up to 56,93x.

This article will target an implementation for Nvidia's Compute Unified Device Architecture (CUDA) [9] focusing on the optimization of problems with large dimensions.

2 Artificial Bee Colony

The ABC algorithm consists of the following steps:

1. **Initialization:** SN random vectors with dimension D are generated through the formula:

$$x_{ij} = LB + rand(0, 1) * (UB - LB) \quad (1)$$

For $i = 1, 2, \dots, SN$ and $j = 1, 2, \dots, D$. Where LB is the lower bound of a solution and UB is the upper bound of a solution. The value of the fitness function is then computed for each solution.

2. **Employed:** Each employed finds a new solution in it's neighbourhood based on the formula:

$$x_{ij} = x_{ij} + \phi_{ij} * (x_{ij} - x_{kj}) \quad (2)$$

Where x_{ij} is parameter j of solution i . ϕ_{ij} is a random number between -1 and 1 , and x_{kj} is a parameter j of randomly selected solution k for which it holds that $k \neq i$. The concrete value of j is chosen randomly. The value of the parameter of a new solution must be modified if necessary so that it lies in the set of admissible values. If the value of the fitness function of a new solution is better than the original's, the old solution is replaced by the new one. Otherwise the number of unsuccessful attempts is increased by one.

3. **Onlookers:** Each onlooker chooses one existing solution based on its quality via the formula:

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (3)$$

Where p_i is the probability of selecting solution i . fit is the value of the fitness function. The algorithm then follows in the same way as in 2).

4. **Scouts:** In this step, solutions which could not be improved even after a certain number of iterations of the algorithm are selected and replaced by new, randomly selected solutions from the formula used during initialization.
5. **Termination:** If the conditions for the termination of the algorithm are met, then it terminates, otherwise it continues from step 2).

All of the steps listed above depend on their predecessors. They thus cannot be parallelized as a whole. The work of individual bees over food sources is however independent. It is thus possible to carry out computations for each bee in a separate block. The computation of the most demanding part of the algorithm, similarly as in other biology-inspired methods, is the computation of the fitness function. To improve scalability and optimization for higher dimensions, the computation of the fitness function for each food source was additionally separated into several blocks. Each thread can thus perform the computation for exactly one parameter of the solution, allowing the full use of the computational power of the GPU.

When implementing the algorithm, it became apparent that the quality of the solution in each iteration of the algorithm depends on the quality of the selection of a modified parameter used for the creation of a new solution. The parallel implementation allows the selection of several potential parameters in a single step, and to compute the value of the fitness function for each of these in case of its modification. The best value out of the computed one is then selected. This procedure on one hand requires a large number of computations of the fitness function, but allows the algorithm to converge faster to the global extreme and reduces the number of iterations of the whole algorithm.

A disadvantage of this solution is the need to synchronize the computations between blocks and the associated exchange of data through global memory, which is very slow. Synchronization between blocks may additionally only be carried out by terminating the kernel. These negative impacts were minimized thanks to the above-mentioned selection of a larger number of parameters for modifying the solution. Since it is necessary to compute the fitness function for each change of the parameter over the same food source, it is not necessary to continuously reload the source from the global memory. The same method is also used to solve the problem of synchronizing blocks. Since one initialization of the kernel leads to a large number of computations, which additionally allow a reduction of the number of iterations of the whole algorithm, synchronization does not need to be carried out as frequently and its impact on the speed of the computation is minimized.

For the purposes of parallelization of the algorithm, it was necessary to change how scouts select sources. In the original version of the algorithm, each scout selects a single solution based on its quality and this is associated with another random solution. This selection method is problematic for the needs of parallelization, since each onlooker must modify another solution, otherwise a collision would occur. Due to this, the selection method was adjusted as follows: each existing source is associated with another source based on its quality. From the results of comparing the serial and parallel implementation in the following section, it is clear that this change did not reduce the optimization capabilities of the algorithm.

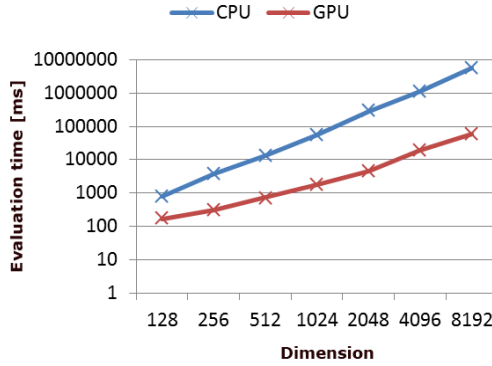


Fig. 1. Comparison of the evaluation time for the Rastrigin function on a CPU and GPU

3 Comparison of the parallel vs serial implementation

The performance of the parallel implementation on the CUDA platform was compared to a serial implementation for the CPU written in C. Experiments on the CPU were carried out on an Intel(R) Core(TM) i5-2500K processor (3,3GHz, 6MB, socket 1155), with 8GB RAM and Windows 7 x64. Experiments on CUDA were carried out on an Nvidia Geforce GTX 660 Ti.

The tests were carried out on four continuous test functions.

The first function is Rastrigin. This is a multimodal function on the interval [5.12, -5.12]. The minimal value of the function is 0, located at [0, 0, ..., 0].

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i) + 10) \tag{4}$$

The second test function is Griewank. This is again a multimodal function on the interval [-600,600]. The minimal value of the function is 0, located at [0, 0, ..., 0].

$$f(x_1, x_2, \dots, x_n) = 1 + \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) \tag{5}$$

The third test function is Schwefel. Similarly to the previous two, this is another multimodal function on the interval [-500, 500]. The minimal value of the function is -n*418.9829, located at [420.9687, 420.9687, ..., 420.9687].

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n -x_i \sin \sqrt{|x_i|} \tag{6}$$

The goal of this experiment was to achieve a lower value than the minimum of the function plus 1. The graphs depict the time required to reach this value.

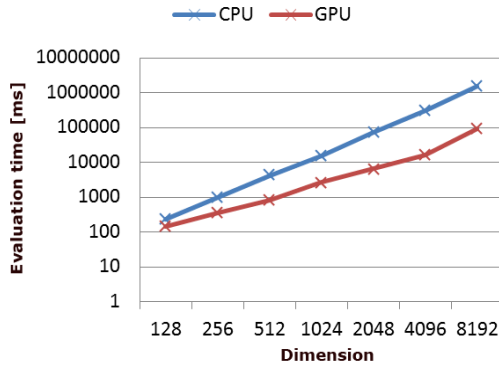


Fig. 2. Comparison of the evaluation time for the Griewank function on a CPU and GPU

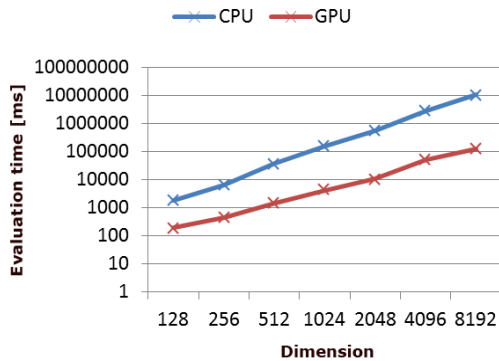


Fig. 3. SrovComparison of the evaluation time for the Schwefel function on a CPU and GPU

When testing the implementation on CUDA, the algorithm was initialized on blocks with 256 threads. All dimensions had the same configuration. The number of blocks was computed automatically based on the current dimension of the problem. The configuration of the serial version of the algorithm was always identical with the configuration of the parallel version.

4 Conclusion

The article demonstrated the possibility of increasing the performance of the ABC algorithm through the use of parallel implementation in the CUDA platform. Test functions demonstrated a speedup of the runtime ranging from a factor of 1.58 (Griewank function at a dimension of 128) to 96.23 (Rastrigin

Table 1. Speedup of the parallel implementation compared to the serial one

Dimension	Test function		
	Rastrigin	Griewank	Schwefel
128	4,69x	1,58x	9,56x
256	12,09x	2,81x	14,28x
512	18,27x	5,21x	27,32x
1024	31,59x	5,96x	38,41x
2048	61,58x	10,72x	53,64x
4093	59,21x	19,19x	53,05x
8192	96,23x	18,17x	86,79x

function at a dimension of 8192). The results indicate that parallel processing provides the greatest contribution at larger dimensions.

Further research will be focused on the improvement of the CUDA implementation to enable algorithm distribution on many GPUs.

References

1. D. Karaboga, "An idea based on honey bee swarm for numerical optimization," tech. rep., Computer Engineering Department, Erciyes University, Turkey, 2005.
2. P.-W. TSai, J.-S. Pan, B.-Y. Liao, and S.-C. Chu, "Enhanced artificial bee colony optimization," *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 12, pp. 5081–5092, 2009.
3. I. Brajevic and M. Tuba, "An upgraded artificial bee colony (abc) algorithm for constrained optimization problems," *Journal of Intelligent Manufacturing*, vol. 24, no. 4, pp. 729–740, 2013.
4. Q.-K. Pan, M. F. Tasgetiren, P. Suganthan, and T. Chua, "A discrete artificial bee colony algorithm for the lot-streaming flow shop scheduling problem," *Information Sciences*, vol. 181, no. 12, pp. 2455 – 2468, 2011.
5. F. Kang, J. Li, H. Li, Z. Ma, and Q. Xu, "An improved artificial bee colony algorithm," in *Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on*, pp. 1–4, 2010.
6. M. Subotic, M. Tuba, and N. Stanarevic, "Different approaches in parallelization of the artificial bee colony algorithm," *International Journal of mathematical models and methods in applied sciences*, vol. 5, no. 4, pp. 755–762, 2011.
7. Y. S. Hong, Z. Z. Ji, and C. L. Liu, "Research of parallel artificial bee colony algorithm based on mpi," *Applied Mechanics and Materials*, vol. 380, pp. 1430–1433, 2013.
8. G.-H. Luo, S.-K. Huang, Y.-S. Chang, and S.-M. Yuan, "A parallel bees algorithm implementation on {GPU}," *Journal of Systems Architecture*, no. 0, pp. –, 2013.
9. D. Kirk and W. Hwu, *Programming Massively Parallel Processors: A Hands-on Approach*. Applications of GPU Computing Series, Elsevier Science, 2010.

Finding Nearest Neighbors in Graph Layout Algorithm

Tomáš Jeřowicz

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

Abstract. There exists many different strategies, how to define the layout of a graph. One of the general and the most used is force-based method. It is very useful to be able to compute layout of a graph as fast as possible. The graph layout computation is in general very time consuming. The requirements are usually both speed and quality. This paper describes a new method to define k-nearest neighbors (KNN) in parallel. KNN was used to simplify the well known Fruchterman-Reingold (FR) graph layout algorithm. The paper also includes performance and quality comparison with the original FR algorithm.

Keywords: graph drawing, KNN, space-filling curve, Fruchterman-Reingold, fast graph layout

1 Introduction

There exist several layout algorithms [11][6][7][5][4] that can be used to create an acceptable graph visualization. Usually, more layout algorithms can be applied on the same graph to provide variant for the readers. The layout should ideally help the reader to better understand the information contained in the graph. Several aesthetic techniques or metrics like the edge crossing, line bends, symmetry, minimum angle or orthogonality are presented in [8]. The [16] shows which of the aesthetic criteria have the greatest influence for human understanding. Note that for some graphs, even if one layout has no edge crossings, the different layout with more crossings may be considered as a better one (showed in [11]). The more aesthetical criteria are required, the more computation time is usually needed.

2 Related work

Some of the graph layout algorithms represent a class of so called force-directed layouts, e.g. Fruchterman-Reingold [6] or Kamada-Kawai [11]. These algorithms iteratively change the positions of vertices/nodes to reduce defined energy function. These layouts are generally considered to be aesthetic. The problem related

to the usage of the force-directed algorithms is that in general they are computationally expensive. The paper [10] describes the speedup of the Fruchterman-Reingold by computing the most expensive part on GPU.

Our research focused on the method of finding the nearest neighbors for every vertex to achieve more effective computation. Thus the solution of the nearest neighbors problem described for example in [12][3] became the primary goal of our research. One of fast methods for nearest neighbors search are based on the space-filling curves (SFC). The space-filling curves used in the graph layout can be found in [15], where a new approach dealing with dense graphs were presented. Moreover, searching a new layout algorithm leads to the consequent challenge that consists of comparison techniques. When a new layout algorithm is found, the natural question is how to compare it with others. The [8] shows several quality measurements for graph layouts. Next, in [16], there were studied quality measurements from the point of view of human reading.

3 Background

3.1 Graph Layout and Fruchterman-Reingold

One of the well known algorithms for the graph layout is the Fruchterman-Reingold, which belongs to the family of force-directed graph layout algorithms. Vertices that are connected by an edge attract each other. It also defines an ideal distance for each vertices. The vertices should be drawn near each other, but not too close. To layout a graph, the vertices are replaced by steel ring and each edge with a spring mechanical system [6]. The algorithm shown above basically consists of three independent parts. The computation of repulsive forces is the bottleneck of the whole algorithm; the remaining parts works with edges and with all vertices again:

1. Calculate repulsive forces ($\Theta(N^2)$).
2. Calculate attractive forces ($\Theta(E)$).
3. Iterate all vertices ($\Theta(N)$).

Space-filling curves was used to solve the nearest neighbors problem. Space-filling curves were used in order to get the nearest neighbors, because they can be computed very fast in parallel way. After reordering of the vertices according to the Hilbert curve [2][13][1], one can obtain ordered list of vertices, where nearest neighbor of each vertex is defined as neighbor in an ordered list.

4 Our Approach

The main goal of our research consists in the speed up Fruchterman-Reingold algorithm graph layout algorithm by avoiding relatively unnecessary operations. Moreover, the algorithm should be well prepared for the further parallelization on graphics processor units (GPUs).

Computation of the repulsive forces is the most time-consuming part of Fruchterman-Reingold algorithm $\Theta(N^2)$. A single repulsive force between two vertices decreasing when the distance between the vertices is increasing. So the forces that modify the positions of vertices can be neglected for all such vertex pairs where the vertices are far each other. Omitting such forces and focusing on the forces between close vertices only significantly reduce the computation time and the time complexity at all. There exist many algorithms dealing with searching all nearest neighbors (NN) or k-nearest neighbors. We refer to [3][12][14] for information on NN problem.

Fruchterman-Reingold is an iterative algorithm and it requires several thousand iterations to complete. It brought us to the idea to compute a different space-filling curve at each iteration. It is based on the fact that the layout algorithm induces the dynamic changes. Thus the set of k-nearest neighbors of a selected vertex in a single iteration is changed in the consequent iteration. Finding the neighbors in each iteration is achieved by displacing the bounding box for the space-filling curve. In every single iteration the bounding box is randomly shifted. Note that the bounding box must be large enough to ensure that all vertices stay inside during the shift operation. The ordered list of vertices that lie on the SFC is splitted into several parts in the next step of the proposed method. The number of parts is one of the internal parameters and can be changed during the computation. Finally, all repulsive forces are computed between vertex pairs in every part of the SFC. The proposed algorithm runs in iterations again and every single iteration consists of the following steps:

1. Random shift of the SFC bounding box.
2. Compute SFC index for every single $v \in V$.
3. Reorder V according to the SFC index (quick sort algorithm was used).
4. Split the set of reordered vertices in V into SFC parts.
5. Within every SFC split parts, the repulsive forces are computed for the set of vertices.
6. Calculate attractive forces and update the positions of all vertices V .

5 Performance and Quality Experiments

Figure 1 contains computation times in milliseconds for different numbers of vertices and different SFC parts. Random graphs were generated to test the behavior of the proposed algorithm. Updating the edges is done in $\Theta(N)$ and it is the same as in case of the original FR algorithm. Thus the required computation times were excluded from the experiments. Measurements include all additional operations that the proposed algorithm requires (space-filling curve computation, quick sort and shifting the bounding box). All measurements were performed 10x and then averaged. For example when 10k vertices is splitted into 500 parts of equal size, the proposed algorithm is 195x faster than the original Fruchterman-Reingold algorithm.

In [16] there was examined the most effective aesthetics criteria which were regarded from the point of view of human reading. The results showed, that

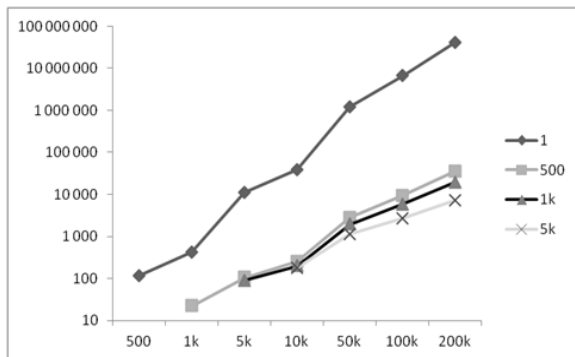


Fig. 1. Computation time for different numbers of vertices and different number of SFC parts

there is a strong emphasis minimize edge crossing and the other criteria appear to have just a little effect. With respect to the mentioned criteria, the comparison targeted to the edge crossing were done for the original and proposed algorithm. Four different type of graphs were tested: Tree graph(|V|=400), Grid graph(|V|=256), Random graph(|V|=128, |E|=512) and Complete graph(|V|=32). The repulsive forces in every graph were computed with particular SFC splits into a given number of parts as it was described above. Table 1 shows the number of edge crosses after 1k, 2k, 4k, 8k and 16k iterations.

Iterations	1k	2k	4k	8k	16k
Grid-PARTS1	12,237	1,283	111	22	0
Grid-PARTS2	12,223	1,490	123	56	0
Grid-PARTS4	12,200	1,651	181	80	0
Grid-PARTS8	12,247	1,832	263	93	21
Grid-PARTS16	12,247	1,959	363	109	48

Table 1. Number of edge crosses after 1k, 2k, 4k, 8k and 16k iterations

Six experiments were created for every graph with different number of SFC parts. In image 2, the first row shows the progress of the computation when the original FR algorithm was used. Next rows illustrate the results of the proposed algorithm, where the set of vertices was divided into 2, 4, 8 and 16 parts to compute the repulsive forces separately. Every part is represented by different color. The first column shows the starting positions of all vertices in the graphs. The next columns show the layout positions after 1 000, 2 000, 4 000, 8 000 and 16 000 iterations. It is evident, that the proposed algorithm requires more iterations to provide the same result as the original algorithm on every single graph. The same conclusion can be made for random graphs. The number of edge

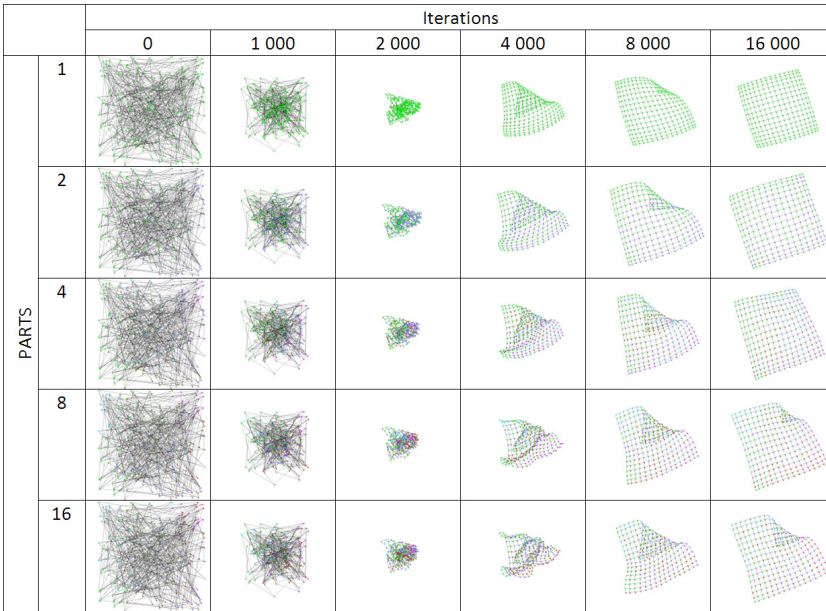


Fig. 2. Grid graph $|V|=256$

crossings in the complete graph was approximately the same in all iterations, as could be expected.

6 Conclusion and Future Work

This paper described the speedup of Fruchterman-Reingold force-based graph layout algorithm, while achieving almost the same results in case of usage of the original algorithm. The quality of the algorithm is now configurable by selecting the number of SFC split parts. The experiments showed that even if the number of SFC split parts is large, the resulting graph is very similar to the outputs of the original algorithm. The proposed algorithm is faster than the current Fruchterman-Reingold and it is also naturally very suitable for further parallelization. Our future work will be focused on speeding up of the algorithm by parallel platform CUDA. Finally, larger graphs could be visualized in acceptable time.

Projects and Publications

The extended version of this paper was published in [9]. 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. SP2014/110.

References

1. Greg Breinholt and Christoph Schierz. Algorithm 781: Generating hilbert's space-filling curve by recursion. *ACM Trans. Math. Softw.*, 24(2):184–189, 1998.
2. Arthur R. Butz. Convergence with hilbert's space filling curve. *J. Comput. Syst. Sci.*, 3(2):128–146, 1969.
3. Michael Connor and Piyush Kumar. Fast construction of k-nearest neighbor graphs for point clouds. *Visualization and Computer Graphics, IEEE Transactions on*, 16(4):599–608, 2010.
4. Emilio Corchado and Bruno Baruaque. Wevos-visom: An ensemble summarization algorithm for enhanced data visualization. *Neurocomputing*, 75(1):171–184, 2012.
5. Emilio Corchado, Michal Wozniak, Ajith Abraham, André Carlos Ponce Leon Ferreira de Carvalho, and Václav Snášel. Recent trends in intelligent data analysis. *Neurocomputing*, 126:1–2, 2014.
6. Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.
7. Apeksha Godiyal, Jared Hoberock, Michael Garland, and John C Hart. Rapid multipole graph drawing on the gpu. In *Graph drawing*, pages 90–101. Springer, 2009.
8. Knut Hartmann, Timo Götzelmann, Kamran Ali, and Thomas Strothotte. Metrics for functional and aesthetic label layouts. In *Smart Graphics*, pages 115–126. Springer, 2005.
9. Tomáš Ježowicz, Petr Gajdoš, Eliška Ochodková, and Václav Snášel. A new iterative approach for finding nearest neighbors using space-filling curves for fast graphs visualization. In *International Joint Conference SOCO 14-CISIS 14-ICEUTE 14*, pages 11–20. Springer, 2014.
10. Tomas Jezowicz, Milos Kudelka, Jan Platos, and Václav Snášel. Visualization of large graphs using gpu computing. In *INCoS*, pages 662–667, 2013.
11. Tomihisa Kamada and Satoru Kawai. An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1):7–15, 1989.
12. Jan Kybic and Ivan Vnucko. Approximate all nearest neighbor search for high dimensional entropy estimation for image registration. *Signal Processing*, 92(5):1302–1316, 2012.
13. Warren M. Lam and Jerome M. Shapiro. A class of fast algorithms for the peano-hilbert space-filling curve. In *ICIP (1)*, pages 638–641, 1994.
14. Swanwa Liao, Mario A Lopez, and Scott T Leutenegger. High dimensional similarity search with space filling curves. In *Data Engineering, 2001. Proceedings. 17th International Conference on*, pages 615–622. IEEE, 2001.
15. Chris Muelder and Kwan-Liu Ma. Rapid graph layout using space filling curves. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1301–1308, 2008.
16. Helen Purchase. Which aesthetic has the greatest effect on human understanding. In *Graph Drawing*, pages 248–261. Springer, 1997.

Formal Software Process Framework

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 Trends in software development tend to improve a control over processes executed in software companies. This approach is known as business process (or rather software process) modeling and except the control and management it also offers additional advantages as process optimization, re-engineering, analysis, and automatized process execution. Our research covers software process modeling activities enhanced with formal (machine-readable) knowledge bases integration that add the advanced features over typical process models.

1 Introduction

This paper is a part of ongoing formal software processes modeling research. We work on an innovative formal software process framework based on knowledge bases and develop a new modeling method over it. The framework is based on existing work and expands it so it meets the following features:

- Utilization of formal reusable software process domain vocabulary and meta-model; (scenarios based on existing research works)
- Intuitive modeling method based on intuitive and easy to use diagrammatic language capable of modeling all necessary aspects of software process based on machine-readable knowledge base; (extended by our research)
- Utilization of pre-modeled process parts in complex process modeling in organization with machine-readable knowledge bases and UML; (proposed by our research)

The paper is organized as follows: after a brief introduction we shall discuss existing work in this area, this section is followed by details of tools used for modeling of software processes in our framework. Fourth section briefly describes a modeling discipline. The paper is completed with conclusion and future work discussion.

2 State of the Art and Related Work

Business processes represent the core of company behavior [1]. The modeling of processes in organization is motivated mostly by process improvement and re-engineering [2, 3] however models could (and apparently should) be used in advanced

ways like simulation [8, 20], process analysis [17], audit purposes [23], and automated process execution (workflow) [16].

2.1 Business Process Meta-Model and Modeling

Capture and management of a software process is a complex activity that contains big set of different activities and requires certain tools [7, 9]. Current trends in software process modeling are based mostly on semi-formal methods based on MOF (Meta-Object Facility) maintained by OMG (Object Management Group) [3, 19, 25]. All of these methods are more or less compatible and quite well accepted in professional groups – UML, BPMN, and ARIS could be named as most spread in software process area [1, 19, 22].

The approach of using semi-formal methods however has some problems [2, 9] – while being user friendly and intuitive, methods listed above are still semi-formal thus not having precise semantics. Of course there are formal methods that provide unambiguous mathematical models like Petri Nets, VPML, but these are quite hard to use and formal methods integration into standard process capture and modeling are still being sidetracked due to their complexity [4, 22].

2.2 Knowledge Bases Integration

Knowledge based systems are getting more favorite in computer science area [5, 16] and such a trend of ontology engineering could not be ignored by software process modeling domain. Last decades after the dawn of formal knowledge bases research have proven many methods to be efficient to support business processes; some well known are Prolog, CYC, Casl, and OWL [15]. While providing almost the same expression power, Web Ontology Language (OWL) became de-facto standard due to its features and characteristics [24]. There are various researches discussing utilization of OWL in software process modeling and execution like [5, 6, 18]. Yet these are focusing on following scenarios that may still be extended and further improved:

- Software process meta-model ontology creation
- Generating or building OWL from UML or other semi-formal sources
- Querying OWL ontologies of software process meta-model
- Model validation based on meta-model ontology

3 Formal Software Process Framework

Our research [10, 11, 13] however aims for a further utilization of the above – combining the approaches to improve the intuitiveness of the modeling, utilize reusable software process domain vocabulary and software process meta-model (see figure 2), support advanced process tools like analysis, simulation, process adaptation and transformation, exceptional states handling and workflow execution in various process

management applications. The framework proposed in our research combines following features:

- Software process modeling based on meta-model compatible with MOF created in OWL, compliant with SPEM [14]; (see figure 1)
- Knowledge based process modeling methodology based on (iterative) bottom-up and top-down process modeling in intuitive and easy to use semi-formal methods compliant with best practices [12];
- Support of models transformation and formal methods support [4, 10];
- Unified process capture and modeling in an organization with machine-readable knowledge base building blocks [11];
- Model validation with explicit rules defined in base-layer knowledge base [14]
- Additionally easy model refactoring, process simulation, workflow support, and process mining [4, 21].

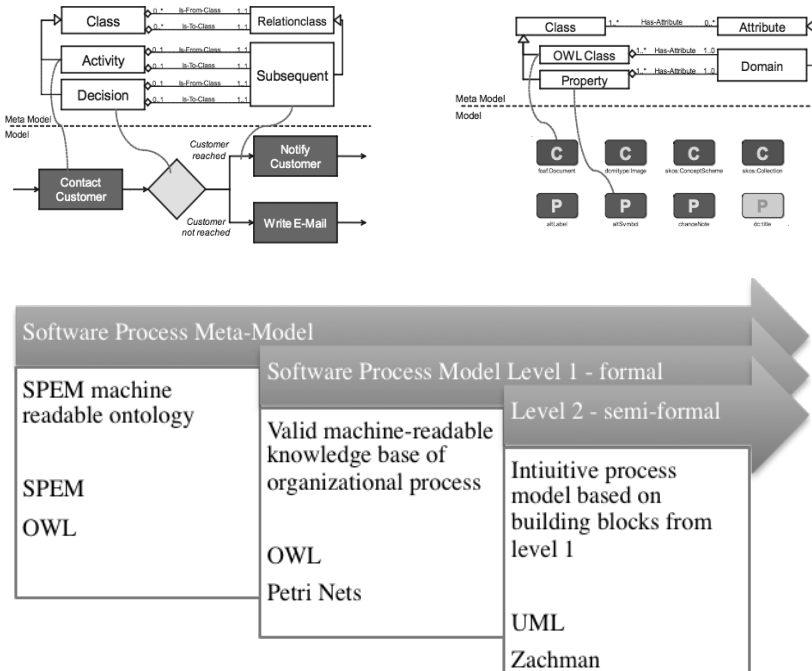


Fig. 1. Software process meta-model to model relation ontology example

With the respect to existing research and state of the art, tools and languages we have selected for our framework are [14]:

- *OWL* as a knowledge layer for meta-model of software process and formal description of static elements in a domain;
- *UML* diagrammatic language to capture the process model in functional, behavioral, informational, and organizational perspectives;

- *Colored Petri Nets* as a formal (and executable) tool behind behavioral perspective;
- *Zachman Framework* as a tool for final enterprise integration of models.

It's obvious that just having tools and languages is not enough to support the software process with domain knowledge. Next section discusses briefly a methodology we propose to use with our formal software process framework [4, 12, 21].

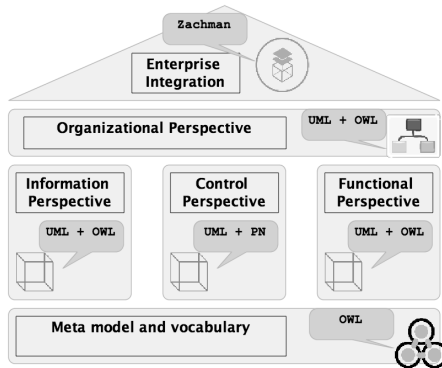


Fig. 2. Formal Software Process Framework reference house

4 Modeling Discipline

The figure 2 shows the formal software process framework reference house. All parts from bottom to top are designed to support process modeling in following way:

1. *Meta-model* and reusable *vocabulary* created in *OWL*; knowledge base includes general concepts, rules, and constraints. This 'basement' of the framework is independent and created by domain specialist well experienced with mathematical methods.
2. *Information*, *functional* and *organizational* perspectives are captured based on terms and rules inherited from meta-model and vocabulary. This is performed by organization management or process auditors with *UML* diagrams. The method generate the toolbox of available terms and rules from knowledge base.
3. *Control* perspective combines the perspectives from second step and could be again performed by management or auditors. Available functions and responsibilities are generated from knowledge defined in steps 1 and 2. The result is captured in *UML* activity diagrams (and exported to *CPN* respectively).
4. *Validation* of created process models based on base knowledge layer's rules
5. Last step is the very *integration* of models created in previous two steps into an organization and brings clear order to all models.

5 Benefits, Conclusion, and Future Work

This paper described formal framework for software processes based on machine-readable knowledge bases. This framework combines UML, OWL, Colored Petri Nets, and Zachman Framework with the respect to existing related works. Our approach focuses on modeling aspects of software processes based on formal knowledge base combined with semi-formal methods, namely UML diagrams. The framework in combination with correct methodology offers sound and intuitive way how to capture software process models and support them with knowledge bases. The benefits of the approach are following:

- separation of semi-formal and formal methods brings the process modeling and understanding to less (mathematically) experienced staff;
- formal definition behind semi-formal models brings additional support of workflow interface based on OWL and Petri Nets models;
- process model validation based on shared formal rules hidden in shared machine-readable knowledge layer
- shared dynamic source of process knowledge brings modeling to next level with building process blocks pre-defined by domain specialist making the process modeling more or less a ‘construction kit’;
- utilization (reusability) of result models in setup of different organization based on behavioral or functional blocks;
- integration of process models into organization with Zachman framework

Future work is aimed to completion of a reference software process model based on software-process meta-model and assembling complete a methodology of ‘knowledge support for software process’ and defining all interfaces to external process support systems.

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 - SP2014/157 ‘Knowledge modeling, simulation and design of processes’ and SP2014/194 ‘Biomedicínské inženýrské systémy X’

1. Košinár, M., A.: Knowledge Modeling of Agile Processes in Healthcare Systems Development in *Lékař a Technika* 04/2013
2. Czopik, J., Košinár, M., Štolfa, J., Štolfa, S.: Formalization of Software Process Using Intuitive Mapping of UML Activity Diagram to CPN, In Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014, ISBN 978-3-319-08156-4
3. Košinár, M., Štrba, R., Černožorský, J.: Knowledge modeling of agile software processes, In *Frontiers in Artificial Intelligence and Applications*, 260, pp. 213-220

4. Košinár, M., Štolfa, J., Štolfa, S.: Knowledge Support for Software Processes, In Proceedings of the 24th European-Japanese Conference on Information Modeling and Knowledge Bases
5. Štrba, R., Štolfa, J., Štolfa, S., Košinár, M.: Intelligent software support of the SCRUM process, In Proceedings of the 24th European-Japanese Conference on Information Modeling and Knowledge Bases
6. Černohorský, J., Košinár, M., Švercová, L., Vlach, K.: Skripta Systémy monitorování a řízení; Úvod do UML, byznys modelování, modelování požadavků, úvod do softwarového inženýrství, Faculty of electrical engineering and computer science, VŠB - Technical University of Ostrava, 2013

References

1. ALLWEYER, T. 2009. Introduction to the Standard for Business Process Modeling, Germany, Herstellung und Verlag: Books on Demand GmbH.
2. BROOKS, F. P., JR. 1987. No Silver Bullet - Essence and Accidents of Software Engineering. *Computer*, 20, 10-19.
3. CURTIS, B., KELLNER, M. I. & OVER, J. 1992. Process modeling. *Commun. ACM*, 35, 75-90.
4. CZOPIK, J., KOŠINÁR, M., ŠTOLFA, J. & ŠTOLFA, S. 2014. Formalization of Software Process Using Intuitive Mapping of UML Activity Diagram to CPN. Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014. Springer International Publishing.
5. FILL, H.-G. & BURZYNSKI, P. Integrating ontology models and conceptual models using a meta modeling approach. 11th International Protégé Conference, Amsterdam, 2009.
6. GASEVIC, D., DJURIC, D., DEVEDZIC, V. & DAMJANOVI, V. Converting UML to OWL ontologies. Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, 2004. ACM, 488-489.
7. HUMPHREY, W. S. 1995. A discipline for software engineering. Addison-Wesley Professional.
8. JENNINGS, N. R., NORMAN, T. J., FARATIN, P., O'BRIEN, P. & ODGERS, B. 2000. Autonomous agents for business process management. *Applied Artificial Intelligence*, 14, 145-189.
9. KELLNER, M. I., FEILER, P. H., FINKELSTEIN, A., KATAYAMA, T., OSTERWEIL, L. J., PENEDO, M. H. & ROMBACH, H. D. Software Process Modeling Example Problem. Software Process Workshop, 1990. 'Support for the Software Process', Proceedings of the 6th International, 1990. 19-29.
10. KOZUSZNIK, J., STOLFA, S., DUZI, M., KOSINAR, M. & CIHALOVA, M. 2011. Know How and Know What for Software Processes. Evaluation of Novel Approaches to Software Engineering, 230, 178-192.
11. KOŠINÁR, M. 2010. Design and Utilization of Knowledge Bases for Software Processes.
12. KOŠINÁR, M. 2013. Knowledge Modeling of Agile Processes in Healthcare Systems Development. *Lékař a technika* 43, 4.
13. KOŠINÁR, M., STOLFA, S. & KOŽUSZNIK, J. Knowledge support for software processes. ENASE 2010 - Proceedings of the 5th International Conference on Evaluation of Novel Approaches to Software Engineering, 2010. 105-111.
14. KOŠINÁR, M., ŠTOLFA, J. & ŠTOLFA, S. Knowledge Support for Software Processes (revisited). In: THALHEIM, B., JAAKOLA, H. & KIYOKI, Y., eds. 24th European-

- Japanese Conference on Information Modeling and Knowledge Bases, 2014 Kiel, Germany. Kiel University.
15. LENAT, D., MILLER, G. & YOKOI, T. 1995. CYC, WORDNET, AND EDR - CRITIQUES AND RESPONSES - DISCUSSION. *Communications of the Acm*, 38, 45-48.
 16. PARK, J., NAM, T. & YEOM, K. OWL Model to Support Business Process and Web Services in SOA Environments. *High Performance Computing and Communications (HPCC)*, 2011 IEEE 13th International Conference on, 2011. IEEE, 680-686.
 17. RAFFO, D. M., ADVISER-KELLNER, M. I., ADVISER-MORTON, T. E. & ADVISER-STURGES, R. H. 1996. Modeling software processes quantitatively and assessing the impact of potential process changes on process performance.
 18. RODRIGUEZ, D., GARCIA, E., SANCHEZ, S. & NUZZI, C. R. S. 2010. Defining Software Process Model Constraints with Rules Using OWL and SWRL. *International Journal of Software Engineering and Knowledge Engineering*, 20, 533-548.
 19. RUMBAUGH, J., JACOBSON, I. & BOOCH, G. 1999. *The unified modeling language reference manual*, Reading, Mass., Addison-Wesley.
 20. SCACCHI, W. 1999. Experience with software process simulation and modeling. *Journal of Systems and Software*, 46, 183-192.
 21. ŠTOLFA, S., KOŽUSZNIK, J., KOŠINÁR, M., DUŽÍ, M., ČÍHALOVÁ, M. & VONDRÁK, I. Building Process Definition with Ontology Background. *Computer Information Systems and Industrial Management Applications (CISIM)*, 2010 International Conference on. 328-334.
 22. VAN DER AALST, W. M. P. 1999. Formalization and verification of event-driven process chains. *Information and Software Technology*, 41, 639-650.
 23. VERGIDIS, K., TIWARI, A. & MAJEED, B. 2008. Business process analysis and optimization: Beyond reengineering. *Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*, 38, 69-82.
 24. W3C. 2009. OWL 2 Web Ontology Language [Online]. W3C. 2014].
 25. WORKFLOW MANAGEMENT COALITION. 1999. *Workflow Management Coalition Terminology & Glossary* [Online]. Workflow Management Coalition Specification. 2014].

Cost-based Optimizations of XML Queries

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 data model is usually understood as an alternative to the relational data model. Its main advantage is a dynamic schema. Querying XML data is a challenging task, many approaches deal with this problem. However, there are still opportunities how to improve the effectiveness of XML query evaluation. In this paper, we propose a new cost-based optimized holistic twig join algorithm for evaluating XML queries and we show that it can outperform the state-of-the-art algorithms on a specific class of queries.

Keywords: XQuery, XML, query processing, TPQ, cost-based optimizations

1 Introduction

XQuery and XPath are the most common languages to represent XML queries, in this paper we use a simplified query model called *twig pattern query* (TPQ) [1, 3], that represents the core functionality of the two mentioned languages.

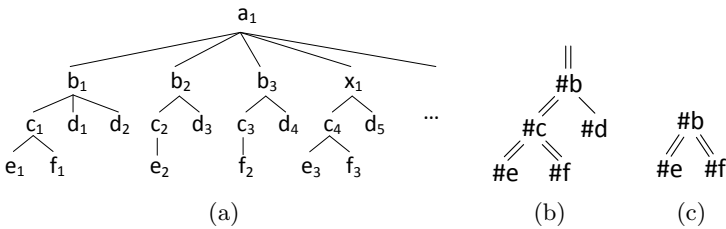


Fig. 1. (a) XML tree (b) TPQ (c) Core subquery

There is an example of XML document tree in Figure 1(a) and a TPQ query in Figure 1(b). TPQ is a rooted ordered labeled tree where nodes represent predicates and edges signify what relationships must be satisfied between the data nodes filtered by the predicates. Single and double edges represent parent-child and ancestor-descendant relationships, respectively.

2 Algorithms

The existing algorithms evaluating TPQs can be divided into two groups: (1) *binary joins* [1, 5] and (2) *holistic twig joins* [3, 4]. The first group works in two steps such that each query edge is considered first and a binary join of the two query nodes connected by the edge is performed, then the intermediate results from the previous step are stitched together. The latter group tends to evaluate the query as a whole. The main advantage of holistic twig joins is that they do not produce large intermediate results when compared to the binary joins.

The binary joins can be based on two principles: a *merge access path* and a *navigational access path*. The first mentioned group is based on reading streams of XML nodes and merging them according to a specified relationship. The other group is based on DOM (Document Object Model) operations.

3 Data structures

For the binary joins based on the merge access path and also for the holistic twig joins there must be a data structure (index) capable to return a *stream* of data nodes according to a predicate. In this paper, by the predicate we understand a node tag, but we can similarly consider for example value-based predicates. For the navigational access path binary joins there must be another data structure that can immediately return all children or descendants of a data data node.

In our prototype native XML database [6] we have the two data structures implemented and we call them *partition index* and *document index*, respectively. The data structures are schematically depicted in Figure 2.

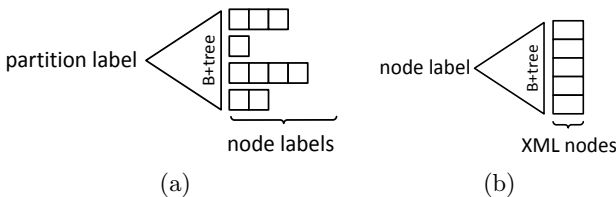


Fig. 2. (a) Partition index (b) Document index

The partition index is a B⁺tree where node tag serves as a key. The leaf B⁺tree nodes point to paged arrays where *node labels* are stored. A node label is a unique identifier of each XML data node and knowing two node labels, we can immediately resolve their structural relationship.

The document index is also a B⁺tree, but with a node label as a key. The leaf B⁺tree nodes point to paged arrays where content of the corresponding data nodes is stored. The pointers in the leaf B⁺tree nodes are sorted according to the document order. Therefore, by performing range queries we can easily obtain a whole subtree of a data node.

4 Contribution

All cost-based optimizations in the existing approaches are based on (1) searching for an optimal order of the binary joins, (2) selecting of an access path of the binary joins or (3) splitting a TPQ on a subquery that is evaluated by the holistic join and the rest that is evaluated by binary joins. In this paper, we show that it is possible to combine the two access paths also when using a holistic twig join and it can be advantageous for highly-selective queries, that are common in practice, since a user is usually interested in a concrete result.

5 Cost-Twig Join

Consider the query from Figure 1(b) that is evaluated over the XML document from Figure 1(a). There is only one subtree rooted at the node b_2 that matches the query. However, the traditional holistic join algorithm has to read all data nodes matching to all nodes of the query. That is, all 3 b data nodes, all 4 c nodes and so on have to be read.

The key idea of the new Cost-Twig Join algorithm (CTJ) is that we choose only some query nodes that will be evaluated using the partition index, while the rest of the nodes will be read from the document index. Following the names of the data structures, we call the query nodes *partition query nodes* and *document query nodes*, respectively.

The subquery, that is formed only of the partition query nodes, we call the *core subquery*. The core subquery is processed by the traditional holistic join algorithm, the rest of the query is also processed in a holistic way, however, *virtual streams* are used instead of true streams of xml nodes matching to the predicates. Virtual streams are constructed whenever the algorithm finds a match for the core subquery.

For example, let $\#b$, $\#e$, and $\#f$ be the only partition query nodes of the query from Figure 1(b). The core subquery for such a situation is depicted in Figure 1(c). The subtree rooted at b_1 is the only subtree of the sample XML document that matches the core subquery. The CTJ finds such a match as a traditional holistic join. Subsequently, it has to make sure that the match is really match of the original query. That is whether (1) the b_1 has a descendant c that is also an ancestor of e_1 and f_1 , and (2) the b_1 has a child d . The CTJ scans the whole subtree of b_1 using the document index and prepares virtual streams for the $\#c$ and $\#d$ query nodes. After that, it can continue as a traditional holistic join, since streams of all query nodes are available. The main advantage is that we have avoided reading many irrelevant c and d data nodes.

6 Cost-based optimization

There is still a problem remaining: how to divide the nodes of the original query into the partition and document query nodes. The partitioning of the nodes we call a *query plan*. If there was a mechanism estimating a cost of a query plan, we

would be able to explore a set of appropriate query plans and select the cheapest one.

6.1 Cost analysis

The cost of a query is usually expressed as a sum of IO and CPU costs. In this paper, we neglect the CPU cost, since the IO cost is usually much more important. The IO cost can be expressed by the following formula:

$$IOcost = \frac{reads}{size_s} + \beta * seeks * height + \alpha * \frac{moves}{size_t} \quad (1)$$

The first term of Equation 1 means the number of pages that have to be loaded from the paged arrays of the partition index. *reads* is a number of node labels read and *size_s* is number of node labels in one page.

The second term represents the number of inner B⁺ tree pages that have to be loaded from the document index. *height* is the height of the B⁺ tree and *seeks* is the number of matches of the core subquery. Therefore, the *seeks* value also represents the number of range queries that have to be performed over the document index to prepare the virtual streams.

The last term is the number of leaf B⁺tree pages that have to be loaded from the document index, where *moves* represents the number of loaded XML nodes and *size_t* the number of XML nodes per one B⁺tree leaf page.

The constants α and β hide some implementation details.

6.2 Estimation methods

The estimation of *reads* is done by summing the sizes of the streams for the partition query nodes, since for each partition query node we have to read the whole stream according to the predicate. The estimation of *seeks* and *moves* values is rather complicated. As we have mentioned before, the *seeks* means the number of matches of the core subquery in the XML document. Therefore, we have to estimate the result size of the core subquery. The situation with *moves* value is similar, we have to estimate the number of all descendants of the data nodes matched by the core subquery.

There are many approaches dealing with the estimation of a query result size. For the purposes of this work, we have used a sampling based estimation method [7].

6.3 Plan enumeration algorithm

There are 2^n possible query plans for a query with n nodes. Clearly, it is not desirable to compute the cost for each possible plan. We instead focus on a number of plans to be good candidates for the best possible plan.

We consider the plan where all nodes are the partition query nodes the base variant. Then we loop through the query nodes trying to change their type to the document query. Whenever the change reduces the cost, the type of the node is fixed to the document query node and it is not changed back anymore.

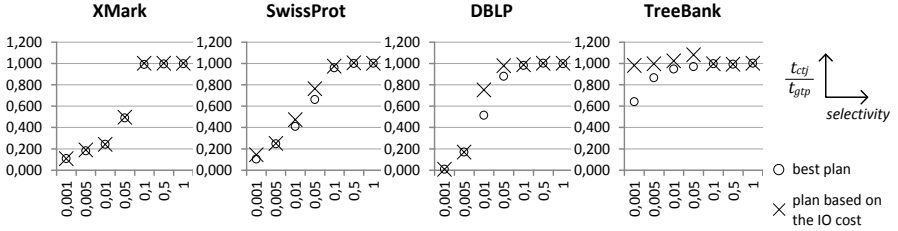


Fig. 3. Experiment results

7 Experiments

In this section we experimentally prove the advantage of the CTJ over GTP Stack state-of-the art holistic twig join algorithm (GTP) [2]. We have tested four XML data collections: XMark¹ (1.11 GB), SwissProt (109 MB), DBLP (127 MB) and TreeBank (82 MB).² We have generated 250 testing XPath queries with varying selectivity for each data collection.

According to the Equation 1, we have empirically obtained the cost model $IOcost = reads + 2 * moves + 130 * seeks$, that we use in our experiments.

Summarized results of the experiment are in Figure 6.3. The values in the charts mean the relative speed-up of the CTJ to the GTP. Horizontal axes mean the selectivity of a group of queries. For each data collection and each selectivity group, there are two values depicted: (1) circle represents the speed up when the best possible query plan is selected and (2) cross represents the speed up when we use the plan that have been obtained by the proposed cost-based approach. The optimization overhead is taken into account.

The CTJ is 10 – 100 times faster than GTP on queries with selectivity up to 0.1 % for all collections except TreeBank. The structure of TreeBank is complicated both for query evaluation and query result size estimation. It can also be observable, that using CTJ is advantageous on queries with selectivity up to approximately 10 %.

8 Conclusions and future work

We have proposed a novel holistic twig join algorithm together with an optimization framework, that can exploit the algorithm’s advantageous properties. We have shown that our approach can outperform the state-of-the-art algorithms for highly-selective queries.

¹ www.xml-benchmark.org/generator.html

² All collections except XMark: cs.washington.edu/research/xmldatasets/www/repository.html

In this paper, we have only considered node predicates that are based on the node tags. The future work can be oriented on extending the approach by the value predicates, that are practically very usefull and commonly used.

Author Publications

1. R. Bača, P. Chovanec, M. Krátký, and P. Lukáš. QuickDB – Yet Another Database Management System?. *Dateso 2014*. Pages 91–99. Písek, April 2014. ISBN 978-80-01-05482-6.
2. 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.
3. P. Lukáš. XQuery Algebra. *Information Sciences and Technologies Bulletin of the ACM Slovakia Vol. 4, No. 4*. Bratislava, 2012. ISSN 1338-1237.
4. R. Bača, P. Lukáš, and M. Krátký. Cost-based Optimizations of Holistic Twig Joins. (in progress)

Author Projects

- SGS SP2013/42 – Škálovatelnost databázových systémů, Ing. Radim Bača, Ph.D.
- SGS SP2014/211 – Efektivní implementace systémů pro detekci plagiovaných dokumentů, Ing. Peter Chovanec

References

1. S. Al-Khalifa, H. V. Jagadish, N. Koudas, J. M. Patel, D. Srivastava, and Y. Wu. Structural Joins: A Primitive for Efficient XML Query Pattern Matching. In *Proceedings of ICDE 2002*, pages 141–152. IEEE CS, 2002.
2. 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*, 22(3):369–393, 2013.
3. N. Bruno, D. Srivastava, and N. Koudas. Holistic Twig Joins: Optimal XML Pattern Matching. In *Proceedings of ACM SIGMOD 2002*, pages 310–321. ACM Press, 2002.
4. S. Chen, H.-G. Li, J. Tatemura, W.-P. Hsiung, D. Agrawal, and K. S. Candan. Twig2Stack: Bottom-up Processing of Generalized-tree-pattern Queries Over XML documents. In *Proceedings of VLDB 2006*, pages 283–294, 2006.
5. T. Grust, M. van Keulen, and J. Teubner. Staircase Join: Teach a Relational DBMS to Watch Its (Axis) Steps. In *Proceedings of VLDB 2003*, pages 524–535, 2003.
6. P. Lukáš, R. Bača, and M. Krátký. QuickXDB: A Prototype of a Native XML DBMS. In *Proceedings of the Dateso 2013 Annual International Workshop*, pages 36–47. FEE CS, 2013.
7. C. Luo, Z. Jiang, W.-C. Hou, F. Yu, and Q. Zhu. A sampling approach for xml query selectivity estimation. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, pages 335–344. ACM, 2009.

Data Structures for Indexing Triple Table

Roman Meca

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

Abstract. Semantic-based approaches are new technologies supported by RDF specifications of W3 Consortium. From physical implementation point of view, several data structures are utilized to handle this data. In many cases, the well-known B-tree is used as a basic index supporting an efficient access for basic operations. Since the semantic data are multidimensional, a common way is to use a number of B-trees for the data. In this article, we review other index data structures; we show that we can create only one index when we utilize a multidimensional data structure like the R-tree. We compare a performance of the B-tree indices with the R-tree and some variants of the R-tree. Our experiments are performed over a huge semantic database. We show advantages and disadvantages of these data structures.

1 Introduction

Semantic-based approaches are modern technologies trying to bring semantic knowledge to computers. W3 Consortium have released some specifications related to semantic technologies¹, e.g. RDF [7] as a model of the semantic data.

Definition 1. [2] *RDF triple* Assume there are pairwise disjoint infinite sets I , B , and L , where I represents the set of terms, B the set of blank nodes, and L the set of literals. We call a triple $(s, p, o) \in (I \cup B)I(I \cup B \cup L)$ an RDF triple, where s represents the subject, p the predicate, and o the object of the RDF triple.

A *triple table* is a set of RDF triples; it is an representation of an RDF graph. A triple store or an RDF database is an engine enabling us to store an RDF graph and performing at least query operations. However, we often require other operations like update operations, transactions and so on.

On W3 Consortium web, we find well-known semantic DBMS². Since the semantic DBMS often utilize a relational DBMS as a storage for the RDF triple table, the B-tree [1] is often used as the main index. The main issue of this physical implementation is that a number of B-trees must be built to support queries over the triple table. However, there are other data structures capable to handle semantic data. In this article, we show it is possible to create only one index if we utilize a multidimensional data structure like the R-tree [4] and some its variants (namely the Signature R-tree [5] or the Ordered R-tree [6]).

¹ <http://www.w3.org/standards/techs/rdf#w3c.all>

² http://www.w3.org/2001/sw/wiki/Category:Triple_Store

2 B-tree

The B-tree [1] is an one-dimensional paged data structure supporting point and one-dimensional range queries. As result, we must create more indices in the case we want to support a general range query without a sequential scan of all leaf nodes.

For example, let us consider a B-tree with the compound key (s, p, o) . In this case, we can utilize range queries $(s, p, *)$ and $(s, *, *)$. On the other hand, the range query $(*, p, o)$ means a sequential scan over all leaf nodes of the B-tree. To support all range queries we must create three B-trees. Consequently, this solution means that the size of indices is higher than the table size.

3 R-tree

Since the multidimensional R-tree [4] supports a general range query we can use it as a solution of this problem.

The R-tree can be thought of as an extension of B-trees in a multidimensional space. It corresponds to a hierarchy of nested n -dimensional *minimum bounding rectangles* (MBR). If \mathcal{N} is an interior node, it contains couples of the form (R_i, P_i) , where P_i is a pointer to a child of the node \mathcal{N} . If R is its 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 its couples of the form (R_i, O_i) , so called *index records*, where R_i contains a spatial object O_i .

The split algorithm has the significant affect on the index performance. Three split techniques (*Linear*, *Quadratic*, and *Exponential*) proposed in [4] are based on a heuristic optimization. The Quadratic algorithm has turned out to be the most effective and other improved versions of R-trees are based on this method.

However, if the dimension of the items is quite large, none of the algorithms can guarantee the split to two nodes without significant overlap. The probability increases linearly with increasing data dimension. This effect is known as *curse of dimensionality* [9].

The R^+ -tree [8] is a variant of the R-tree which allows no overlap between regions corresponding to nodes at the same tree level and an item can be stored in more than one leaf node. R^+ -trees are considered to be one of the most efficient indexes for supporting point and range queries.

Since some intervals of a range query include only one value in the case of the triple table, we call the query as the narrow range query. Therefore, we utilize the Signature R-tree [5] which is capable to handle the range query more efficiently than the R-tree and its variants.

The Signature R-tree [5] contains MBRs in inner nodes and signatures related to the MBRs. These signatures are created for tuples inserted in the subtree related to each MBR. Therefore, we can use two types of filtering when a range query scans the tree: the first filtering method tests if an MBR is intersected by a query and the second filtering method tests if a signature can include tuples of

the query. As result, the Signature R-tree reads a lower number of nodes during the range query processing. This R-tree variant is however usable only for point data and the narrow range query.

Moreover, we use the Ordered R-tree [6] since we can define an ordering for triples. The Ordered R-tree [6] is a simple combination of the R-tree and the B-tree. It means, we can use a general range queries, however we can define an ordering for tuples inserted in the tree. Evidently, we can define only one ordering for the tree. There are two consequences:

1. For some range queries, all leaf nodes intersected by the query rectangle include only result tuples. It is not generally true for the R-tree and its variants, but the range query of the B-tree provides the same behaviour.
2. We get result tuples sorted and it is not necessary to sort them after the range query is processed.

In this article, we utilize mainly the first improvement.

4 Experiments

In our experiments³, we compare the B-tree, as the main index data structure utilized in semantic Database Research Group, with the R-tree, Signature R-tree, and Ordered R-tree. All index data structures are implemented in C++⁴. We utilize a generated synthetic data collection called LUBM including 133,573,856 triples [3].

We test a performance of point and range queries processed over the index data structures when a SPARQL query is evaluated. We use 5 groups of queries related to the selectivity (GP1 range query and result size in interval $\langle 1;1 \rangle$, GP2 range query and result size in interval $\langle 2;1,000 \rangle$, GP3 range query and result size in interval $\langle 1,001;1,000,000 \rangle$, GP4 range query and result size in interval $\langle 1,000,001;\infty \rangle$, GP5). GP5 represents a sequence of point queries processed during a join operation. In the case of GP1 and GP2, it is necessary to repeat a sequence of queries since the processing time is unmeasurable. The number of iterations is 10,000 for GP1 and 50 for GP2.

We built the B-trees, the R-tree, the Signature R-tree, and the Ordered R-trees for the test data collection. The page size is 2,048 B for all data structures. In Figure 1, we see basic characteristics of these index data structures. In the case of the Ordered R-tree we do not need more trees like in the case of the B-tree, however, in this article, we want to test whether it is possible to find an optimal ordering for the Ordered R-tree. We can see that the B-tree size is up-to $3\times$ higher than the size of the R-tree-based indices. The R-tree is build in 58% of the B-tree build time. On the other hand, the build time for other R-tree-based indices is up-to $2\times$ less efficient compared to the B-tree.

³ We run our experiments on 2 x Intel Xeon E5 2690 2.9GHz and 300GB RAM memory, OS Windows Server 2008.

⁴ A part of the $\mu\mu$ DB framework developed by DBRG – <http://db.cs.vsb.cz/>.

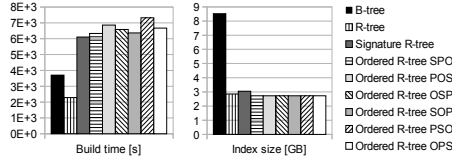


Fig. 1. Index build times and sizes

In Figure 2, we can see the query processing time for all query groups. The processing time is the average time of all queries in the group. Evidently, the B-tree provides the most efficient performance especially in the case a higher selectivity. The reason of this result is the minimal DAC of the B-tree. In the case of the lower selectivity (see GP4 in Figure 2), results of all index data structures are similar.

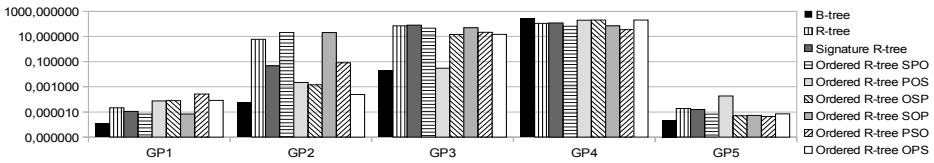


Fig. 2. Processing time for all query groups [s]

We see that the Signature R-tree and the Ordered R-tree outperform the R-tree in most cases. Although the average processing time of the Signature R-tree is lower compared to the Ordered R-tree, we can find a query in each query group where it exists an ordering of the Ordered R-tree such that the Ordered R-tree outperforms the Signature R-tree. Let us consider query processing times in Figure 2. In the case of QG1 where is typical query ($S='AssociateProfessor'$, $P='type'$, $P=*$), the Ordered R-trees SPO and SOP outperform the Signature R-tree and other Ordered R-trees, however in the case of QG2 with typical query ($S=*$, $P='PublicationAuthor'$, $O='AssistantProfessor'$) the performance of these Ordered R-trees is the lowest. Also in the case of QG3 with typical query ($S=*$, $P=*$, $O='Course2'$), the Ordered R-tree OPS give to us similar performance like the B-tree. However, we must keep in mind that this effect depends on a query and concrete ordering of the Ordered R-tree.

Although, it is clear that the B-tree provides the most efficient processing time, there are some improvements of multidimensional data structures. The first one, the index size of a multidimensional data structure is up to $3\times$ lower the B-tree index size. The second one, in the case of the B-tree it is necessary to change ordering of values in a triple when a query processor want to use an index with different ordering than another index returns.

5 Conclusion

In this article, we compared the performance of the B-tree with the R-tree, the Signature R-tree, and the Ordered R-tree for RDF data and point and range queries processed during an evaluation of a SPARQL query. The Signature R-tree and the Ordered R-tree outperform the R-tree for the most queries. Although the average processing time of the Signature R-tree is lower compared to the Ordered R-tree, we can find a query in each query group where it exists an ordering of the Ordered R-tree such that the Ordered R-tree outperforms the Signature R-tree. The B-tree provides the most efficient processing time; the average processing time is 74% of the B-tree compared to the Signature R-tree. However, there are some improvements of multidimensional data structures. The first one, the index size of a multidimensional data structure is up to $3\times$ lower the B-tree index size. The second one, in the case of the B-tree it is necessary to change ordering of values in a triple when a query processor want to use an index with different ordering than another index returns.

Author Publications

- Data Structures for Indexing Triple Table (2014) - In progress
- Join less SPARQL query execution (2014) - In progress

Author Projects

- Škálovatelnost databázových systémů (SP2014/211), Ing. Radim Bača, Ph.D.
- Efektivní implementace systému pro detekci plagiovaných dokumentů (SP2013/42), Ing. Peter Chovanec

References

1. D. Comer. Ubiquitous B-tree. *ACM Computing Surveys (CSUR)*, 11(2):121–137, 1979.
2. S. Groppe. *Data management and query processing in semantic web databases*. Springer, 2011.
3. Y. Guo, Z. Pan, and J. Heflin. LUBM: A benchmark for OWL knowledge base systems. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(2):158–182, 2005.
4. A. Guttman. *R-trees: a dynamic index structure for spatial searching*, volume 14. ACM, 1984.
5. M. Krátký, V. Snášel, J. Pokorný, and P. Zezula. Efficient processing of narrow range queries in multi-dimensional data structures. In *Database Engineering and Applications Symposium, 2006. IDEAS'06. 10th International*, pages 69–79. IEEE, 2006.
6. F. Křížka, M. Krátký, and R. Bača. On support of ordering in multidimensional data structures. In *Advances in Databases and Information Systems*, pages 575–578. Springer, 2011.

7. O. Lassila, R. Swick, et al. Resource description framework (RDF) model and syntax specification. 1998.
8. Timos K. Sellis and Nick Roussopoulos and Christos Faloutsos. The R^+ -Tree: A Dynamic Index for Multi-Dimensional Objects. In *Proceedings of the International Conference on Very Large Data Bases (VLDB 1997)*, pages 507–518. Morgan Kaufmann, 1997.
9. C. Yu. *High-Dimensional Indexing*. Lecture Notes in Computer Science. Springer-Verlag, 2002.

Chaos Level Measurement in Logistic Map Used as the Chaotic Numbers Generator in Differential Evolution

Lenka Skanderova and Ivan Zelinka

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{lenka.skanderova, ivan.zelinka}@vsb.cz

Abstract. In present time some researchers use chaotic numbers generators in evolutionary algorithms like differential evolution, SOMA or particle swarm optimization. These chaotic numbers generators are based on chaotic discrete systems which replace pseudorandom numbers generators like Mersenne Twister, Xorshift etc. In this paper we will investigate the influence of chaos level in logistic map which is used as chaotic numbers generator to the convergence speed of differential evolution to the global minimum of testing functions.

1 Introduction

Differential evolution (DE) uses pseudorandom numbers generators in many steps of the algorithm. At first pseudorandom numbers generator (PRNG) is used when the first population is created – parameters of individual are generated randomly in lower and upper bounds. Then DE needs PRNG in random choosing of three different parents, then PRNG is needed in crossing, etc. In this paper PRNG is replaced by chaotic discrete system – Logistic map. We know that the level of chaos is given by Lyapunov exponents. If Lyapunov exponent is greater than zero, system behavior can be described as chaotic. This paper will deal with influence of chaos level to convergence speed of DE. In section 2 we explain motivation of this paper, in section 3 experiments design is described. In section 4 we can see results of experiments and section 5 summarize findings.

1.1 Differential evolution

In DE we will see all principles of evolutionary algorithms – natural selection, crossing and mutation. In this paper DE/rand/1/bin is used.

In present time DE appears in many areas of research. In [3] authors deal with multi objective optimization by an adaptive DE. In [4] DE plays an essential role in identification time-delayed fractional order chaos. In [5] authors present a novel Particle Swarm Optimization (PSO) based on a non-homogenous Markov chain and DE and in [6] authors analyze the behavior of DE algorithm applied to the objective function, which are transformed by means of local searches.

Authors of [7] use distributed DE in detecting moving objects from a video sequence. In [8] authors describe repairing the crossover rate in adaptive DE.

DE in connection with chaos and chaotic systems is mentioned for example in [9] – [13].

1.2 Chaos level

Chaos level is given by Lyapunov exponent's value. Lyapunov exponent is computed for an orbit. We know that the Lyapunov exponent can be undefined for some orbits. In [2] authors says: "*In particular, an orbit containing a point x_i with $f'(x) = 0$ causes the Lyapunov exponent to be undefined.*"

1.3 Chaos level in Logistic map

Logistic map is defined by Eq. (1).

$$x_{n+1} = ax_n(1 - x_n) \quad (1)$$

In our research we will change value of parameter a and for each value of a the Lyapunov exponent will be computed. Then we will observe influence of the computed Lyapunov exponent value to the DE convergence speed to the global minimum.

Logistic map appears for example in [14], where period 3 and chaos for unimodal maps are studied. In [15] Logistic map is mentioned in connection with chaos optimization algorithms based on chaotic maps with different probability distribution. In [16] authors describe logistic neural networks and their chaotic pattern recognition properties and in [17] discrete fractional Logistic map and its chaos is investigated.

2 Motivation

As it was mentioned above, the main goal of this research was to investigate DE convergence speed reliance on Lyapunov exponent's values. Chaos is defined by a Lyapunov exponent greater than zero [2]. In this paper we observe differential evolution convergence speed when Lyapunov exponent acquires different values.

3 Experiment design

Precise setting of DE parameters is mentioned in the Table 1, where NP means number of individuals in population, D dimension (number of parameters of the individual), $Generations$ means number of generation cycles, F mutation constant and CR crossing probability. In our research Schwefel's, Griewangk's, Rastrigin's, Egg Holder's and Rana's functions have been used as cost functions (the equation can be found in [1]). Schwefel's global minimum is $f(x) = -415.9829D$

where D denotes dimension, for Rastrigin’s and Griewangk’s the global minimum is $f(x) = 0$. For Egg Holder’s and Rana’s functions there is not common formula for easy calculation of global minimum value. For experiments HP Pavilion dv7-6050 with processor Intel Core i7 with frequency 2 GHz, 4 GB RAM and graphic card AMD Radeon HD 6770M and Microsoft Visual Studio 2010 have been used. The experiments have been processed by Mathematica 8.

Table 1: DE setting.

Parameter	Value
NP	50
D	20
$Generations$	1800
F	0.9
CR	0.4

At first parameter a of Logistic map had been set to the beginning value $a = 3.5$ and then it was increased by 0.01. For each value of parameter a one hundred experiments have been generated. For each cost function 5100 experiments have been generated. Initial value of x has been set to $x = 0.02$. This value has been chosen randomly.

4 Results

In Table 2 we can see resultant values of parameter a , Lyapunov exponent, average fitness and median fitness for all cost functions. These results are mentioned in connection with the highest convergence speed of DE. In Tables 3 and 4 we can find cost functions median fitness values intervals for $a \in [3.50, 3.60]$ and $a \in [3.61, 4.00]$.

Table 2: The fastest convergence of DE for Schwefel’s, Griewangk’s, Rastrigin’s, Egg Holder’s and Rana’s functions.

	Schwefel	Griewangk	Rastrigin	Egg Holder	Rana
Parameter a	3.94	3.77	3.76	3.98	3.6
Lyapunov exp.	0.540	0.398	0.383	0.596	0.178
Average fitness	-6554.91	0.843	68.53	-6477.80	-4177.43
Median fitness	-6541.40	0.848	61.25	-8920.75	-4198.46

Table 3: Cost functions (Schwefel’s, Griewangk’s and Rastrigin’s) median fitness values interval.

Interval of a	Schwefel	Griewangk	Rastrigin
$a \in [3.50, 3.60]$	[-6399.96, -6307.58]	[0.849,0.862]	[68.58, 70.66]
$a \in [3.61, 4.00]$	[-6541.40,-6186.59]	[0.843, 0.864]	[61.25, 71.37]

Table 4: Cost functions (Egg Holder’s, Rana’s) median fitness values interval.

Interval of a	Egg Holder	Rana
$a \in [3.50, 3.60]$	[-6507.70, -6412.98]	[-4211.54, -4142.86]
$a \in [3.61, 4.00]$	[-6587.57, -6390.79]	[-4223.99,-4105.34]

5 Conclusion

From results mentioned in section 4 we can make some conclusions:

- **Schwefel’s function:** DE has reached global minimum in all cases. When parameter a had been set to $a = 3.94$ DE convergence speed was the highest. In this case Lyapunov exponent gained the value 0.540 and average fitness value gained value -6554.91, median fitness value was then -6541.40. When $a \in [3.50, 3.60]$ median fitness values moved in interval [-6399.96, -6307.58]. When $a \in [3.61, 4.00]$ median fitness values moved in interval [-6541.40,-6186.59], see Table 3.
- **Griewangk’s function:** In this case DE has not reached the global minimum for any setting of the logistic map. It is probably caused by DE’s setting. The highest convergence speed was observed when $a = 3.77$ and Lyapunov exponent gained the value 0.398, average fitness value was 0.843 and median fitness value was 0.848. When $a \in [3.50, 3.60]$ median fitness values moved in interval [0.849,0.862]. When $a \in [3.61, 4.00]$ fitness values moved in interval [0.843, 0.864], see Table 3. The smallest fitness value was reached when $a = 3.81$, fitness value gained the value 0.483.
- **Rastrigin’s function:** In the case of Rastrigin’s function DE has not reached global minimum. It is probably caused by DE’s setting. The smallest fitness value was reached when $a = 3.74$, its value was 28.44. When $a = 3.76$ Lyapunov exponent gained value 0.383 and DE’s convergence speed was the highest, the average fitness value was 68.53 and median fitness value 61.25. When $a \in [3.50, 3.60]$ median fitness values moved in interval [68.58, 70.66]. On the other hand when $a \in [3.61, 4.00]$ median fitness values moved in interval [61.25, 71.37], see Table 3.
- **Egg Holder’s function:** We know that for Egg Holder’s function there is not described global minimum in the literature. The smallest fitness value was reached when $a = 3.59$. Lyapunov exponent for this value of parameter a has the value 0.138. When $a = 3.98$ DE’s convergence speed was the highest, average fitness value gained the value -6477.80 and median fitness value -8920.75. Lyapunov exponent for $a = 3.98$ is 0.596. When

$a \in [3.50, 3.60]$ median fitness values moved in interval $[-6507.70, -6412.98]$. When $a \in [3.61, 4.00]$ median fitness values moved in interval $[-6587.57, -6390.79]$, see Table 4.

- **Rana's function:** As well as Egg Holder's function for Rana's function there is not described global minimum in the literature. The smallest fitness value -5588.24 was reached when $a = 3.92$. In this case Lyapunov exponent gained the value 0.517 for $a = 3.92$. When $a = 3.6$, Lyapunov exponent gained the value 0.178 and the convergence speed of DE was the highest. Average fitness value was -4177.43 and median fitness value -4198.46. When $a \in [3.50, 3.60]$ median fitness values moved in interval $[-4211.54, -4142.86]$. When $a \in [3.61, 4.00]$ median fitness values moved in interval $[-4223.99, -4105.34]$, see Table 4.
- When we look at the results mentioned above, we can make conclusion that DE convergence speed was the highest when Lyapunov exponent had gained values greater than zero for all testing functions. For Schwefel's function its value was 0.540 ($a = 3.94$), for Griewangk's 0.398 ($a = 3.77$), for Rastrigin's 0.383 ($a = 3.76$), for Egg Holder's 0.138 ($a = 3.98$) and for Rana's 0.517 ($a = 3.6$). For all functions DE convergence speed was the highest when $a \geq 3.6$.

In the future, we would like to extend our research by adding other cost functions especially CEC2013 benchmark. We would like to try these experiments to other evolutionary algorithms like PSO, Self-organizing migrating algorithm or genetic algorithms. This paper has been presented at the conference Nos-tradamus 2014.

6 Acknowledgement

The following grants are acknowledged for the financial support provided for this research: Grant Agency of the Czech Republic - GACR P103/13/08195S, is partially supported by Grant of SGS No. SP2014/42, VB - Technical University of Ostrava, Czech Republic, 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.

References

1. Zelinka, I. et al: Evolutionary Algorithms and Chaotic Systems, ISBN 978*3-642-10706-1, 2010 Springer-Verlag Berlin Heidelberg
2. Alligood, KT., Sauer, TD and Yorke, JA: Chaos - an introduction to dynamical systems, Textobooks in Mathematical Sciences, 1197 Springer - Verlag New York, Inc., ISBN 0- 987-94677-2, 1996
3. Venske, SM et al: ADEMO/D: Multiobjective optimization by an adaptive differential evolution algorithm, NEUROCOMPUTING, Vol. 127, 65-77, MAR 15 2014

4. Gao, F. et al: Identification time-delayed fractional order chaos with functional extrema model via differential evolution, *EXPERT SYSTEMS WITH APPLICATIONS*, Vol. 41,1601–1608, MAR 2014
5. Zeng, NY et al: A novel switching local evolutionary PSO for quantitative analysis of lateral flow immunoassay, *EXPERT SYSTEMS WITH APPLICATIONS*, Vol. 41, 1708–1715, MAR 2014
6. Locatelli, M. et al: Differential evolution methods based on local searches, *COMPUTERS & OPERATIONS RESEARCH*, Vol. 43, 169–180, MAR 2014
7. Ghosh, A. et al: Moving object detection using Markov Random Field and Distributed Differential Evolution, *APPLIED SOFT COMPUTING*, Vol. 15, FEB 2014
8. Gong, WY et al: Repairing the crossover rate in adaptive differential evolution, *APPLIED SOFT COMPUTING*, Vol. 15, 149–168, FEB 2014
9. Oplatkova, ZK et al: Analytic programming in the task of evolutionary synthesis of a controller for high order oscillations stabilization of discrete chaotic systems, *COMPUTERS & MATHEMATICS WITH APPLICATIONS*, Vol. 66, 177–189, AUG 2013
10. Senkerik, R. et al: Investigation on the Differential Evolution Driven by Selected Six Chaotic Systems in the Task of Reactor Geometry Optimization, 2013 *IEEE CONGRESS ON EVOLUTIONARY COMPUTATION (CEC)*, 3087–3094, 2013
11. Senkerik, R. et al: Chaos Driven Differential Evolution with Lozi Map in the Task of Chemical Reactor Optimization, *ARTIFICIAL INTELLIGENCE AND SOFT COMPUTING, PT II*, Lecture Notes in Artificial Intelligence, Vol. 7895, 56–66, 2013
12. Senkerik, R.: On the Evolutionary Optimization of Chaos Control - A Brief Survey, *NOSTRADAMUS: MODERN METHODS OF PREDICTION, MODELING AND ANALYSIS OF NONLINEAR SYSTEMS*, Advances in Intelligent Systems and Computing, Vol. 192, 35–48, 2013
13. Senkerik et al: Evolutionary Synthesis of Control Rules by Means of Analytic Programming for the Purpose of High Order Oscillations Stabilization of Evolutionary Synthesized Chaotic System, *NOSTRADAMUS: MODERN METHODS OF PREDICTION, MODELING AND ANALYSIS OF NONLINEAR SYSTEMS*, Advances in Intelligent Systems and Computing, Vol. 192, 191–201, 2013
14. Cheng, KJ et al: Period 3 and Chaos for Unimodal Maps, *DISCRETE AND CONTINUOUS DYNAMICAL SYSTEMS*, Vol. 34, 1933–1949, MAY 2014
15. Yang, DX et al: Chaos optimization algorithms based on chaotic maps with different probability distribution and search speed for global optimization, *COMMUNICATIONS IN NONLINEAR SCIENCE AND NUMERICAL SIMULATION*, Vol. 19, 1229–1246, APR 2014
16. Ke, Q and Oommen, BJ: Logistic Neural Networks: Their chaotic and pattern recognition properties, *NEUROCOMPUTING*, Vol. 125, 184–194, FEB 11 2014
17. Wu, GC and Baleanu, D.: Discrete fractional logistic map and its chaos, *NONLINEAR DYNAMICS*, Vol. 75, 283–287, JAN 2014

Familiarization with Distributed Defence System for Public Networks

Zbyněk Složil and Daniel Stříbný

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{zbynek.slozil.st, daniel.stribny}@vsb.cz

Abstract. Any network attacks can be performed in a distributed manner, from gaining access to various types of DoS attacks. This article describes general characteristics of distributed attacks and their impact on the public network. Furthermore, the article discusses the various ways of defense against distributed attacks and introduces a new kind of system to defend against these types of attacks. The aim of the proposed system is to create a defence system, by means of which long-term distributed attacks on a particular infrastructure in cooperation with IDS would be actively prevented. Proposed system is intended for use in independent infrastructures, which require only a minimum exchange of information.

Key words: distributed attack, defence system, public network, message, node, zone

1 Introduction

Any network attacks can be performed in a distributed manner, from gaining access (searching through the infrastructure, password cracking) to various types of DoS attacks. The basis is a large amount of seized stations (zombies, bots), which expand the leaves of the attackers tree. A major part of the public network is considered unsafe (or already under attack) and pertains to a botnet. Due to its distributed nature, both detection of an attack and defence itself are very complicated, as botnet is usually expanded through various parts of the world public network, i.e. through various AS or ISP in various geographic locations [3].

1.1 Possibilities of defense against distributed attack

Infrastructure of the public network is very open because of its very nature, both for common customers and for attackers. That is why some providers and peering companies try to protect their customers, but they themselves do not want to accept responsibility for proper data filtering.

For example, if it is discovered that the attack comes from one autonomous system, the easiest manner would be redirecting the given compromised data flow. However, by removing it from the BGP table, the problem could get worse,

the BGP itself can find an alternative way. A solution for this could be Remotely-Triggered Black Hole (RTBH) [1], which sends attack into the so-called black hole. However this solution is incomplete and slow.

Another solution may be an automatic filtering manner, which would not affect the common traffic, which would preserve independency of autonomous systems and which would be quite simple. This problem was outlined in the article about PushBack technology [2], which serves for distributed limitation of throughput, both in case of an attack and in case of an increase in common incoming flows. This approach consists in expansion of QoS policies. The traffic is classified and suspicious flow groups have limited throughput, i.e. they are sorted into QoS queues.

2 Distributed defence system

The aim of the proposed system of Distributed Access Control List (DACL) is to create a defence system, by means of which long-term distributed attacks on a particular infrastructure in cooperation with IDS would be actively prevented. The system communicates through supported infrastructures and attempts to block an attacking flow as close as possible to the source without limiting the common traffic. Its efficiency depends to a great extent on IDS of the given client, which is the only one to control the given filtering of attacks directed towards it. Transit infrastructures do not detect attacks heading to other infrastructures, since they cannot determine, if the given flow is compromised or not. The total size of the defence area is defined on the basis of support and interconnection of transit companies and their interconnecting via the system.

2.1 Terminology

The list of basic terms for the system DACL [Fig. 1]:

Node is the basic element of the system DACL, which secures filtering and forwarding of messages.

Source Node is Node, which created defence messages in cooperation with IDS.

Domain is a logical area, on the borders of which one and more Nodes are located (fullmesh). Each Node pertains to one or two Domains.

Zone is a particular Domain, which is under a particular administration. Each Node pertains to just one Zone.

Control Message (CM) is a message send between Nodes. It contains one and more filtering rules.

Control Message Flow (CMF) is a flow, which was sent from the Source Node through other Nodes and contains Defence Messages. It can create a tree structure.

Normal Data Flow (NDF) is a flow, which has not been checked yet, or has not been marked as compromised in any Source Node.

Compromised Data Flow (CDF) is a flow heading to the Zone, in which it was marked as compromised. If its source is distributed, a tree with a root is created in this Zone.

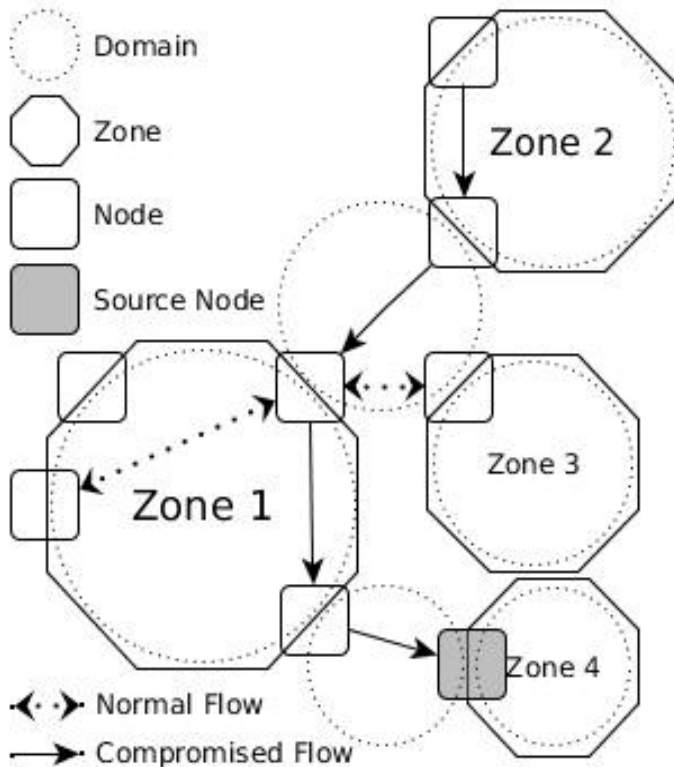


Fig. 1. Terminology of DACL

2.2 Description of the system

The defence system DACL attempts to fulfil requirements of the use in independent infrastructures (e.g. autonomous systems) and to increase their safety against cyber attacks. The proposed system DACL is intended for use in independent infrastructures, which require only a minimum exchange of information. That is why just a minimum communication is necessary to activate the service and then the system controls itself (autonomously) on the basis of clearly defined rules. Control of the given filtering is always at the charge of the Source

Node in cooperation with the used IDS, so other Nodes just filter the passing Compromised Data Flows. Zones and their Nodes, through which Compromised Data Flows pass, cannot affect filtering in any way, they only adjust it according to the accepted defence information and then attempt to pass this information on in the opposite direction of the attack. Normal Flow through Zones is not limited in any way.

2.3 Proposal

The proposal consists of 3 phases, which are gradually performed during the run of the system.

Attack detection and creation of Control Message Detection can be performed by any Node, which is connected to IDS. This IDS checks only the traffic heading to the given Zone, in which this IDS is located (it cannot control transit traffic). Any IDS can be used, but it must be able to discover Compromised Data Flows and to use defined API of DACL system used in the individual Nodes. If IDS complies with these conditions, it sends a request for filtering of a certain Compromised Data Flow to the Node. This node then becomes the Source Node for the particular Control Message Flow, which will be sent in the opposite direction of the attack. Source Node will create Control Message including rules for every particular filtering (filtering is performed at the 3rd and 4th layer of ISO OSI model) and adds lifetime to each rule (i.e. period of inactivity, after which the rule will be automatically removed from the filters). These rules are immediately applied in the filter in the Source Node.

Filtering according to Control Message After receiving a message into the Node (with the help of IDS or by accepting message from another Node), it will unpack the Control Message and add the rules contained in it into the filter. If a Compromised Data Flow comes across the rule corresponding to it, the lifetime of the rule will be activated to the original value, while this lifetime gradually decreases since the time of its creation. If the rule is inactive for a longer period of time and its lifetime decreases to zero, the rule will be erased from the filter. Rule lifetime determines the maximum time delay between two received PDU of an attack and it is detected by means of IDS.

The selected final manner of rule lifetime determination is the hybrid method. Life of each rule accepted into the Node is compared with the limit value of the given Zone and if it exceeds this limit value, it is modified to comply with this value. However, lifetime is sent unchanged to other Nodes (with the original value received from the Source Node), so that a potential gradual degradation does not occur. We still expect an exact determination of lifetime with a slight deviation of intensity of the attack messages. Another phase, forwarding of the given rule, is initiated, only if the given rule was used at least once.

Forwarding of a Control Message The compiled Control Message is sent in the opposite direction to the Compromised Data Flow. The Source node sends the

Control Message always in a direction outside of its Zone. Other Nodes send Control Message always into another Domain, that the one, from which it was received. However, before the particular Control Message is sent, the closest connected Node, through which the Compromised Data Flow passes, must be found. There are several possibilities of how to solve this problem, to prevent asynchronous routing or potential falsification of addresses of the source (the so-called spoofing):

1. Forwarding of Control Messages to all Nodes in the Node Domain (use of multicast broadcasting). If the given Compromised Data Flow does not pass through the Node, the rule will be automatically removed after its lifetime expires.
2. Marking of all Flows, which pass through the given Domain. Each Node has a unique marking in this Domain, which can be added into the header of the IP packet or added as a new header between the individual layers.

Forwarding of rules to other Nodes should have a certain delay, so that an excessive fragmentation of the Control Message is prevented. The phases 2 and 3 keep repeating, until there is a neighbour Node for the next forwarding of the Control Message. Filters are automatically erased in the opposite direction to the Compromised Data Flow and if the attack is still going on, only those filters, which are the closest filters to the attackers, will remain. If the given Compromised Data Flow disappears, all rules in filtering are removed as well. If the shape of the attack changes (e.g. its direction, intensity), the attack will pass to IDS again, where the phase 1 will be applied again and the other mentioned phases will follow.

3 Conclusion

Cyber attacks have become a big current threat for many companies. Defence against these attacks is underestimated and in many cases certain parts of infrastructure become a part of an attack. There are many types of local defence against these attacks, such as status and non-status filters, Honeypots, IDS and others. With regard to the growing number of connected users, the number of attacks is increasing as well, and thereby also the load on transit infrastructures increases. DACL system attempts to protect these areas in cooperation with the victim of the attack.

Another improvement may be compression of transferred rules, because if more distributed attacks occur, there may be a great amount of created rules. However, summarisation is not possible, as the system would lose precision of filtering. This weakness should be solved in further research.

DACL can become an efficient active defence against distributed attacks. Its greatest advantage is simplicity of both the protocol and the expected implementation. If a sufficient support from providers is achieved, defence will be so efficient, that it can often block attackers at the very source.

References

1. Mahajan, Ratul and Bellovin, Steven M. and Floyd, Sally and Ioannidis, John and Paxson, Vern and Shenker, Scott Controlling High Bandwidth Aggregates in the Network, SIGCOMM Comput. Commun. Rev., July 2002.
2. Turk, D Configuring BGP to block Denial-of-Service attacks, 2004.
3. Daniel Stříbný. Implementation of Distributed Defence System for Public Networks, WOFEX 2014, 2014.

Approximation of molecular energies using neural networks

Martin Stachoň and Aleš Víték

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

Abstract. When performing molecular dynamics or Monte Carlo simulations of molecular systems, a method to compute intramolecular energies is crucial. Fully quantum ab-initio computations are accurate, but very costly. Practical simulations of larger systems (in either number of atoms or time scale) thus often use simplified formulas based on physical insight, with parameters fitted to experimental data. We have explored a different approach, using neural networks.

Keywords: Neural networks, molecular dynamics

1 Introduction

We tested the approach on a system of six water (H_2O) molecules. On input, we are given coordinates of atoms in cartesian coordinates. Expected output is the intramolecular energy. A common method is to use n -site ($n = 2 \dots 6$) water model, in which charge is assumed to be placed in specific points around the water molecule. Electrostatic interaction between the sites is modelled by Coloumb's law and repulsion forces by the Lennard-Jones potential. This results in a simple analytic formula with several parameters, which are fitted to the situation which is being modeled (simulation at given pressure, temperature etc.)

We tried a new approach, described eg. in works of [3], which uses neural networks to approximate the energy calculation, without explicitly constructing an interaction formula. The neural network is trained by using data from very accurate quantum chemical computation or from experimental measurements.

2 Neural network inputs

We used a classic multi layer feed-forward neural network with sigmoid activation function, trained using backpropagation on the training set.

Standard cartesian coordinates (x - y - z) are not well suited as inputs for neural networks, because intermolecular energies are invariant under translation and rotation of the whole system. Training NN with all possible translations and rotations would be impractical, therefore some transformation must be used on

the cartesian coordinates. We have adopted the approach of symmetric functions by J. Behler [1].

Symmetric functions are defined for every pair and triplet of atoms with distance of R_{ij} :

$$G_i^2 = \sum_j e^{-\eta(R_{ij}-R_s)^2} f_c(R_{ij})$$

$$G_i^3 = 2^{1-\zeta} \sum_{j,k \neq i} (1 + \lambda \cos \Theta_{ijk}) \zeta e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk})$$

Where λ, η, ζ are parameters, Θ_{ijk} is the angle between the atoms and f_c is dampening function defined as follows :

$$f_c(R_{ij}) = 0.5 \left(\cos \left(\frac{\pi R_{ij}}{R_c} \right) + 1 \right) \text{ for } R_{i,j} < R_c$$

$$f_c(R_{ij}) = 0 \text{ otherwise}$$

Where R_c is cutoff distance. The coordinates have no special meaning, they are just a way of sampling of intramolecular distances and angles. One can see a similarity to using moments in pattern recognition with NNs.

3 Results

We have trained the NN on 907 configurations of water hexamer, with accurate energies calculated using Density Functional Theory method (exchange potential M06L, basis aug-cc-pVTZ on oxygens, cc-pVTZ on hydrogens). The results are shown in Fig. 2. Using the NN, we have achieved a variance of 2.0416e-5 Hartree². For reference, compare to the TIP6P model (a six-site water model) on Fig. 1, which achieved a variance of 7.42331e-7 Hartree², two orders better.

4 Conclusions and Future work

The precision of energy approximation using neural networks is not good enough yet. We see issues that potentially need to be addressed :

- We need a better transformation of input coordinates. The symmetric coordinates may be blurring the image of the system too much.
- It might be a problem to obtain large enough testing set. We used a set of 902 inputs, which might not be enough to cover the whole configuration space.

There are some issues common to all uses of bio-inspired methods - the NN works as a black box, and we cannot extract any physical model back from it.

Some other approaches other than NN could be used. We have tried genetic programming without success, reports of an experiment with SVM are in [2].

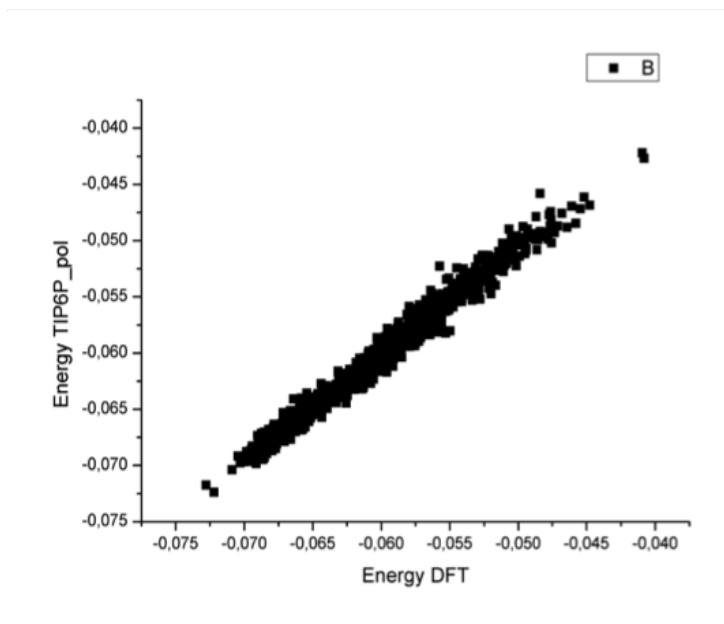


Fig. 1. Expected energy from DFT vs. energy obtained from TIP6P model.

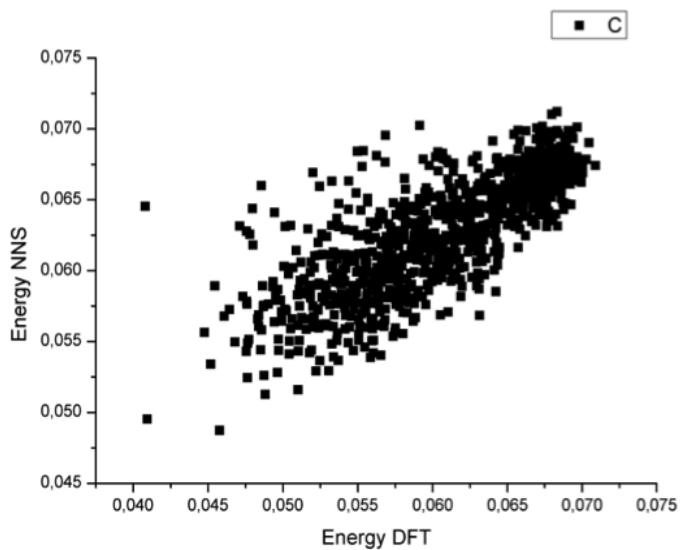


Fig. 2. Expected energy from DFT vs. energy obtained from Neural Network.

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. J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *The Journal of Chemical Physics*, vol. 134, no. 7, p. 074106, 2011.
2. A. Vítek, M. Stachoň, P. Krömer, V. Snášel, "Towards the modeling of atomic and molecular clusters energy by support vector regression," *5th IEEE International Conference on Intelligent Networking and Collaborative Systems, INCoS 2013*
3. J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potentialenergy surfaces," *Phys. Rev. Lett.*, vol. 98, p. 146401, Apr 2007.

An Application of Process Mining to Analysing the User Process on the Web Page

Jakub Stolfa and Michal Holis

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{[jakub.stolfa](mailto:jakub.stolfa@vsb.cz), [michal.holis](mailto:michal.holis@vsb.cz)}@vsb.cz

Abstract. In today’s highly competitive environment of e-commerce and web applications it is necessary to innovate and enhance the core user experience. One of the ways to achieve this is to thoroughly analyse user base, find patterns in their behaviour and understand differences between valuable buying customers and ordinary visitors. Data from such analysis then acts as cornerstone in decision making in applications further development. In this paper usage of process mining methods is presented applied to the problem of main customer workflow throughout the web application focused on selling custom hand-made chocolates.

Keywords: Process mining, process improvement, web

1 Introduction

In modern world, when building applications targeted to common end-users, it is necessary to optimize the user experience and applications main workflow in order to retain customers interest and make application viable in context of it’s competition.

Optimization of such process is never-ending task. As the competition and user’s needs evolve, iterative workflow optimization is a must. In order to be able to effectively perform such optimizations, it is necessity to understand the behaviour of applications users. Therefore thorough analysis of their actions in such application is a key to understanding the user’s needs. Also throughout this analysis it is possible to divide users into categories according to the business value they provide to the owner of the application. Usage of this data then allows to understand how users with higher business value interact with the application so the general user experience of the application can be fine-tailored to their needs. Also it allows to work with the users with lower business values in order to increase their value.

This paper focuses on analysis of users workflow process in web application focused on selling custom hand-made chocolate to the user.

The paper is divided into four main sections. In section 1 the problem is introduced. In section 2 state of the art is presented to the reader. In section 3 obtained data and analyses are described. Final section 4 concludes the whole paper and provides discussion over obtained results and summarizes planned future research in given subject.

2 State of the Art

In last decades, systems started to be more and more process oriented [1]. The shift to process oriented systems was motivated by the idea of supporting systems to the daily business, to shift the knowledge about operations that could be described as processes from humans to systems. Process oriented systems started to be worshipped as the only way to control the processes and activities that has to be enacted. The knowledge about the processes and their enactment was transferred to the systems.

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. Modelling and simulations can help you to adjust the process, find weaknesses and bottleneck during the design phase of the process. Sometimes you guess or know the patterns and occurrence probabilities of variations that are used during the execution phase. However, not even modelling and simulation of the processes can tell you, how processes are really enacted in the system, what is e.g. the perceptual usage of the variations and whether some variations are enacted at all. If you want to analyse 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 analyse the process itself. Process mining can answer you the question, how the process was really executed, which variations were used and what are the probabilities of the enactment of each process variation. Process mining can be seen as a supportive method for the BP and BPI (Business process intelligence)[2] analysis and from the perspective of BPM(Business process management)[3], can be used as a feedback to the BPM methods [4].

Our work is focused to the process mining, estimation of the future process parameters (process data) and process formalization. We are working on holistic process view that can bring us complex view of the intended processes, real process execution, process estimation and all of this supported by process formalization [7–16].

3 Possibilities of process analysis

This paper analyses web application focused on selling custom hand-made chocolate. As the application is quite new and the general user experience is still undergoing frequent changes and improvements, the need to analyse behaviour of customers arose. Main question in mind was to find patterns in their behaviour in order to understand the difference between user that orders the product from the website and user that visits the website without ordering anything.

Additionally it is planned to use the data and results of the analysis to optimise the user interface.

3.1 Process context

As stated before, focus of this paper is web application designed to allow users to buy custom hand-made chocolate. It provides the user with an ability to pick their favourite chocolate bar, pick their favourite toppings and finally choose the box that the chocolate should be packed in.

As a second product the user is able to purchase box of chocolate. He is presented with choice of the box, choice of up to five types of pralines and is given an option to upload their own picture that will decorate the box.

3.2 Data description

We have loaded the log with dates between 7. 11. 2013 - 28. 7. 2014, totally we have loaded *25 356* records for adjustment.

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.

The process in the process mining consists of at least of:

- Case - one pass of the process
- Event - one step of the process
- Start time - start time of the task

Our log is mapped in the following way:

- Case - IP + number of occurrence.
- Event - Page
- Start time - Datetime

We had to split cases from one IP address to the several cases. User from one ip address can visit the web several times and it represents several performed processes cases. We had to decide where the borderline between these processes is. We had set up that if the time between performing particular activities is longer than 2 hours we consider it a new process. For example if the user performs 6 activities and 3 of them are performed on 1.1.2014 from 10 to 11 AM and rest of the activities are performed on 1.1.2014 from 3 to 4 PM, these are two separated processes of one user IP address. In the log we have found 30 activities, i.e. Vytvorit/bonbonieru, Vytvorit/cokoladu and so on.

3.3 Summary of the log

The process mining tool ProM 6.2 was used for all analysis attempts [5].

Table 1 depicts top five events ordered by occurrences. We can see that most used events are /vytvorit/cokoladu and /vytvorit/bonbonieru.

Table 1. Occurrences of top 5 events

Class	Occurrences (absolute)	Occurrences (relative)
/vytvorit/cokoladu	25770	50,818%
/vytvorit/bonbonieru	9296	18,332%
/kosik	4564	9,0%
/vytvorit	3236	6,381%
/MainPage	2682	5,289%

The Table 2 depicts start-event classes of the process. That means start classes that are starting events for at least one of process enactments. The Table 3 depicts end-event classes of the process. That means task classes that are end for at least one of process enactments.

We can see that 42 percent of process executions start on the main page and 48 percent of process executions end by /vytvorit/cokoladu. It means users mostly come to the main page and then try to create and customize own chocolate product.

If the process starts with either one of the events - /vytvorit/cokoladu or /vytvorit/bonbonieru it means that users come directly to this pages. The reason might be that they click on some kind of advertisement of these chocolates on another web page (facebook, etc.).

Table 2. Top 3 Start events

Class	Occurrences (absolute)	Occurrences (relative)
/MainPage	587	42,352%
/vytvorit	329	23,737%
/vytvorit/cokoladu	313	22,583%
/vytvorit/bonbonieru	69	4,978%

Table 3. Top 3 End events

Class	Occurrences (absolute)	Occurrences (relative)
/vytvorit/cokoladu	671	48,413%
/vytvorit/bonbonieru	225	16,234%
/MainPage	140	10,101%

3.4 Analysis of the process characteristics

We have reconstructed process model from the log by BPMN Analysis (using Casual Net Miner) [6]. Process model is depicted in the figure Fig. 1. The process model shows also 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 that the process model is not easily readable and contains lot of events. It is because of that the domain of the web process mining is quite different from the process mining of ordinary informational systems. Ordinary informational system usually has to follow at least some basic workflow. On the web page user usually don't have to follow the process so much. But we can observe there are some patterns that are followed. For example pattern of ordering chocolate, etc. In this area we can also use sequence alignment methods [16].

However the process model can replay all the cases obtained from the log.

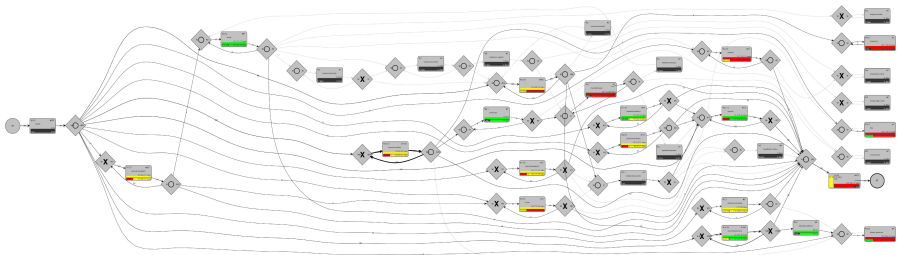


Fig. 1. Process model visualisation in BPMN

4 Conclusion and future work

This paper shows ability to map and analyse processes of the web pages using process mining techniques. Not all options of such techniques are presented in this paper, but the results seem very interesting and promising. Using process mining techniques in such scenarios could lead to substantial improvements in user experience and could become another option for companies looking for feedback that could validate and/or improve their product's workflows.

For example we can focus more on the process executions that start by clicking on some kind of advertisement. We would like to observe the effectiveness of such advertisement and compare the number of customers that end up buying some product through advertisement to number of users that buy the product spontaneously without any such advertisement. Also it might be interesting to combine this output with heat-map analysis. We would like to continue in this domain and follow on the work presented in this paper.

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) 321343.
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) 18.
4. W.M.P. van der Aalst, A.H.M. ter Hofstede, M. Weske, *Business process management: a survey*, *Business Process Management* (2003) 112.
5. 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.
6. Van Der Aalst, W., Adriansyah, A. & Van Dongen, B. 2011, *Causal nets: A modeling language tailored towards process discovery*.
7. Stolfa, J., Stolfa, S., Kobersky, O., Kopka, M., Kozusznik, J., Snašel, V.: *Methodology for Estimating Working Time Effort of the Software Project*, In: 2012 Databases, Texts, Specifications, and Objects (DATESO), pages 25-37, 2012
8. Jakub Stolfa, Svatopluk Stolfa, Jan Kozusznik, Tereza Moudra: *Business Process Formal Modeling in Graphical Ontology Tool - Functional View*, In: *The International ACM Conference of Emergent Digital EcoSystems (MEDES)*, 2012
9. Stolfa, J., Kobersky, O., Kopka, M., Kromer, P., Stolfa, S., Kozusznik, J., Snašel, V.: *Value estimation of the use case parameters using SOM and fuzzy rules*, In: *The International ACM Conference of Emergent Digital EcoSystems (MEDES)*, 2012
10. Jakub Stolfa, Ondrej Kobersky, Pavel Kromer, Svatopluk Stolfa, Martin Kopka and Vaclav Snašel: *Comparison of Fuzzy Rules and SVM Approach to the Value Estimation of the Use Case Parameters*, *IFSA World Congress NAFIPS Annual Meeting*, 2013
11. Svatopluk Stolfa, Jakub Stolfa, Pavel Kromer, Ondrej Kobersky, Martin Kopka and Vaclav Snašel: *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
12. Jakub Stolfa, Martin Kopka, Svatopluk Stolfa, Ondrej Kobersky and Vaclav Snašel: *An Application of Process Mining to Invoice Verification Process in SAP*, *The 4th International Conference on Innovations in Bio-Inspired Computing and Applications*, Ostrava 2013
13. Martin Kopka, Milos Kudelka, Jakub Stolfa, Ondrej Kobersky and Vaclav Snašel: *Extraction and Analysis Social Networks from Process Data*, *5th International Conference on Computational Aspects of Social Networks (CASoN)*, 2013
14. Michael Alexander KOSINAR, Jakub STOLFA, Svatopluk STOLFA, *Knowledge Support for Software Processes*, *24th International Conference on Information Modelling and Knowledge Bases*, June 3-6, 2014, Kiel, Germany
15. Radoslav STRBA, Jakub STOLFA, Svatopluk STOLFA, Michal KOSINAR, *Strategic Mission of Intelligent Support of the Scrum Software Process*, *24th International Conference on Information Modelling and Knowledge Bases*, June 3-6, 2014, Kiel, Germany
16. J. Stolfa, S. Stolfa, K. Slaninova, J. Martinovic, *An Application of Process Mining by Sequence Alignment Methods to the SAP Invoice Process Example*. In *Dateso*, Roudnice nad Labem, Czech Republic, 81-90, 2014

Implementation of Distributed Defence System for Public Networks

Daniel Stříbný 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. Modern network infrastructure suffer greatly from ongoing DDoS attacks. These attacks targeting not only servers and services running on them, but also in many cases the infrastructure itself. Proposed Distributed Defense System is targeting these DDoS attacks and tries to effectively and quickly protect their victims. Implementation of this system could be tricky in modern heterogenous network infrastructure. This article describes possible ways of implementation of this system on network devices of most commons manufacturers. And also shows prototype implementation on Linux system which serves as proof of concept. In conclusion are described possible future improvements based on testing of the prototype.

Key words: distributed attack, DDoS, OpenFlow, defense system, Mono

1 Introduction

Any network attacks can be performed in a distributed manner, from gaining access (searching through the infrastructure, password cracking) to various types of DoS attacks. The basis is a large amount of seized stations (zombies, bots), which expand the leaves of the attackers tree. A major part of the public network is considered unsafe (or already under attack) and pertains to a botnet. Due to its distributed nature, both detection of an attack and defence itself are very complicated, as botnet is usually expanded through various parts of the world public network, i.e. through various AS or ISP in various geographic locations. Proposed system is intended to deal with these attacks [5].

2 Implementation Options

Implementation options are limited by heterogeneity of network devices in role of Node (typically routers) [5]. There isn't tool for uniform creation, modification and checking of firewall rules on Nodes.

This problem can be solved in many ways:

Network containing only Cisco Systems devices using Cisco One PK

- One PK is Software Defined Network (SDN) technology. Provides common API for managing services on network devices of Cisco Systems. There are libraries for accessing these APIs for most common programming languages - C++, Java, Python. Functionality of this technology is provided on devices of Cisco Systems so overall deployability of system based on this technology is very limited [1].

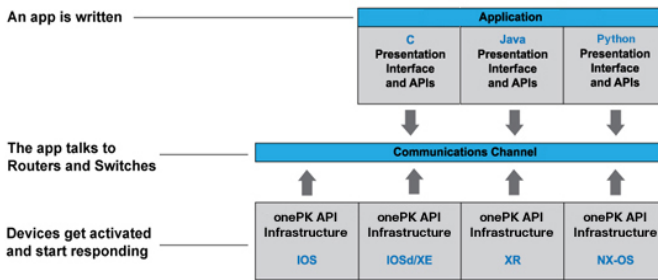


Fig. 1. Design schematic of Cisco One PK [1]

OpenFlow - is also SDN technology. Developed as University research project, it defines new approach to network devices. These devices has separated Control plane and Data plane. Control plane is distributed so implementation of this defense system could be very effective. Only drawback is low commonness of devices based on OpenFlow protocol [2].

Custom framework - custom modular framework for settings firewall rules on devices of different manufacturers. This can be enabled via APIs or CLI modules for each manufacturer. Base for this framework is described in previous diploma work on VSB-TUO [3]. This base code could be used for future development.

In all cases is used some kind of central server for managing single Nodes. This approach is rather centralized in contrast with systems distributed and decentralized proposal [5].

3 Prototype Implementation

Prototype is implemented in Mono framework via C# language. Operation System used is Ubuntu 14.04. Its main purpose is to demonstrate functionality of proposed concept and to create platform for future testing.

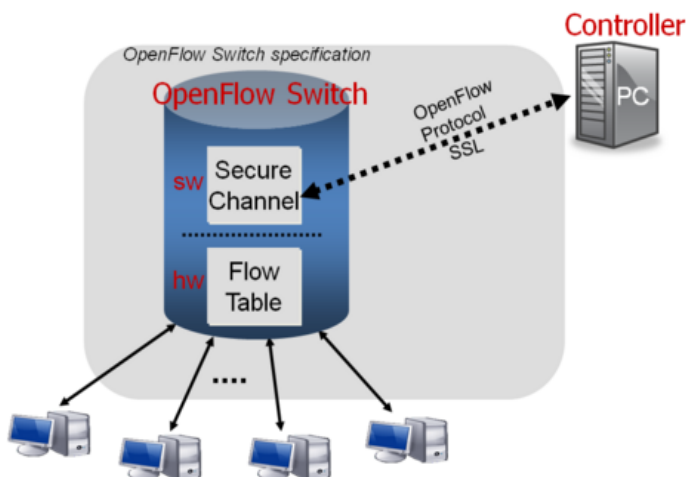


Fig. 2. Design schematic of OpenFlow [2]

In this implementation is testing computer used in role of Node and also management server. Operation System provides routing and firewall functionality. Firewall functionality is provided via Iptables tool. For managing rules in this tool is included library - Sharpknocking [4] - which itself is wrapper for Iptables CLI.

Implemented prototype reacts on incoming messages, sets up firewall rules and also sends messages in opposite direction to attack flow as is described in proposal of whole system [5]. Basic schematic of prototype is described on figure 4.

Control messages contains parameters for created firewall rules as seen on figure 5.

Usage of this prototype is very limited - in real networks is nearly impossible to find software router based on Linux operation system. Future development will therefore be focused on custom modular framework - as described in previous chapter.

4 Conclusion

Testing prototype proved functionality of proposed system. And also necessity of taking heterogenous network environment into account in future implementations. This could be done via modern SDN technologies. But due to legacy devices is custom framework for managing firewall rules only possible implementation way.

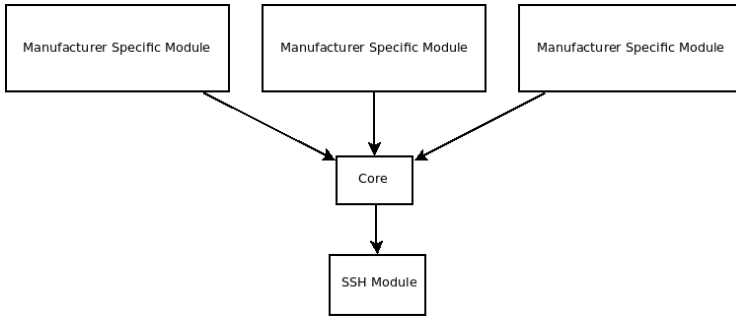


Fig. 3. Design schematic of possible custom framework implementation

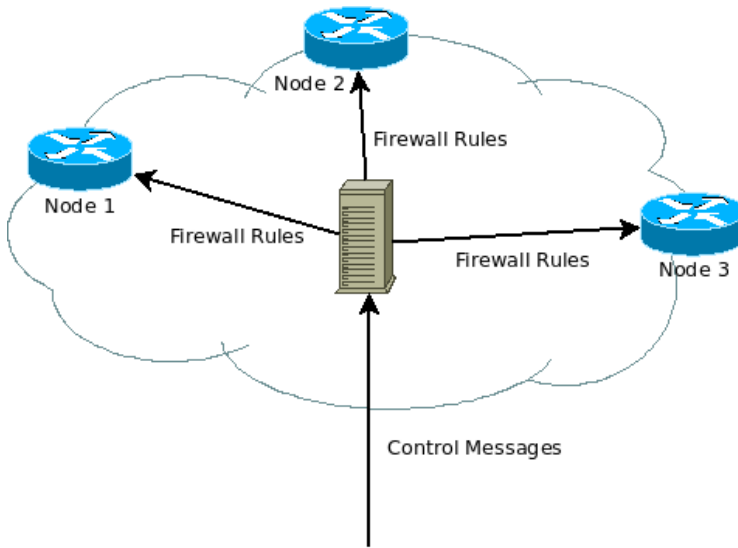


Fig. 4. Use case of prototype

It is also necessary to focus on securing whole system and his hardening against abuse. One fact which must also be taken in account is amount of centralization in this by proposal decentralised system.

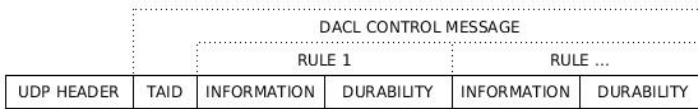


Fig. 5. Control message [5]

References

1. Cisco Systems, Inc. Cisco One PK, <http://www.cisco.com/c/en/us/products/ios-nx-os-software/onepk.html>, 2014.
2. Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, Jonathan Turner OpenFlow: Enabling Innovation in Campus Networks, <http://archive.openflow.org//documents/openflow-wp-latest.pdf>, 2008.
3. Jan Bednár. Centralized Environment for Design and Distribution of Firewall Rules in Local Area Networks, <http://dspace.vsb.cz/handle/10084/103808>, 2014.
4. mangelp. Sharpknocking Library, <http://code.google.com/p/sharpknocking/>, 2014.
5. Zbynk Složil. Familiarization with Distributed Defence System for Public Networks, WOFEX 2014, 2014.

Simulation of Agile Software Process with the aid of Artificial Neural Network

Radoslav Štrba

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

Abstract. Company success in a highly competitive agile environment depends on the speed of right decisions making process. The quality of those decisions depends on availability of statistical reports, results of simulation and other information gathered from various supporting software systems. This calls for establishment method that allows combining developed intelligent software tools. This paper presents a method for the support of agile software processes. We proposed the method for modeling and simulation of the software process model with the aid of a neural network. In particular, we aim at improving software process development using results of the simulation and effort estimation.

Keywords: Agile Methods, Software Process, SCRUM, Software Process Simulation, Neural Networks, Classification, Effort Estimation

1 Introduction

This paper provides an overview of work being constructed in field of intelligent support of agile software processes. Our work is focused on neural network based effort estimations and simulations of the scrum processes in commercial companies. It identifies three important questions: “Why simulate and estimate”, “What is the scope of the estimation and simulation model” and “How to simulate and estimate” – suitable simulation and effort estimation techniques in the specific agile environment.

We developed a method and intelligent software tools as a part of our research on agile methods, support of software processes modeling, estimation, simulation and executing. These tools allow modeling of the software process using formal methods. They also allow simulation and effort estimation for development of commercial software products.

2 Reasons for Simulation

There is a wide variety of reasons for undertaking simulations of software process models. In many cases, a simulation is an aid to decision making. It also helps in risk analyze and reduction and helps management at the strategic and operational levels. We have divided reasons for using simulations of software processes into three categories of purpose:

- Planning of the software project: The Project management planning can be supported by the simulation. It can help estimate the effort and cost of products. The Simulation can also help analyze a risk in initial planning phase. Project managers can use the simulation to predict a possible outcome if a proposed action.
- Improvement of the software process: In contrast to planning, here is an improvement of the process. A Simulation can be used to identify bottlenecks and result can be used to calibrate the model.
- Understanding of the software process: Using simulation tools with the output graphical interface we can visualize a process flow and help people to understand effect of changes in the process configuration. The simulation can helps people understand the inherent uncertainty in forecasting agile process outcomes and the likely variability in actual results seen.

In previous paragraphs we've provided examples of practical issues, which can be addressed with the simulation. When developing software process simulation models, identifying the purpose and the questions project management would like to address is important to defining the model scope and data that need to be collected. [4], [5]

3 The Scope and Structure of the Simulation Model

Proposed approach for the simulation and the effort estimation in SCRUM environment starts with an architecture. The architecture involves effort estimation components and learning rule used here is the Back-propagation algorithm. The effort estimation consists of the three inputs components which get inputs from the design document. The three components are: [5], [6]

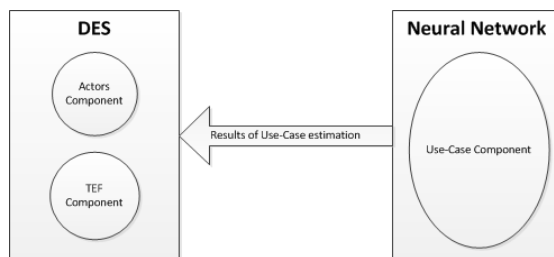


Fig. 1. The structure of simulation components

- Actor components: This takes information about the actors involved in the system.
- Use-case component: This takes information about the Use-Case involved in the design document.
- TEF component: This takes the information regarding the technical and environmental factors involved in the system.

Very important is thinking of time span of the simulation process model. We can think of a short time span, (it means that software product is developing) less than 12 months and organization simulation breath of one scrum-team. If we want to answer the key questions, we need the result variables. Depends on the key question being asked, there can be many different variables devised as the results of a software process simulation. [4], [5], [6]

Result variables for software process simulation of SCRUM include the following:

- An effort (cost)
- An ideal duration of one iteration (planning of iterations)
- A schedule (developers-time management)
- A defect level (quality assurance)

Input Parameters:

- Manually set parameters (number of actors – developers, developers properties, amount of tasks – User Stories)
- Application log-data (historical information from project development)

4 Model for Discrete-Event Simulation

We've created 4 basic types of entities that are fundamental for the SCRUM process models: Project, Iteration, User Story and Team involving developers. A Project contains a set of Iterations. Iteration contains a set of User Stories. The 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 User Stories. The User Story is divided into particular tasks for developers. Developers are the main actors/roles of an agile software development process. They are responsible for develop the whole system. In the model, each developer is characterized by a set of attributes: experience, availability and average lines code. [9], [10], [11]

5 Design of the Neural Network and Input Parameters

For our experiment, we choose one of the most common neural network architectures, the feed-forward back-propagation neural network. Term, “feed-forward” describes how this neural network processes and recalls patterns. In a feed-forward net-

work neurons are only connected forward. Back-propagation is a form of supervised training. The back-propagation and feed-forward algorithms are used together.

Our feed-forward network begins with input layer. The input layer is connected to a hidden layer. Hidden layer is connected directly to the output layer. There is one hidden layer. The output layer of the neural network is what actually presents a pattern to the external environment. The number of input and output neurons is directly related to the intended use of neural network. This neural network is used to classify items into groups. We have output neuron for each group. [4], [5], [8]

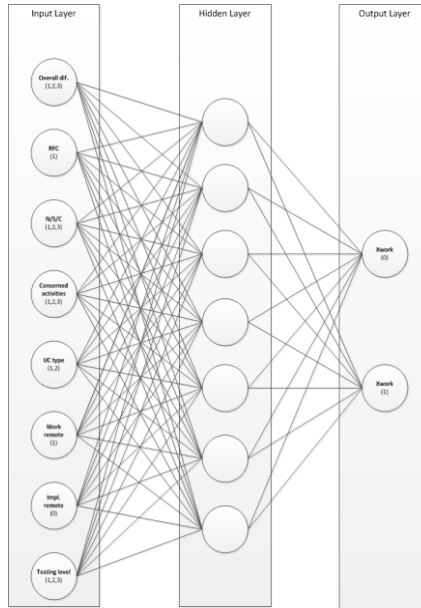


Fig. 2. Feed-forward neural network for classification of use-cases

The input layer consists of eight neurons. Each neuron is related to one of Use-Case parameters. There are eight input parameters: Overall difficulty, RFC, N/S/C, Concerned activities, Use-Case type, Work remote, Implementation remote and the Testing level. These inputs to the neural network are integer values transformed to floating point numbers.

We've decided to use one hidden layer, which can approximate any function that contains a continuous mapping from one finite space to another. Number of seven neurons has been determined by using of experiment results and two advised rule-of-thumb methods for determining the correct number of neurons. The simulation executions were performed using input parameters given in (Table 1.) These values are based on real data taken from development. Number of tasks and typical iteration duration are constant values. A function for generation time duration of every single task considers two input parameters: Minimum Estimated Hours and Maximum Estimated Hours. Collection of tasks is automatically generated depending on the parameters number of tasks and estimated hours.

Table 1. Input parameters

Parameter	Value	Other
Number of Tasks	100	
Number of Developers	5	Table 4.
Minimum estimated hours	4 hrs.	
Maximum estimated hours	14 hrs.	
Typical iteration duration	10 days	

Table 2. Actors: List of developers

Id	Experience	ALC	Availability
1	1	120	140
2	2	130	150
3	3	160	250
4	4	190	180
5	4	210	180

One or more actors execute each activity of the process. The team whose activities are simulated consist of developers and those stands as actors within the simulation platform (Table 2). These actors are described by three main attributes: first property, experience means practice in years, second one is ALC (Average Lines of Code per day) and second property availability means monthly working time. Some rules can utilize more attributes like developers' certifications, specializations etc. Once the static structure of the system to be simulated is fed into simulation platform, we can execute the simulation itself based on dynamic rules.

6 Simulation execution and results

Results of the simulated process (Table 4.) flow combining DES and the neural network based approach, including experience, availability and ALC of team's developers. Important factor that influences the process flow is generation of obstacles, which are taken from real software process execution and could be e.g. blackout or internet connection outage. Obstacles are generated in periods following an exponential distribution. Duration of typical development events follows normal (Gaussian) distribution.

Table 4. Result values of the SCRUM process simulation

Output parameter name	Count	Number of developers / 4 weeks
Task-estimated hours	852	4
Developer-lines code	8726	4
Task-remaining hours	948	4

7 Conclusion and future work

In this paper we've introduced an approach for simulation of the SCRUM software process. We also estimated the software project effort with the aid of neural network. The simulation can help to management with estimation of an effort needed for development of some parts of the product, which are easy to get. They can optimize the process or plan better. So far the tests of simulation tools and formal model we've build prove that this approach works even it still has some insufficiencies and not all simulations are precise enough compared to reality. However these deviations are usually caused by external factors like unpredicted obstacles or changes to the team.

Future works should lead to design of the neural networks for classification and pattern recognition in the context of an agile software process.

Recent publications.

1. Štrba, R., Svatopluk, Š., Jakub, Š., Košinár, M.,: Intelligent software support of the SCRUM process; In Proceedings of the 24rd European-Japanese Conference on Information Modeling and Knowledge Bases; Editors: Yasushi Kiyoki,Takehiro Tokuda; 2014, Kiel, Germany
2. 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
3. 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

Other references.

4. Marc I. Kellner, Raymond J. Madachy, and David M. Raffo. Software process simulation modeling: Why? What? How? *The Journal of Systems and Software*, 46(2–3):91–105, April 1999.
5. Heaton Jeff. *Introduction to neural networks*, ISBN: 1-60439-008-5
6. Beer M, Spanos P.D. *Neural network based Monte Carlo simulation of random processes*, 2005 Millpress, Rotterdam, ISBN 90 5966 040 4
7. Vondrák, I.: *Neural networks (czech)*. VŠB-TUO, Czech Republic, Ostrava, revision 2009.
8. Madachy RJ (2008) *Software Process Dynamics*. 2nd edn. Wiley-IEEE Press.
9. Scacchi W (1999) Experience with software process simulation and modeling. *Journal of Systems and Software* 46 (2-3):183-192.
10. George S. Fishman. *Discrete-Event Simulation: Modeling, Programming, and Analysis*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag, Berlin, 2001.
11. J. Highsmith and A. Cockburn. Agile software development: the business of innovation. *Computer*, 34(9):120–127, 2001

Kaira: Analysis and Verification

Martin Šurkovský and Ondřej Meca

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

Abstract. The research of a group which the authors are members is focused on simplifying the process of developing parallel programs, especially programs intended to systems with distributed memory. The process of developing programs – not necessarily parallel programs – is not only writing a code. During this process a programmer usually comes across some other tasks such as debugging or profiling. These activities are also common in the field of parallel programs and because of their complexity some new ones may appear. This article is focused on two specific activities: analyses of data obtained from programs runs and formal verification of parallel programs. Both of them have been implemented into a tool which we are developing, called Kaira.

Keywords: Parallel programs, MPI, visual programming, analysis, verification, state-space exploration

1 Introduction

Every year we can see growing use of parallel computers. As technology is developing these computers are more and more common. From huge supercomputers to smaller clusters using quite a big number of processors, or even multi-core processors inside our current personal computers and/or mobile devices. Creating programs for such a wide range of devices require an appropriate amount of programmers.

In comparison with developing sequential programs, programming the parallel ones is much more difficult. Firstly, parallelism brings a new sort of software bugs, e.g. race conditions. Synchronization and communication are also new tasks with which a programmer must count during the design of their parallel programs. Secondly, it is important to realize that the process of developing programs is not only writing a code. It consists of other activities such as finding bugs (*debugging*), tuning a performance or memory usage (*profiling*), and so forth.

Our research is focused on simplifying the process of developing parallel programs. We want to build a unifying prototyping framework for creating, debugging, analyzing, and formally verifying parallel applications. Hence, a programmer can easily implement their ideas in a short time and experiment with them, create a real running program, and verify its performance and scalability. The

presented ideas are implemented in Kaira¹, an open source tool which we are developing.

The key property of Kaira is strict separation of communication and computational parts of developed programs. While a computational parts are written in standard programming language (C/C++), the communication part of programs is described visually by a defined visual language. The visual language is based on Coloured Petri nets [4] – a standard formalism for modeling and analysing concurrent systems. The interconnection of communication and computational parts is mostly done by inserting of a source code, written in C/C++, into transitions. Such inserted code is executed whenever a transition is fired. More information about semantics of the visual language and its interconnection with a C/C++ code can be found in [2].

This paper is about some of the *supportive activities* which Kaira covers. Currently, Kaira supports the following four supportive activities: *visual debugging*, *performance analysis*, *performance prediction*, and *formal verification*. There will be discussed two of them, a new part of performance analysis and formal verification, so the structure of the paper is divided into two main parts. The first part deals with analyses of parallel programs runs. There is introduced an API which serves for extending the tool about new functionality. Then those new functions may be used just for such analyses. The second part is focused on a verification framework. Parallel programs are non-deterministic in general, hence standard testing methods do not guarantee a full coverage of all possible behaviors even for fixed input. In order to guarantee the correctness of a program, one has to systematically explore an entire program's state space.

This paper is based on the joint work with Stanislav Böhm and Marek Běhálék that was published in article [3] and will be published in journal paper [1] and article [5].

2 Analyses

As was said analyses of parallel programs runs are included in the part of performance analyses. They are realized by a series of operations that can be added through an API which serves for extending Kaira about new features. Kaira, very early after its creation, has been extended about the tracing mode, so generated applications could have started record some events from their runs. Firstly, a visual debugger was introduced which uses a data from traced versions, since then it is possible to replay the run of an application. Secondly, there has been added a possibility to present the performance data graphically such as showing a utilization of processes in time. These activities are common for all applications created in Kaira, but with a possibility to record also some other data (e.g. values of some variables in time) there has appeared a requirement to process them somehow. Because, those data are related to a specific application it have not made sense to add a special chart/analysis case by case. Something more general must have come, and this is a new API for so-called *user operations*.

¹ <http://verif.cs.vsb.cz/kaira>

User operations represent more or less complicated functions that may be combined together to achieve a desired result. They are based on a principle that one operation takes some data and produces a new ones which can be used by other operations or by the original operation itself. Of course, there can be created operations that either do not take any input data or do not produce any outputs. These inputs and outputs are represented as *sources* – non trivial data like tables or tracelogs – different kinds of data formats. Therefore, if a new operation is needed a programmer follows only a few simple rules, create a new operation, and the whole integration into Kaira is performed automatically. The same thing holds for adding a new data format – a new source. An example of operation usage can be seen in Figure 2.

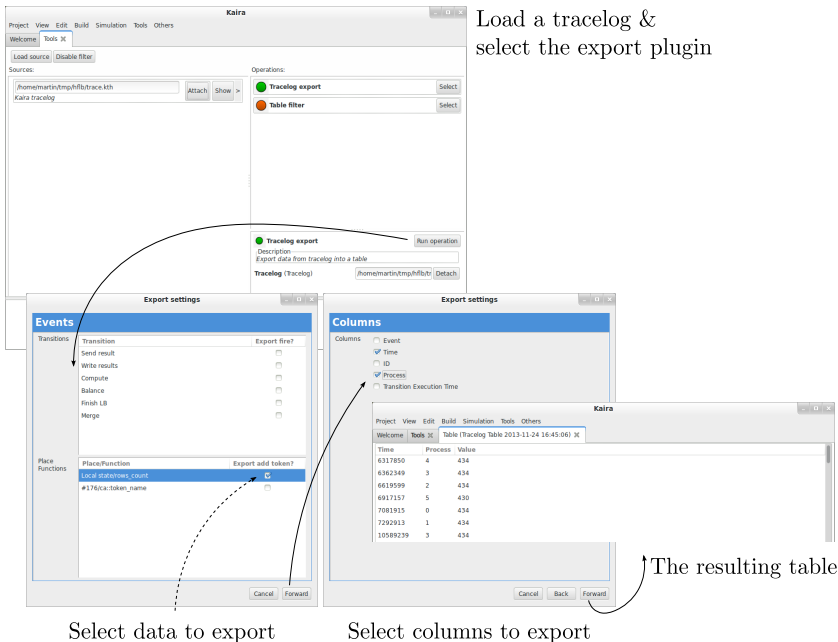


Fig. 1. User operations, the process of exporting data from a tracelog

The basic concept of operation were adopted from tool PROM². This tool is intended to process mining, so it provides a lot of functions and data formats for dealing with problems in this area. What is interesting for us is their simple graphical user interface. It is able to combine all these functions and data formats altogether, so they are able to cooperate. In fact, PROM provides merely this interface, most of its functions are work of third parties. It has been a great inspiration for us, because it represents almost limitless way for making new extensions which can collaborate with those existing.

² <http://www.promtools.org/prom6>

At this time, Kaira supports two (stable) operations intended to export data from tracelogs and their filtering. There is also a prototype of an operation³ which tries to interconnect our tool with R⁴ – a statistical tool which provide a variety of analyses and data visualization. Currently, this operation supports only drawing an x-y chart, but for proving the concept it is enough.

3 Verification

This section introduces verification tools offered a state-space exploration and briefly describes their pros and cons. For tools that are publicly accessible, there is provided a simple performance comparison and it is shown that the approach of Kaira is comparable with others, although our implementation is still in a very immature state.

3.1 Verification tools

Historically the first verification tool in the area of MPI applications was MPI-SPIN⁵. It can check whether a program has a property that can be described by a temporal logic formula. However, the input of MPI-SPIN is a model in PROMELA⁶. Therefore, a user has to provide and maintain the model of the application and verification is performed on this model.

The input of the next tool named TOOLKIT FOR ACCURATE SCIENTIFIC SOFTWARE⁷ (TASS) is a source code written in C with MPI calls. It eliminates problems of creating and maintaining a model. But, TASS is not able to run C++ applications and it does not support non-blocking operations.

A tool IN-SITU PARTIAL ORDER⁸ (ISP) accepts applications written in C/C++ and supports almost all MPI functions. ISP verifies a program only with respect to a used implementation of MPI, not the full standard of MPI. Because of the usage of *stateless* analysis ISP is not able to detect a presence of a live lock. Similar to ISP is DISTRIBUTED ANALYZER OF MPI⁹ (DAMPI). The main advantage of DAMPI is ability to perform a verification in parallel in a distributed memory system. Other features and limitations remain the same as for ISP.

A recent tool in this area is MPISE. MPISE offers symbolic execution and verifies directly C/C++ programs. However, MPI functions are restricted only to blocking variants. Unfortunately, DAMPI and MPISE are not publicly available; therefore, they are not included in the benchmark section.

³ This operation is available from: <https://github.com/sur096Kaira/kaira.git>, r-chart branch

⁴ <http://r-project.org>

⁵ <http://vsl.cis.udel.edu/mpi-spin/>

⁶ <http://www.dai-arc.polito.it/dai-arc/manual/tools/jcat/main/node168.html>

⁷ <http://vsl.cis.udel.edu/tass/>

⁸ http://www.cs.utah.edu/formal_verification/ISP-release/

⁹ http://www.cs.utah.edu/formal_verification/DAMPI/

3.2 Comparison with Kaira

This section presents a comparison of verification tools on two MPI programs that contain common patterns that occur in distributed programming. In both programs, a deadlock analysis was performed. Resulting times are shown in Table 1. All experiments were executed on Intel i5-3570 3.40GHz x 4 with 8GB RAM.

Table 1. Running times of verification tools for selected examples. The time limit was 3600s.

	# p	MPI-SPIN	TASS	ISP	Kaira
workers (5 jobs)	3	0.131s	runtime error	18.758s	0.001s
workers (5 jobs)	4	8.459s	runtime error	560s	0.037s
workers (10 jobs)	3	0.287s	runtime error	1040s	0.113s
workers (10 jobs)	4	18.966s	runtime error	time exceeded	12.467s
heat flow	2	0.025s	not supported	0.421s	0.006s
heat flow	4	5.979s	not supported	0.470s	0.010s
heat flow	6	time exceeded	not supported	0.592s	0.016s

The first example, *workers*, represents a standard problem where one “master” node divides jobs to “slave” nodes. The second example, *heat flow*, represents other standard problem of spreading the heat on a cylinder by an iterative process. The detailed information for both examples can be found in [2] sections: 4.1 and 4.2. Applications for ISP and TASS has been prepared in a form of standard source code. The same computation code uses the versions for Kaira, however the communication part have been programmed in the visual language. For MPI-SPIN have been created only a PROMELA model of communication without computational parts.

We have not been able to verify any of these programs in TASS. Firstly, because of runtime error in workers example and secondly because of non-blocking operations used in heat flow example. From results, it can be seen that both, MPI-SPIN and ISP, exceed the time limit (one hour) even for a small number of processes. Kaira has verified both programs with all settings and with the best execution times.

The state-space size grows exponentially. Hence, only a small example could be checked by an exhaustive exploration. Fortunately, there are *partial order reduction* (POR) methods that may reduce the state space (e.g. [6]) with preserving checked properties. Therefore, it is not necessary to explore the entire state space.

Some variants of POR methods are used by all mentioned tools. In specific examples (e.g. heat flow) the state space is significantly pruned and verification can be performed on large instances of programs. However, MPI standard is too complex and despite the POR methods there remains examples (e.g. manager-workers) where verification is manageable only for small inputs. We hope that Kaira’s approach can be helpful in this situation, because of its abstract computational model. Even though programs created in Kaira are translated into MPI,

the semantics of Kaira is different and it does not cover the full semantics of MPI. This computational model is easier to verify and provides more opportunities to reduce the state space.

4 Conclusion

This text introduced some of the new functions implemented to our tool Kaira. These functions thus extended the set of already existing supportive activities. We believe that the coverage as much supportive activities as possible by a one abstract computational model might be the manner of simplifying the process of developing parallel programs.

The new API for extending Kaira about new functionality represents an extension of performance analysis part. In fact, its potential is bigger and in the future it may roof more activities. On the other hand, it is the fact that current version does not support many operations, but we are working on them. We want also to improve the environment of operations about possibility to make a new operation by composition of those already existing.

Verification is a completely new feature which has been added. Currently, this is a hot topic which is intensively developed. Despite its immature state it can be compared with existing tools. In the near future we are going to parallelize the algorithm. Moreover, we would like to extend the rules of partial order reduction algorithm which are specific to our visual language in order to reduce the state space more. Thus, we could verify bigger instances of program.

References

1. Böhm, S., Běhalek, M., Meca, O., Šurkovský, M.: Visual Programming of MPI Applications: Debugging, Performance Analysis, and Performance Prediction. *Computer Science and Information Systems* (2014), to appear
2. Böhm, S.: Unifying Framework For Development of Message-Passing Applications. Ph.D. thesis, FEI VŠB-TUO Ostrava, 17. listopadu 15, Ostrava (11 2013), <http://verif.cs.vsb.cz/sb/thesis.pdf>
3. Böhm, S., Běhalek, M., Meca, O., Šurkovský, M.: Kaira: Development Environment for MPI Applications. In: Ciardo, G., Kindler, E. (eds.) *Application and Theory of Petri Nets and Concurrency*, Lecture Notes in Computer Science, vol. 8489, pp. 385–394. Springer International Publishing (2014)
4. Jensen, K., Kristensen, L.M.: *Coloured Petri Nets - Modelling and Validation of Concurrent Systems*. Springer (2009)
5. Meca, O., Böhm, S.: Verification of MPI Applications. In: *12th International Conference of Numerical Analysis and Applied Mathematics*. Rhodes, Greece (2014), to appear
6. Valmari, A.: Stubborn sets for reduced state space generation. In: Rozenberg, G. (ed.) *Advances in Petri Nets 1990*, Lecture Notes in Computer Science, vol. 483, pp. 491–515. Springer Berlin Heidelberg (1991), <http://dx.doi.org/10.1007/3-540-53863-1.36>

Automatic Detection and Segmentation of the Carotid Artery in Ultrasound Images

Jan Tomeček and Radim Farana

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

Abstract. An ischemic stroke is the third most frequent cause of death and the most frequent reason for disability of our population. Then its treatment is very important, but so far, there has not been discovered any convenient therapeutic method to treat this disease. In this case, prevention and related monitoring of the disease development plays so important role. The best indicator of the progress of the disease appears to be the arteria carotis communis. Due to a large expansion of ultrasound diagnostic techniques and possible repeatability, this method is currently being used to monitor the disease. Images, however, are significantly affected by noise, and therefore their correct evaluation by the physician can be difficult. The aim of this work is to design and implement software tools that would reliably replace the picture analysis and allow doctors to quickly and effectively measure the size of the carotid artery constriction.

Keywords: active contour, picture analysis, atherosclerosis, FOTOM^{NG}, gradient vector flow, Hough transform, threshold, speckle noise, image processing.

1 Introduction

A stroke is the third most frequent cause of death. Specifically an ischemic stroke represents the largest group of this kind of diseases. Despite all the advances in medical therapeutic methods, no methods that would reliably reduce mortality from this disease have been found. Prevention is still the most significant way to combat this disease. As the frequent cause of an ischemic stroke are atherosclerotic plaques in the carotid artery, its exploration can help to determine the development of disease.

Our goal was to adapt and reconstruct a carotid artery ultrasound image acquired on an acquisition unit for automatic scanning of carotid arteries and to create a 3D model. In addition, to create tools which will be able to automatically detect the artery in the image and perform segmentation for a quantitative analysis of its potential constriction due to atherosclerotic plaque.

2 Picture acquisition

Carotid artery picture acquisition is realized with a positioning device, developed at the VSB-Technical University of Ostrava between years 2006 and 2008. The device is based on automatic linear ultrasonic probe movement, during which a cross sectional carotid artery video signal is recorded. The base part of the movement mechanism is fixed to the patient's bed. Automatic movement of the probe within the range of 2-3

cm is realized by a stepping motor. One step moves the probe by 0.025 mm. In terms of positioning accuracy of the ultrasonic probe (considering the will of the thread), we have achieved accuracy better than 0.5 mm.

An important factor for correct picture scanning is the carotid deformation due to heart activity. This situation complicates the subsequent reconstruction of the carotid arteries from images. Patient's ECG is measured and scanned to eliminate this influence. According to the measured ECG, all images are taken at the same part of the heart rhythm. This procedure is shown in Fig. 1.

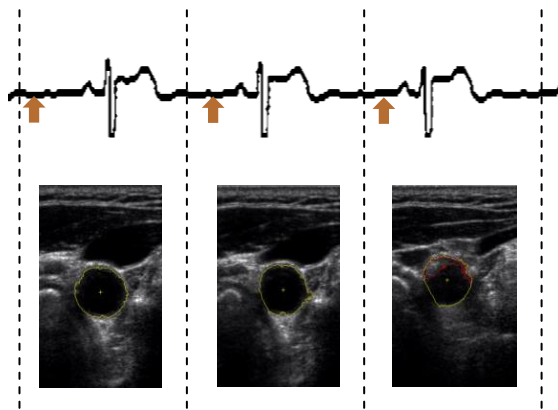


Fig. 1. Data acquisition depending on ECG signal

3 Pre-processing and image reconstruction

3.1 Image pre-processing

Despite the fact that the ultrasonograph determines the dynamic range, and due to the amplification of all incoming echoes (Gain) and TGC, the analyzed image brightness and contrast is very low. One of the first steps of picture pre-processing is to adjust the image contrast. Although the contrast adjustment is used mainly before the image visual assessment, image adjustment with help of the CLAHE method (contrast limited adaptive histogram enhancing)[1] led to significantly better results in the final arteries detection.

CLAHE is a method for histogram equalization, which prevents noise highlighting, typically caused by classical histogram equalization methods.

Next pre-processing step is filtering to suppress speckle noise. Many filters were tested, for example SRAD [2], anisotropic diffusion, Mean shift, etc. The best results were obtained by SRAD both under subjective evaluation as well as by SNR and PSNR. Good edge retention was seen after the use of SRAD or Mean shift filters. But characteristics in the following picture processing are more important. It has been experimentally discovered that the best is to use a classical Median filter, especially for subsequent edge detection and Hough transformation. Hough transform has been chosen for artery recognition in the picture, especially Hough transform for circle detection. Although the artery does not always has a circular shape, in comparison

with a generalized Hough transform, the results of artery detection obtained by the HT circle detection were much better and with less computational complexity since the accumulator had only three dimensions.

After finding the maximum in the accumulator (Fig. 2), we work with information about the radius of the found circle and its center.

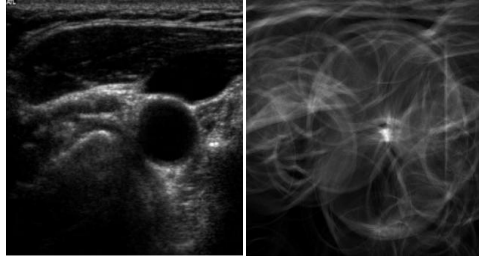


Fig. 2. original image left, accumulator for Hough transform given radius right

3.2 Image reconstruction

An artifact of an acoustic shadow is shown in most of the analyzed ultrasound artery images, particularly in the presence of calcified plaque at the top part of the artery. In this case, we miss information about a large part of the artery wall, and this situation results in unwanted contour development outside the region of interest.

. Proposed method for the missing artery part reconstruction uses only the current image. An algorithm that converts the edge image in polar coordinates, when the middle of the found circle serves as the beginning point, is used in this case. Minimum distance for a given angle is only applied.

Two values are subsequently selected, one as the threshold for maximum distance, when the artery is not completed, and also the distance for adding new pixels, if no artery side is found for a given angle. The first value was experimentally set to $t = r + 4$, where r is the found circle radius in pixels. The second value was set to r .

The next step is to prepare the image for parametric active contour segmentation. The image after the median filter application is cropped around the found centre to maintain only the artery surrounding for further processing. Then an adaptive iterative threshold method is applied, in this time with a modified threshold boundary shifted by brightness value 10 towards higher intensity values. This value was also set up experimentally. Subsequently, the edges are detected.

4 Segmentation

Inner artery side detection is realized by a parametric active contour [3]. Two-dimensional parametric active contours, or deformable objects, are primarily used for interactive segmentation.

.Contour deformation takes place in order to minimize the functional:

$$E = E_{extern} + E_{intern} = E_{extern} + \int_0^1 \left(\alpha(p) \left| \frac{\partial \mathbf{s}(p)}{\partial p} \right|^2 + \beta(p) \left| \frac{\partial^2 \mathbf{s}(p)}{\partial p^2} \right|^2 \right) dp \quad (1)$$

Minimizing equation (1) goes to Euler-Lagrange equation:

$$\frac{\partial \mathbf{s}}{\partial t} - \alpha(p) \frac{\partial^2 \mathbf{s}(p)}{\partial p^2} + \beta(p) \frac{\partial^4 \mathbf{s}(p)}{\partial p^4} + \frac{\partial E_{vnějsí}}{\partial p} = 0(2)$$

whose iterative numerical solution leads to sequential deformation and thus the object segmentation in the image.

In 1997 Xu and Price [4] suggested defining a new external force, the gradient vector flow (GVF), which is calculated from the edge image.

The following functional based on energy minimization is used to determine this field:

$$E_{gvf} = \iint \mu \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 \right] + |\nabla f|^2 |\mathbf{v} - \nabla f|^2 dx dy \quad (3)$$

where $\mathbf{v}(x, y) = (u(x, y), v(x, y))$ is GVF, f is the edge image, μ is the weighting factor of the first term in the functional and $|\nabla f|$ is the weighting factor of the second term in the functional. Minimizing the functional (3) using variational calculus leads to the following Euler-Lagrange equations determining the minimum conditions:

$$\mu \nabla^2 u - \left(u - \frac{\partial f}{\partial x} \right) \left(\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right) = 0(4)$$

It is possible to see that it is a generalized diffusion equation. If at some point the gradient and thus the entire second term is zero, the equation goes to form isotropic diffusion.

The calculated contour parameters are set to $\alpha = 1$, $\beta = 0$, the number of iteration in GVF calculation is 50 and μ is 0.1. The initial contour is placed into the found HT centre and its radius is 25 pixels.

The result of the segmentation is a contour that defines the inner side of the artery. When determining the faultless artery outline, it is then possible to determine the content and size of the plaque by the difference between the found outline of the artery and the contour (Fig. 3).

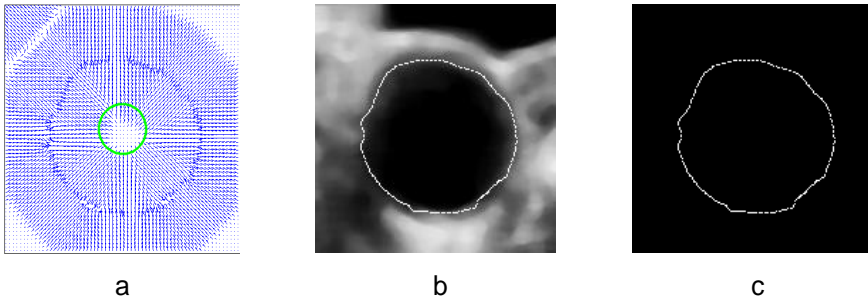


Fig. 3. a) GVF field with initial contour in green colour, b) the resulting contour in the image, c) resulting contour only

5 Results

5.1 Artery detection results

The 1100 images obtained from videos taken by the acquisition unit were tested. The artery was correctly found in 870 images. The artery was identified incorrectly in 230 images.

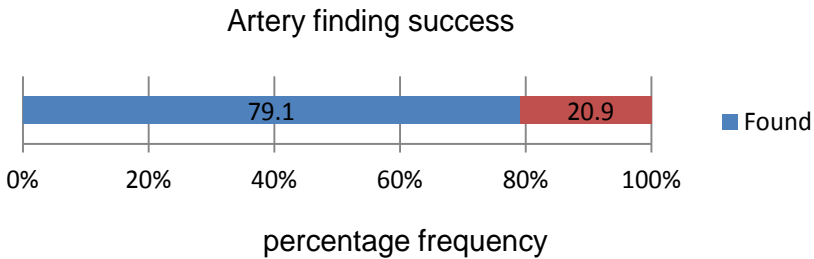


Fig. 4. Sample result of the proposed algorithm for the artery detection and segmentation

Proper detection was significantly dependent on the time of testing, the detection success dropped down with time, and thus with the probe shifting. Given the character of the test images, it is obvious that this happened because the ultrasound probe put out of tune. At the beginning of the measurement, the probe is positioned by a doctor so that the artery is the most pronounced in the image. In these cases the detection success rises up to 100%. Even the presence of large calcified sclerotic plaques does not often lead to greater error detection. Over the time as the probe is moved automatically, the detection capability of the algorithm degrades with degrading visual quality of the object.

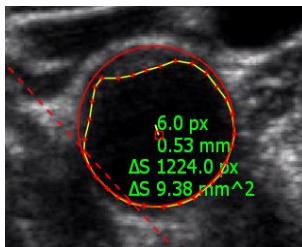


Fig. 5. Quantitative evaluation example in FOTOM^{NG} system

5.2 Artery segmentation results

Segmentation results were compared with manually obtained artery sides. They were marked out directly by a doctor, or based on their training in image processing. Static ultrasound picture evaluation is very subjective in many cases. It is often impossible to differentiate between the plaque and the noise, the obtained result rather depends on doctor's experience. Another complication is the fact that during a classical examination realized by a doctor, the parameters of the ultrasound probe could be changed, such as amplification (gain), frequency, etc. This is impossible when a static record is processed. Then we must work on limited information from the initial setup values. Despite these facts, segments considered as correct (reference) to measure the deviation of automatic segmentation were manually marked. The area that does not overlap with the reference regions, was measured in mm². Although it might not be only the plaque, these deviations were observed, mainly because FOTOM allows to easily measure the area based on the difference between two object areas. It means the area of the contour of the artery side and inside the artery.

It is evident that using the semi-automatic detection, when the user manually marks the area of the artery side, more accurate results can be significantly achieved. An average value of the defective marked area is 2.816 mm² with variance 2.593 mm⁴ for automatic detection, whereas for the semi-automatic detection the average value is 1.105 mm² and variance 0.585 mm⁴.

6 Literature

1. Pizer, S.M., Amburn, E.P. and Austin, J.D.: Adaptive Histogram Equalization and Its Variations. *Computer Vision, Graphics, and Image Processing*. 1987, č. 39, pp. 355-368.
2. Yongjian, Y. and Acton, S.T.: Speckle reducing anisotropic diffusion. *IEEE Trans. on Image Proces.* 2002, year 11, volume 11, pp. 1260-1270.
3. Kass, M., A.Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331, 1988.
4. Xu, C. and Prince, J. L.: "Gradient vector flow: A new external force for snakes," in *IEEE Proc. Conf. on Computer Vision and Pattern Recognition*, 1997, pp. 66-71.

7 Publications presenting the achievements:

1. Ličev, L., Tomeček, J. a Farana, R.: ADJUSTING THE INPUT ULTRASOUND IMAGE DATA AND THE ATHEROSCLEROTIC PLAQUE DETECTION IN THE CAROTID ARTERY BY THE FOTOMNG SYSTEM Impact Factor (2012): 0,622, ISSN 1310-2818, Journal title: BIOTECHNOLOGY & BIOTECHNOLOGICAL EQUIPMENT, Pages XXX-XXX, Journal Country/Territory: BULGARIA, Publisher: DIAGNOSIS PRESS LTD, Publisher Address: 67 DONDUKOV BLVD, 1504 SOFIA, BULGARIA, Impact Factor (2012): 0,622, subject categories: BIOTECHNOLOGY & APPLIED MICROBIOLOGY, počet časopisů v oboru: 160, pořadí: 137
2. Ličev, L., Tomeček, J. a Šivic, J.: Systém FOTOM-NG, Architektura, funkce a použití. In Sborník referátů XX. konference SDMG a IGDM, VŠB TU Ostrava. Milin, 2-4.10.2013, vol. XX, čís. 1, p. 117-132, XX. konference SDMG, ISBN 978-80-248-3190-9
3. Ličev, L., Hudeček, T. a Tomeček, J.: Rozpoznávání objektů na snímcích v hornictví v systému FOTOM-NG. In Sborník referátů XX. konference SDMG a IGDM, VŠB TU Ostrava. Milin, 2-4.10.2013, vol. XX, čís. 1, p. 133-142, XX. konference SDMG, ISBN 978-80-248-3190-9 Editor: Ing. R. Dandoš
4. Ličev, L., Kuzník, R. a Tomeček, J.: Analýza snímků za účelem vytvoření kvalitního snímku pro následně vyhodnocení. In Sborník referátů XX. konference SDMG a IGDM, VŠB TU Ostrava. Milin, 2-4.10.2013, vol. XX, čís. 1, p. 143-152, XX. konference SDMG, ISBN 978-80-248-3190-9 Editor: Ing. R. Dandoš
5. Ličev L., Babiuch M., Farana R. a Tomeček J.: Software Analysis of Mining Images for Objects Detection: Acta Montanistica Slovaca, Košice, 18/2013, str. 59-66, 8 str., ISSN 1335-1788, Impact Factor (2012): 0,094, Scimago Journal Ranking (2011): 0,182, Hirsh Index: 6. Subject categories: GEOSCIENCES, MULTIDISCIPLINARY, počet časopisů v oboru: 172, pořadí: 4, (Editor-in-Chief: Igor Podlubny).

Visualisation of volumetric data on the GPU

Vojtěch Uher

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

Abstract. Today, volumetric data are increasingly encountered. They are primarily obtained through medical equipment or industrial scanners. This paper summarises our method, which we implemented, of displaying medical data, and evaluates its performance. The integrated renderer is based on the principle of direct volume rendering (DVR) and uses modern Monte Carlo techniques for a more efficient run. The DVR algorithm is a rather complex calculation-intensive task. A key subtask is traversing the ray through the volume which determines the speed of the algorithm. The paper includes a comparison of the Ray Marching and Woodcock tracking methods of traversing, describing their parallelisation on the GPU using the CUDA architecture.

Keywords: Direct volume rendering, Volumetric Ray Casting, CUDA, Ray Marching, Woodcock tracking, Global Lighting Model

1 Related Work

In direct volumetric rendering, mainly procedures based on Levoy's [4] or Drebin's (see e.g. [8]) integration algorithm are used, which accumulate the colour of the collected samples along the rays with transparency. Today, there are several algorithms for volume ray traversing. Usually, methods of direct tracing of intersected voxels are used, such as 3DDA [2] or the Ray Marching sampling method [7]. Ray Marching (RM) samples the volume by constant increments of the ray line parameter. Apart from RM, there is also a stochastic method of Woodcock tracking (WT) described e.g. in paper [7], which samples the environment based on a calculated distribution function. An example of Monte Carlo (MC) ray casting utilising WT is the renderer described in paper [3]. For faster passages through large empty or homogeneous spaces, tree structures of sub-spaces are often used.

2 Our algorithm

Our Monte Carlo ray casting includes the following steps: 1) Loading the data and creating data representation, 2) Creating the camera, 3) Traversing ray through volume, 4) Transfer function, 5) Calculation of shading and lighting model, 6) Integration of values along the ray

Volumetric data can be taken e.g. from CT or MRI imaging. The images represent slices through the human body showing densities of the materials. Other variables (attenuation, components of the lighting model) must be assigned to densities using the transfer function. Data representation is based on the edge model. Voxels thus represent values at the vertices of a grid and the entire volume is described by cells formed by eight neighbouring voxels (the value inside is determined by interpolation). During data loading we calculate a minimum bounding box and a maximum attenuation of the volume (needed later). As a camera we use a standardized pinhole model (described e.g. in [8]). Aliasing which arises with discrete data is reduced by the super-sampling (SS) algorithm. The Global Lighting Model (see [6]) used in the application is an extended Phong's lighting model. The classic Phong's model prefers solid shiny surfaces and is not unbiased. Our model is therefore composed of a spot light which contributes to sharp reflections and a planar emitter simulating ambient light. The surface of the emitter is sampled by the IS method (pap. [1]) based on the Gaussian distribution. The resulting image will then appear more vivid and realistic. The integration of sampled values is done in FTB manner. The resulting image is adjusted by gamma correction.

2.1 Traversing ray through volume

In this section we discuss some possibilities of the ray traversing through the volume. Due to its complexity, passage through the volume is the most demanding of the entire renderer's operations. There are several major problems:

- we want to avoid accessing the sites through which the ray does not pass
- passage of large empty spaces and homogeneous areas
- numerical errors

As mentioned above, basically, two families of algorithms are available. We need a rapid display of data and the possibility of easy parallelisation on the GPU. Therefore, we preferred sampling methods whose implementation is straightforward and which provide a high quality image.

Sampling methods We sample the ray parameter within the bounding box. The sample is localised in the cell grid and its value is obtained. In the application we tested: Ray Marching (RM) and Woodcock tracking (WT). For constant stepping of RM, we assume that the value between two samples keeps being the same. RM is suitable for fast data display, without special graphic effects. When more advanced shading models are used, the time of calculation increases significantly. RM also does not take account of the data character and does not address passages through empty and homogeneous spaces. But all the problems mentioned are solved by WT. Unlike RM, this is an unbiased MC sampling algorithm which passes a volume by stochastically determined jumps (pap. [7]). Materials with a high degree of opacity are preferred, but at the same time even less distinctive materials are contained in the result. The next step for WT

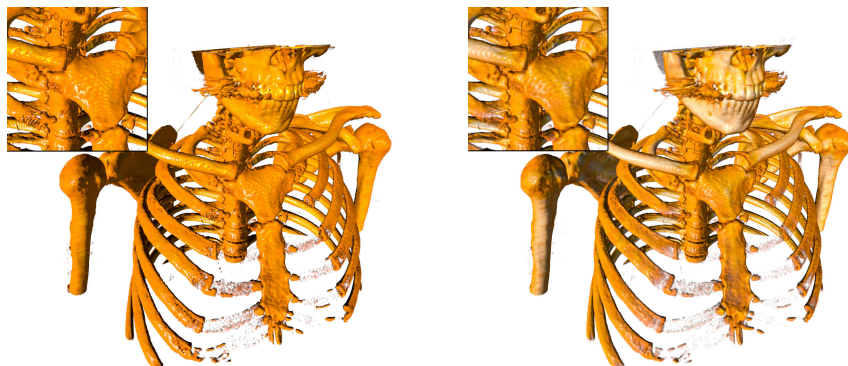


Fig. 1. Comparing the results achieved by the RM (left) and WT technique (right). With WT we obtained an image with fewer artefacts in half of the time (Chest dataset).



Fig. 2. Achieved image quality on the cthead dataset. Left: RM method without shading. Center: WT method with shading displaying softer tissues. Right: WT method with shading displaying areas with good normals.

is calculated as $t = (-\ln(1-r)/A_{max})$, where A_{max} is the maximum attenuation value of the data set and r is a normalized random number. The proposed sample is then accepted with the probability At/A_{max} , where At is the attenuation factor of the current sample. Compared with the earlier mentioned methods, WT is far easier to implement, does not require extensive preprocessing and comprehensively addresses most of the issues discussed. WT reduces the number of samples that need to be computed, thereby saving time. Typically, WT is used for stochastic determination of a single representative sample along the ray. This method is thus quick but produces extremely noisy image. Furthermore, displaying more layers of a volume with transparency poses problems. However, in our implementation, as in the case of RM, we are looking for several samples across the volume by repeated WT tracing. The calculation is thus still faster than RM, while producing less noisy image than when determining a single

scattering point. Noise is reduced by SS. Taking into account all the associated calculations (shading, calculation of colour, casting shadows, etc.) which are done in vain in RM, then WT is still a better alternative. Section 3 deals with the comparison of the performance of both methods. The image differences between both methods can be seen in fig. 1

3 Parallelisation on the GPU

In this section, we describe our particular way of implementation on the CUDA architecture. We developed two implementations and compared them. The parallelisation poses several problems:

- data sets take up quite a lot of memory and must be uploaded to the GPU
- reading order of voxels from the memory is view-dependent
- the numbers of samples per ray vary
- the entire rendering pipeline uses a large number of registers
- generating of random numbers on the GPU

Several acceleration techniques were developed for parallelisation on the GPU. Paper [5] describe bricking and slab-based techniques. A disadvantage of these algorithms is that they themselves are very complex. We decided to use linear sampling techniques whose implementation is far easier.

3.1 Basic procedure

The maximum resolution of the tested data was $512^2 \times 220$. All the data sets can therefore be seamlessly stored in the GPU global memory. The camera parameters, scenes and pointers to global memory are stored in the constant memory. In the first case, we easily paralleled the whole Ray Casting. The image is divided into sub-images which correspond to individual kernel calls (these are further divided into 2D blocks). Each thread calculates one ray. The entire pipeline of one ray is thus solved simultaneously. The colour accumulator is a local variable. We generated random numbers using the cuRAND library. Each thread maintains its own state of the random generator in the device memory. A single initialisation of the cuRAND is performed in advance using a special kernel. Values of the super-sampling sub-rays are stored in the shared memory (SM). Finally, they are averaged and stored in the global memory of the image. In this case, each SS sub-ray corresponds to one thread. The final phase therefore requires the block synchronisation.

3.2 Optimisation

In the optimized implementation some changes significantly accelerated the calculation. The CUDA architecture on today's graphics cards typically supports 3D textures with the 2048^3 resolution (enough for datasets), they are cashed by blocks (this speeds up the gradient calculation too) and they support hardware

interpolation, which is fast and reduces number of registers. Given the view-dependent random accesses for RC, these properties come useful. To reduce the use of the local memory, we divided the calculation into multiple kernels. First, the rays and their intersections with the bounding box are pre-calculated. In the next kernel, these pre-calculated data are loaded into the local memory and used to trace the volume. To simplify the procedure, the SS is calculated at the level of the entire frame rather than a single ray. The results are accumulated in the image memory and finally averaged. The advantage of this approach is that SS can be virtually arbitrary (not bound by the block size), the shared memory is not used, and synchronisation is not done. The MC ray casting is frequently working with the random generator. It is therefore advantageous to upload statuses into the SM at the start of the calculation and then store them back after its completion.

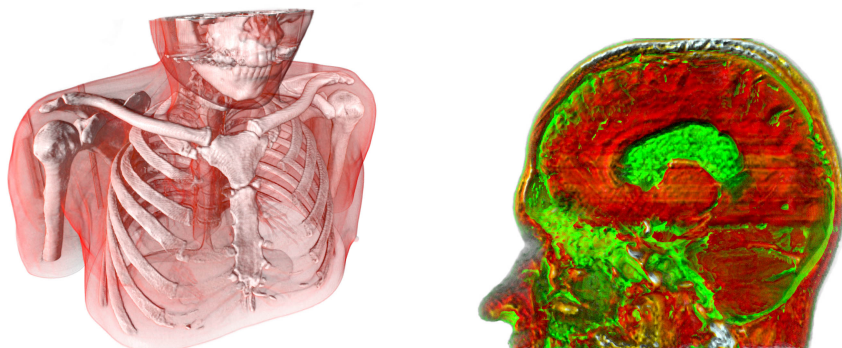


Fig. 3. Left: Display of multiple layers of volume over each other with transparency (Chest dataset). Right: Brain section with ventricles highlighted.

4 Measurement and results

Graphs 4 show a comparison of calculation times for RM and WT depending on SS and a comparison of the basic method and the method we optimised. WT in our case calculates several samples along the ray to achieve a standard accumulation effect. The basic step of tracing was the same in all cases of measurement. The WT reduces the number of samples, which can be seen mainly in cases where a lighting is used (dashed curves). But random tracing leads to greater divergence in the warp. However, applying the optimizations, WT performs better than RM. RM is actually faster only in the simple case of accumulation of colour (solid curves). To generate real time pre-views, RM is more easily parallelisable. In the cthead data set and the res. 1024x768px, we achieved about 27 FPS (see

Fig. 2, left). By optimising the RM method, we achieved a 10 to 40 times better performance, depending on the data set. For WT, the acceleration is about fivefold. Pictures 1, 2, 3 show the achieved quality of visualisation. Images with no indication otherwise are rendered using WT. To achieve a finer shading, one spot light and one planar light is used in the scene.

The measurement was performed using the following configuration: Intel Core i5 760 (4 x 2.8GHz), 16 GB RAM, Geforce GTX 460 (1 GB GDDR5), Windows 7 64-bit.

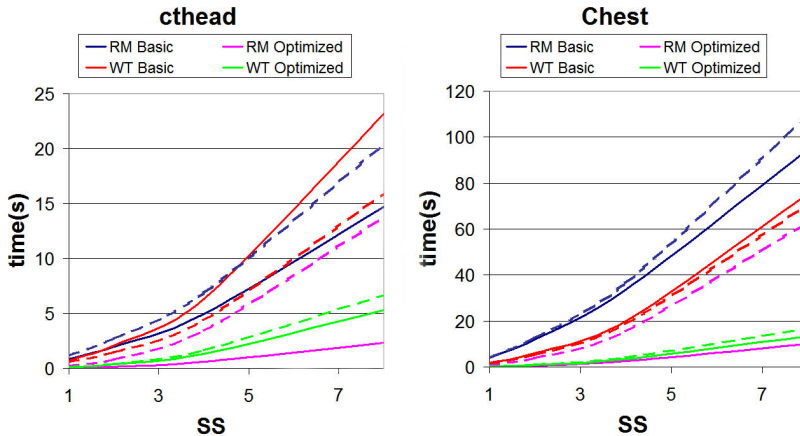


Fig. 4. Comparing calculation times for RM and WT, with shading (dashed) and without shading (solid) on the data sets cthead (256x256x99) and Chest (512x512x220).

5 Conclusion

We introduced a stochastic algorithm, which we implemented, used for the calculation of photorealistic visualisation of volumetric data. We showed how we can efficiently parallelise Volumetric Ray Casting on the CUDA architecture. We verified that the texture memory is clearly more convenient to store the volume slices than the global memory and that by limiting the use of local memory and registers the occupancy of GPU significantly increases. We tested and compared the options of parallelising the Ray Marching and Woodcock tracking sampling algorithms. We found that the RM is useful only for fast and simple visualisation, while WT effectively reduces the number of samples and related calculations along the ray. WT is clearly suited for photorealistic imaging of volume. With the WT multiple sampling, we achieve an accumulative effect with transparent materials more efficiently than using the RM algorithm. The improvements in parallelisation that we implemented enable lower resolution real-time rendering, even when using standard graphics cards.

References

1. David Cline, Parris K. Egbert, Justin F. Talbot, and David L. Cardon. Two stage importance sampling for direct lighting. In *Proceedings of the 17th Eurographics Conference on Rendering Techniques*, EGSR'06, pages 103–113, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association.
2. A Fujimoto, T. Tanaka, and K. Iwata. Arts: Accelerated ray-tracing system. *Computer Graphics and Applications, IEEE*, 6(4):16–26, April 1986.
3. Thomas Kroes, Frits H. Post, and Charl P. Botha. Exposure render: An interactive photo-realistic volume rendering framework. *PLoS ONE*, 7(7):e38586, 07 2012.
4. M. Levoy. A hybrid ray tracer for rendering polygon and volume data. *Computer Graphics and Applications, IEEE*, 10(2):33–40, March 1990.
5. Jörg Mensmann, Timo Ropinski, and Klaus H. Hinrichs. An advanced volume raycasting technique using gpu stream processing. In *GRAPP: International Conference on Computer Graphics Theory and Applications*, pages 190–198, Angers, 2010. INSTICC Press.
6. Matthias Raab, Daniel Seibert, and Alexander Keller. Unbiased global illumination with participating media. In Alexander Keller, Stefan Heinrich, and Harald Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 591–605. Springer Berlin Heidelberg, 2008.
7. Lszl Szirmay-Kalos, Balzs Tth, and Milan Magdics. Free path sampling in high resolution inhomogeneous participating media. *Comput. Graph. Forum*, 30(1):85–97, 2011.
8. Jiří Žára, Bedřich Beneš, Jiří Sochor, and Petr Felkel. *Moderní počítačová grafika*. Computer Press, Praha, 2. edition, 2005.

Dependencies Extraction by Grammatical Evolution approach

Tomas Vantuch

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

Abstract. Looking for dependencies and relations between variables is proceeded to gain better view on some complex problem, that seems to be unclear. With this knowledge we can do more precise classification of its behavior or even the prediction. This article covers the topic about extracting this information by the set of operations known as grammatical evolution. Adjusting the grammar, the input variables, the evolution algorithm and the evaluation of the candidates will be the key requirements for success.

1 Introduction

The term of Symbolic Regression covers the set of algorithms, that are able to build the complex solution of small elementary blocks as an answer for the given problem [1, 3]. These blocks can be represented by mathematical operations, commands of an programming language, parts of an electric circuit, or other elementary artifacts.

The symbolic regression is the tool to transform generated string from evolutionary algorithm to individual solution for its next evaluation. This approach can be called "generate solution and test it". The final solution breded by genetic algorithm should be the closest candidate to best solution for given problem, for example the polynomial for fitting the curve, simple source code containing needed behavior or a device able to handle some special conditions.

From the set of symbolic regression's algorithms, there was the grammatical evolution chosen for these experiments.

1.1 Grammatical evolution

The grammatical evolution, introduced by O'Neill in 2003 [2], is a kind of genetic programming based on defined grammar. This grammar is in BNF described by four parts $\{N, T, P, S\}$, where N stands for the set of non-terminal symbols, T means the set of terminal symbols, P is for the set of transfer rules and S is the starting symbol of the grammar.

The main point of the GE is the substitution of non-terminal symbols from left side by chosen terminals until the result does not contain any other non-terminal [1, 2]. The first individual's encoding is represented by binary vector

N = { expression, operation, pre-operation, variable }
 T = { sin, cos, tan, +, -, *, /, 1, X_t }
 S = expression

Table 1. Subsets of the grammar’s symbols

expression =	{ expression operation expression, (0)	
	pre-operation(expression), (1)	
	variable } (2)	
operation =	{ sum, (0)	
	difference, (1)	
	multiply, (2)	
	divide } (3)	
pre-operation =	{ sinus, (0)	
	cosines, (1)	
	tangent } (2)	
variable =	{ 1, (0)	
	X_t } (1)	

Table 2. The grammar in Bacus-Naur Form used in this experiment

divided to eight-bit parts, so called kodons. These kodons are in the next phase transformed to integer numbers and each of them is integer divided (MOD) to decide the item of grammar for substitution.

All these steps of the substitution and the result itself can be represented by the tree structure for the better view. In this structure, the leafs from left to right will represent the final polynomial.

The grammatical evolution is the way of transforming the genotype to phenotype, but nothing more. Some kind of evolution algorithm has to be used for generating individuals with different genotype and evaluating its rightness by adjusted fitness function. For this purposes, there was used the genetic algorithm in one experiment and in the other one, there was used the particle swarm optimization.

1.2 Genetic algorithm (GA)

The GA is a kind of evolution algorithms [4]. This set of algorithms imitates the evolution theory described by Darwin. The GA purpose is to evolve the best individual for given problem.

As a first step of the GA there is a population of random individuals. Each of the individuals is represented by the binary vector with the defined length. The transfer from genotype to phenotype is made by grammatical evolution.

There is a defined fitness function for the candidate’s evaluation. It has a purpose to mark candidates value in the population to keep better individuals

according to the fitness and create the new generation by their crossover. The one-point crossover [6] was used for crossbreeding in this experiment.

The breded individuals are affected by an added mutation. It means that in the random time, there can be changed one bit of an individual genotype. This process brings to this algorithm some added randomness.

The fitness function of this GA covers the requirement of the minimal difference between known time set values and computed time set values from individuals polynomial.

1.3 Particle swarm optimization (PSO)

The PSO is algorithm inspired by swarm intelligence [7]. Each individual is described by its position in n-dimensional hyperspace, velocity and memory for the latest best position. This algorithm is not divided into generations, because individuals are not dying and creating again, they are just moving during iterations in this space. Their directions are affected by its previous best position or the best position in its neighborhood [8]. The quality of the position is evaluated by the objective function, and during every iteration, all the current positions of all individuals are confronted with its best positions.

The used PSO was able to use the item of neighborhood [8] which means that individual is not affected by the best position of all individuals, but only by the best position of smaller amount of closer individuals, its neighbors. The neighborhood is not defined by the distance between individuals, but randomly. The present of the neighborhood brings to this system the opportunity to find the best global maximum from more observed local maximums.

The objective function in this optimization method was used the same as in GA approach.

2 Experiment Design

These experiments are focused to gain some results from implemented grammatical evolution in Java, connected once to the PSO and second to the GA. The PSO algorithm is supported by Java library JSwarm and GA was obtained from the ECJ framework [9]. The resulted polynomial has to be executed and evaluated. For this purpose there will be used connection to Matlab, where by simple loop we can compute the sum of all differences between compared time set and time set made by polynomial.

The algorithm of symbolic regression is not able to gain any result, so the final judgement will be made about the combination of algorithms, input variables or fitness function with some ideas of improvement.

Target is to approximate the input vector by polynomials, that does not come from the same vector, but only from correlated time series. In this case, there was applied two tests. The first one is to generate the the polynomial with the same behavior as $\cot g(X)$ without the definition of operation cotangent. As we all know the result has to be $1/\tan(X)$. In the second test, there will be the task

to find relations between variables, where probably is not. The fitted variables comes from the time set of close prices of the Microsoft stock prices and the variables for polynomial will be the its open prices and daily volumes.

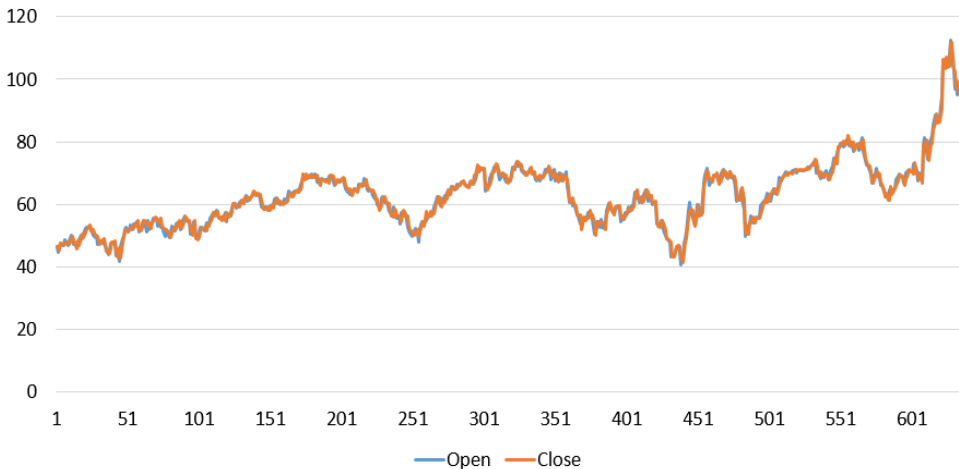


Fig. 1. Correlated open and close prices in MSFT stock data

The chart of open and close prices shows, that this two variables are highly correlated and this can leads our algorithm the bad result that $close_t = open_t$. To prevent this mistake, there was added the checking of the length of the polynomial into the fitness function and all polynomials of size 1 are handled as individuals with zero fitness value.

	GA	PSO
individuals / particles	350	100
generations / iterations	100	100
genotype length	1000	1000
mutation probability	100%	
max velocity		50

Table 3. Adjustment of the optimization algorithms

3 Conclusions

As it was said before, to find good solution strictly belongs to the kind of used evolution algorithm and its fitness function. One of the weakest part of the genetic algorithm is its crossover and in case of PSO it is the orientated moving.

In case of GA, there was used one-point crossover, that does not guarantee that the created individual made from two chosen parents is even valid. On the other hand, the good impact has the 100% probability of added mutation and relatively high number of individuals per generation. Because of the condition, that no individual can be in one generation twice, the fine candidate can be found in relatively small count of generations. The PSO has the same problem as GA about validity of iterated particles. The changes in particle vector does not guarantee that the change will make some impact (in case of changing the last numbers of vector) or event if the particle in next iteration will be valid.

The other weakness of this concept is the fitness (in case of PSO it is objective) function. If it tests only for minimize the difference between time series (generated and real), it will quickly converge to insufficient results like $close_t = close_{t-1}$ or $close_t = open_t$.

The strength of the solution will be based on the choosing and preprocessing of the inputs, difficulty of the fitness and of course the quality of the crossover algorithm.

The final quantity of all possible polynomials does not depend on the grammar, because there can be made the polynomial of infinite length with infinite variations of sinus and cosines functions. The count of possible polynomials depends on the length of the candidate's vector that represents its genotype, because one genotype can be transformed to the right one polynomial. In our case, that we used genotype of length 1000 and the maximal options for each kodon is five, we have the maximal possible count of polynomials no more than 5000.

The Grammatical evolution and genetic algorithm was able to create polynomial $1/\tan(Y)$, which is the answer for the first test where was the task to find out the equally good solution to relation $Y = \text{cogt}(X)$. The algorithms converged to this solution quite quickly because of low number of input constants.

The other question was to find out the relation between open price, close price and daily volumes in the market data of the Microsoft company. As a set of constant for our grammar, there was used $volume_{t-2}$, $volume_{t-1}$, $open_{t-1}$ a $open_t$. The polynomials with best fitness in PSO and GA case are shown right there.

$$close_t = open_{t-1} + \tan(volume_{t-1})/volume_{t-1} * \cos(\sin(\cos(open_t))) * open_{t-1} \quad (1)$$

$$close_t = open_{t-1} - \cos(open_t)/volume_{t-2} \quad (2)$$

In both cases we can clearly see that the compound part contained the division is in every time very close to zero, so it has no significant impact to the result

of the polynomial. In both cases the used algorithm says, that $close_t = open_{t-1}$ which is as the result insufficient.

3.1 Future work

The next steps in this work will be the implementation of the crossover algorithm base on symbolic regression [1, 3] with crossing the blocks of genotype that represents the valid polynomial parts. These blocks can be weighted by its influence on resulted polynomial to build up the polynomial made of parts that has some real impact.

To increase the number of constants, we can in the experiments make some analysis of the stock data to find some correlated market indexes. To find these dependencies we can use some kind of self-organizing maps [10] based on neural networks.

ACKNOWLEDGMENT

The following grants are acknowledged for the financial support provided for this research: Grant Agency of the Czech Republic - GACR P103/13/08195S, is partially supported by Grant of SGS No. SP2014/159, VB - Technical University of Ostrava, Czech Republic, 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.

References

1. Ivan Zelinka, Zuzana Oplatkov, Milos Seda, Pavel Osmera, Frantisek Vcelar, 2008, Evolucni vypocetni techniky principy a aplikace.
2. Michael O'Neil, Conor Ryan, 2003, Grammatical Evolution
3. Nguyen Quang Uy, Nguyen Xuan Hoai, Michael O'Neill, R. I. McKay, Edgar Galvn-Lpez, 2010, Semantically-based crossover in genetic programming: application to real-valued symbolic regression
4. John R. Koza, 1992, Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, Massachusetts.
5. K. E. Kinnear, Jr., Ed., 1994, Advances in Genetic Programming, MIT Press, Cambridge, Massachusetts.
6. Riccardo Poli and W.B. Langdon, Genetic Programming with One-Point Crossover
7. J. Kennedy and R. C. Eberhart, Particle swarm optimization, in Proc. IEEE Int. Conf. Neural Networks, Perth, Australia, Nov. 1995, pp. 19421948.
8. J. Kennedy, The particle swarm: Social adaptation of knowledge, in Proc. 1997 Int. Conf. Evolutionary Computation, Indianapolis, IN, Apr. 1997, pp. 303308.
9. David R. White, Software review: the ECJ toolkit, March 2012
10. Kohonen, T., 1990, The self-organizing map

Context Transformations

Michal Vašínek

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

Abstract. Data compression is a broad scientific area. It covers compression of texts, images, videos, sounds and has many other applications. In this paper is presented different approach to data compression. The approach is based on a context information that could be used to reduce entropy of messages. The proposed method that turns data into one with lower entropy is called the context transformation.

Keywords: compression, context, transformation, entropy

1 Introduction

There are two main classes of algorithms employed in the data compression. The first class of algorithms deals directly with the compression and their purpose is to decrease the size of the input message. Examples[3] of such algorithms could be Huffman coding, Arithmetic coding, Lempel Ziv algorithms or PPM. The second class of algorithms behaves more like preprocessors for the first class, these algorithms are usually called transformations, examples are Burrows-Wheeler transformation[1] or MoveToFront[2] transformation that are used in bzip2 file compressor.

The purpose of transformations is not to decrease the message size, but to change the internal message structure that could be more easily handled by some of the first class algorithms. The algorithm presented in this paper belongs to the second class.

In [5] and [6] was proposed a reversible transformation method called a ‘Context transformation’, that could be used to reduce entropy of input messages. The transformed data then could be more efficiently compressed using entropy coding algorithms, like Huffman coding.

2 Context transformations

The context transformation is a function that swaps occurrences of two different bigrams, the bigram uv that is present in the message and the bigram uw that is not present in the message. Firstly, we present some basic properties of context transformations and then we show how to form transformations to be reversible.

Definition 1. *Context transformation is a function $T : \Sigma^n \rightarrow \Sigma^n$, where Σ is the alphabet of the input message and n is the length of the input message.*

Let $T^{-1} : \Sigma^n \rightarrow \Sigma^n$ is an inverse function to the function T , that returns the input message N from the transformed message $T(N)$.

$$T^{-1}(T(N)) = N \quad (1)$$

To evaluate context transformations, we use unconditional Shannon entropy[4] in the form:

$$H = - \sum_{i \in \Sigma} p_i \log(p_i) \quad (2)$$

We will search for transformations for which holds that entropy of the transformed message is lower than the one of the input message. We have to count also with metadata M that contains descriptions of transformations. Let $L(N)$ is a function that returns the length of the message N , then:

$$H(T(N))L(N) + L(M) < H(N)L(N) \quad (3)$$

Let's consider an example string 'kakaoo', such a string could be represented by its matrix, we call such matrix a 'Context matrix'. Entries of a context matrix are non-negative integers that represents frequencies of digrams represented by a row symbol followed by a column symbol. For our example string, the context matrix is shown in Table 1.

Table 1. The context matrix that represents the string 'kakaoo'

	k	a	o
k	0	2	0
a	1	0	1
o	0	0	1

Since the probability distribution of symbols is uniform, the entropy of our example string is maximal. From the context matrix we see that there are several accessible transformations, we select a transformation that replaces all digrams 'ka' for a digram 'kk', so the condition of zero and non-zero matrix entry is fulfilled.

The resulted string is 'kkkkoo' and its context matrix is shown in Table 2. We can see two different impacts of this transformation:

- the alphabet size decreased,
- the entropy decreased.

Table 2. The context matrix that represents the transformed string ‘kkkkoo’

	k	a	o
k	3	0	1
a	0	0	0
o	0	0	1

Let $T_{d(s)}^i(N)$ be a transformation, where $i \in \{r, c\}$, $d \in \{\rightarrow, \leftarrow\}$, r stands for a context relation between two different rows, c stands for a context relation between two different columns. The arrow represents the direction in which the input message is transformed, so \rightarrow means from the beginning to the end and \leftarrow means from the end to the beginning of the message. Symbol s stands for transformation description, i.e. if we are replacing occurrences of uw for uv then $s = uvw$. For each transformation type(row, column) there is only one possible combination of transformation and its inverse.

Theorem 1. *Let $T_{\leftarrow}^c(N)$ is a context transformation, then the context transformation $T_{\rightarrow}^c(N)$ is its inverse transformation.*

Proof. We have to prove that $T_{\rightarrow}^c(T_{\leftarrow}^c(N)) = N$. We can perform transformation only if two nodes v, w are locally context related through some third node u . Since we know that bigram uv has no occurrence in message N and there is at least one occurrence of uw (in the case that both bigrams are not in message N it is trivial since change nothing for nothing leaves message unchanged) we can replace occurrence of uw for uv .

We have to show what bigrams will be created or destroyed by transformation. We will use a direct proof. Suppose an input string $\alpha uw\beta$ and $\alpha, \beta \notin \{u, v\}$, then after replacing uw for uv we get a string $\alpha uv\beta$ that can be simply reversed back.

The next case we have to deal with is the case when $u = w = \alpha$, then the transformation will be performed in the following way: $uuu \rightarrow uvv \rightarrow uvv$, then the inverse transformation is able to revert the resulted string: $uvv \rightarrow uvv \rightarrow uuu$.

The final case is when uwu and $v = u$ then the string transformation is following: $uwu \rightarrow uuu$ and its inverse is $uuu \rightarrow uwu$. We showed that for each case when bigram uv is created there is an inverse transformation T_{\rightarrow}^c reverting it back to its initial state.

Theorem 2. *Let $T_{\rightarrow}^r(N)$ is a context transformation, then transformation $T_{\leftarrow}^r(N)$ is its inverse transformation.*

Proof. We should realize that $T_{\rightarrow}^r(N)$ is the same as $T_{\leftarrow}^c(N_R)$ performed on the mirror message N_R , then the proof of Theorem 1 is exactly the same, like the one for the case of Theorem 2.

3 Algorithm

We present here the most successful algorithm so far. Other tested algorithms could be found in [5] and [6]. All algorithms expect a context matrix in a special form:

Let $f_c(w, N) : \Sigma \times \Sigma^* \rightarrow \mathbb{N}$ is a function returning for symbol w its frequency in the input message N . We sort nodes of the context map according to f_c and form the context matrix so that the following relations hold:

$$f_{c,i}(w_i, N) \geq f_{c,i+1}(w_{i+1}, N) \quad (4)$$

and

$$f_{c,j}(w_j, N) \geq f_{c,j+1}(w_{j+1}, N) \quad (5)$$

Indices i, j represent rows respectively columns of the context matrix. The goal of algorithms is to find transformations for the input message, that reduces the amount of entropy.

The presented algorithm is based on a greedy approach, algorithm iterates through rows of the matrix from the most probable one to the least probable one. Each time the transformation is found and performed we check if entropy of the message decreased, if not then we perform inverse transformation and algorithm continues with a next context matrix entry.

In the algorithm we use a variable *LIMIT* to avoid usage of transformations that increase entropy. In this paper we work with *LIMIT* = 24, as each transformation needs three symbols of 8 bits length to save its description.

Fig. 1. Entropy reduction test transformation(ERTT)

```

cm ← input_context_matrix
for k = 1 to dim(cm) do
  for i = 1 to dim(cm) do
    zero_index = find_zero_assoc(k)
    max_index = find_max_assoc(k)
    entropy = current_entropy()
    perform_transformation(k, i, zero_index)
    perform_transformation(k, i, max_index)
    if (entropy - current_entropy()) * file_size < LIMIT then
      perform_inverse_transformation(k, max_index, i)
      perform_inverse_transformation(k, zero_index, i)
    end if
  end for
end for

```

Function *perform_transformation* performs substitution of bigrams according Theorem 1. Let t be a number of performed transformations and n be a

length of the message, then complexity of *perform_transformation* is $O(tn)$. The function *find_zero_assoc* serves to find the first zero valued entry in the context matrix row for particular node k and its complexity is $O(d)$ and similarly the function *find_max_assoc* that searches for maximum value in the same row. Both functions start searching in row k from column i to the end of the row. Impact of the algorithm on the distribution of symbols is shown in Fig. 2 for files bib and paper1 from Calgary corpus.

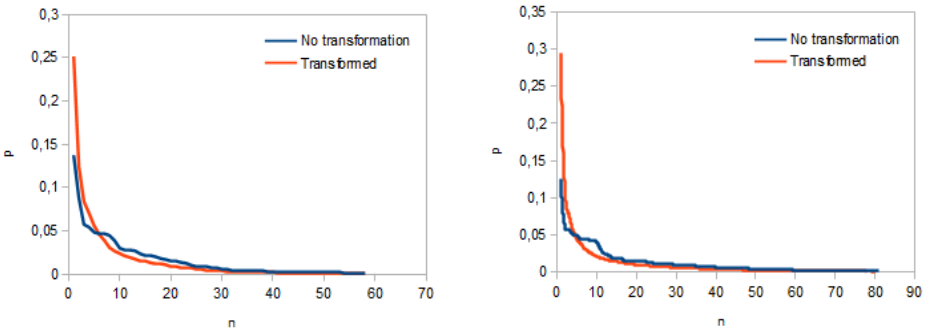


Fig. 2. Distribution of symbols for files bib and paper1 from Calgary corpus: for each sorted node n there is probability p .

The resulted entropy decrease is presented in Table 3 for different files from Calgary corpus. Results show that there are types of files that cannot be successfully transformed. These types are binary and image files. On the other hand, we see that text files could be transformed and with the resulted decrease of entropy up to 13%.

Table 3. Percentual decrease of entropy using context transformations for selected files in Calgary corpus.

File	ERTT(%)
bib	13.02
book1	2.53
book2	9.21
news	7.02
obj1	0.00
paper1	7.09
pic	0.00
progc	8.49
progl	11.63

4 Conclusion and future works

In this paper was presented algorithm based on the concept of context transformations. There were showed some basic properties, that has to be considered, to correctly design algorithms. There was showed, that using transformations, we are able to decrease entropy in files. This decrease is a result of a change in a symbol distribution throughout an input file. Algorithm was the most successful in the case of text files, there was no decrease of the entropy in binary and image files since there were no accessible transformations.

Ideas in this paper were presented at Data Compression Conference 2014 in Salt Lake City. Author prepares another publication that targets a more complete theoretical description of the concept.

5 Acknowledgment

This work was supported by the Grant of SGS No. SP2014/110, VŠB - Technical University of Ostrava, Czech Republic.

References

1. Burrows, M. and Wheeler D.J., *A block sorting lossless data compression algorithm*, 1994
2. Bentley, Jon Louis and Sleator, Daniel D. and Tarjan, Robert E. and Wei, Victor K. *A locally adaptive data compression scheme* Commun. ACM, vol 4, pp. 320-330, 1986
3. Salomon, David, *Data Compression: The Complete Reference*, Springer, NewYork, 2007.
4. Shannon, Claude E., *A mathematical theory of communication* Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, 1948.
5. Vašínek, Michal, *Kontextové mapy a jejich aplikace* Master Thesis, VŠB - Technical University of Ostrava, 2013
6. Vašínek, Michal and Platoš, Jan, *Entropy reduction using context transformations* Data Compression Conference(DCC), pp. 431-431, 2014

Comparison of Arrhythmia Classification Methods in ECG Series

Ľukáš Zaorálek

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

Abstract. The main goal of the work is to test two well-known algorithms for feature transform such as Singular Value Decomposition and Principal Component Analysis in the task of arrhythmia recognition in ECG records. The original signal were transformed by these two techniques and a neural network, linear discriminant analysis and support vector machine were used for classification. Values of sensitivity and accuracy were observed and consequently compared for each transformation. Unlike in other similar works, our experiments were performed on a high number of beats and the tested database included over 47 000 experimental heart beats with different diseases.

Keywords: SVD, PCA, Neural network, LDA, SVM, Arrhythmias, PVC

1 Introduction to ECG and problem definition

Electrocardiography (ECG) is a very-well known diagnostic tool allowing observation of an actual state of the cardiovascular system. In the present time ECG is able to function by the means of small electronic devices. The improvement of the ECG brings advantages in observing patients while they are performing regular daily activities [7]. Unfortunately, these long-term signals have to be properly processed and evaluated by a sophisticated algorithm because it is not possible to evaluate these long-term records by manually. There are many types of arrhythmias which have different symptoms. Each arrhythmia type changes the shapes of regular beats and is important to know how many arrhythmias occurred in the signals during a period. In this paper we focus on three basic arrhythmias: left bundle branch block (LBBB), right bundle branch block (RBBB) and premature ventricular contraction (PVC) [8].

2 Dataset and previous research

The MIT-BIH database was used for experiments. The MIT-BIH contains 48 2-leads ECG records and was created by cooperation between the Massachusetts

institute of Technology and Boston's Beth Israel hospital. The data were independently annotated by three cardiologists. The database was sampled at 360 samples per second with 11-bit resolution and 10mV range. The database is used as a reference database for comparison of different approaches in the task of arrhythmia recognition or QRS detection. The dataset contains 97468 beats in total (7128 PVC, 7253 RBBB, 8072 LBBB and 75015 normal beats). There are many algorithms which were used in similar tasks. Fractal features and a neural network were used in [1]. Adaptive wavelet network was used in [2]. S-transform and a neural network were used in [3].

3 Methods

3.1 Principal Component Analysis and Singular Value Decomposition

The Principal Component Analysis (PCA) is a statistical method which uses orthogonal transformation to transform a set of possible correlation features to a set of linearly uncorrelated variables called principal components. The main goal of this algorithm is to reduce the size of the original feature space. Let matrix X be an input to the PCA algorithm. The size of X is $n \times p$, where n is the number of samples and p is the number of original features, then:

$$Z = XA \quad (1)$$

where X has to be centered, Z consists of principal components which are linear combinations of columns of X . The matrix A is an orthogonal matrix [5].

The Singular Value Decomposition (SVD) is similar to the PCA and the main goal of the algorithm is a factorization of the input matrix X . The SVD calculates significant properties of the original feature space and represents them as linear combinations of base vectors. The SVD decomposes the matrix of original features, calculating singular values and singular vectors of the feature matrix. Let X be the same matrix as in the previous chapter. The input matrix X can be factorized according to equation:

$$X = U\Sigma V \quad (2)$$

Now the new training feature set $RS = U\Sigma$ and the new reduced testing set $QS = TV$ can be calculated. The matrix T is the original testing set. Now, these two matrices can be used just like their original forms [6].

3.2 Neural Network

A neural network (NN) is a distributed computing environment which consists of partial subsystems (neurons). The NNs are inspired by knowledge about the structure and activity of the human brain and attempt to mimic its abilities and functions. There are many steps in a NN and the multilayer perceptron with

backpropagation algorithm is used. The resulting neuron in a higher layer, x_j , is a linear function of the outputs, y_i , of the neurons of the lower layer and of the weights, w_{ji} , on their connections. Each neuron can have extra input. This input is called a bias. A neuron has a real valued output, y_i , which is a non-linear function of its total input:

$$y_i = \frac{1}{1 + e^{-x_j}} \quad (3)$$

We have to find a set of weights that ensure that for each input vector, the output vector produced by the network is the same or very close to the desired output vector. For the purpose finding optimal value of the weights we choosed backpropagation algorithms. As our description is greatly simplified, we encourage the reader to study the matter in detail for example in [4].

3.3 Linear Discriminant Analysis

The Linear Discriminant Analysis (LDA) is a technique for pattern recognition. The main aim of this algorithm is to find linear combinations of features, which provide the best separation between classes. The basic principle is the measurement of metric or cosine distances between new instances and the centroid of classes. The new instances are classified according to the expression:

$$\arg \min d(z\Phi, \bar{x}_k\Phi). \quad (4)$$

3.4 Support Vector Machine

Let N be the number of training samples, where each sample belongs to class $y_1 = -1$ or to class $y_1 = 1$. In cases where data is linearly separable, we can use a hyperplane for recognition of these classes. A perpendicular distance between the hyperplane and the zero point of the feature vector is possible to express by Equation 5

$$\frac{b}{\|w\|} \quad (5)$$

Now we can define two new hyperplanes, $H1$ and $H2$, which are parallel to the separation hyperplane and which separate both classes and, additionally, there are no samples between them. The main aim of this classifier is to find the hyperplane which maximizes the margin such that the distance between both hyperplanes, $H1$ and $H2$, is the same. This can be achieved by minimizing $\|w\|$. The data can not lie in the margin. Therefore, the following conditions have to be met:

$$y_i = +1 \Rightarrow x_i \cdot w - b \geq 1 \quad (6)$$

$$y_i = -1 \Rightarrow x_i \cdot w - b \leq -1 \quad (7)$$

4 Experiments and results

The chapter describes the experiments conducted with feature extraction using Singular Value Decomposition and Principal Component Analysis on ECG signals and also evaluates their results with a Support Vector Machine and Neural Network as the classifier as well as Linear Discriminant Analysis. The following section is dedicated to the configuration of the Neural Network, Principal Component Analysis as well as Singular Value Decomposition which we used for our experiments. In the final section we evaluate impact of these algorithms in the task of arrhythmia pattern recognition and compare these results to each other.

4.1 Configurations

As the type of the neural network was chosen the feedforward neural network (described in section 3.2). The neural network had following the configuration: training function was set to the scaled conjugate gradient method, maximum validation error was set to 20. The learning rate of the neural network was chosen several times to achieve the best results in the experiments. Specifically, we chose the following values as the learning rate of the neural network: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8.

In the experiments where the SVD and PCA were used we set the rank of eigenvalues to the following numbers: 5, 7, 9, 11, 12, 15 (same for both algorithms). For each experiment conducted with these algorithms, the number of eigenvalues was chosen so that the sum of their values exceeds 95% of the sum of all eigenvalues. We have performed the experiments 48 times for the SVD and PCA algorithms.

4.2 Results

The experiment that uses the SVD with the neural network is shown in Table 2. The sensitivity values in this case are 98.38, 99.31, 99.42 and 96.16. The results are slightly worse than the sensitivity values in the first experiment. Table 1 presents the results of the best configuration for the combination of the PCA and the neural network. The PCA rank included 15 eigenvalues and the learning rate of the neural network was set to 0.7. The table shows that the sensitivity values are slightly lower in comparison with the first experiment: 98.73, 99.17, 98.29, 95.26. Table 4.2 shows result experiment with Linear Discriminant Analysis as classifier. In this case we choosed SVD for features extraction from ECG signal. The sensitivity of this experiment are 71.07, 96.52, 89.03 and 73.07. In other words, values of the sensitivity represent significant worse result in compare with previous two experiments. The experiment that uses Support Vector Machine with Singular Value Decomposition represents table 4.2. Values of the sensitivity are 96.71, 99.09, 93.05 and 90.75. It means that the sensitivity is higher than previous experiment but slightly worse then first two experiments. The best results were obtain by using neural network as ECG arrhythmia classifier together with SVD method which represents ECG signal transformation.

	LBBB	Normal	RBBB	PVC
LBBB	3986	172	0	63
Normal	30	37209	26	89
RBBB	4	46	3566	17
PVC	17	91	36	3399
Accuracy	94.43	99.61	98.15	95.93
Sensitivity	98.73	99.17	98.29	95.26
Specificity	99.47	98.70	99.85	99.68

Table 1. Confusion matrix of the PCA+NN

	LBBB	Normal	RBBB	PVC
LBBB	3972	121	0	51
Normal	40	37260	19	68
RBBB	0	39	3607	18
PVC	25	98	2	3431
Accuracy	95.84	99.66	98.44	96.48
Sensitivity	98.38	99.31	99.42	96.16
Specificity	99.61	98.86	99.87	99.72

Table 2. Confusion matrix of the SVD+NN

	LBBB	Normal	RBBB	PVC
LBBB	2869	766	0	263
Normal	1123	36214	397	626
RBBB	1	60	3230	72
PVC	44	478	1	2607
Accuracy	73.60	94.41	96.05	83.29
Sensitivity	71.07	96.52	89.03	73.07
Specificity	97.70	80.90	99.71	98.84

Table 3. Confusion matrix of the SVD+LDA

	LBBB	Normal	RBBB	PVC
LBBB	3904	276	0	32
Normal	126	37177	247	293
RBBB	0	16	3376	5
PVC	7	49	5	3238
Accuracy	92.69	98.24	99.38	98.15
Sensitivity	96.71	99.09	93.05	90.75
Specificity	99.31	94.07	99.95	99.86

Table 4. Confusion matrix of the SVD+SVM

5 Conclusion

In this paper, we compare the performance of the Singular Value Decomposition, Principal Component Analysis in the task of arrhythmia recognition in ECG

records. As a classifier, a feedforward neural network was chosen as well as support vector machine and linear discriminant analysis. Concretely, we compare the performance of arrhythmia recognition with using the SVD, PCA and NN as a signal preprocessing. We measured the performance by three validating methods: accuracy, sensitivity and specificity. Also, confusion matrices belonging to each experiment were shown. The focus of the future work will be classifying ECG arrhythmia on intra-individual level.

References

1. Chia-Hung Lin; Chao-Lin Kuo; Jian-Liung Chen; Wei-Der Chang, "Fractal features for car-diac arrhythmias recognition using neural network based classifier," *Networking, Sensing and Control*, 2009. ICNSC '09. International Conference on , vol., no., pp.930,935, 26-29 March 2009 doi: 10.1109/ICNSC.2009.4919405
2. Chia-Hung Lin; Pei-Jarn Chen; Yung-Fu Chen; You-Yun Lee; Tainsong Chen, "Multiple Cardiac Arrhythmia Recognition Using Adaptive Wavelet Network," *Engineering in Medicine and Biology Society*, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the , vol., no., pp.5655,5659, 17-18 Jan. 2006 doi: 10.1109/IEMBS.2005.1615769
3. Das, M.K.; Ari, S., "ECG arrhythmia recognition using artificial neural network with S-transform based effective features," *India Conference (INDICON)*, 2013 Annual IEEE , vol., no., pp.1,6, 13-15 Dec. 2013 doi: 10.1109/INDCON.2013.6726153.
4. Masters, Timothy. *Practical Neural Network Recipes in C++* . Academic Press, New York, Year 1993.
5. Peterek, Tomas; Krohova, Jana; Smondrk, Maros; Penhaker, M., "Principal component analysis and fuzzy clustering of SA HRV during the Orthostatic challenge," *Telecommunications and Signal Processing (TSP)*, 2012 35th International Conference on , vol., no., pp.596,599, 3-4 July 2012 doi: 10.1109/TSP.2012.6256366
6. Moravec, P., Gajdoš, P., Snášel, V., Saeed, K.: Normalization Impact on SVD-Based Iris Recognition. In *Proceedings of International Conference on Biometrics and Kansei Engineering*, 2009.. Ed. Khalid Saeed, Ajith Abraham, Piotr Porwik. IEEE press. 2009. 60-64. ISBN 978-0-7695-3692-7
7. Zaořlek, L., Peterek, T., Dohnlek, P., Gajdo, P.: Comparison of feature reduction methods in the task of arrhythmia classification. In: Kmer, P., Abraham, A., Snel, V. (eds.) *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014, Advances in Intelligent Systems and Computing*, vol. 303, pp. 375–382. Springer International Publishing (2014)
8. Peterek, T., Zaořlek, L., Dohnálek, P., Gajdoš, P.: Arrhythmias classification using singular value decomposition and support vector machine. In: *Intelligent Data analysis and its Applications, Volume II*, pp. 591–600. Springer (2014)

Mapping of a quantum system of particles onto a classical system of cyclic polymers

Rajko Čosić

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
rajko.cosic@vsb.cz

Abstract. In the present paper a scheme which maps quantum system of structureless particles onto a classical system of cyclic polymers is described. Since thermodynamical properties of quantum systems are hard to compute, this mapping is used to transform a quantum system onto a classical one the thermodynamical properties of which can be simulated via Monte Carlo methods. In first section the benefits of this mapping are discussed. In following two sections theoretical background and implementation are shown, and in the last section we give some outlooks for the future.

Keywords: statistical thermodynamics, quantum systems, classical systems, path-integral Monte Carlo

1 Motivation

In quantum mechanics, the system in a mixed state is described by so-called *density matrix*. The density matrix is an operator on Hilbert space (in our case of N structureless particles the space is $\mathcal{H} = L^2(\mathbb{R}^{3N})$). For systems in thermal equilibrium given by the canonical (NVT) ensemble the density matrix (or statistical operator on \mathcal{H} , $\hat{W} : \mathcal{H} \rightarrow \mathcal{H}$) is

$$\hat{W} = e^{-\frac{\hat{H}}{kT}}, \quad (1)$$

where k is the Boltzmann constant, T is temperature, self-adjoint operator \hat{H} is so-called *Hamiltonian operator*, $\hat{H} : \mathcal{H} \rightarrow \mathcal{H}$ and

$$\hat{H} = \hat{T} + \hat{V}, \quad (2)$$

where for $\psi \in \mathcal{H}$, $\hat{T}(\psi) = -\lambda\Delta\psi$ is kinetic energy operator ¹ and $\hat{V}(\psi) = V\psi$ is potential energy operator.

¹ The constant $\frac{\hbar^2}{2m}$, where \hbar is reduced Planck constant and m is the mass of the particle, is denoted by λ to follow notation stated in [2] and to avoid confusion since the letter m is used in text as summation index.

Since L^2 is an infinite-dimensional separable Hilbert space there exists a countable orthonormal basis (Schauder basis) in which the Hamiltonian operator can be represented by a hermitian matrix. Although the basis is infinite the representation exists and for the given orthonormal basis is unique (see [1]).

Let us remind how operator exponentials are defined

Definition 1. Let \hat{A} be self-adjoint operator on a separable Hilbert space, φ_1, \dots and a_1, \dots are eigenvectors and eigenvalues of \hat{A} . Then

$$e^{\hat{A}} = \sum_{i=1}^{\infty} e^{a_i} |\varphi_i\rangle\langle\varphi_i|.$$

The notation used in this paper is so-called *bra-ket* notation. In this notation the term $|\varphi_i\rangle\langle\varphi_i|$ means the projective operator on the one-dimensional space generated by vector φ_i . For more see [3].

Problem is that for computing the density matrix \hat{W} one needs the spectrum of Hamiltonian operator \hat{H} . Since \hat{H} is infinite-dimensional matrix the spectral decomposition is, in general case, impossible to find in real time. The sequence of finite matrices the spectra of which would converge to the spectrum of \hat{H} is also hard to find (for creating such a sequence one again needs to know the eigenvalues of \hat{H}) so some other techniques which will yield good result are needed.

One of these techniques takes the quantum system and assigns to each particle some number of so-called *ghosts*. The particle with its ghosts forms a cyclic polymer which can be described in language of classical physics. The potential between the particles in the quantum system transforms into a sum of potentials between corresponding ghosts of each polymer. The kinetic energy in the quantum system is now interpreted as interaction between neighboring ghosts in the polymer.

In next section we show that with this mapping the problem of finding the spectrum of the Hamiltonian matrix can be evaded by approximating the potential and kinetic parts by a many-dimensional integral.

2 Theory

We already know that the density matrix of the studied system is

$$\hat{W} = e^{-\frac{\hat{H}}{kT}}. \tag{3}$$

Let us now introduce the *position space* notation of this matrix (see [2]). The position space density matrix is ²

$$\begin{aligned} \rho(R, R', \beta) &= \langle R | e^{-\beta\hat{H}} | R' \rangle \\ &= \sum_i \Phi_i^*(R) \Phi_i(R') e^{-\beta\mathcal{H}}, \end{aligned} \tag{4}$$

² In bra-ket notation the term $\langle R | e^{-\beta\mathcal{H}} | R' \rangle$ represents the dot product of vector R (R' respectively) and the image of vector R' (R respectively) under the action of operator $e^{-\beta\mathcal{H}}$.

where Φ_i are eigenvectors of the Hamiltonian operator.

Then we can use the following identity (see [3])

$$\int dR_2 |R_2\rangle\langle R_2| = \mathbb{I} \tag{5}$$

to rewrite equality

$$e^{-(\beta_1+\beta_2)\hat{H}} = e^{-\beta_1\hat{H}}e^{-\beta_2\hat{H}} \tag{6}$$

into the position space notation

$$\rho(R_1, R_3, \beta_1 + \beta_2) = \int dR_2 \rho(R_1, R_2, \beta_1) \rho(R_2, R_3, \beta_2). \tag{7}$$

The idea of the path-integral approach is to apply this product M times so that

$$e^{-\beta\hat{H}} = \left(e^{-\tau\hat{H}} \right)^M, \tag{8}$$

where $M \geq 1$ and $\tau = \beta/M$. In position representation:

$$\rho(R_0, R_M, \beta) = \int \dots \int dR_1 \dots dR_{M-1} \rho(R_0, R_1, \tau) \dots \rho(R_{M-1}, R_M, \tau). \tag{9}$$

Purpose of this multiplication will be clear after introducing the second property used in path-integral approach. Since the Hamiltonian can be split into its kinetic and potential parts (see [2]), we can write

$$e^{-\tau(\hat{T}+\hat{V})+\frac{\tau^2}{2}[\hat{T},\hat{V}]} = e^{-\tau\hat{T}}e^{-\tau\hat{V}}. \tag{10}$$

In the limit $\tau \rightarrow 0$ the comutator term can be neglected because it is much smaller than other terms and we can write

$$e^{-\tau(\hat{T}+\hat{V})} \approx e^{\tau\hat{T}}e^{-\tau\hat{V}}. \tag{11}$$

The limit $\tau \rightarrow 0$ means that $M \rightarrow \infty$. From [4] we get that the approximation error will not grow as $M \rightarrow \infty$ and hence the limit could be applied to equation (8)

$$e^{-\tau(\hat{T}+\hat{V})} = \lim_{M \rightarrow \infty} \left(e^{-\tau\hat{T}}e^{-\tau\hat{V}} \right)^M. \tag{12}$$

Let M be finite but sufficiently large. Then (11) written in position space representation will read

$$\rho(R_0, R_2, \tau) = \int dR_1 \left\langle R_0 \left| e^{-\tau\hat{T}} \right| R_1 \right\rangle \left\langle R_1 \left| e^{-\tau\hat{V}} \right| R_2 \right\rangle. \tag{13}$$

Now let us evaluate each of the two parts in the integral on the right hand side of (13). Firstly let us have a look at the potential part. The potential operator \hat{V} is diagonal in the position representation and hence the matrix elements are

$$\left\langle R_1 \left| e^{-\tau\hat{V}} \right| R_2 \right\rangle = e^{-\tau V(R_1)} \delta(R_2 - R_1), \tag{14}$$

where $\delta(R_2 - R_1)$ is called *Dirac (generalised) function* (see [5]) and satisfies the following conditions

$$\left. \begin{aligned} \int dR_2 \delta(R_2 - R_1) &= 1 \\ \delta(R_2 - R_1) &= 0 \text{ for } R_1 \neq R_2. \end{aligned} \right\} \tag{15}$$

To evaluate the kinetic part we have to find the eigenvalues and eigenvectors of operator \hat{T} . Since

$$\hat{T} = -\lambda \Delta \psi$$

the eigenvalue problem corresponds to finding the solution of stationary Schroedinger equation with zero potential part.

Consider the stationary Schroedinger equation with zero potential part in a cube of side L with periodic boundary conditions.

$$\left. \begin{aligned} -\frac{\hbar^2}{2m} \Delta \psi &= E \psi \\ \psi(\mathbf{x} + \mathbf{t}L) &= \psi(\mathbf{x}), \quad \mathbf{x} \in \langle 0, L \rangle^3, \\ &\mathbf{t} \in \mathbb{Z}^3. \end{aligned} \right\} \tag{16}$$

From the Fourier method we get a complete set of solutions

$$\begin{aligned} \psi_{\mathbf{k}} &= \left(\frac{1}{L}\right)^{\frac{3N}{2}} e^{i\frac{2\pi}{L}(\mathbf{k}R)}, \quad \mathbf{k} \in \mathbb{Z}^{3N}, \\ E_{\mathbf{k}} &= \lambda \frac{4\pi^2 \mathbf{k}^2}{L^2}, \quad \mathbf{k} \in \mathbb{Z}^{3N}, \end{aligned} \tag{17}$$

where N is the number of particles.

Using the completeness of $\psi_{\mathbf{k}}$, one gets for the kinetic part of (4) (see [2])

$$\left\langle R_0 \left| e^{-\tau \hat{T}} \right| R_1 \right\rangle = \sum_{\mathbf{k} \in \mathbb{Z}^{3N}} L^{-3N} e^{-\lambda \frac{4\pi^2 \mathbf{k}^2}{L^2} \tau - i \frac{2\pi \mathbf{k}}{L} (R_0 - R_1)} \tag{18}$$

We can approximate the sum in (18) by integral if the following statement holds:

$$\lambda \tau \ll L_2.$$

This condition can be obtained from error estimate of the multidimensional rectangular rule for numerical quadrature. Since $\tau = \beta/M$ for sufficiently large M , the condition is satisfied and the sum can be approximated by the following integral

$$\int_{\mathbb{R}^{3N}} L^{-3N} e^{-\lambda \frac{4\pi^2 \mathbf{k}^2}{L^2} \tau - i \frac{2\pi \mathbf{k}}{L} (R_0 - R_1)} d^{3N} \mathbf{k}. \tag{19}$$

The resulting expression of the kinetic part of (4) is then

$$\left\langle R_0 \left| e^{-\tau \hat{T}} \right| R_1 \right\rangle = (4\pi \lambda \tau)^{-\frac{3N}{2}} e^{-\frac{(R_0 - R_1)^2}{4\lambda \tau}}. \tag{20}$$

If we put (14) and (20) together, we get the density matrix of the system

$$\rho(R_0, R_M, \beta) = \int dR_1 \cdots dR_{M-1} (4\pi\lambda\tau)^{-\frac{3NM}{2}} \cdot e^{-\sum_{m=1}^M \left[\frac{(R_{m-1} - R_m)^2}{4\lambda\tau} + \tau V(R_m) \right]}. \quad (21)$$

The main thermodynamical parameter, the *partition function* (see [7]), is defined by

$$Z = \text{Tr } \rho, \quad (22)$$

where $\text{Tr } \rho$ is the trace of the operator ρ . If we know this parameter, other thermodynamical properties can be obtained easily. Partition function is in position representation written as

$$Z = \int dR \rho(R, R, \beta), \quad (23)$$

which can be, using (21) with $R_M = R_0$, written as

$$Z = \int dR_0 \cdots dR_{M-1} (4\pi\lambda\tau)^{-\frac{3NM}{2}} \cdot e^{-\sum_{m=1}^M \left[\frac{(R_{m-1} - R_m)^2}{4\lambda\tau} + \tau V(R_m) \right]}. \quad (24)$$

The condition $R_M = R_0$ appears in the sum $-\sum_{m=1}^M \left[\frac{(R_{m-1} - R_m)^2}{4\lambda\tau} + \tau V(R_m) \right]$ when $m = M$. It is a classical description of cyclic polymers.

3 Implementation

As we can see from (24), the partition function can be computed as many-dimensional integral so the idea of computing it via Monte Carlo is obvious. We generate a set of random configurations of the model system of ring polymers via *Metropolis algorithm* (for a more detailed introduction into the Metropolis algorithm see [6]). This algorithm works with so-called *acceptance ratio*³

$$p_{acc} = \max \left\{ \frac{\pi(x_{new})}{\pi(x_{old})}, 1 \right\}, \quad (25)$$

which is the probability that the new randomly chosen configuration will be accepted. The probability density function π is computed as

$$\pi(R_0, \dots, R_{M-1}) = e^{-\sum_{m=1}^M \left[\frac{(R_{m-1} - R_m)^2}{4\lambda\tau} + \tau V(R_m) \right]}. \quad (26)$$

The points x_{old} and x_{new} represent old and new configurations. As we can see for generating a new configuration the algorithm has to know its predecessor so the parallelisation is difficult.

³ x_{old} is the configuration from n^{th} step of Metropolis algorithm (x_0 can be arbitrary point in configuration space) and x_{new} is randomly generated point in the configuration space

The exponent of the probability distribution π is given above in (24) and consists of two parts. The first one is potential between particles.

$$E_p = \sum_{m=1}^M \tau V(R_m).$$

In the present version of the program it is represented by Lennard–Jones potential between corresponding ghosts but can be replaced by any model by changing the respective piece of the code or replacing the current module implementing the potential energy. The second one is quadratic potential within the chain of ghosts representing each particle,

$$E_k = \sum_{m=1}^M \frac{(R_{m-1} - R_m)^2}{4\lambda\tau}.$$

Since the computation of the potential energy can be generally very complex (depending on used model), the parallelisation via MPI is involved.

4 Outlooks

The main aim of this project is to model the behavior of snowballs of He_n^+ . These snowballs develop when the superfluid helium cluster at is ionised low temperature. Then the crystallic core develops surrounded by the rest of the cluster which stays superfluid. Path-itnegral Monte Carlo simulations will help us to better understand what happens on the boundary between the core and the rest of the cluster. Such simulations can also be used for modelling of ions of another elements embeded in a superfluid helium cluster (see [8]) which is used in experimental physics to study their low temperature chemistry.

References

1. Blank, J., Exner, P., Havlíček, M.: Lineární operátory v kvantové fyzice. Karolinum, Prague (1993)
2. Ceperley, D. M.: Path integrals in the theory of condensed helium. Rev. Mod. Phys. 67, (1995)
3. Formánek, J.: Úvod do kvantové teorie. Academia, Prague (1983)
4. Trotter, H. F.: On the product of semi-groups of operators. Proc. Am. Math. Soc. 10, 545 (1959)
5. Dirac, P. A. M.: The principles of Quantum Mechanics. Oxford, (1930)
6. Lewerenz, M.: Monte Carlo Methods: Overview and Basics. Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, Lecture Notes, John von Neumann Institute for Computing, Julich, NIC Series, Vol. 10, ISBN 3-00-009057-6, 1 (2002)
7. Kvasnica, J.: Statistická fyzika. Academia, Prague (1983)
8. Theisen, M., Lackner, F., Ernst, W. E.: Forming RB^+ snowballs in the center of He nanodroplets. Phys. Chem. Chem. Phys. 12, 14861 (2010)

SPN as a tool for risk modeling 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. This article presents a new possible approach to modeling the risk to safety of personnel by Stochastic Petri Nets (SPN). Stochastic Petri Nets model for small leak occurrence on an offshore platform is shown, based on a realistic example from the offshore industry. The probabilities of fatalities were obtained from the simulation by using the GRIF's Petri Nets module and compared to probabilities obtained by direct Monte Carlo simulation methods.

Keywords: Risk modeling; Stochastic Petri Nets; Offshore industry; GRIF;

1 Introduction

Modeling the risk to safety of personnel in offshore industry is often realized by the application of Event Trees. However, fatality probabilities in the offshore industry often depend on personnel reactions and their consequences. As the Event Tree is a steady state method, Stochastic Petri Nets were considered a suitable tool for creating a model of the risk to safety of personnel.

First, the steady state representation of a time-dependent event was modeled using SPN. A small hydrocarbon leak occurrence was chosen as an example. Detailed description of Small Leak Scenario is shown in [2]. Construction of steady state SPN model and result comparison between SPN and ET methods was presented in [1].

In this paper, the dynamic representation of the same event was modeled, demonstrating the potential of SPN-based risk modeling.

Probabilities of fatalities due to small leak occurrence obtained by the GRIF's Petri Net module were compared to results obtained by the application of direct Monte Carlo methods, presented in [2].

2 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 [2]).

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.

2.1 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.

2.2 Dynamic SPN model of Small Leak Scenario

For the construction of time-dependent SPN model of Small Leak Scenario, Personnel and Installation Actions table from [2] was used as a moot point. The table describes escape of personnel to lifeboat station at the Accommodation Platform. Personnel has 15 minutes to arrive at the lifeboat station. If any personnel fails to arrive, search party is formed and given 15 minutes to search for and rescue missing personnel. 30 minutes after leak occurrence, lifeboats are launched and personnel evacuates.

All possible events during escape of the personnel are divided in six time intervals, every possible event is described and linked with corresponding ignition case and occurrence time. However, for creating SPN model of the Small Leak Scenario, quantification of contributions of fatality probabilities is required.

With realistic data from offshore industry available, tables containing different scenarios with corresponding probability of fatality contributions in Small Leak Scenario were made, one for each ignition case. As an example, table for 2 to 5 minutes ignition case is shown in Fig. 1.

In the dynamic SPN model, three blocks of scenarios were made, one block for each ignition case. No ignitions cases were not modeled.

When token leaves starting place in each block, immediate transition fires and token is send to place distributing tokens amongst all available scenarios for given ignition case. It should be noted that mentioned immediate transition is substituted in model by timed transition with delay determined by Dirac law with parameter $\delta = 0$.

If suitable delay law is chosen, SPN model allows to 'split' the arriving token amongst multiple transitions. In this way, token is divided between all enabled

No.	Time T (mins)	Person Location and Activity at Time T	Prob of Fatal at Time T
17	0-1	In Zone 8.	0
	1-25	Escaping from Zone 8 to Muster Area on AP.	0
	25-30	Embarkation to and evacuation by lifeboats from AP.	0
18	0-1	In Zone 8.	0
	1-25	Escaping from Zone 8 to Muster Area on AP.	0
	25-30	Lifeboats on AP not available. Route to secondary evac not available. Evacuation to the sea.	0.5
19	0-1	In Zone 8.	0
	1-25	Escaping from Zone 8 to Muster Area on AP.	0
	25-30	Embarkation to and evacuation by lifeboats from AP.	0
20	0-1	In Zone 8.	0
	1-25	Escaping from Zone 8 to Muster Area on AP.	0
	25-30	Lifeboats on AP not available. Route to secondary evac not available. Evacuation to the sea.	0.5
21	0-1	In Zone 8.	0
	1-25	Escaping from Zone 8 to primary and secondary means of evacuation.	0.1
	25-30	Embarkation to and evacuation by primary and secondary means of evacuation.	0.1
22	0-1	In Zone 8.	0
	1-15	Escaping from Zone 8 to primary and secondary means of evacuation.	0.2
	15-30	Lifeboats on AP not available. Route to secondary evac not available. Evacuation to the sea.	0.5
23	0-1	In Zone 8.	0
	1-15	Trapped during escape from Zone 8 to primary and secondary means of evacuation.	0.2
	15-30	No escape to any lifeboats or life rafts available. Evacuation to the sea.	0.7

Fig. 1. Contributions of probabilities of fatality for Small Leaks with ignition between 2nd and 5th minute.

transition. This division is then based on occurrence probability for each available scenario. In presented SPN model, exponential distribution was used with rates λ equal to the occurrence probability of each available scenario. Each enabled transition then fires and sends 1800^1 tokens to computation blocks.

In computation blocks (Figure 2 shows computation block for scenario 23²), time-dependent probability of fatality of given scenario is calculated. Each computation block has to simulate its own time for its own probability calculation. In this SPN example, time flow is represented by variable *Time*, incremented by a set of transitions with Dirac delay law with $\delta = 1$. When a transition is enabled by a token, transition fires, variable *time* is raised by 1 and distribution function values $F(\text{time})$ and $F(\text{time} - 1)$ are computed. The probability of fatality value is then adjusted by difference $F(\text{time}) - F(\text{time} - 1)$. This approach was chosen due to internal logic of the GRIF's Petri Net module, which evaluates all variables once a transition fires.

In this paper, it is assumed that the probability of fatality follows exponential distribution. Therefore:

$$F(t) = 1 - e^{-\lambda t}, \quad (1)$$

¹ As mentioned in part 2.2, personnel has 30 minutes to evacuate the platform, therefore each token represents one second in simulation.

² Description of scenario 23 can be viewed in Fig. 1

where t is a value of variable *time* and λ is a rate parameter for current phase of evacuation. Correct phase of evacuation for λ computation is determined by a set of conditions utilizing time thresholds given by ignition case tables (shown in 1 for 2 to 5 minute ignition case).

Let x be length of current phase in seconds and $F(x)$ be probability of fatality at the end of the current phase. Then rate parameter λ is given by following equation:

$$\ln(1 - F(x)) = -\lambda x . \tag{2}$$

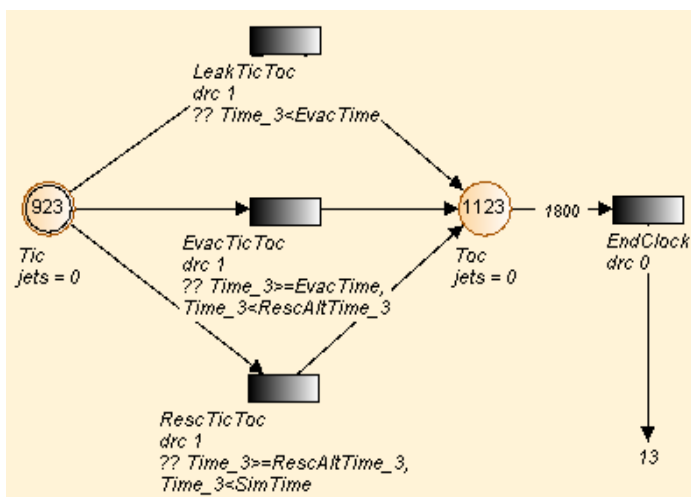


Fig. 2. SPN model : Computation block for scenario 23.

In this way, probability of fatality is computed for each available scenario. However, this does not reflect occurrence probability of those scenarios, thus the previously mentioned token 'splitting' was used. For each scenario, computed probability of fatality was multiplied by scenario occurrence rate (i.e. portion of token received by corresponding computation block). By adding these modified probabilities of fatality, the overall probability of fatality was constructed, resulting time-dependent probability of fatality graph is shown in Fig. 3.

3 Results

For computation of the probabilities of fatalities, GRIF's Petri Net module was used. For comparison with the results presented in [2] by application of the Event Tree and Monte Carlo methods, same sample was chosen, consisting of 609 occurrences of hydrocarbon small leaks.

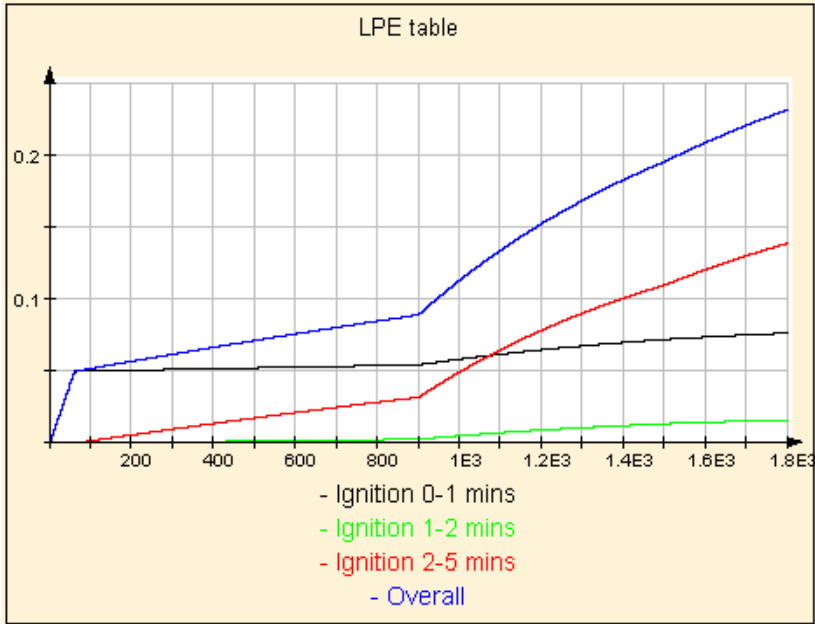


Fig. 3. SPN model : Overall time-dependent probability of fatality of Small Leak Scenario.

3.1 Dynamic SPN model

Probability values at the time of 1800 seconds were used and compared to values obtained by DMC model presented in [2]. Probability values obtained by SPN model were multiplied by ignition case frequencies, resulting in comparison shown in Table 1.

Table 1. Resulting probabilities of fatalities obtained by the SPN and the MC model.

Ignition Case	SPN model	DMC method
0 and 1 minute	3.02e-4	2.6e-4
1 and 2 minutes	8.82e-5	8.5e-5
2 and 5 minutes	9.59e-4	1.1e-3
Total	1.349e-3	1.423e-3

Considering results of SPN and DMC methods in Table 1, the most significant difference measured between ignition cases was $1.71e - 4$ (measured between 2 to 5 minute ignition cases). Cumulative probabilities of fatality are differing by

a margin of $7.4e - 5$. However, despite the differences, resulting probabilities of fatality could be regarded as fairly accurate and thus allowing SPN to be considered a viable alternative to other time-dependent probability modeling methods.

4 Conclusion

In this article, a new approach for modeling risk to safety of personnel in process industries was presented in the application of SPN.

Dynamic 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 Direct Monte Carlo method. As the differences were low, the SPN can be considered a suitable tool for risk modeling in process industries. In comparison with DMC method, the SPN method results were less accurate, though by a small margin. Thus, the SPN method could be regarded as a viable alternative to DMC method in process industry. Proposed further work in this area is construction of SPN models for Medium and Large Leak Scenarios.

Acknowledgements

This article was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

1. Ondřej Grunt.: Risk Modeling of Accidental Events and Responses in Process Industries. WOFEX 2013, pp. 327-332.
2. Radim Briš, Sava Medonos, Chris Wilkins, Adam Zdráhala.: Time-Dependent Risk Modeling of Accidental Events and Responses in Process Industries. Reliability Engineering and System Safety 125(2014)54-66.
3. 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.
4. M. Ajmone Marsan, Gianfranco Balbo, Gianni Conte, Susanna Donatelli, Giuliana Franceschinis.: Modeling with Generalised Stochastic Petri Nets. Università degli Studi di Torino, 1994.

Optimization Process for Searching Thermal Conductivity of Mill Scales

Martin Hasal

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
martin.hasal@vsb.cz

Abstract. Mill scale is material, which is formed on the surface of material under rolling process. Mill scales must be removed from the surface of the rolled material and they are removed by high pressured water jet. After removing they fall down into reservoir with water. Great deal of oil is used in the industry till rolling process is completed. Hence in the reservoir there are mill scales, water and oil. Mill scales is a material which can be recycled. But mill scales can not be given into blast furnace with oil and water, so mill scales must be dried from the water. But for setting of water drying process, it is necessary to know thermal conductivity of the mill scales with oil.

Keywords: Mill scales, thermal conductivity, inverse optimization.

1 Introduction

Trinecké železářny - Moravia Steel is one of the biggest developer of rails. In the beginning there are some standard raw materials for creating a steel. From the steel an ingot is created. This ingot goes into gas furnace. After this molten ingot goes onto rough rolling and finish rolling. These are two moments when mill scales are created. Mill scales are formed on the outer surfaces of rails (commonly also on plates, sheets or other profiles) when it are being produced by rolling red hot iron or steel billets in rolling or steel mills, [4]. They are composed of iron oxides mostly ferric and is bluish black in colour. It is usually less than a millimeter thick and initially adheres to the steel surface and protects it from atmospheric corrosion provided no break occurs in this coating, [5].

When the mill scale stays on the surface, it will increase the corrosion of steel exposed at the break. It is a nuisance when the steel is to be processed, i.e. rail goes trough rolling-mill, mill scale is created on the surface, it is not removed from the surface and rail goes again trough rolling-mill. Then mill scale is pushed into material and it creates an asperities on the surface and inhomogeneity on the surface. Due to that the mill scale must be removed from the rail after rolling process.

Removing of mill scale can be done by several ways, e.g. by flame heating, acid pickling or grit/sand blasting. During rolling process, when the steel is molten, the steel is rinsed by water under pressure of 10 to 100 MPa, commonly amount 20 to 30 MPa. Water jet has a shape of wall, which covers one side of rail under 15° , analogous to chisel. There are two mechanisms, how mill scale is removed, mechanical and heat mechanism.

Mechanical mechanism effects on surface by crushing force and cutting by direction of the angle. Heat mechanism causes that mill scales are shrunken and separated from material. Separating is caused by different thermal expansion mill scale from rail. Separation is supported by a fact, that mill scales are not homogeneous material, but it is created by layers of different chemical composition of different thermal expansion. Another fact is that, the small holes are present on surface of mill scales. When water goes inside these holes, it is suddenly evaporated by small steam explosion and this helps with mill scales removing. Removed mill scales falls into water reservoir together with oil, which lubricate rolling-mill stand.

1.1 Possibilities of material recycling

Mill scale is material which can be recycled under several conditions, [6]. As it was mentioned above, mill scales are inside water reservoir with oil. Simply, such mill scales can not be taken and given inside the blast furnace, due to one significant reason. There is high temperature in the blast furnace. Oil can cause fast increasing of temperature. Water change phase into steam, so the volume of water increases. Blast furnace is closed device, this fact can cause explosion of blast furnace.

For removing water and oil from mill scales special device is developed. But for precise setting of such device the knowledge of thermal conductivity of mill scales is needed.

2 Determination of isotropic thermal conductivity

It is said that several ways of determining the thermal conductivity exist, most of them were determined by experiment. We decided to use mathematical modelling, because it can save a time. We can do a lot of computations in a while and find out precise value of thermal conductivity. Generally, the thermal conductivity k ($W \cdot m^{-1} \cdot K^{-1}$) is defined like amount of heat Q (joule), which goes through surface S (m^2) into deep d (m) for a given time t (s),

$$Q = kS \frac{\Delta T}{d} t, \quad (1)$$

where T is the temperature ($^\circ C$ or K), which is measure of the kinetic energy of molecular motion; and ΔT is a temperature change i.e. difference between the final and initial temperatures for a process on a distance d .

2.1 Experiment

Mill scales were placed inside steel cylinder with very thin wall, as can be seen in Figure 1. This cylinder was isolated on the boundary, see Figure 2. Cylinder with isolation was given inside oven, see Figure 1. Temperature-sensitive elements were placed inside mill scales, on the cylinder wall, into oven and outside of the oven. There were three temperature-sensitive elements inside mill scales, the first one in the center of the cylinder, the second one in the middle of the cylinder and wall; and the last one was on the wall.

The oven was warmed up to 550° and this temperature was kept until inner temperature of the mill scales was the same as temperature of oven for all temperature-sensitive elements. The temperature loss during 1800 minute on the temperature-sensitive elements was measured.

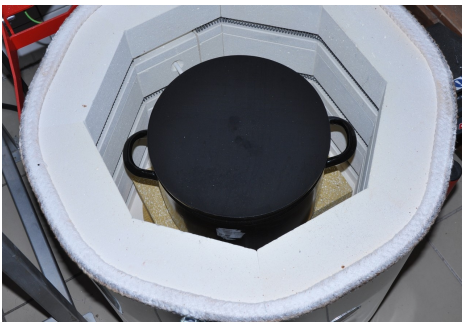


Fig. 1. Mill scales in steel cylinder within oven without isolation.



Fig. 2. Steel cylinder with isolation with temperature-sensitive elements.

3 Numerical simulations

Numerical simulation of previous experiment was done using COMSOL Multiphysic 4.2a, [1]. This software provides heat transfer equation (2), [2],

$$\rho C_p \frac{\partial T}{\partial t} = \nabla \cdot (k \nabla T) + Q, \quad (2)$$

where $C_p = 630 (J \cdot kg^{-1} \cdot C^{-1} \text{ or } J \cdot kg^{-1} \cdot K^{-1})$ is the heat capacity at constant pressure and $\rho = 2000 (kg \cdot m^{-3})$ is the density. The density was measured by weighing.

We created an axisymmetric 2D model. As a boundary condition we set analytical equation, which is polynomial interpolation of temperature from temperature-sensitive elements at the wall of steel cylinder from experiment, see Figure 2. We found out the inner temperature at given points of mill scales, see Section 2.1. Our aim was to determine the thermal conductivity from experiment

data. We used Nelder-Mead optimization method, [3]. This method needs limits for minimal and maximal value of thermal conductivity. This was derived from tabular values and our initial shoot was $k = 3$.

3.1 Results

Main aim was to find the parameter k . Secondary aim was to repeat experiment by numerical simulation, but it did not have to correspond with measured data. Success of secondary task gave us information about convergence and right setting of mathematical model. Computation time was approx. 38s.

Main goal was done by Nelder-Mead optimization method. The final value of thermal conductivity of mill scales was determined as $k = 0.12$. Recomputation of numerical model (Figure 3) with this value gave us a pretty nice correspondence with measured data.

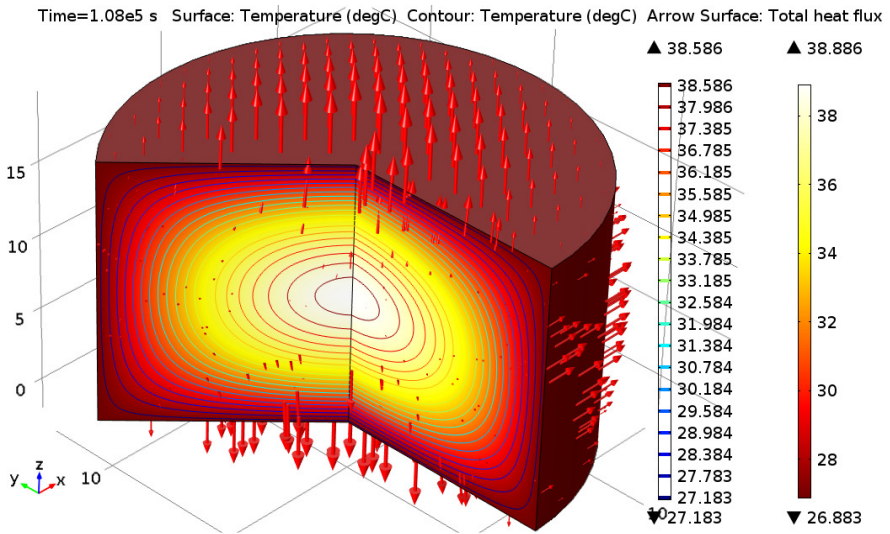


Fig. 3. Temperature field ($^{\circ}C$) at last time step $t = 1800s$. Contours at intersection show temperature field ($^{\circ}C$) inside mill scales. Arrows show total heat flux ($W \cdot m^{-2}$).

4 Conclusion

In the paper the determination of the thermal conductivity by repeating experiment by numerical simulations was introduced. We found the thermal conductivity of mill scales with water and oil. This parameter of mill scales was very important for following research of re-using mill scales for iron and steel industry. The proposed method of finding values can be used for many different coefficients at different physical models.

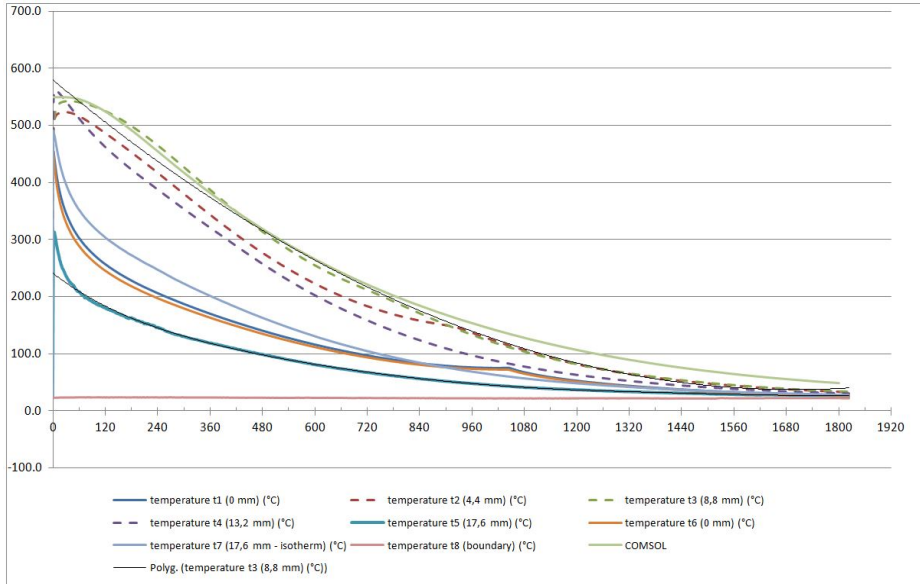


Fig. 4. Graph of dependency between the temperature ($^{\circ}C$) on the time t (s) at different elevations of the temperature-sensitive elements. Important is correspondence between COMSOL and temperature-sensitive elements 8.8mm (central point of the cylinder).

My dissertation thesis is aimed at fluid flow in porous medium. Mill scales with water and oil create a model of multiphase fluid flow in porous medium. Knowledge of parameters in equation (2) gives the final temperature field. If I know dependency between density and temperature of liquid inside porous medium, I can model fluid flow in porous media more precisely.

Acknowledgements This work has been supported by Project SP2014/204, "Parallel solving computationally intensive tasks II", of the Student Grant System, VŠB - Technical University of Ostrava.

References

1. Kittel, C.; Kroemer, H. *Thermal Physics*. W. H. Freeman and Company. Freeman and Company (1980). Chapter 14. ISBN 978-0716710882.
2. Introduction to Comsol Multiphysics (section Heat transfer). COMSOL 4.2a
3. J.E. Dennis, R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* SIAM, Philadelphia 1996
4. Degarmo, E. Paul; Black, J T.; Kohser, Ronald A. (2003), *Materials and Processes in Manufacturing* (9th ed.), Wiley, ISBN 0-471-65653-4.
5. M. C. Bagatini, V. Zymła, E. Osrio and A. C. F. Vilela, *Characterization and Reduction Behavior of Mill Scale*, Isij International, Vol. 51, No. 7, 2011, pp. 1072-1079. iew.

6. N. A. El-Hussiny, F. M. Mohamed and M. E. H. Shalabi, *Recycling of Mill Scale in Sintering Process*, *Science of Sintering*, Vol. 43, No. 1, 2011, pp. 21-31.

Optimization methods for inverse problems with energy norm

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 deals with the application of the energy norm in the inverse problem which is based on a minimization of a nonlinear least squares function. Three different parameter distribution examples are solved by the steepest descent, nonlinear conjugate gradient and Newton methods. Efficiency comparison is also presented.

1 Setting of the problem

We consider a **state problem**

$$\begin{cases} -\operatorname{div}(p\nabla u) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \quad (1)$$

where p is a parameter vector and $\Omega = \langle 0, 1 \rangle \times \langle 0, 1 \rangle$ is a unit square which is decomposed into subdomains $\Omega_i, i = 1, \dots, r$ such that $\bar{\Omega} = \bigcup_{i=1}^r \bar{\Omega}_i, \Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$.

The set \mathcal{U}_{ad} of admissible parameters is defined as follows:

$$\mathcal{U}_{ad} = \left\{ p \in L^\infty(\Omega) \mid p_{min} \leq p \leq p_{max}, p|_{\Omega_i} \in P_0(\Omega_i), i = 1, \dots, r \right\},$$

i.e. \mathcal{U}_{ad} is the set of piecewise constant functions on the partition of Ω into $\{\Omega_i\}_{i=1}^r, 0 < p_{min} < p_{max}$ are given and $p = (p_1, \dots, p_r), p_i = p|_{\Omega_i}$. In addition, material interfaces align the decomposition of Ω .

The state problem (1) is solved using the Finite Element Method. After the discretization by piecewise linear elements we obtain a linear system of algebraic equations

$$A(p)u(p) = b, \quad (2)$$

where A is a symmetric positive definite matrix, $p \in \mathbb{R}^r, b \in \mathbb{R}^n$. Inverse problem will be based on the minimization of a cost function \mathcal{J} which depends on the solution of the discretized state problem $u(p)$ and measurements $\hat{u} \in \mathbb{R}^n$ of the state variable. We suppose that the measurements are at our disposal at each node of the discretization, therefore we are able to use the energy norm (3). Now

we are ready to formulate the inverse problem in the discrete form which reads as follows:

$$(\mathbb{P})_h \begin{cases} \text{find } p^* \in \mathcal{U}_{ad} \text{ such that} \\ \mathcal{J}(p^*) \leq \mathcal{J}(p) \quad \forall p \in \mathcal{U}_{ad}, \end{cases}$$

where

$$\mathcal{J}(p) = \frac{1}{2} \langle A(p)(u(p) - \hat{u}), u(p) - \hat{u} \rangle \tag{3}$$

For the existence of a solution to the inverse problem $(\mathbb{P})_h$, we refer to [1]. Function \mathcal{J} is the energy norm and its application in identification problems is thoroughly studied in [3] and [4].

2 Numerical methods

In what follows optimization Algorithm 1 will be exploited. There are two important steps: sensitivity analysis which relies on a differentiation of the cost function and enables to find a descend direction d^k in the $k - th$ iteration and finding a step length with a line search method in order to obtain a proper parameter α (in Newton method this step is omitted).

```

Data:  $p^0$ 
Result:  $p^*$ 
 $k = 0;$ 
while  $\|p^k - p^*\| > \textit{precision}$  do
    | sensitivity analysis  $\rightarrow$  set the direction  $d^k;$ 
    | sufficient decrease of  $\mathcal{J}(p^k + \alpha d^k)$  subject to  $\alpha \in (0, 1)$  (line search) ;
    | set  $p^{k+1} = p^k + \alpha d^k;$ 
    |  $k = k + 1;$ 
end
 $p^k = p^*;$ 
    
```

Algorithm 1: The optimization algorithm

In case of the energy norm defining \mathcal{J} one can derive the gradient of \mathcal{J} for steepest descent method (SDM) so that $d^k = -\nabla \mathcal{J}(p^k)$, see e.g. [5].

$$\nabla \mathcal{J}(p^k) = \left(\frac{\partial \mathcal{J}}{\partial p_1}(p^k), \dots, \frac{\partial \mathcal{J}}{\partial p_m}(p^k) \right)^T,$$

where

$$\frac{\partial \mathcal{J}}{\partial p_j}(p^k) = -\frac{1}{2} \left\langle \frac{\partial A}{\partial p_j}(p^k)(u(p^k) + \hat{u}), u(p^k) - \hat{u} \right\rangle.$$

In the case of Newton method, the Hessian matrix $H^k := H(p^k)$ at p^k of \mathcal{J} has the following form:

$$H^k = \frac{1}{2} W^T A(p^k)^{-1} W \quad W = [w_1, \dots, w_m] \in R^{n \times m}, \quad w_j = \frac{\partial A}{\partial p_j}(p^k)u(p^k)$$

The Newton step is given by a solution of the linear system: $H^k d^k = -\nabla \mathcal{J}(p^k)$ which is solved by the preconditioned conjugate gradient method with incomplete Cholesky factorization of the matrix A .

Nonlinear conjugate gradient (NCG) methods can take various forms, in our case Fletcher-Reeves formula was used, see e.g. [5]. NCG generates orthogonalized directions d^k .

Wolfe line search was chosen in case of SDM and NCG, for $\alpha^k \in (0, 1)$ and $c_1 = 10^{-4}$, $c_2 = 0.4$ (similar values as are suggested in [5]):

$$\mathcal{J}(p^k + \alpha^k d^k) \leq \mathcal{J}(p^k) + c_1 \alpha^k \nabla \mathcal{J}(p^k)^T d^k, \quad \nabla \mathcal{J}(p^k + \alpha^k d^k) \geq c_2 \nabla \mathcal{J}(p^k)^T d^k$$

3 Numerical results

Three cases of a decomposition of $\Omega = \langle 0, 1 \rangle \times \langle 0, 1 \rangle$ marked by different colours can be seen in Fig 1. 1) domain A on the left corresponds to two parameters $p = (p_1, p_2)$, where p_1 is for the blue part and p_2 for the red part. 2) Checkerboard partition of the middle domain B. 3) The partition of C in addition contains small L-shaped regions of p_3 material. We present numerical results corresponding to these decompositions using a regular discretization 64×64 with 8192 elements.

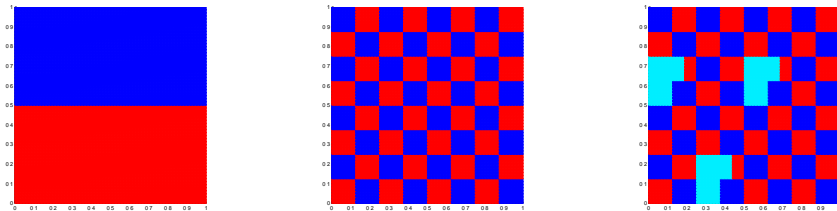


Fig. 1. Domain A on the left, domain B in the middle and domain C on the right side

Inverse problem in our examples consists of computation of the state problem in A and B with $p^* = (10, 40)$ and for C with $p^* = (10, 40, 17)$. The measurements \hat{u} were taken in every node in the domain. The aim is to find the original values p^* based on \hat{u} with initial values $p^0 = (7, 10)$ for the domains A and B, and $p^0 = (7, 10, 36)$ for the domain C.

Presented results are computed with the stopping criterion: $\|p^k - p^*\| \leq 4 \cdot 10^{-2}$ to make the graphs visually clearer. Moreover, this accuracy turns out to be sufficient for more complicated numerical simulations.

Results in each figure are arranged in a way that all x-axes represent iterations and illustrate (a) the cost function minimization with its value at the last iteration above, (b) convergence of $\|p^k - p^*\|$ with the value at the last iteration above, (c) visualizes convergence of parameters $p_i^k, \forall i$, and (d) displays cost function evaluations needed at every iteration to satisfy Wolfe line search conditions.

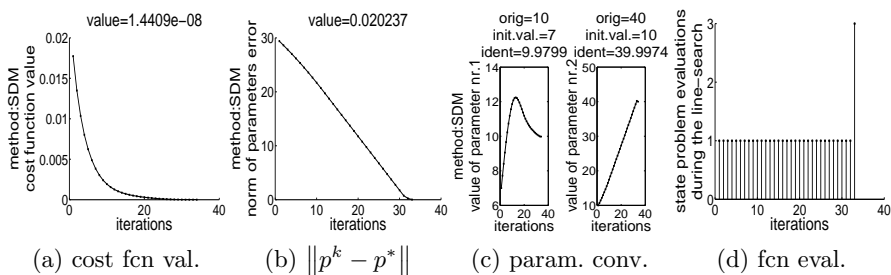


Fig. 2. Steepest descent method on Domain A

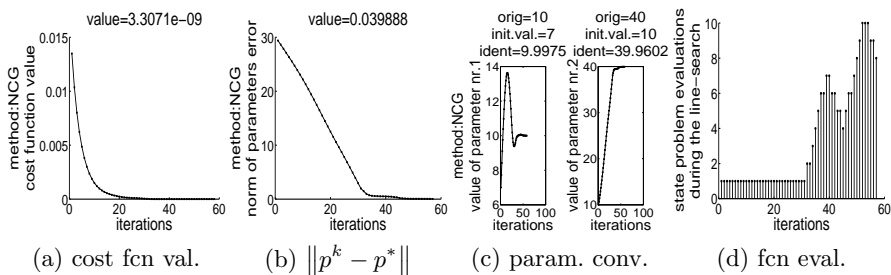


Fig. 3. Nonlinear conjugate gradient method on Domain A

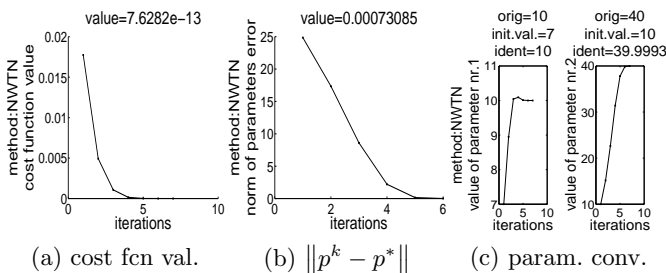


Fig. 4. Newton method on Domain A

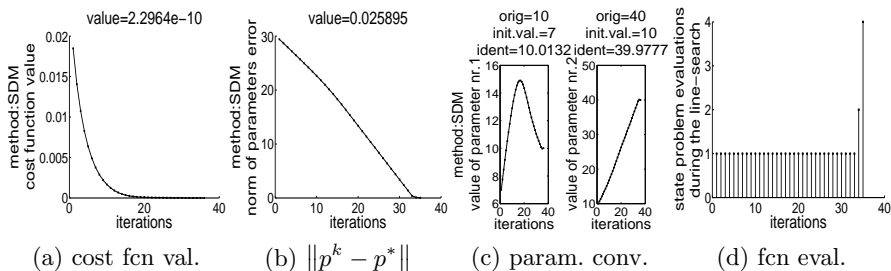


Fig. 5. Steepest descent method on Domain B

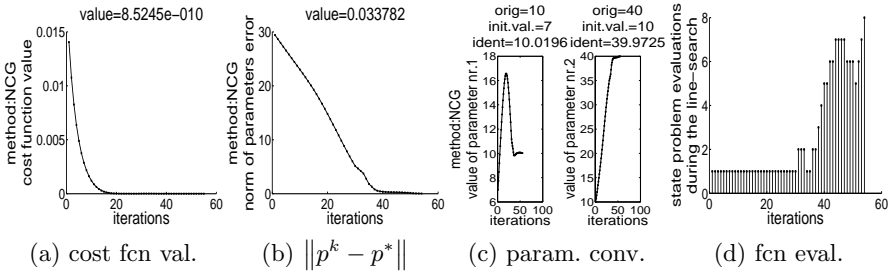


Fig. 6. Nonlinear conjugate gradient method on Domain B

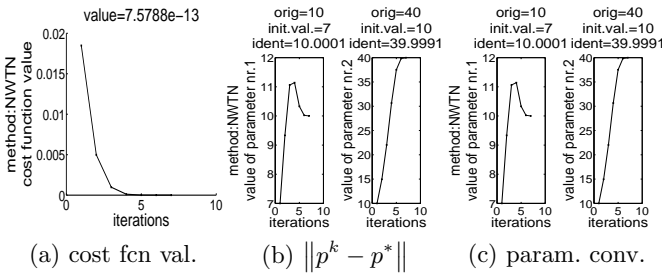


Fig. 7. Newton method on Domain B

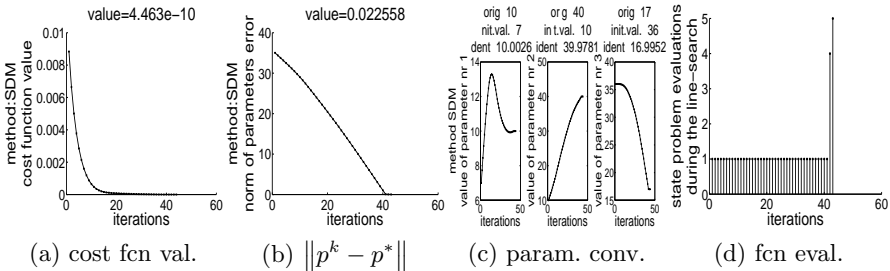


Fig. 8. Steepest descent method on Domain C

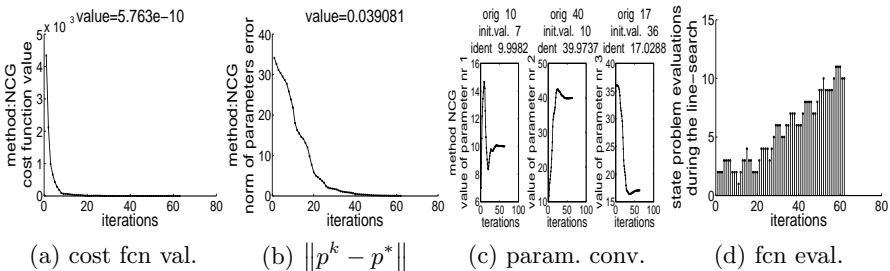


Fig. 9. Nonlinear conjugate gradient method on Domain C

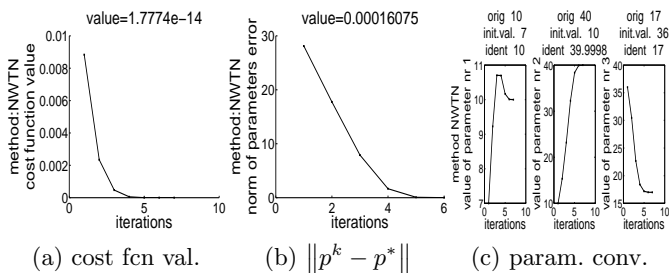


Fig. 10. Newton method on Domain C

Time to convergence [s]	domain A	domain B	domain C
number of elements	8192	8192	8192
SDM	151	162.3	269.6
NCG	617.7	750	1163
Newton	16	15.3	21

3.1 Conclusions

The contribution is focused on the solution of inverse problems based on the energy norm and compares different optimization methods. The energy norm enables cheap computations of the gradient and Hessian of the cost function. SDM gave uniform parameter error convergence, smooth parameter convergence without significant oscillations and low number of line search rule violations. NCG is not so stable, it is more line search settings dependent. Both SDM and NCG methods deteriorate their efficiency when more complicated structure containing three parameters is computed. Newton method gave fast convergence which according to the theory seems to be quadratic even if the initial guess is not close to the solution.

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) funded by Structural Funds of the European Union and state budget of the Czech Republic.

References

1. R. Hrtus, R. Blaheta, J. Haslinger: Identification problems with a priori given material interfaces (2013)
2. Haslinger J.: Identification Methods. Manuscript (2014)
3. T.F. Chan, X. Tai: Identification of discontinuous coefficients in elliptic problems using total variation regularization (2003)
4. M. S. Gockenbach, A. A. Khan: An Abstract Framework for Elliptic Inverse Problems: Part 1. An Output Least-Squares Approach (2005)
5. J. Nocedal, S.J. Wright, Numerical Optimization, Springer 2006

LiDAR data and eigenvalues

Milan Jaroš

Centre of Excellence IT4Innovations and
Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
milan.jaros@vsb.cz

Abstract. In this paper we use a statistical method the principal component analysis (PCA) to obtain the necessary information from the LiDAR data. We are able to determine the plane or the line through the use of eigenvalues and eigenvectors of the covariance matrix consisting of a small group of neighbouring points. The results of this work will be used for the numerical modelling tasks.

1 Introduction

The numerical modelling tasks, such as air pollution, river floods, sound propagation, fire spread requires the relatively accurate model of the Earth's surface. This model can be created by the data processing of the LiDAR data (= Light Detection And Ranging). The laser scanning works in the infra-red region with a wavelength of 1064 nm. The LiDAR is described in [Wehr99]. The vertical and horizontal accuracy of the measured data depends on the used technology. The vertical and horizontal accuracy is calculated to 5.1 cm and 33 cm in [Xharde06]. We need to process of the large amounts of the data to create of the simplified model.

In this work we use statistical method the principal component analysis (PCA) to obtain the necessary information from the measured data. We are able to determine the plane or the line through the use of eigenvalues and eigenvectors of the covariance matrix consisting of a small group of neighbouring points. It is also necessary to identify objects and choose an appropriate approximation. Then the new triangulation is created. The finished model is smaller and suitable for the further use.

This paper is organized as follows. In Section 2 we present the model example. In Section 3 we describe PCA. In Section 4 we identify objects. In Section 5 we make the face clustering.

2 Model example

In this paper we consider a model example, which simulates the scanned data and contains all interesting objects. The 1D object is represented by a railing. The 2D objects are the roof and the floor and the 3D objects are trees and bushes (see Fig. 1).

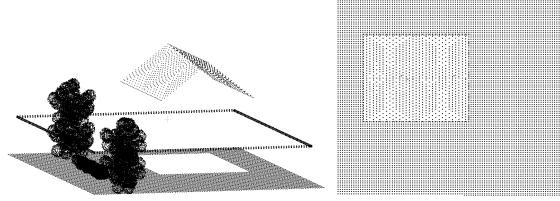


Fig. 1. Model example: LiDAR data simulation - the view from the side (left), the view from the above (right).

3 Covariance matrix and principal component analysis

Let us introduce the basic terms. We need to know the relationship between points to identify the specific object. Let $P \subset \mathbb{R}^3$ is the set of all data points. We find a set of the nearest points for the each point $p \in P$

$$N_p = \left\{ q \mid p, q \in P, |pq| = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2} < d_{3D} \right\}. \tag{1}$$

The example of point sets is in Fig. 2.

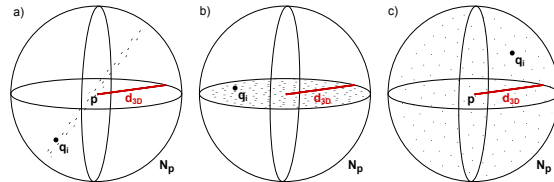


Fig. 2. The three types of the sets of points in the neighbourhood of p : a) linearity, b) planarity, c) sphericity.

Now we will show how to build the covariance matrix for each N_p . In 1962 Hu [Hu62] introduced the recognition of geometric patterns in an image by using the invariant moment. In [Maas99,Gross02] a two-dimensional moment was extended to the third dimension.

Let we consider the set N_p is solid $B \subset \mathbb{R}^3$. Then $(i + j + k)$ -th moment is defined by the equation

$$m_{ijk} = \iiint_B x^i y^j z^k f(x, y, z) dV, \tag{2}$$

where $i, j, k \in \mathbb{Z}_0^+$, $dV = dx dy dz$ and $f(x, y, z)$ is continuous function which describes the probability density of solid B . Let $f(x, y, z) = 1$ for $\forall p \in P$.

Let $(i + j + k) \leq 2$. We can define the coordinate of the center of gravity \bar{p} of the solid B

$$\bar{p} = \left(\bar{x} = \frac{m_{100}}{m_{000}}, \bar{y} = \frac{m_{010}}{m_{000}}, \bar{z} = \frac{m_{001}}{m_{000}} \right), \tag{3}$$

where m_{000} is the mass of the solid B and $m_{100}, m_{010}, m_{001}$ are the static moments with respect to the coordinate planes yz, xz, xy .

The $(i + j + k)$ -th centralized moment of the solid B is defined by

$$\mu_{ijk} = \iiint_B (x - \bar{x})^i (y - \bar{y})^j (z - \bar{z})^k f(x, y, z) dV. \tag{4}$$

After the normalization we have

$$\widehat{\mu}_{ijk} = \frac{\mu_{ijk}}{\mu_{000}} = \frac{\iiint_B (x - \bar{x})^i (y - \bar{y})^j (z - \bar{z})^k f(x, y, z) dV}{\iiint_B (x - \bar{x})^0 (y - \bar{y})^0 (z - \bar{z})^0 f(x, y, z) dV}. \tag{5}$$

After the numerical approximation of the normalized centralized moment we get

$$\widehat{\mu}_{ijk} = \frac{\sum_{n=1}^{|N_p|} (x_n - \bar{x})^i (y_n - \bar{y})^j (z_n - \bar{z})^k}{\sum_{n=1}^{|N_p|} (x_n - \bar{x})^0 (y_n - \bar{y})^0 (z_n - \bar{z})^0}, \tag{6}$$

$$\widehat{\mu}_{ijk} = \frac{1}{|N_p|} \sum_{n=1}^{|N_p|} (x_n - \bar{x})^i (y_n - \bar{y})^j (z_n - \bar{z})^k. \tag{7}$$

Finally we can put together the covariance matrix for each point $p \in P$

$$C_p = \begin{pmatrix} \widehat{\mu}_{200} & \widehat{\mu}_{110} & \widehat{\mu}_{101} \\ \widehat{\mu}_{110} & \widehat{\mu}_{020} & \widehat{\mu}_{011} \\ \widehat{\mu}_{101} & \widehat{\mu}_{011} & \widehat{\mu}_{002} \end{pmatrix}. \tag{8}$$

We can simplify the matrix C_p by substituting the moments

$$C_p = \frac{1}{|N_p|} \sum_{n=1}^{|N_p|} \begin{pmatrix} (x_n - \bar{x})^2 & (x_n - \bar{x})(y_n - \bar{y}) & (x_n - \bar{x})(z_n - \bar{z}) \\ (x_n - \bar{x})(y_n - \bar{y}) & (y_n - \bar{y})^2 & (y_n - \bar{y})(z_n - \bar{z}) \\ (x_n - \bar{x})(z_n - \bar{z}) & (y_n - \bar{y})(z_n - \bar{z}) & (z_n - \bar{z})^2 \end{pmatrix}, \tag{9}$$

$$C_p = \frac{1}{|N_p|} \sum_{n=1}^{|N_p|} \begin{pmatrix} [x_n - \bar{x}] \\ [y_n - \bar{y}] \\ [z_n - \bar{z}] \end{pmatrix} [x_n - \bar{x} \ y_n - \bar{y} \ z_n - \bar{z}], \tag{10}$$

$$C_p = \frac{1}{|N_p|} \sum_{n=1}^{|N_p|} (q_n - \bar{p})^T (q_n - \bar{p}), \tag{11}$$

where $q \in N_p$ and \bar{p} is the center of gravity of the set N_p .

The covariance matrix C_p contains only real numbers. It has dimension $n = 3$ and it is symmetric positive semi-definite, i.e. the all its eigenvalues are real with a complete set of the orthonormal eigenvectors. Then there exists an orthogonal matrix Q and a diagonal matrix D such that

$$C_p = QDQ^T. \tag{12}$$

The diagonal elements of the matrix D are the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and columns of the matrix Q are the orthonormal eigenvectors v_1, v_2, v_3 of the matrix C_p

$$C_p = (v_1^T \ v_2^T \ v_3^T) \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \tag{13}$$

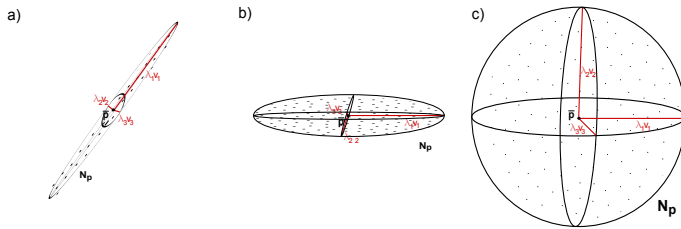


Fig. 3. Ellipsoid: a) linearity (a_{1D}), b) planarity (a_{2D}), c) sphericity (a_{3D}).

4 Objects recognition

We can represent the shape of the set N_p by the ellipsoid [Gross02, Demantke11], which shows linear (a_{1D}), planar (a_{2D}) or spheric (a_{3D}) behaviour of the neighbourhoods L_{N_p} (see Fig. 3)

$$a_{1D} = \frac{\sigma_1 - \sigma_2}{\sum_d \sigma_d}, \quad a_{2D} = \frac{\sigma_2 - \sigma_3}{\sum_d \sigma_d}, \quad a_{3D} = \frac{\sigma_3}{\sum_d \sigma_d}, \tag{14}$$

where $\sigma_j = \sqrt{\lambda_j}$ is the standard deviation and $\sum_d \sigma_d$ is the normalization coefficient.

We divide all points into the three groups by a assigned value

$$L_{N_p} = \max \{a_{1D}, a_{2D}, a_{3D}\}. \tag{15}$$

In Fig. 4 we can see points divided into the groups. If $\sigma_1 \gg \sigma_2, \sigma_3 \cong 0$, a_{1D} will be greater than the two other values and the point p is marked as linear

and it is assigned to the group P_{line} . If $\sigma_1, \sigma_2 \gg \sigma_3 \cong 0$ (a_{2D}), the point p is assigned to the group $P_{surface}$. If $\sigma_1 \cong \sigma_2 \cong \sigma_3$ (a_{3D}), the spheric point p belongs to $P_{vegetation}$. We can write

$$P_{line} = \{p|p \in P : L_{N_p} = a_{1D}\}, \tag{16}$$

$$P_{surface} = \{p|p \in P : L_{N_p} = a_{2D}\}, \tag{17}$$

$$P_{vegetation} = \{p|p \in P : L_{N_p} = a_{3D}\}. \tag{18}$$

5 Face clustering

Let $P_{surface} = \{p|p \in P : L_{N_p} = a_{2D}\}$. We find the two-dimensional $N_{p_{surface}}$ for the each point $p \in P_{surface}$

$$N_{p_{surface}} = \left\{ q \mid p, q \in P_{surface}, |pq| = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} < d_{2D} \right\}. \tag{19}$$

After assembling the covariance matrix C_p and the spectral decomposition, we get the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$, the corresponding eigenvectors v_1, v_2, v_3 and the standard deviation $\sigma_1 \geq \sigma_2 \geq \sigma_3$. We divide the points into two groups P_{planar} and $P_{boundary}$ using the criteria a_{planar}

$$a_{planar} = \frac{\sigma_3}{\sum_d \sigma_d}, \tag{20}$$

$$P_{planar} = \{p|p \in P_{surface} : a_{planar} \leq \epsilon\}, \tag{21}$$

$$P_{boundary} = \{p|p \in P_{surface} : a_{planar} > \epsilon\}, \tag{22}$$

where $\sum_d \sigma_d$ is the normalization coefficient and ϵ is the tolerance parameter. If $a_{planar} > \epsilon$, the point is on the edge and it is assigned to the group $P_{boundary}$. In the opposite case the point is on the plane and it is assigned to the group P_{planar} .

We will use the eigenvector v_3 with the smallest eigenvalue λ_3 to the face clustering. The eigenvector v_3 corresponds to the normal vector $v_3 = n_p = (n_x, n_y, n_z)$. We select any point $p_{planar} = (p_x, p_y, p_z)$ from the set P_{planar} and calculate a parameter d using the equation of plane [Sampath08]. The parameter d assigns the points to the face (see Fig. 4)

$$d_{planar} = -(n_x p_x + n_y p_y + n_z p_z), \tag{23}$$

$$P_{cluster_planar} = \{q|q \in P_{planar} \cup P_{boundary} : |n_x q_x + n_y q_y + n_z q_z + d_{planar}| < \epsilon\}, \tag{24}$$

where ϵ is the tolerance parameter.

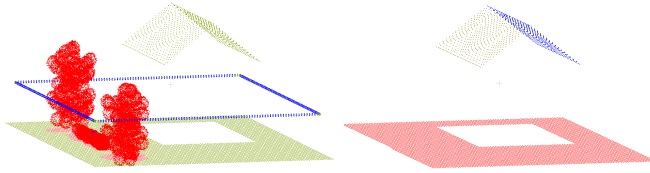


Fig. 4. (left) recognised 1D (blue), 2D (yellow) and 3D (red) objects. (right) clustered faces $P_{cluster_planar}$.

6 Future work

To create a simplified model we need to determine the boundaries of the each face. We can use a method Concave Hull (Alpha Shapes). This method is described in [Edelsbrunner83]. We can find the simplified edge using the boundaries of the face. Each edge is projected along the axis z and we get the wall. We need to fix the position of all new points by snapping. All points of the all edges serves as input of the Delaunay triangulation. We will replace all 3D objects ($P_{vegetation}$) by cubes in the next step. This gives a complete watertight model ready for numerical modelling tasks.

References

- [Wehr99] Aloysius Wehr, Uwe Lohr: *Airborne laser scanning—an introduction and overview*. ISPRS Journal of Photogrammetry & Remote Sensing, 54, 1999, p. 68–82.
- [Xharde06] R. Xhardé, B.F. Long, D.L. Forbes: *Accuracy and Limitations of Airborne LiDAR Surveys in Coastal Environments*. Geoscience and Remote Sensing Symposium, IGARSS 2006, p. 2412-2415.
- [Hu62] Ming-Kuei Hu: *Visual pattern recognition by moment invariants*. Information Theory, IRE Transactions, Volume:8, Issue:2. 1962, p. 179-187.
- [Maas99] Hans-Gerd Maas, George Vosselman: *Two algorithms for extracting building models from raw laser altimetry data*. ISPRS Journal of Photogrammetry & Remote Sensing, 54, 1999, p. 153–163.
- [Gross02] Hermann Gross, Ulrich Thoennessen: *Extraction of lines from laser point clouds*. ISPRS Commission III Photogrammetric Computer Vision PCV, 2006.
- [Sampath08] Aparajithan Sampath, Jie Shan: *Building roof segmentation and reconstruction from lidar point clouds using clustering techniques*. ISPRS Commission III/ThS-7, 2008, p. 279-284.
- [Demantke11] Jérôme Demantké, Clément Mallet, Nicolas David, Bruno Vallet: *Dimensionality based scale selection in 3d lidar point clouds*. ISPRS Commission III - WG III/2, 2011.
- [Edelsbrunner83] Edelsbrunner Herbert, Kirkpatrick David G., Seidel Raimund: *On the shape of a set of points in the plane*. Information Theory, IEEE Transactions, 29, 1983, p. 551-559.

Parallel Matrix-vector Multiplication, its Memory Requirements and Optimizations

Michal Kravčenko

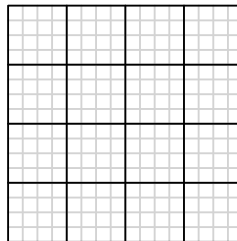
Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michal.kravcenko@vsb.cz

Abstract. The paper addresses problems arising while performing large scale parallel operations with matrices, namely a core operation of many iterative solvers: matrix-vector multiplication. In the first chapter, the parallel matrix-vector multiplication is described and the memory required to perform it is analysed. It is shown that during multiplication it matters how we divide the work among processors. The second chapter offers a particular way how to optimize the memory required by such multiplication. The third chapter shows and analyses the results of the proposed optimization. The fourth and last chapter concludes this paper.

Keywords: matrix-vector multiplication, parallelism, optimization, cyclic graph decomposition, rho-labeling.

1 Introduction

The task to parallelly compute $y = Ax$, when A is dense, has a very high memory demands. Thus it makes sense to focus on either approximating A as B while keeping the result $z = Bx$ sufficiently close to $y = Ax$ and to assign the workload among the processors in a way which minimizes the memory transfers among the processors. In this paper, we will focus on the latter. The particular algorithm to compute Ax analysed here will be such that A is represented as a block matrix. Let $A \in \mathbb{R}^{n \times n}$ and for simplicity assume that the number of processors N divides n without a remainder. Then A can be represented as $N \times N$ block matrix. Example is below.



A

Figure 1. Example of block representation of A when $N = 4$

The parallel computation of Ax can then be performed block-wise. Each processor will be assigned N such blocks and will compute its part of Ax by performing multiplication of blocks assigned to it with the required parts of x , as can be seen in figure 2.

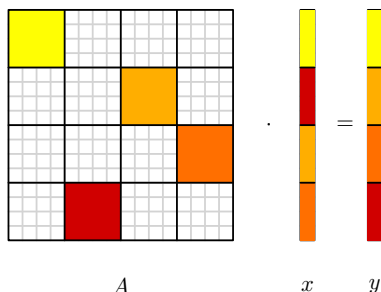


Figure 2. Illustration of a block assignment for one processor and the required parts of x and influenced parts of Ax .

During such parallel computation of Ax the processors communicate. Each processor obtains the required part of the vector x and after the part of Ax is computed it is collected. However, the way the blocks are assigned to processors is important, because it influences the amount of data required to communicate during the computation.

Particularly, figure 2 shows block assignment to a single processor which causes the maximal data transfers during the computation (it can be observed, that such block assignment requires the whole vector x and influences the whole vector Ax). On the other hand, figure 3 shows block assignment of a single processor which causes the minimal data transfer during the computation.

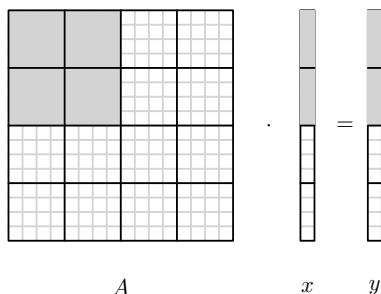


Figure 3. Example of a block assignment of A to a processor causing minimal data transfers required to compute partial Ax .

To define the 'quality' of block assignment we need to look at the block representation of A and check how many different block parts of x and Ax are required/influenced during the computation of Ax . Further in the text we will refer to this number as the number of block indices, and the lower it is the lower the memory transfers are during the computation.

2 Optimization

This chapter explains some concepts in graph theory and how they can be used to optimize the memory transfers during parallel matrix-vector multiplication. First we will define three terms required to form a final theorem, which we will not prove here, but is proven in [4].

Definition 1 (ρ -labeling). Let $G = (V, E)$, where $|E| = m$, and let there be an injective non-surjective $f : V \rightarrow \{0, 1, \dots, 2m\}$. Let's define the labeling d of an edge $\{u, v\} \in E$ induced by f as

$$d(u, v) = \min \{|f(u) - f(v)|, 2m + 1 - |f(u) - f(v)|\}. \tag{1}$$

We call d a ρ -labeling of G when $\{d(u, v) : \{u, v\} \in E\} = \{1, 2, \dots, m\}$.

Definition 2 (Rotation of a graph). Let there be graphs G and H such that $V(G) \subseteq V(H)$ and $E(G) \subseteq E(H)$. Then we define the rotation of H as a bijective function $r : V(H) \rightarrow V(H)$ such that $\forall u \in V(G) : u \rightarrow r(u)$ and $\forall \{u, v\} \in E(G) : \{u, v\} \rightarrow \{r(u), r(v)\}$.

Moreover we define an i -th rotation of G as G_i (note that $G = G_0$), where $G_i = (\{r(u) : u \in V(G_{i-1})\}, \{\{r(u), r(v)\} : \{u, v\} \in E(G_{i-1})\})$.

A graphical explanation of a graph rotation can be seen in figure 4 below.

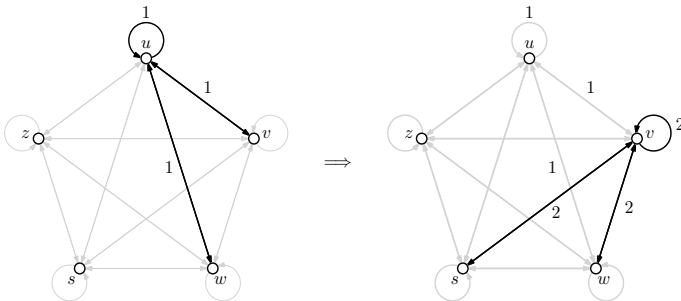


Figure 4. Illustration of a graph rotation where $H = K_5$, G is bolded out.

Definition 3 (Cyclic decomposition of a graph). Let $G \subseteq H$ be such that there $\exists k \in \mathbb{N} : k|E(G)| = |E(H)|$ and $\exists r : \bigcup_{i=0}^{k-1} E(G_i) = E(H) \wedge E(G_k = G_0 = G)$ and for any and all $i \neq j : E(G_i) \cap E(G_j) = \emptyset$. Then we call G a cyclic decomposition of H .

We defined all sufficient terms needed to form the following theorem, which is a core of the optimization algorithm described further in this chapter.

Theorem 1 (Using a ρ -labeling to cyclically decompose K_N). For simplicity let's assume that N is odd. Let $f : V(K_N) \rightarrow \{0, 1, \dots, N-1\}$ be a bijective function. Now pick any $G \subseteq K_N$ such that $|E(G)| = \frac{N-1}{2}$, G has some ρ -labeling d induced by f and define a particular rotation r such that $f(r(u)) = (f(u) + 1) \pmod N$.

Then G together with d and r define a cyclic decomposition of K_N .

Proof. Please, see [4].

To see how we can use cyclic decomposition of K_N to assign workload among processors, see figure 5.

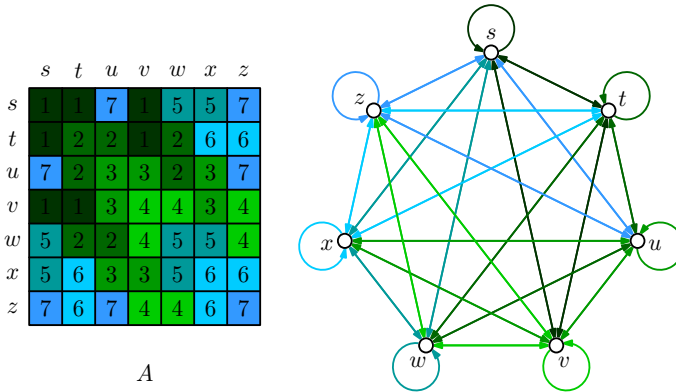


Figure 5. An example of cyclic decomposition of K_7

Figure 5 pictures a complete graph on 7 vertices together with its adjacency matrix, where the adjacency matrix can be viewed as a block representation of A for 7 processors.

Each edge has one of 7 colours/numbers and each set of edges of one colour represent one particular G_i from the cyclic decomposition. Edges of the same colour can then be looked at as the workload assignment for processors.

Since this assignment is cyclic, all processors will have the same number of block indices to work with, thus limiting a chance a bottleneck appears during communication while computing.

To minimize the number of block indices of any processor, we need to find such G which has the minimal number of vertices. The graph representation of the problem also implies that $|V(G)| \geq \sqrt{N - \frac{3}{4}} + \frac{1}{2} = \mathcal{O}(\sqrt{N})$.

2.1 Optimization algorithm

The task to find G with the minimal number of vertices for any given N is an NP-hard problem. However, a very simple heuristic algorithm is able to find G with $\mathcal{O}(\sqrt{N})$ vertices (*yet to be proved*, so far true for $N < 10000$). The algorithm follows:

1. Begin with empty G . Pick any edge in K_N and add it to G .
2. Choose a $u \in V(K_N)$ such that if u was added to G together with all its incident edges, then $|\{d(v, w) : \{v, w\} \in E(G)\}|$ would be maximal.
3. Go to point 2, unless G contains a subset of its own edges of cardinality $\frac{N-1}{2}$ which can be ρ -labelled.
4. Pick any subset of edges of G of cardinality $\frac{N-1}{2}$ which can ρ -labelled and use it for cyclic decomposition of K_N .

3 Experiments & Results

In this chapter the optimized block assignment is compared to non-optimized **random** block assignment of workload. It merely contains a list of figures representing measurements with commentaries.

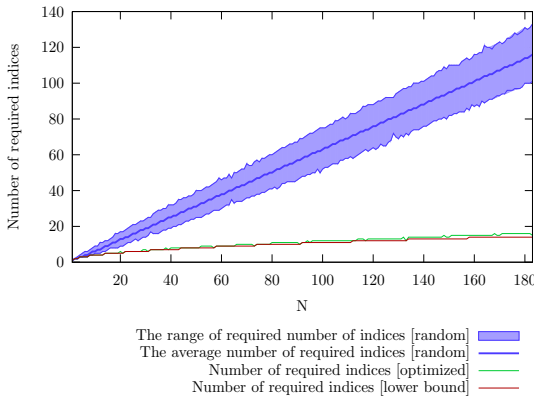


Figure 6. Ranges obtained by random block assignment is pictured in blue colour. Results from optimized algorithm using Graph Theory is in green colour. Theoretically minimal number of indices is pictured in the red colour.

The blue area in figure 4 describes the range of number of block indices obtained by random block assignment among the processors. It can be seen that the average number of indices roughly lies along the line $\frac{N}{2}$ and thus it causes the average memory transfers of $\frac{n}{N} \cdot \frac{N}{2} = \frac{n}{2}$ per processor during the computation.

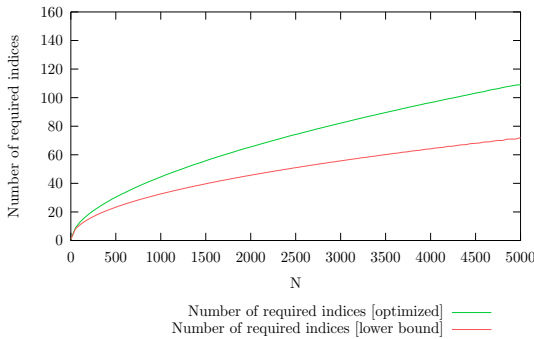


Figure 7. Results from optimized algorithm using Graph Theory is in green colour. Theoretically minimal number of indices is pictured in the red colour.

Figure 5 describes the properties of optimized block assignment more clearly than figure 4. One can observe that the optimized number of indices can be bounded from above by $2 \cdot$ (lower bound). As was stated in the previous chapter, the lower bound on the number of block indices is $\mathcal{O}(\sqrt{N})$, thus it can be concluded that the described optimization technique is **asymptotically** optimal.

4 Conclusion

With the evolution of computational power in recent years we gained the ability to effectively work with large scale matrices, however a new bottleneck has been reached and that is the memory required. In the first chapter, this paper established that it is not futile to think of ways how to decrease the required communication between processors during a parallel computation of a matrix-vector multiplication.

The second chapter reminded us of graph theory and that it can be used to decide the workload assignment among the processors. The particular optimization algorithm was not described in its full detail, but readers can easily use it as a foundation for their own experiments and implementations.

The third chapter presented some measured results obtained from our own experiments. Described optimization algorithm was compared to a random workload assignment and to a theoretical lower bound which we know cannot be surpassed. The optimization algorithm shows a valuable property, which is that for a number of processors $N \leq 5000$ it produces asymptotically optimal results and leads to a communication between processors with a factor of $\mathcal{O}\left(\frac{n}{\sqrt{N}}\right)$ during a parallel matrix-vector multiplication.

The described optimization algorithm is currently being implemented into fast Boundary Element Method solvers in an IT4I project.

References

1. Jakl, O.; *Paralelní a Distribuované systémy*, lecture slides, VŠB-TU Ostrava, 2008.
2. Kovář, P.; *Teorie grafů*, Graph Theory script, VŠB-TU Ostrava, 2012, available at http://home1.vsb.cz/~kov16/files/skriptum_teorie_grafu_rozsirene.pdf.
3. Lukáš, D.; *Superpočítání s vtípem*, lecture composed for the anniversary of the department of Applied Mathematics, VŠB-TU Ostrava, 2012.
4. Kravčenko, M.; *Parallel Boundary Element Methods*, diploma thesis, VŠB-TU Ostrava, 2013. Supervisor doc. Ing. Dalibor Lukáš Ph.D., available at <http://www.fe.i.vsb.cz/export/sites/fei/k470/cs/theses/mgr/2013/KRA568.pdf>.

Parallel k -Means Clustering in Computational Astrophysics

Lukáš Malý

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.maly@vsb.cz

Abstract. We introduce problems of computational astrophysics where during the simulations of plasma using particle-in-cell model we need to reduce the total number of particle which growth in consequence of collisions. To the purpose of merging of particles we introduce the k -means clustering algorithm. We present the classical Lloyd's algorithm in brute-force version and also suggest version with kd -tree structure. This approach brings significant speed up into the clustering process. Next we optimize the algorithm for demands of astrophysics simulations using vectorization, OpenMP parallelization and GPU acceleration. At the end we present numerical results where compare timing for different implementations.

1 Introduction

Astrophysical plasmas are ionized gasses with physical behaviours, which are studied as part of computational astrophysics. They contains charged particles, ions and electrons. When the charges move they generate electrical currents with magnetic fields, and as a result, they are affected by each other's fields. This governs their collective behaviour. Under the influence of magnetic field, the particles may form structures such as filaments, beams and double layers. There are several types of common astrophysical plasmas as Solar wind, Solar corona, Solar core and others. Basic different between these plasmas is in density of particles.

To simulate behaviour of plasma, one can use two different approaches. Fluid dynamics or particle-in-cell (PIC) model. When we decide to use PIC model, it is necessary to work with hundreds of billion of computational particles, which consist of many elementary particles - electrons, ions, or photons. During the physical simulations with particles come about splitting of them in consequence of collisions. Splitting of the particles leads to exponential growth of their number and to memory overflow. Since we want to avoid the memory overflow and we also want to keep a physical correctness of the simulation we have to reduce the total number of particles in a sensible way. To do it, we will merge particles which are close together. For this purpose we will touch k -means clustering method.

Since we want to simulate the astrophysical plasmas we think about computational particles in 6-dimensional space, $\mathbf{p} \in \mathbb{R}^6$, where each particles is

defined by its spatial components, vector $\mathbf{s} \in \mathbb{R}^3$, by its velocity components, vector $\mathbf{v} \in \mathbb{R}^3$, and by its weights, scalar w . The weights determine number of elementary particles. So we have $\mathbf{p} = (\mathbf{s}, \mathbf{v}) = (s_x, s_y, s_z, v_x, v_y, v_z)$ and w . To maximal utilization of computer memory, we want to keep constant number of particles during whole physical simulation. And from the essence of the simulation, it means that will want to keep about 90 percent of the original number of particles during the merging.

2 k -Means Clustering

Generally, k -means clustering is a method for partitioning set of points in d -dimensional space. The method aims to partition n data-points into k clusters in which each data-point belongs to the cluster with the nearest mean/center/-centroid. The cluster is then represented by its mean and can be viewed as a single point. This results in a partitioning of the data space into Voronoi cells (see Fig. 1).

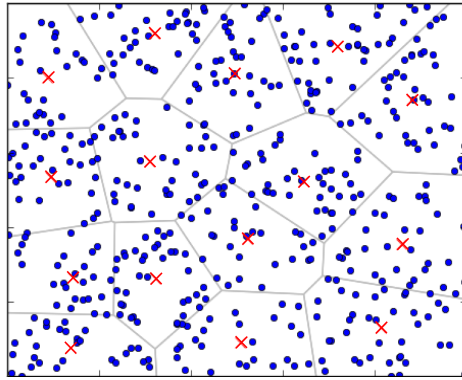


Fig. 1. Example of Voronoi cells in 2D, 100 data points clustered by k -means clustering into 15 clusters.

The partitioning into the clusters is done to minimize the residual sum of squares (RSS). So given a set of n data points $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$ the k -means clustering aims to partition data points into k sets $S = \{S_1, S_2, \dots, S_k\}$. Then the RSS is defined as

$$\arg \min_S \sum_{i=1}^k \sum_{\mathbf{p}_j \in S_i} \|\mathbf{p}_j - \boldsymbol{\mu}_i\|^2, \quad \boldsymbol{\mu}_i = \frac{1}{|S_i|} \sum_{\mathbf{p}_j \in S_i} \mathbf{p}_j,$$

where $\boldsymbol{\mu}_i$ is the mean of points in S_i and $|S_i|$ denotes the number of points in set S_i . RSS is also sometimes called as loss quality measure. In practice we can use some other methods how to measure the quality of clustered data depending on application needs, [6].

In general, the problem of clustering is computationally difficult, in fact it is NP-hard. However, there are efficient heuristic iterative algorithms that are commonly employed and converge quickly to a local optimum.

2.1 Lloyd's Algorithm

Lloyd's algorithm is standard heuristic algorithm which is spread used. It uses an iterative refinement technique. In Algorithm 1 we can see modified Lloyd's algorithm for data-points with weights, where the weights influence positions of centres, means of the clusters. This algorithm is used within the astrophysics simulations for merging of particles.

Algorithm 1 Lloyd's algorithm with weights.

```

function K-MEANS( $(\mathbf{p}_1, \dots, \mathbf{p}_n), \mathbf{w}^{in}, k$ )
   $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k) \leftarrow \text{Forgy}((\mathbf{p}_1, \dots, \mathbf{p}_n), k)$            (initial set of centroids)
  while relative RRS > threshold do
    for  $i \leftarrow 1, \dots, k$  do
       $S_i \leftarrow \{\}$ 
       $w_i^{out} \leftarrow 0$ 
    end for
    for  $j \leftarrow 1, \dots, n$  do
       $i \leftarrow \arg \min_{\bar{i}} \|p_j - \mu_{\bar{i}}\|$ 
       $S_i \leftarrow S_i \cup p_j$                                        (reassignment of data-points)
       $w_i^{out} \leftarrow w_i^{out} + w_j^{in}$ 
    end for
    for  $i \leftarrow 1, \dots, k$  do
       $\boldsymbol{\mu}_i \leftarrow \frac{1}{w_i^{out}} \sum_{\mathbf{p}_j \in S_i} \mathbf{p}_j \times w_j^{in}$            (recomputation of centroids)
       $w_i^{out} = \sum_{\mathbf{p}_j \in S_i} w_j^{in}$ 
    end for
  end while
  return  $((\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k), \mathbf{w}^{out})$ 
end function

```

The algorithm eventually converge, although the result is not necessarily the minimum of the RSS. That is because of non-convexity of the problem and because the algorithm is heuristic, converging to local minimum. However, practice shows that the results are reliable and suitable all the time.

In our implementation we measure the relative change of RSS, nevertheless there exist more alternatives [3]. Very important part of the algorithm is the initial set of clusters. Again, there are more different techniques how to distribute the initial clusters, [3]. For our purpose, the best choice is Forgy method, where the clusters are set randomly over all input data points.

Lloyd's algorithm works well for demands of astrophysics simulations. Since we consider the cluster ratio about 90 percent, the data-points in 6 dimensional space, and use Forgy method for initial distribution, the algorithm converge in about 4 to 10 iteration under reasonable relative accuracy. The only trouble is an enormous computation. In the algorithm, the distance between each pair point-cluster has to be computed within each iteration. In our set up with mentioned parameters, to compute the distance between 2 points,

$$d = \|\mathbf{x} - \mathbf{y}\|^2 = (x_1 - y_1)^2 + \dots + (x_6 - y_6)^2,$$

and to define new nearest cluster,

$$d_{min} > d,$$

gives 18 FLOP (FLoating-point OPERations) – 6 subtraction, 6 multiplication, 5 sums, 1 comparison – without I/O etc. For our tested numerical examples, where we cluster $2.4 \cdot 10^6$ points into $2 \cdot 10^6$ clusters, the mentioned operations will cost 86.4 TFLOP for one iteration of k -means. This is the reason why we need to use different approach or at least parallel implementation of this so-called naïve brute force algorithm.

2.2 kd -Tree

The kd -tree is a space-partitioning data structure for organizing points in a space, it is a special case of binary space partitioning tree, [7]. When the tree is constructed, we partition the space according to some, mostly axis-orthogonal, hyperplane. Typically we iterate the dividing axis on each level of the tree and use the median of i th coordinate of the associated points to determine the dividing hyperplane. Given n data points, this produces a balanced binary tree with $O(n)$ nodes and $O(\log n)$ depth.

The kd -tree data structure is very useful for several applications, such as searches involving a multidimensional search key, e.g. range searches and nearest neighbour searches. Using it within the k -means algorithm we can prune some clusters during the assignment process, so we don't have to compare each pairs point-cluster, which distinctively decreases computation of distances compare to brute force algorithm. The significant benefit of this approach can be observed in numerical results. You can find more about kd -tree in clustering method at [7].

3 Parallel Implementation

Our aim was to produce as fast method as possible to make astrophysical simulation faster. So we tested and implemented several improvements to reach the maximal efficiency of our algorithms. For k -means clustering using kd -tree structure we based our implementation on kmpp library [5]. For GPU accelerated runs we used only naïve brute force algorithm, which is very easy for parallelization.

The core of each k -means clustering algorithm is computing distances between two points. We vectorized this operation using SSE/AVX instructions for Intel processors. The vectorization allows to operate with 2 numbers in the same time and brings the speed up even a little more than 2 because of better memory throughput.

Next optimization step was to use OpenMP parallelization on kd -tree version of the algorithm. The main part of the algorithm is tree traversing and pruning the clusters and these operations are not very suitable for parallel implementation. We are limited, because the data in the tree are dependent and we need to have a barrier on each level. But still, there is a possibility to use openmp tasks together with openmp taskwait, which brings very nice improvement. Effective, but not so significant, can be also to parallelize all loops over clusters and data-points within the code.

Another approach is to try to use GPU acceleration. Unfortunately, the kd -tree version of k -means is quite unsuitable for such a parallelization because of its complexity. That is the reason why only the brute-force algorithm was implemented to use GPU. It was done using OpenCL.

4 Numerical Results

We tested our implementations on example, which is typical for demands of astrophysics simulations, where we cluster $2.4 \cdot 10^6$ points into $2 \cdot 10^6$ clusters, in 6-dimensions and with weights. All numerical experiments were run at NBI cluster which is composed of several hundreds of CPU cores AMD Opteron 6272, where each node has supporting GPU device AMD Radeon HD 7850.

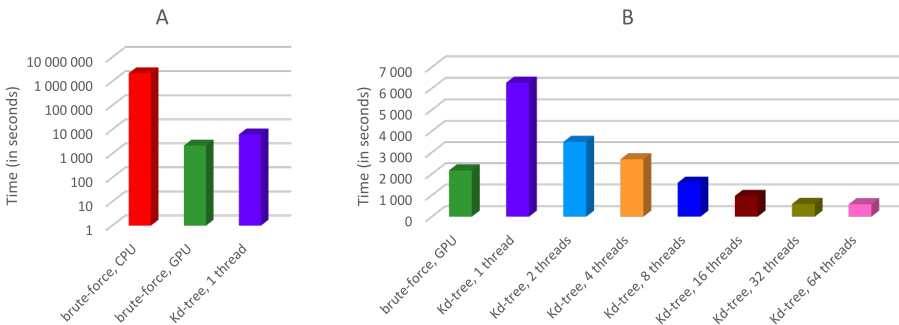


Fig. 2. Timing results for clustering of $2.4 \cdot 10^6$ of data-points into $2 \cdot 10^6$ clusters using different implementations of k -means. A: Comparison between the brute-force algorithm on CPU and GPU, and the kd -tree algorithm in sequential case. Time axis is in logarithmic scale. B: Comparison between GPU parallel implementation and CPU kd -tree implementation with OpenMP parallelization.

In Figure 2A we present time results, where in Figure 2A is a comparison between 3 basic types of implementations, brute-force CPU, brute-force GPU and kd -tree, where we can see the significant benefit of kd -tree approach.

In Figure 2B we can see comparison between GPU implementation and kd -tree implementation with OpenMP. kd -Tree algorithm is surprisingly well scalable in spite of complex parallelization of tree traversing. This scalability works only up to 64 threads, than overhead stops the speed up and algorithm starts to be a little slower.

Anyway, to conclude our work, we reached great improvements just by implementation of kd -tree version of the algorithm, where we got speed up of factor about 10^3 , compared to original brute force algorithm. And the parallelization using OpenMP brought more. Both parallel implementations are time-comparable.

Acknowledgements

The work on this topic was initiated during the author's stay at the Niels Bohr Institute at University of Copenhagen from July 2013 to August 2013 within PRACE Summer of HPC. Together with the colleague Mordechai Butrashvily from Tel-Aviv University we worked on the project Parallel k -Means Clustering in Computational Astrophysics during the summer school and most of the above presented results are joint work. For the connection with real application of plasma simulations are responsible our mentor Mads Kristensen and Jacob Trier Frederiksen from NBI. We did not presented our results so far, but now we are working on paper where we will include our implementations into real simulations of plasmas and we plane to present gained result at a conference.

References

1. Haugboelle, T., Frederiksen, J. T., Nordlund, A.: Photon-Plasma: a modern high-order particle-in-cell code. *Phys. Plasmas*, vol. 20, 2013
2. Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, Ch. D., Silverman, R., Wu, A. Y.: An Efficient k -Means Clustering Algorithm: Analysis and Implementation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, 2002.
3. Faber, V.: Clustering and the Continuous k -means Algorithm, Los Alamos Science, vol. 22, pp. 138-144, 1994.
4. Moore, A.: Very Fast EM-Based Mixture Model Clustering Using Multiresolution kd -Trees, *Proc. Conf. Neural Information Processing Systems*, 1998.
5. Arthur, D., Vassilvitskii, S.: k -means ++ : The Advantages of Careful Seeding, in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, vol. 8, pp. 111.
6. Ackerman, M., Ben-David, S.: Measures of clustering quality: A working set of axioms for clustering, in *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*, 2009, pp. 1211-128.
7. Bentley, J. L.: Multidimensional binary search trees used for associative searching, *Commun. ACM*, vol. 18, no. 9, pp. 509-517, Sep. 1975.

Parallel solution of the time dependent wave equation

Michal Merta

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`michal.merta@vsb.cz`

Abstract. This paper deals with a parallel solution of time dependent wave equation using Galerkin boundary element method. The formulation of the problem is based on the retarded potential boundary integral equations. For its discretization we use a novel approach presented in [9]. This paper describes its efficient parallel implementation in our library BEM4I.

1 Introduction

Efficient modelling of the acoustic and electromagnetic wave scattering plays an important role in many industrial areas, e.g., in nondestructive testing, seismology, or ultrasonic imaging. The boundary element method (BEM) is well suited for the solution of this kind of problems, since it reduces a problem on an unbounded domain to the boundary of the scatterer. However, application of the BEM on the governing hyperbolic time dependent wave equation is not as straightforward as in the case of the elliptic problems.

Besides the collocation methods, which are hard to analyse mathematically, there are two major approaches in the field of the time dependent boundary integral equations [2, 3]: the convolution quadrature method introduced by Lubich [5] and the Galerkin method presented by Bamberger and Ha Duong [1]. In this paper we deal with the latter one. The main drawback of the Galerkin formulation is a need for a special quadrature when computing elements of the system matrices, since the integration domains are intersections of boundary elements with a discrete light cone. Therefore we implement the new approach introduced by Sauter and Veit [9] which uses compactly supported, infinitely smooth basis functions to overcome this problem.

The outline of this paper is as follows: in Section 2 we describe the model problem and the associated boundary integral formulation, in Section 3 we present the implementation in the BEM4I library. Finally, in Section 4 the results of the numerical experiments are presented.

2 Boundary integral formulation of the sound scattering problem

Let us consider the following wave scattering problem on an unbounded domain. Let $\Omega^i \in \mathbb{R}^3$ be a open bounded domain representing a scatterer and $\Omega^e := \mathbb{R}^3 \setminus \bar{\Omega}^i$ be an exterior to the domain. Let $\Gamma := \partial\Omega^i = \partial\Omega^e$ be a common boundary of Ω^i and Ω^e , and let u^{inc} be the incident wave. The scattered wave u^{sc} can be found by solving the following initial boundary value problem for the wave equation

$$\left\{ \begin{array}{ll} \frac{1}{c^2} \frac{\partial^2 u^{sc}}{\partial t^2}(t, x) - \Delta u^{sc}(t, x) = 0 & \text{in } [0, T] \times \Omega^e, \\ u^{sc}(0, x) = 0 & \text{in } \Omega^e, \\ \frac{\partial u^{sc}}{\partial t}(0, x) = 0 & \text{in } \Omega^e, \\ \frac{\partial u^{sc}(t, x)}{\partial n} = -\frac{\partial u^{inc}(t, x)}{\partial n} & \text{on } [0, T] \times \Gamma. \end{array} \right. \tag{1}$$

Here $T \in \mathbb{R}_+$, c is the speed of sound in a given medium (which is assumed to be equal to one in what follows), and n is an unit outward normal to the $\partial\Omega^i$. The Neumann boundary condition corresponds to the acoustically hard scatterer. In the case of Dirichlet boundary conditions, a scatterer is said to be acoustically soft.

The value of the solution u^{sc} in an arbitrary point $x \in \Omega^e$ can be obtained using the double layer representation

$$u^{sc}(t, x) = \mathcal{D}\Phi(t, x) = -\frac{1}{4\pi} \int_{\Gamma} \frac{n_y(x-y)}{\|x-y\|} \left(\frac{\Phi(t-\|x-y\|, y)}{\|x-y\|^2} + \frac{\partial\Phi(t-\|x-y\|, y)}{\|x-y\|} \right) d\Gamma_y, \quad (t, x) \in [0, T] \times \Omega,$$

where Φ is an unknown density function. We search for the Φ using the hyper-singular operator

$$Wv(t, x) := \lim_{x^* \in \Omega \rightarrow x} n_x \nabla_{x^*} Dv(t, x^*), \quad (t, x) \in [0, T] \times \Gamma,$$

obtaining the boundary integral equation

$$W\Phi = g, \quad [0, T] \times \Gamma.$$

For the discretization we use the Galerkin method with piece-wise linear basis and testing functions. A discrete representation of the unknown density function Φ then reads as

$$\Phi_h(t, x) := \sum_{i=1}^L \sum_{j=1}^M \alpha_i^j \varphi_j(x) b_i(t), \quad (t, x) \in [0, T] \times \Gamma,$$

where $\alpha_i^j \in \mathbb{R}$, and $\varphi_j, j \in \{1, \dots, M\}$ are spatial basis functions and $b_i, i \in \{1, \dots, L\}$ are temporal basis functions.

After discretization we obtain the linear system $A\alpha = g$, with elements of the matrix A given by

$$\begin{aligned}
 A_{k,i}(j,l) := & \int_0^T \int_{\Gamma} \int_{\Gamma} \frac{n_x \cdot n_y}{4\pi|x-y|} \varphi_j(y) \ddot{b}_i(t-|x-y|) \varphi_l(x) \dot{b}_k(t) \\
 & + \frac{\text{curl}_{\Gamma} \varphi_j(y) \cdot \text{curl}_{\Gamma} \varphi_l(x)}{4\pi|x-y|} b_i(t-|x-y|) \dot{b}(t) d\Gamma_y d\Gamma_x dt,
 \end{aligned}
 \tag{2}$$

and elements of the vector g given by

$$g_k(l) := \int_0^T \int_{\Gamma} g(x,t) \varphi_l(x) \dot{b}_k(t) \Gamma_x dt,$$

for $i, k \in \{1, \dots, L\}, j, l \in \{1, \dots, M\}$. The resulting system matrix possesses a block-Hessenberg structure with $L \times L$ sparse blocks of order M . GMRES algorithm is used to solve the linear system.

An evaluation of the integral in Equation (2) is not straightforward as it involves an integration over an intersection of mesh elements and discrete light cone [3, 4]. We treat this problem using a recently introduced method that employs smooth temporal basis functions. A detailed description of this approach is beyond the scope of this article, therefore we refer a reader to [9].

3 Implementation in the BEM4I library

A parallel solver for scattering problems based on the time dependent wave equation has been implemented in the BEM4I library [8]. A library is written in C++ in an object oriented manner. Some of the techniques used for its development are treated, e.g., in [6, 7]. The core of the library consists of set of classes responsible for assembly of system matrices. The `BESpace` class keeps the information about the spatial as well as temporal basis functions. The main

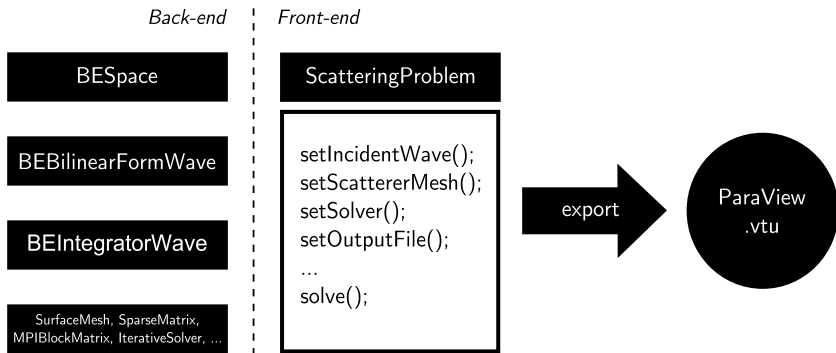


Fig. 1. Structure of the solver for wave scattering problems in the BEM4I library

purpose of the `BEBilinearFormWave` class is the discretization of boundary operators. The assembly is parallelized using MPI and OpenMP at this level. The class `BEIntegratorWave` carries out the assembly of element system matrices. The user controls the process of solution using the interface provided by the `ScatteringProblem` class. The result can be exported to the ParaView `.vtu` format (see Fig. 1).

4 Numerical experiments

The numerical experiments were carried out on Anselm cluster at IT4Innovations National Supercomputing Centre. The cluster consists of 180 non-accelerated nodes, each equipped with two Intel Xeon E5-2665 8-core processors and 64 GB of RAM. The nodes are interconnected by the InfiniBand network.

Dependence of the L_2 -error of numerical solution on the temporal discretization is depicted in the left-hand side of Fig. 2. Using piece-wise linear temporal basis functions we obtain a linear convergence.

The right-hand side of the Fig. 2 depicts the strong scalability of the system matrix assembly. The test was performed using the problem discretized to 5120 surface elements and 30 time-steps. We obtain almost optimal scalability up to 256 cores. In Fig. 3 the wave scattered off the L-shaped domain is depicted.

5 Conclusion

We have presented a parallel solver for sound hard scattering problems modelled by the time-dependent wave equation using the novel approach to overcome the discretization problems introduced by Sauter and Veit [9]. Among possible future applications we list, e.g., the shape optimization based on the wave equation.

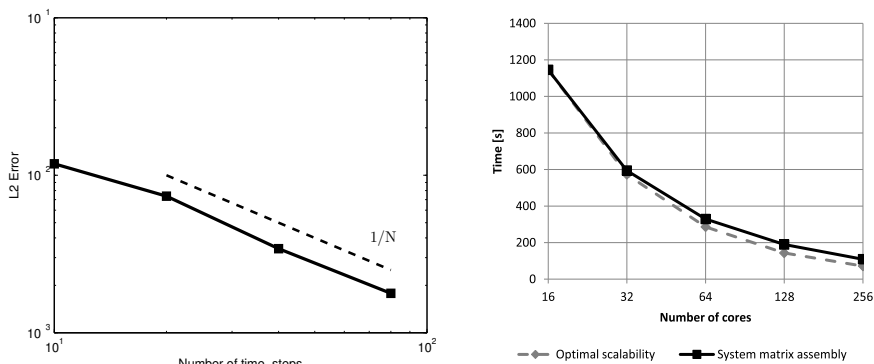


Fig. 2. Convergence of the numerical solution (left), scalability of system matrix assembly (right)

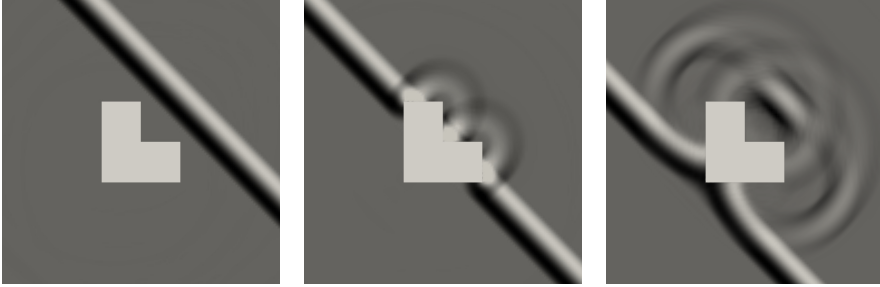


Fig. 3. Sound wave scattering off the L-shaped domain

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.

The presented results will be summarized in the paper Veit, A., Merta, M., Zapletal, J. *Numerical Solution of Time-Domain Boundary Integral Equations Arising in Sound-Hard Scattering* which is currently in preparation.

Development of the BEM4I library is a joint work with Jan Zapletal.

References

1. Bamberger, A., Ha Duong, T. *Formulation variationnelle espace-temps pour le calcul par potentiel retardé de la diffraction d'une onde acoustique (I)*. Math. Methods Appl. Sci. 8, 1986, pp. 405–435.
2. Costabel, M.: Time-Dependent Problems with the Boundary Integral Equation Method. In Stein, E., De Borst, R., Hughes, T.: *Encyclopedia of Computational Mechanics*, John Wiley & Sons, Ltd., 2004.
3. Ha-Duong, T.: On Retarded Potential Boundary Integral Equations and their Discretisation. In *Topics in Computational Wave Propagation*, Springer, 2003, pp. 395–405.
4. Glaefke, M., *Adaptive Methods for Time Domain Boundary Integral Equations*. Diploma thesis. Brunel University, London, 2012.
5. Lubich, C.: *On the multistep time discretization of linear initial-boundary value problems and their boundary integral equations*. Numer. Math., 67, 1994, pp. 365–389.
6. Lukáš, D., Kovář, P., Kovářová, T., Merta, M.: *A parallel fast boundary element method using cyclic graph decompositions*. Submitted to Numerical Algorithms.
7. 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.

8. Merta, M., Zapletal, J.: *BEM4I*. <http://industry.it4i.cz/en/products/bem4i/>.
9. Sauter, S., Veit, A.: *A Galerkin method for retarded boundary integral equations with smooth and compactly supported temporal basis functions*. Numer. Math., 123, 2013, pp. 145–176.

Semi-Smooth Newton Method Combined with TFETI for Contact Problems with Coulomb Friction

Kristina Motyčková

FEECS,

VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
kristina.motyckova@vsb.cz

Abstract The contribution deals with contact problems for two elastic bodies with friction. The semi-smooth Newton method is used to find the solution. We present active sets algorithm combined with TFETI. A real word problem concerning the clamp joint demonstrates the behaviour of the algorithm.

Keywords: contact problems, semi-smooth Newton method, TFETI, clamp joint

1 Introduction

Let us start with the discrete formulation of the problem, for the continuous setting see [4,5]. Consider

$$\mathbf{K}\mathbf{u} - \mathbf{f} + \mathbf{N}^\top \boldsymbol{\lambda}_\nu + \mathbf{T}^\top \boldsymbol{\lambda}_t = \mathbf{0}, \quad (1.1)$$

$$\mathbf{N}\mathbf{u} - \mathbf{d} \leq \mathbf{0}, \boldsymbol{\lambda}_\nu \geq \mathbf{0}, \boldsymbol{\lambda}_\nu^\top (\mathbf{N}\mathbf{u} - \mathbf{d}) = \mathbf{0}, \quad (1.2)$$

$$\left. \begin{aligned} |\lambda_{t,i}| &\leq \mathcal{F}_i \lambda_{\nu,i} \\ |\lambda_{t,i}| < \mathcal{F}_i \lambda_{\nu,i} &\Rightarrow u_{t,i} = 0 \\ |\lambda_{t,i}| = \mathcal{F}_i \lambda_{\nu,i} &\Rightarrow \exists c_t \geq 0 : u_{t,i} = c_t \lambda_{t,i} \end{aligned} \right\} \quad i = 1, \dots, m. \quad (1.3)$$

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.

Next we use the equivalent formulation of the previous problems as the systems of non-smooth equations. We introduce the projection mappings

$$\mathbf{P}_{\mathbb{R}_+^m} : \mathbb{R}^m \mapsto \mathbb{R}_+^m \quad \mathbf{P}_{\mathbf{A}(\mathbf{r})} : \mathbb{R}^m \mapsto \mathbf{A}(\mathbf{r})$$

with $\mathbb{R}_+^m := \{\boldsymbol{\mu} \in \mathbb{R}^m : \boldsymbol{\mu} \geq \mathbf{0}\}$, $\mathbf{A}(\mathbf{r}) := \{\boldsymbol{\mu} \in \mathbb{R}^m : |\boldsymbol{\mu}| \leq \mathbf{r}\}$, respectively and $\mathbf{r} \in \mathbb{R}^m$, $\mathbf{r} \geq \mathbf{0}$. The definitions of the components of $\mathbf{P}_{\mathbb{R}_+^m}$, $\mathbf{P}_{\mathbf{A}(\mathbf{r})}$ are based on

the max-function in \mathbb{R}^1 :

$$(\mathbf{P}_{\mathbb{R}_+^m})_i(\boldsymbol{\mu}) = \max\{0, \mu_i\}, \tag{1.4}$$

$$(\mathbf{P}_{\Lambda(r)})_i(\boldsymbol{\mu}) = \max\{0, \mu_i + r_i\} - \max\{0, \mu_i - r_i\} - r_i. \tag{1.5}$$

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

$$\mathbf{G}(\mathbf{y}) = \mathbf{0} \tag{1.6}$$

where $\mathbf{G} : \mathbb{R}^{2n+2m} \mapsto \mathbb{R}^{2n+2m}$ is defined by

$$\mathbf{G}(\mathbf{y}) := \begin{pmatrix} \mathbf{K}\mathbf{u} - \mathbf{f} + \mathbf{N}^\top \boldsymbol{\lambda}_\nu + \mathbf{T}^\top \boldsymbol{\lambda}_t \\ \boldsymbol{\lambda}_\nu - \mathbf{P}_{\mathbb{R}_+^m}(\boldsymbol{\lambda}_\nu + \rho(\mathbf{N}\mathbf{u} - \mathbf{d})) \\ \boldsymbol{\lambda}_t - \mathbf{P}_{\Lambda\mathcal{F}\mathbf{P}_{\mathbb{R}_+^m}}(\boldsymbol{\lambda}_\nu + \rho(\mathbf{N}\mathbf{u} - \mathbf{d}))(\boldsymbol{\lambda}_t + \rho\mathbf{T}\mathbf{u}) \end{pmatrix}.$$

1.1 Algorithm

As we use the semi-smooth Newton method (SSNM), our algorithms are based on the following iterative scheme:

$$\mathbf{F}^o(\mathbf{y}^{(k-1)})\mathbf{y}^{(k)} = \mathbf{F}^o(\mathbf{y}^{(k-1)})\mathbf{y}^{(k-1)} - \mathbf{F}(\mathbf{y}^{(k-1)}), \quad k = 1, 2, \dots, \tag{1.7}$$

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 (1.7) 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 $\mathcal{S} \subseteq \mathcal{M}$ we introduce the diagonal matrix

$$\mathbf{D}_{\mathcal{S}} = \text{diag}(s_1, \dots, s_m) \in \mathbb{R}^{m \times m}, \quad s_i = \begin{cases} 1 & \text{for } i \in \mathcal{S}, \\ 0 & \text{for } i \notin \mathcal{S}. \end{cases}$$

Now, 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:

$$\begin{aligned} \mathcal{I}_t^+(\mathbf{y}) &= \{i \in \mathcal{M} : \lambda_{t,i} + \rho(\mathbf{T}\mathbf{u})_i > \mathcal{F}_i(\lambda_{\nu,i} + \rho(\mathbf{N}\mathbf{u} - \mathbf{d})_i)^+\}, \\ \mathcal{I}_t^-(\mathbf{y}) &= \{i \in \mathcal{M} : \lambda_{t,i} + \rho(\mathbf{T}\mathbf{u})_i < -\mathcal{F}_i(\lambda_{\nu,i} + \rho(\mathbf{N}\mathbf{u} - \mathbf{d})_i)^+\} \end{aligned}$$

Let us denote $\mathbf{B}_2 = (\mathbf{D}_{\mathcal{A}_\nu} \mathbf{N}, \rho \mathbf{D}_{\mathcal{A}_t} \mathbf{T}, \mathbf{B})^\top$, $\mathbf{B}_1^\top = (\mathbf{N}^\top, \mathbf{T}^\top, \mathbf{B}^\top)$, $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_\nu, \boldsymbol{\lambda}_t, \boldsymbol{\lambda}_e)^\top$, $\mathbf{h} = (\mathbf{D}_{\mathcal{A}_\nu} \mathbf{d}, \mathbf{0}, \mathbf{0})^\top$ and

$$\mathbf{C} = \left(\begin{array}{c|c|c} -\mathbf{D}_{\mathcal{I}_\nu} & \mathbf{0} & \mathbf{0} \\ \mathcal{F}(\mathbf{D}_{\mathcal{I}_t^-} - \mathbf{D}_{\mathcal{I}_t^+}) \mathbf{D}_{\mathcal{A}_\nu} & \mathbf{D}_{\mathcal{I}_t^+ \cup \mathcal{I}_t^-} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right),$$

which changes (2.8) into

$$\begin{pmatrix} \mathbf{K} & \mathbf{B}_1^\top \\ \mathbf{B}_2 & -\mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{h} \end{pmatrix}. \tag{2.9}$$

Moreover, we denote by $\mathcal{N}(\mathbf{K})$ and $\mathcal{R}(\mathbf{K})$ the null-space and the range-space of \mathbf{K} and by \mathbf{R} we denote full rank matrix which columns span $\mathcal{N}(\mathbf{K})$. Because we do not know the null-space of \mathbf{K} à-priori, we can identify it by using the Cholesky factorization of \mathbf{K} . We assume that the linear dependent rows are ordered at the end of \mathbf{K} . The Cholesky factorization exhibits the following block structure:

$$\mathbf{K} = \begin{pmatrix} \mathbf{L}_1 & \mathbf{0} \\ \mathbf{L}_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{L}_1^\top & \mathbf{L}_2^\top \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

The generalized inverse to \mathbf{K} is defined by:

$$\mathbf{X} = \begin{pmatrix} \mathbf{L}_1^{-\top} \mathbf{L}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

It is easy to show that

$$\mathbf{R} = \begin{pmatrix} -\mathbf{L}_1^{-\top} \mathbf{L}_2^\top \\ \mathbf{I} \end{pmatrix}.$$

As proved in [2] the first equation in (2.9) is satisfied iff

$$\mathbf{f} - \mathbf{B}_1^\top \boldsymbol{\lambda}^* \in \mathcal{R}(\mathbf{K}) \quad \text{and} \quad \mathbf{u}^* = \mathbf{X}(\mathbf{f} - \mathbf{B}_1^\top \boldsymbol{\lambda}^*) + \mathbf{R}\boldsymbol{\alpha}, \tag{2.10}$$

where the pair $(\boldsymbol{\lambda}^*, \mathbf{u}^*)$ is the solution to (2.8), \mathbf{X} is arbitrary generalized inverse to \mathbf{K} and $\boldsymbol{\alpha} \in \mathbb{R}^n$. The pair $(\boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*)$ satisfies

$$\begin{pmatrix} \mathbf{B}_2 \mathbf{X} \mathbf{B}_1^\top + \mathbf{C} & -\mathbf{B}_2 \mathbf{R} \\ -\mathbf{R}^\top \mathbf{B}_1^\top & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_2 \mathbf{X} \mathbf{f} - \mathbf{h} \\ -\mathbf{R}^\top \mathbf{f} \end{pmatrix}, \tag{2.11}$$

after substituting the first component of the solution \mathbf{u}^* into (2.9). The system matrix in (2.11) is nonsingular.

When we solve (2.8) we can compute first $(\boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*)$ by solving (2.11) and then obtain \mathbf{u}^* using the formula in (2.10).

To continue, we denote by $\mathbf{F} = \mathbf{B}_2 \mathbf{X} \mathbf{B}_1^\top + \mathbf{C}$, $\mathbf{G}_1 = -\mathbf{R}^\top \mathbf{B}_2^\top$, $\mathbf{G}_2 = -\mathbf{R}^\top \mathbf{B}_1^\top$, $\mathbf{d} = \mathbf{B}_2 \mathbf{X} \mathbf{f} - \mathbf{h}$, and $\mathbf{e} = -\mathbf{R}^\top \mathbf{f}$, which changes (2.11) into

$$\begin{pmatrix} \mathbf{F} & \mathbf{G}_1^\top \\ \mathbf{G}_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ \mathbf{e} \end{pmatrix}. \tag{2.12}$$

It is easy to show that the *orthogonal projector* onto $\mathcal{N}(\mathbf{G}_i)$ is given by

$$\mathbf{P}_{\mathbf{G}_i} = \mathbf{I} - \mathbf{G}_i^\top (\mathbf{G}_i \mathbf{G}_i^\top)^{-1} \mathbf{G}_i, \quad i = 1, 2. \quad (2.13)$$

As $\mathcal{N}(\mathbf{G}_2)$ and $\mathcal{R}(\mathbf{G}_2^\top)$ are orthogonal complements, we can uniquely decompose $\boldsymbol{\lambda}^*$ as

$$\boldsymbol{\lambda}^* = \boldsymbol{\lambda}_{\mathcal{N}}^* + \boldsymbol{\lambda}_{\mathcal{R}}^*, \quad (2.14)$$

where $\boldsymbol{\lambda}_{\mathcal{N}}^* \in \mathcal{N}(\mathbf{G}_2)$ and $\boldsymbol{\lambda}_{\mathcal{R}}^* \in \mathcal{R}(\mathbf{G}_2^\top)$. Such $\boldsymbol{\lambda}^*$ is the first component to the solution to (2.12) iff

$$\boldsymbol{\lambda}_{\mathcal{R}}^* = \mathbf{G}_2^\top (\mathbf{G}_2 \mathbf{G}_2^\top)^{-1} \mathbf{e} \quad (2.15)$$

and it is easy to verify that such $\boldsymbol{\lambda}_{\mathcal{R}}^*$ fulfills $\mathbf{G}_2 \boldsymbol{\lambda}^* = \mathbf{G}_2 \boldsymbol{\lambda}_{\mathcal{R}}^* = \mathbf{e}$. Applying $\mathbf{P}_{\mathbf{G}_1}$ on the first equation in (2.12), we get the problem

$$\mathbf{P}_{\mathbf{G}_1} \mathbf{F} \boldsymbol{\lambda} = \mathbf{P}_{\mathbf{G}_1} \mathbf{d}, \quad \mathbf{G}_2 \boldsymbol{\lambda} = \mathbf{e}. \quad (2.16)$$

Substituting (2.14) into (2.16) we arrive at

$$\mathbf{P}_{\mathbf{G}_1} \mathbf{F} \boldsymbol{\lambda}_{\mathcal{N}}^* = \mathbf{P}_{\mathbf{G}_1} (\mathbf{d} - \mathbf{F} \boldsymbol{\lambda}_{\mathcal{R}}^*). \quad (2.17)$$

The second component of the solution to (2.12) is given by

$$\bar{\boldsymbol{\alpha}} = (\mathbf{G}_1 \mathbf{G}_1^\top)^{-1} \mathbf{G}_1 (\mathbf{d} - \mathbf{F} \boldsymbol{\lambda}^*), \quad (2.18)$$

as follows from the first block equation in (2.12). One can show that the singular matrix $\mathbf{P}_{\mathbf{G}_1} \mathbf{F}$ is the symmetric, positive definite operator on $\mathcal{N}(\mathbf{G}_2)$ and that is why we can use the *projected* BiCGSTAB for solving (2.17).

3 Numerical experiments

Real word problem: clamp joint This problem arises from the mining industry. Let us consider the clamp joint of the support of mines. These joints are placed in the tunnels to prevent buckling. The arch of the support of mines consists of several segments that are tied by clamp clips. In the centre of attention is the place of contact, where these segments overlap.

For the computations we interpret the situation as the planar problem, see Figure 3.1. See [5] or [4] for more info about the problem.

Example We compare the algorithms *ActiveSetCoulombTFETI* and QPCE (a variant of the SMALBE, see [1]) (from MatSol) for different numbers of subdomains. The numbers *iter/n_K* (outer (Newton) iterations/matrix-vector multiplications by \mathbf{K}^{-1}) are written down. *Ncut* stands for the number of subdomains, see the last table.

4 Conclusion

We have analyzed the SSNM applied to the solution of contact problems with Coulomb friction. Our algorithm was extended using TFETI. We have experimentally compared SSNM with QPCE on the real word problem. SSNM seems more promising.

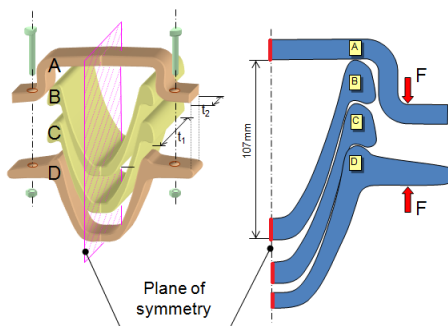


Figure 3.1. Planar problem of the clamp joint.

N_{cut}	n/m	QPCE	$ActiveSetCoulombTFETI$
10	67454/1942	23/697	29/1317
20	68236/2724	53/2170	24/898
40	69332/3820	13/766	30/1424
100	71936/6424	39/962	27/1865
150	73716/8204	33/779	22/690
200	75376/9864	31/890	23/499
260	77262/11750	57/1112	19/575
360	79814/14302	97/4674	21/459
400	80548/15036	56/1310	27/929
460	81908/16396	18/1268	36/982
500	82742/17230	31/1374	34/3168
600	84632/19120	31/995	19/335
900	89660/24148	80/2010	37/1389
1000	9108/25568	102/2020	26/994

References

1. Z. Dostál. *Optimal Quadratic Programming Algorithms, with Applications to Variational Inequalities*. Springer, 2008, New York, 1st ed.
2. R. Kučera, T. Kozubek, A. Markopoulos. *On large-scale generalized inverses in solving two-by-two block linear systems*. *Linear Algebra and its Applications*, 438 (2013), pp. 3011-3029.
3. R. Kučera, T. Kozubek, A. Markopoulos, J. Haslinger, L. Mocek. *Projected methods for solving non-symmetric two-by-two block linear systems arising from fictitious domain formulations*. Preprint submitted to *Journal of Computational and Applied Mathematics*, (2013).
4. R. Kučera, K. Motyčková, A. Markopoulos. *The R-linear convergence rate of an inexact semi-smooth Newton method*. Preprint sent to the journal "Computational Optimization and Applications", 2014.
5. K. Motyčková. *The Semi-Smooth Newton Method for Solving Contact Problems with Friction*. Ph.D. thesis, in preparation.

Active-set algorithm with spectral steps for semicoercive quadratic problems in particle dynamics

Lukáš Pospíšil

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.pospisil@vsb.cz

Abstract. This paper shows the possibilities of iterative solution of inner quadratic programming problems with positive semidefinite Hessian matrix on special convex sets which arise in the particle dynamics simulations. We focus on the active-set methods based on previously developed optimal quadratic programming algorithms. We suggest new type of step to deal with non-trivial kernel and propose the usage of projected Barzilai-Borwein method for decreasing the number of projection steps. The efficiency of our algorithms is demonstrated on the solution of simple simulation with hundreds of moving spherical particles and static box obstacles.

1 Introduction

For the last two decades, in the Department of Applied Mathematics in VŠB-TU Ostrava, the strictly convex quadratic programming problems on special convex sets and algorithms for solving these problems were studied intensively [10], [1]. The motivation for the development is based on the practical applications, mostly in the parallel numerical solution of the linear elasticity contact problems with or without friction. Nowadays, we are standing in front of new challenge - the particle dynamics problems [8], [9]. In this case, the object quadratic function is only semi-definite, so new theory and modification of our active-set algorithms is necessary. In this paper, we shortly review the basics of the simulation dynamics contact problems with particles. The theory can be found in [14] and the usage of the modification of our active-set algorithm is motivated by [15]. In next sections, we suggest some modification of current algorithms and present the results of numerical experiments.

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 project SGS SP2014/204.

2 Time-stepping schema and formulation of optimizing problem

Let us consider the system of $nb \in \mathbb{N}$ rigid bodies (*particles*) in vector space $\{(x, y, z) \in \mathbb{R}^3\}$. Each particle has 6 degrees of freedom - location of centre of gravity $[r_x, r_y, r_z]^T$ and unit quaternion of rotation $[e_0, e_1, e_2, e_3]^T$. For every body $T_{(i)}$ in the system in given time t , let us denote $q_{(i)}^t \in \mathbb{R}^7$ as a *vector of generalized position* and $v_{(i)}^t \in \mathbb{R}^6$ as a *vector of generalized velocities*. In our simulations, the rotation of the body is represented by the unit quaternion of rotation and the angular velocity is represented by Euler angles. The position of bodies in the next time-step can be evaluated using simple time-stepping schema

$$q^{(t+h)} = q^{(t)} + h.Qv^{(t)} ,$$

where h is *sufficiently small* time step. Here, Q denotes the matrix of linear mapping between derivative of position vector and vector of velocities. This equation can be considered as a discretized numerical solution of the first Newton law using Euler method.

The same method is also used for the computation of velocities. The increment in the next time-step depends on the mass of each body, affecting external forces $F_{ext}(t, q, v)$, and contacts and other limiting conditions. This situation is described by second Newton law, i.e.

$$v^{(t+h)} = v^{(t)} + hM^{-1}(F_{ext} + F_C) , \tag{1}$$

where M is generalized mass matrix, F_C is a vector of forces induced by contact constraints, and F_{ext} is a vector of external forces. In our simple simulation, the vector of external forces represents the gravity force affecting each body.

The contact between two bodies T_A and T_B constitutes forces and torques

$$\begin{aligned} F_A &:= -\tilde{\gamma}n_A(C) , & F_B &:= \tilde{\gamma}n_A(C) , \\ M_A &:= C^A \times F_A , & M_B &:= C^B \times F_B , \end{aligned}$$

where $n_A(C)$ is unit outer normal to the body T_A at the point of contact C in global coordinate system, and $\tilde{\gamma} \geq 0$ is unknown size of the force. This force evokes the change of the position of the body T_B (the components of the generalized velocity vector corresponding to the position of gravity center).

Force F_B evoking the change of the position of the body T_B (the components of the generalized velocity vector corresponding to the position of gravity center) and the change of rotation of the body T_B (the components of the generalized velocity vector corresponding to the rotation) is effected by associated torque. Analogically, the forces F_A and M_A change the position and rotation of body T_A . All forces can be expressed by one vector

$$F_C = \begin{bmatrix} F_A \\ M_A \\ F_B \\ M_B \end{bmatrix} = \begin{bmatrix} -\tilde{\gamma}n_A(C) \\ C^A \times F_A \\ \tilde{\gamma}n_A(C) \\ C^B \times F_B \end{bmatrix} = \begin{bmatrix} -\tilde{\gamma}n_A(C) \\ -\tilde{\gamma}\tilde{C}^A n_A(C) \\ \tilde{\gamma}n_A(C) \\ \tilde{\gamma}\tilde{C}^B n_A(C) \end{bmatrix} = \underbrace{\begin{bmatrix} -n_A(C) \\ -\tilde{C}^A n_A(C) \\ n_A(C) \\ \tilde{C}^B n_A(C) \end{bmatrix}}_{=:D \in \mathbb{R}^{12,1}} \tilde{\gamma} ,$$

where we used the matrix representation of the vector product.

The unknown size of the force $\tilde{\gamma}$ is constrained by the non-penetration conditions of the bodies, which can be described by *gap function* $\Phi : \mathbb{R}^{7+7} \rightarrow \mathbb{R}$. It holds

- $\Phi([q_A, q_B]) = 0, \tilde{\gamma} \geq 0$ if the bodies are in contact,
- $\Phi([q_A, q_B]) > 0, \tilde{\gamma} = 0$ if the bodies are not in contact,
- $\Phi([q_A, q_B]) < 0$ if the bodies penetrate each other.

Merging these observations, we obtain the *complementarity condition*

$$0 \leq \Phi(q) \perp \gamma \geq 0 .$$

Instead of these conditions, we can consider more numerically stable conditions (see [16])

$$0 \leq \frac{1}{h} \Phi(q) + D^T v^{(t+h)} \perp \gamma \geq 0 . \tag{2}$$

This problem can be reformulated to the quadratic programming problem with bound constraints.

If we consider a problem with friction, then the contact forces consist of additional tangential components, i.e.

$$F = F_n + F_T = \gamma_n n + \gamma_u u + \gamma_w w ,$$

where $F_n = \gamma_n n \in \mathbb{R}^3$ is normal component of the friction force, $F_T = \gamma_u u + \gamma_w w \in \mathbb{R}^3$ is tangent component of the friction force, $\gamma_n > 0$ is a size of normal component of the friction force, and $\gamma_u, \gamma_w \in \mathbb{R}$ are sizes of tangent components of friction force. Relation between components of the friction force can be described by *Coulomb friction model*

$$\begin{aligned} \gamma_n &\geq 0, \quad \Phi(q) \geq 0, \quad \Phi(q)\gamma_n = 0, \\ \sqrt{\gamma_u^2 + \gamma_w^2} &\leq \mu\gamma_n, \\ \|v_T\| \left(\mu\gamma_n - \sqrt{\gamma_u^2 + \gamma_w^2} \right) &= 0, \\ \langle F_T, v_T \rangle &= -\|F_T\| \cdot \|v_T\|. \end{aligned} \tag{3}$$

Here, $\{n, u, w\}$ is orthonormal basis of tangent space at contact point.

Similarly, the problem (3) can be reformulated as quadratic programming problem.

Theorem 1. *The solution of the optimizing problem*

$$\min_{\gamma \in \Omega} \frac{1}{2} \gamma^T N \gamma + r^T \gamma , \tag{4}$$

where

$$\begin{aligned} N &:= D^T M^{-1} D, \quad r := \left[\frac{1}{h} \Phi, 0, 0 \right]^T + D^T M^{-1} k, \quad k := M v^{(t)} + h \cdot F_{ext} \\ \Omega &:= \Omega_1 \times \dots \times \Omega_{nc}, \quad \Omega_j := \{[x, y, z]^T \in \mathbb{R}^3 : \sqrt{y^2 + z^2} \leq \mu_i x\} \end{aligned}$$

is equivalent to the solution of original problem (3).

If the problem consists of more contacts, then the matrices are defined as block matrices.

3 Active-set methods

For solving the problem (4), we are using active-set algorithms. These algorithms are based on the decomposition of the set of indices of all constraints \mathcal{M} into two disjoint subsets based on the value of constraint function

$$\mathcal{F}(x) := \{j \in \mathcal{M} : h_j(x_{\mathcal{I}_j}) < 0\}, \quad \mathcal{A}(x) := \{j \in \mathcal{M} : h_j(x_{\mathcal{I}_j}) = 0\},$$

and using this decomposition, we decompose the gradient of object function $g(x) = \nabla f(x) = Ax - b$ into *free* and *chopped* gradient

$$\begin{aligned} \varphi_{\mathcal{I}_j}(x_k) &= g_{\mathcal{I}_j} \text{ for } j \in \mathcal{F}(x_k), & \varphi_{\mathcal{I}_j}(x_k) &= 0 \text{ for } j \in \mathcal{A}(x_k), \\ \beta_{\mathcal{I}_j}(x_k) &= 0 \text{ for } j \in \mathcal{F}(x_k), & \beta_{\mathcal{I}_j}(x_k) &= g_{\mathcal{I}_j} - \min\{n_j^T(x_k)g_{\mathcal{I}_j}, 0\}n_j(x_k) \\ & & & \text{for } j \in \mathcal{A}(x_k), \end{aligned}$$

where $n_j(x)$ is outer normal of j -th constraint $h_j(x_{\mathcal{I}_j})$.

Algorithm 1. Modified Proportioning with Gradient Projection (MPGP).

```

Choose  $x_0 \in \Omega$ 
for  $k = 0, 1, 2, \dots$  (while stopping criterion is not achieved)
  if  $\|\varphi(x_k)\| \gg \|\beta(x_k)\|$  (proportioning condition)
    Kernel step
      if  $\|P\varphi\|$  is large, make feasible step to boundary using  $-P\varphi$ 
    CG step or CG halfstep
      make one CG step to solve problem on free set
      if this step means leaving  $\Omega$ , do only a half-step and restart CG
  else
    Gradient projection step.
      make one CG step to solve problem on active set (non-friction)
      make projected Barzilai-Borwein step (friction)
      restart CG on free set
  endif
   $k := k + 1$ 
endfor

```

Our algorithm is based on the usage of these gradients to minimize the object function on free set and afterwards on active set, see Algorithm 1.

Moreover, we present also new kind of step to deal with non-trivial kernel of semidefinite Hessian of object function. If the norm of the projection of free gradient onto kernel is large, then we make maximal feasible step to boundary. Using this technique, we regularize the object function on free set and we can use standard conjugate gradient method. The theory will be published in [5].

In the case of problem with friction, we suggest to use projected Barzilai-Borwein method [12], [6], [4]. This modification was inspired by Spectral projected gradient method, which uses the similar type of steps [13]. In original MPGP algorithm, it is used projection step with constant step-length [11] which always induces the descend of object function. However, using non-monotone algorithms,

such as projected Barzilai-Borwein, usually decreases the number of projection steps. This phenomena was first time presented in [3].

As a stopping criteria, we are using the value of *projected gradient* $g^P := \varphi(x) + \beta(x)$ or *scaled projected gradient* $\tilde{g}_\alpha^P := \frac{1}{\alpha}(x - P_\Omega(x - \alpha g(x)))$. The equivalency between these gradients was proved in [2] and discussed in [7].

4 Numerical experiments

In this section, we present the numerical results of efficiency of our algorithm on the simulation of 480 spherical particles with friction. These particles are scattered into charging equipment represented by 13 box obstacles, see Fig. 1. All algorithms were implemented in Matlab environment. For contact detection, we are using our own implementation of *Moving Bounding-Box algorithm* [17].

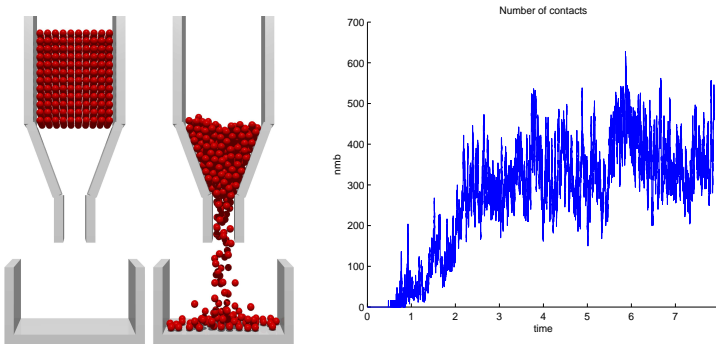


Fig. 1. The state of testing benchmark is in $t = 0$ s and $t = 5$ s (left). The varying number of contacts during the simulation (right).

References

1. Dostál Z., Pospíšil L.: *Optimal iterative QP and QPQC algorithms*, Ann. Oper. Res., issn 0254-5330, (2013).
2. Bouchala J., Dostál Z., Kozubek T., Pospíšil L., Vodstrčil P.: *On the solution of convex QPQC problems with elliptic and other separable constraints with strong curvature*, accepted (2014).
3. Pospíšil L.: *An optimal algorithm with Barzilai-Borwein steplength and superrelaxation for QPQC problem*, conference Programs and Algorithms of Numerical Mathematics 16, isbn 978-80-85823-62-2, pp. 155-161, (2012).
4. Dostál Z., Pospíšil L.: *The Projected Barzilai-Borwein method for solving Quadratic Programming problems with Separable Elliptic Constraints*, conference MOD-ELLING 2014, Roznov pod Radhostem, (2014).
5. Dostál Z., Pospíšil L.: *Fast conjugate gradients for symmetric positive semidefinite least square problems*, in preparation (2014).

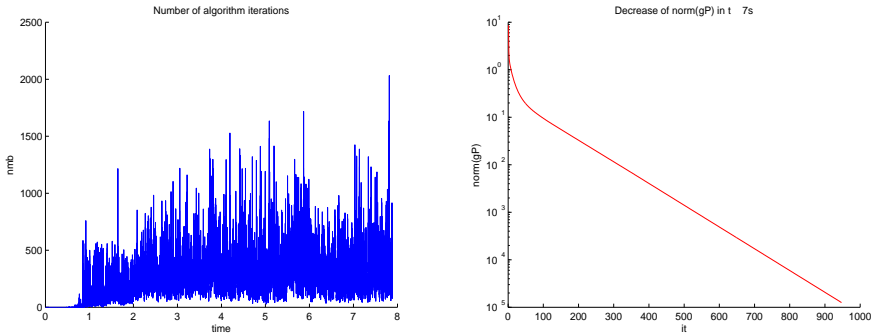


Fig. 2. The number of iterations necessary for solving inner optimizing problem to obtain accuracy $\epsilon = 10^{-4}$ during the simulation (left). The decrease of the norm of stopping gradient during iterations for the problem in $t = 7$ s (right).

6. Pospíšil L., Menšík M.: *Faster Gradient Descent Methods*, conference SNA12, Liberec, (2012).
7. Pospíšil L., Dostál Z., Kozubek T.: *Optimal active-set and spectral algorithms for the solution of 3D contact problems with anisotropic friction*, conference MAFE-LAP 2013, London, (2013).
8. Pospíšil L., Dostál Z.: *The solution of Quadratic Programming problem with Separable Conical Constraints in Granular dynamics*, conference Programs and Algorithms of Numerical Mathematics 17, (2014).
9. Dostál Z., Pospíšil L.: *Minimization of quadratic function with semidefinite Hessian subject to bound constraints*, conference Parallel Matrix Algorithms and Applications 14, Lugano, (2014).
10. Dostál Z.: *Optimal Quadratic Programming Algorithms, with Applications to Variational Inequalities*, 1st edition. SOIA 23. Springer US, New York, (2009).
11. Dostál Z., Schöberl J.: *Minimizing quadratic functions subject to bound constraints with the rate of convergence and finite termination*, *Comput. Optim. Appl.*, 30, 1, pp. 23-44, (2005).
12. Barzilai J., Borwein J. M.: *Two point step size gradient methods*, *IMA Journal of Numerical Analysis* 8, pp. 141-148, (1988).
13. Birgin E.G., Martinez J.M., Raydan M.: *Nonmonotone spectral projected gradient methods on convex sets*, *SIAM Journal on Optimization* 10, pp. 1196-1211, (2000).
14. Heyn T.: *On the Modeling, Simulation, and Visualization of Many-Body Dynamics Problems with Friction and Contact*, Ph.D. Thesis, (2013).
15. Heyn T., Anitescu M., Tasora A., Negrut D.: *Using Krylov Subspace and Spectral Methods for Solving Complementarity Problems in Many-Body Contact Dynamics Simulation*, *Int. J. Numer. Meth. Engng*, pp. 541561, (2012).
16. Anitescu M.: *Optimization-based simulation of nonsmooth rigid multibody dynamics*, *Mathematical Programming*, 105(1), pp. 113-143, (2006).
17. Schinner A.: *Fast Algorithms for the Simulations of Polygonal Particles*, *Granular Matter* 2, Springer-Verlag, pp. 35-43, (1999).

Traffic Flow Modeling Based on Sparse Data

Lukas Rapant

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.rapant@vsb.cz

Abstract. In this article we present numerical approach to macroscopic traffic modeling of a highway traffic. Our goal is to model traffic on the D1 highway with focus on traffic jam dissolution. Most problematic part of traffic modeling is usually lack of accurate data concerning a traffic. We have two sources of data available. First source is floating car data (FCD). This source provides data for the entire D1 but can be inaccurate and contains only information about speed. The other source are ASIM detectors which are present on certain toll gates. We have chosen Cell transmission model - velocity as the best fitting model for modeling the D1 traffic from the available data. This article comprises of problem statement, theoretical background and experimental results.

Keywords: traffic flow modeling, Lighthill-Whitham-Richards PDE, cell transmission model - velocity, floating car data, ASIM sensors

1 Introduction and problem statement

During our work on RODOS project there has arisen a need for a traffic flow model, that will be capable of macroscopic highway traffic modeling and short term prediction. Main aim was to model problematic situations like traffic column dissolution after traffic accidents and traffic jams. The model was also required to be able to run on the available data. These data come from two sources: Floating car data and ASIM sensors.

Floating car data (FCD) collection is an approach to determine the traffic speed on the road network. This approach is based on the measurement of location, speed, travel direction and time information from certain vehicles in the traffic (mostly fleet vehicles like trucks, couriers, ...). These informations are obtained from the GPS receiver inside the car and broadcast by radio unit or cell phone. Unlike the other traffic data collectors like cameras, number plate recognition systems, toll gates or induction loops this method is quite cheap as it does not require any other road hardware. Therefore this method is very popular for many applications like traffic congestion detection or delay calculation. In our case FCD have specific format. D1 highway is divided into 56 sections and we are provided with the traffic speeds for each of these sections every minute. These

sections have various lengths from 300m to 12km. The traffic speed is calculated each minute as a mean of speeds of all floating cars that passed through that section in the last minute. This approach has several drawbacks. If only few cars pass through the section, the resulting speed can be inaccurate. Also sections with different length (some of them are very long) produce only coarse speed profile of the highway because there is only one speed known for entire section. Another problem is that FCD do not provide any information regarding traffic intensity or density.

ASIM sensors are placed on certain toll gates. They comprise of various sensors like passive infrared detectors and radars. They are able to distinguish individual vehicle types and measure their speed and intensity. They provide very reliable information, but they are deployed very sparsely. Only 25 sensors are placed between Brno and Prague on the D1 highway. Their measurements are aggregated every five minutes and mean speed and intensity are calculated. They provide very broad and reliable informations about traffic but their network is so sparse that it is impossible to run the model with ASIM data only.

Our task is to propose a macroscopic traffic flow model, which can be based on these sparse data sources (as we have only 76 speed values from the D1 highway between Brno and Prague and 25 of these cover the same places). Usually this problem is solved by application of some form of kinematic wave model [6] because it accurately describes physical behavior of traffic flow. However, absence of traffic intensity information in FCD rules out most of these traffic flow models. Solution was found in a form of Cell transmission model - velocity (CTM-v) [2, 3]. CTM-v is special version of CTM [1]. It has advantage, that it does not need information about the traffic density or intensity and it requires only information about the traffic speed.

2 Cell transmission model - velocity

CTM-v is based upon classical Lighthill-Whitham-Richards (LWR) partial differential equation[6]. LWR PDE for modeling traffic on highways is:

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0 \quad (1)$$

where $q(x, t)$ is flow and $\rho(x, t)$ is density of the vehicles at location x and time t . This equation is derived from hydrodynamics theory and expresses the conservation of mass for a fluid of density ρ and of flux q . Empirical relation called fundamental diagram $q(x, t) = Q(\rho(x, t))$ is used for expression of q as a function of ρ . Q is flux function and is considered to be independent of variables. For transformation of LWR PDE to velocity LWR (LWR-v) PDE we will use specific flux function called Greenshields [8]:

$$v = v_{max} \left(1 - \frac{\rho}{\rho_{max}} \right) \quad (2)$$

where v_{max} and ρ_{max} denote respectively the maximal velocity and the maximal density allowed by the model. This function expresses relation between density

and velocity. Greenshields flux function can be easily inverted to express ρ as a function of v . Now we can substitute this expression into 1 to obtain LWR-v PDE:

$$\frac{\partial v}{\partial t} + \frac{\partial R(v)}{\partial x} = 0 \quad (3)$$

where $R(v) = v^2 - v_{max}v$. Variable change $v = v - \frac{v_{max}}{2}$ transforms 3 into its final form:

$$\frac{\partial v}{\partial t} + \frac{\partial v^2}{\partial x} = 0 \quad (4)$$

on the domain $(x, t) \in \langle a, b \rangle \times \langle 0, T \rangle$. The initial and boundary conditions are too long to be presented here and can be found for example in [4]. For practical implementation, the LWR-v PDE is discretized using a Godunov numerical scheme [7]. By application of Godunov scheme we obtain CTM-v [3]. Let $N \in \mathbb{N}$ be number of time steps of length $\delta t = \frac{T}{N}$, $M \in \mathbb{N}$ number of space cells of length $\delta x = \frac{b-a}{M}$. Then v_i^n is called discrete value of v at time step n in cell i ($0 \leq n \leq N$ and $0 \leq i \leq M$). At each time step v_i^{n+1} is computed from the previous time step. This is done by formula:

$$v_i^{n+1} = v_i^n - \frac{\Delta t}{\Delta x} (g(v_i^n, v_{i+1}^n) - g(v_{i-1}^n, v_i^n)) \quad (5)$$

where the numerical flow function g is:

$$g(v_1, v_2) = \begin{cases} R(v_2) & \text{if } v_1 \leq v_2 \leq v_c \\ R(v_c) & \text{if } v_1 \leq v_c \leq v_2 \\ R(v_1) & \text{if } v_c \leq v_1 \leq v_2 \\ \max(R(v_1), R(v_2)) & \text{if } v_1 \geq v_2 \end{cases} \quad (6)$$

where $v_c = \frac{v_{max}}{2}$, $R(v) = v^2 - v_{max}v$ and v_{max} is maximum speed allowed in the model. Note that Δt and Δx must satisfy Courant-Friedrichs-Lewy (CFL) condition:

$$\left| \frac{\Delta t}{\Delta x} v_{max} \right| \leq 1 \quad (7)$$

For implementation of boundary conditions, we use ghost cells placed at each side of the domain defined by the boundary conditions we would like to be satisfied [3, 4].

3 Experimental results

Our CTM-v model was implemented in MATLAB. It was tested on two traffic situations. The first was experimental highway 60 km long. Simulation was set to model next 30 minutes. Spatial discretization Δx was chosen to be 100 m and temporal discretization Δt was chosen to be 1 s. These values are satisfying CFL condition (maximum speed of the model was set to be $150 \frac{\text{km}}{\text{h}}$). For the initial condition, we have generated values of speed for each cell from Gaussian distribution with $\mu = 100$ and $\sigma = 20$. Traffic in the cells from 25th to 30th km of the

highway was set to have a speed of $40 \frac{\text{km}}{\text{h}}$ to simulate a traffic column. Boundary conditions were represented by ghost cells with random walk beginning in the free flow speed ($100 \frac{\text{km}}{\text{h}}$). Results of the model can be seen in Figure 1. Axis x represents time in seconds, axis y represents 100m long cells and scale represents speed in individual cells. Figure 1 shows fluent dissolution of traffic jam.

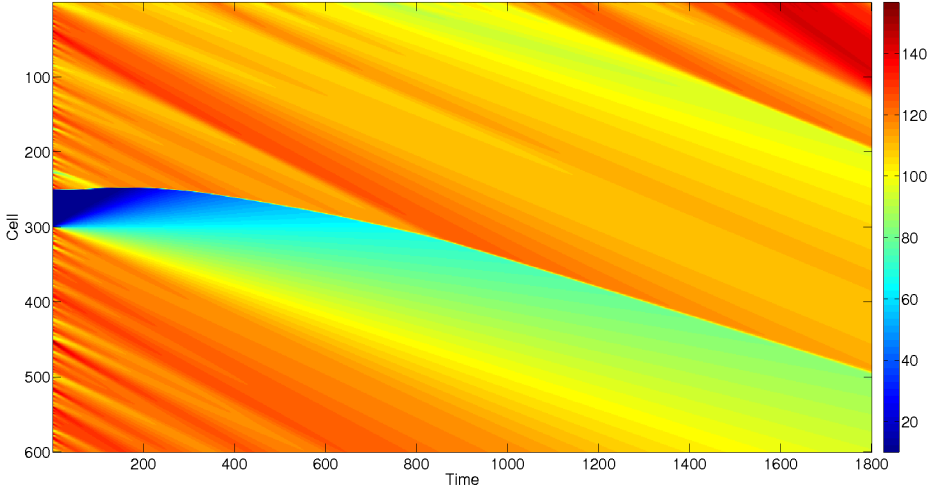


Fig. 1. Model of experimental 60km long highway

After testing the model on experimental data, we have proceeded test it on the real data from D1 highway. Our model will be tested on stretch from Brno to Praha which is 197km long. Discretization was set to be the same as in previous experiment. As it was mentioned earlier, we have two data sources available: FCD and ASIM. The first of all we have aggregated FCD measurements into 5 minute intervals to have their aggregation period consistent with the ASIM measurements. Then we have assigned cells their initial speed based on FCD. This created the coarse speed profile with 56 parts (each part was further divided into cells) each having its speed determined from FCD aggregated over last 5 minutes. Then we have replaced values from FCD with the values from ASIM in the cells containing ASIM sensors. ASIM sensors are considered to be much more accurate than FCD. This has sometimes created difficulty as values from ASIM were quite different from the FCD ones. To avoid these "spikes" and to make FCD measurements closer to reality we have replaced FCD values on parts with ASIM sensor with values from Gaussian distribution with $\mu = (v_{\text{ASIM}} - v_{\text{FCD}}) + v_{\text{FCD}}$ and $\sigma = (v_{\text{ASIM}} - v_{\text{FCD}})/6$. Variables v_{ASIM} and v_{FCD} represent speeds obtained from these sensors. This method was found to be preferable to

simply setting the speed for entire part to ASIM speed value. While speed can be very different in the cell with ASIM sensor, in other cells it can be closer to FCD because some parts are very long and probably do not have the same speed along their entire length. Parts without ASIM sensor were left as they are because we do not have any information whether FCD measurements are accurate or not. Such speed profile is used as initial condition for our model of the D1 highway. Boundary conditions are again handled by random walk from free flow speed. In Figure 2 you can see results from our model run on data from 24.3.2014 at 17:35. Axis x represents time in seconds, axis y represents 100m long cells and scale represents speed in individual cells.

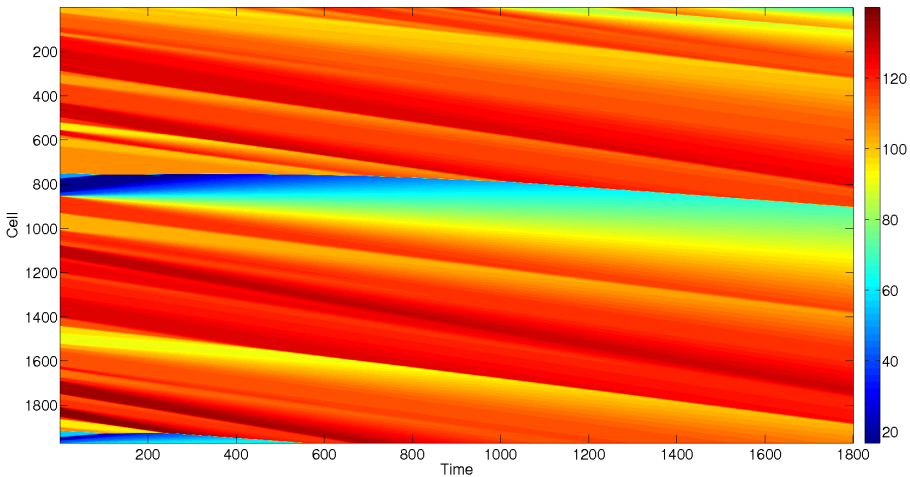


Fig. 2. Model of D1 highway

You can see dissolution of two traffic jams. Their dissolution time roughly corresponds to the real one. However speed profile after 30 minutes of simulation is only very roughly following real one (although it is difficult to check it exactly as we again only have FCD data which can be inaccurate). Therefore with available data, CTM-v alone can simulate trends but not the exact speed. This is because of sparsity of the data and FCD inaccuracy. Even combined data sources have only 76 measurements with 25 of them being duplicate. Due to this fact, CTM-v model is suitable only for very rough modeling of the D1 highway and would require either additional data or some more elaborate statistical approach.

4 Conclusion

In this article we have described CTM-v model and applied it on experimental data and real data from the D1 highway. Results have shown that CTM-v suitably represents traffic jam dissolution. However with the available data from the D1, its exact speed prediction capability is not good. It catches most of the trends but speeds received from the model and speeds from the measurements are sometimes quite different. This is because of sparsity and inaccuracy of data as CTM-v model is proven to be well representing the traffic flow. Accuracy of our model can be improved by application of some form of Kalman filter [4, 5] or Bayes network which can bring additional information into the model. We are planning to try these approaches in near future.

5 Publications

- Tomala, K., Plucar, J., Dubec, P., Rapant, L., & Voznak, M. (2013). The Data Extraction Using Distributed Crawler Inside Multi-Agent System. *Advances In Electrical And Electronic Engineering*, 11(6), 455 - 460.

References

1. C. F. Daganzo., "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory", *Transportation Research Part B* 28, no. 4 (1994): 26987.
2. D. B. Work et al. "A Traffic Model for Velocity Data Assimilation", *Applied Mathematics eXpress*, Vol. 2010, No. 1, 2010, pp. 135.
3. V. Coric, N. Djuric, S. Vucetic. "Traffic State Estimation from Aggregated Measurements with Signal Reconstruction Techniques", *Transportation Research Record: Journal of the Transportation Research Board*, No. 2315, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 121130.
4. Work, D., O. Tossavainen, S. Blandin, A. Bayen, T. Iwuchukwu, K. Tracton. "An Ensemble Kalman Filtering Approach to Highway Traffic Estimation Using GPS Enabled Mobile Devices", *Proc., 47th IEEE Conference on Decision and Control*, Cancun, Mexico, Dec. 9-11, IEEE, New York, 2008, pp. 50625068.
5. X. Sun, L. Munoz, R. Horowitz, "Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control", *Proc. of the 42nd IEEE Conference on Decision and Control*, vol. 6, Maui, HI, 2003, pp. 6333-6338.
6. M. Lighthill and G. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads", *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 229, no. 1178, 1955, pp. 31745.
7. J. P. Lebacque, "The godunov scheme and what it means for first order traffic flow models", In *13th International Symposium on Transportation and Traffic Theory*, 1996, pp. 64777.
8. B. D. Greenshields, "A study of traffic capacity", *Highway Research Board* 14, 1935, pp. 44877.

Use of the Simplex Algorithm for Solving the Multiple Traveling Salesman Problem

Tomáš Režnar

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`tomas.reznar@vsb.cz`

Abstract. This paper deals with the use of parallelism for solving the multiple traveling salesman problem (MTSP). MTSP is solved using the simplex algorithm in combination with branch and bound method (BB). The main objective of this paper is to experimentally verify the scalability of the BB algorithm.

Keywords: Multiple Traveling Salesman Problem, Integer Linear Programming, The Simplex Algorithm, Branch-and-Bound Algorithm, Parallel Algorithms, Parallel Implementation

1 Introduction

Various companies engaged in the delivery of goods to the customer or the collection of goods from the customer to the headquarters may need to design the optimal route of their collection / distribution, in order to reduce the time needed to visit all the destinations to the minimum.

At the same time, these companies may have a requirement that a route of each employee does not exceed a certain price. For example, if we consider time as the price of labor per employee, then this constraint can ensure that an employee has the time to visit all the destinations of their route within given working time.

We formulate this problem as integer linear programming problem which can be solved using the simplex algorithm and the BB method. In the actual implementation of the BB algorithm we use parallel computing of individual solutions for incurred branches. The scalability of solution method has been verified on a model problem.

2 Definition of the problem

Traveling Salesman Problem (TSP) is finding the cheapest route that passes through all the given cities. This route begins and ends in the same city. Each city is visited exactly once [1]. There are two basic types: the first type is symmetric when the price of the route between two cities is the same in both directions, and

the asymmetric type, with the price of route that may differ depending on the direction (oriented graph). In this paper, we further consider only asymmetric variant because it reflects the real world better (such as one-way roads, closures, etc.).

The input to this problem is a graph $G = \{N, E\}$, where N is the index set of all nodes. Node 0 is first and last node of the resulting route. The E is the set of all edges of the graph. For each edge of the graph, there is cost of transition through the edge $c_{i,j}: (i, j) \in E$.

The result of the problem is the vector \mathbf{x} whose elements are variables $x_{i,j} \in \mathbf{x} : (i, j) \in E$ and each element of the vector \mathbf{x} can have a value 0 or 1. If $x_{i,j} = 0$, then the cheapest route does not cross the edge (i, j) and if $x_{i,j} = 1$, then the cheapest route crosses the edge (i, j) .

If a situation with more routes with the same cheapest price arises, then the solution is any of these routes.

2.1 Multiple traveling salesman problem

Multiple traveling salesman problem (MTSP) is an extension of TSP problem with a given number of travelers who may visit each city under the condition that each city is visited by just a single traveling salesman. The exception is the starting city that is visited by all the travelers. The solution to the problem are then routes of each traveler, which together cover all the cities. Not all travelers must leave the starting city.

For certain situations it may be necessary to add an additional condition to limit the maximum cost of the traveler’s route, so it is guaranteed in the real world, for example, that the travelers could make their entire route within their working time.

2.2 The linear programming of MTSP

The MTSP can be formulated as a minimization problem

$$\text{minimize} \quad \sum_{(i,j) \in E \wedge t \in T} c_{i,j} \cdot x_{t,i,j} \tag{1}$$

where T is the set of all travelers. Vector \mathbf{x} , which is the solution to the problem contains elements: $x_{t,i,j} \in \mathbf{x} : (i, j) \in E \wedge t \in T$. The variable $x_{t,i,j}$ contains t index, indicating the number of the traveler. It can be easily defined which traveler passes which edge from the solution of the problem.

Constraints (2) and (3) ensure that each node is visited exactly by one traveler and only once, except the initial node because the node is visited by each traveler.

$$\forall i \in N \setminus \{0\} : \sum_{j \in N \wedge j \neq i \wedge t \in T} x_{t,i,j} = 1 \tag{2}$$

$$\forall i \in N \setminus \{0\} : \sum_{j \in N \wedge j \neq i \wedge t \in T} x_{t,j,i} = 1 \quad (3)$$

To ensure that the first node will be entered and exited by every traveler no more than once, constraints (4) and (5) were added.

$$t \in T : \sum_{j \in N \wedge j \neq 0} x_{t,0,j} \leq 1 \quad (4)$$

$$t \in T : \sum_{j \in N \wedge j \neq 0} x_{t,j,0} \leq 1 \quad (5)$$

Constraints (2) and (3) only ensure that each node (except the first one) is visited exactly once, however they do not guarantee that the node is entered and exited by the same traveler. Therefore, we need to add additional constraint (6) to ensure that the node will be entered and exited by the same traveler. For each node n number of visited input edges of the node must equal number of visited output edges of the node by each traveler t .

$$(\forall n \in N)(\forall t \in T) : \sum_{i \in N \wedge i \neq n} x_{t,i,n} - \sum_{j \in N \wedge j \neq n} x_{t,n,j} = 0 \quad (6)$$

Further constraint (7) ensures that the maximum travel cost of each traveler is not exceeded. Vector \mathbf{m} assigns the maximum cost of route for each traveler.

$$\forall t \in T : \sum_{(i,j) \in E} c_{i,j} \cdot x_{t,i,j} \leq m_t \quad (7)$$

The vector x is binary, i.e. the traveler passes (value 1) or does not pass (value 0) through each edge of the graph. It is ensured by constraint (8).

$$x_{t,i,j} \in \mathbf{x} : x_{t,i,j} = 0 \vee x_{t,i,j} = 1 \quad (8)$$

Last constraint (9) ensure, that each route passes through the first node and each traveler has exactly one route [3].

$$(\forall t \in T)(\forall i \in N)(\forall j \in N) : \\ (1 \leq i \neq j \leq (|N| + 1)) \rightarrow u_{t,i} - u_{t,j} + (|N| + 1) \cdot x_{t,i,j} \leq |N| \quad (9)$$

3 Solving the problem

MTSP is solved by the two phase simplex algorithm described in [2]. The simplex algorithm itself is for finding a relaxed solution. To find a binary solution the BB algorithm described in [4] is used.

4 Parallel implementation of BB algorithm

Parallelization is used at the point of branching of tree. Whenever variables of calculated relaxed solution are not integers, there is a branching into two new branches. Those two branches can be solved in parallel. This can be repeated for each branching until all available processes are not used.

4.1 Description of parallelization

The number of processes is given and invariable. The number of processes must be higher than two. The first process is the master, the remaining processes are slaves. The master process is responsible for controlling slave processes, collects information about the busy processes and updates the current best solution. Once all slave processes end their tasks, it means that a final solution has been found and then the master process parses it.

Slave processes wait to receive a job (node of tree). Once they receive a job, they compute relaxed solution of the problem. If the solution does not meet the integer conditions, there is a branching into two new problems. At the moment of branching, the slave process sends information about its local best solution to the master process and receives the current global best solution (synchronization). The slave process then sends information to the master process that there is a branching and receives the number of the slave process that currently does not perform any work or it is informed that all other processes are busy.

If the current process has received information that another process is available, then one branch is sent to the available process, the second branch is resolved by the current process itself. In case that no process is available, both branches will be resolved by the current process itself. At each branching the checking for available process occurs, and therefore the involvement of all available slave processes occurs earliest in the depth l of the tree, where 2^l is the number of children processes.

4.2 Sample problem of parallelization

Consider a problem where the result of simplex algorithm always violates constraints of binary variables. We also assume, that there are three variables. This means that the depth of the tree is also three, because each level is restricting one variable. The given number of processes is four.

The figure 1 shows a tree of solution of the sample problem. The nodes are numbered $N1$ to $N3.4$. Each node has calculated the price of relaxed solution and whether relaxed solution satisfies all the constraints of all binary variables. $P1$ to $P4$ indicates the number of the process that will solve the node.

4.3 Scalability of testing problems

Three random problems (T1, T2 and T3) were created. Problem T1/T2/T3 had 12/14/16 nodes in the graph which means 143/195/255 variables as simplex

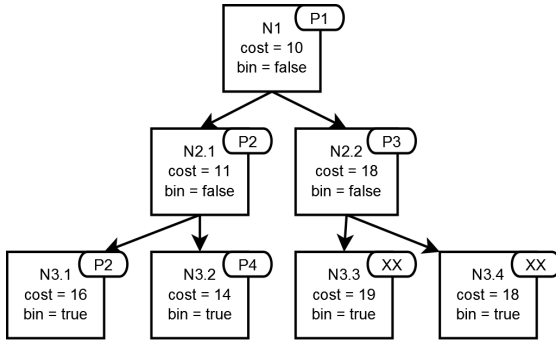


Fig. 1. Tree of solution of the sample problem

algorithm input. The problems were solved on the supercomputer Anselm. 32 nodes were used for computing on Anselm. Each node consists of two eight-core processors. Table 1 displays measured average times needed for the solution of the problem. Figure 2 shows scalability graph.

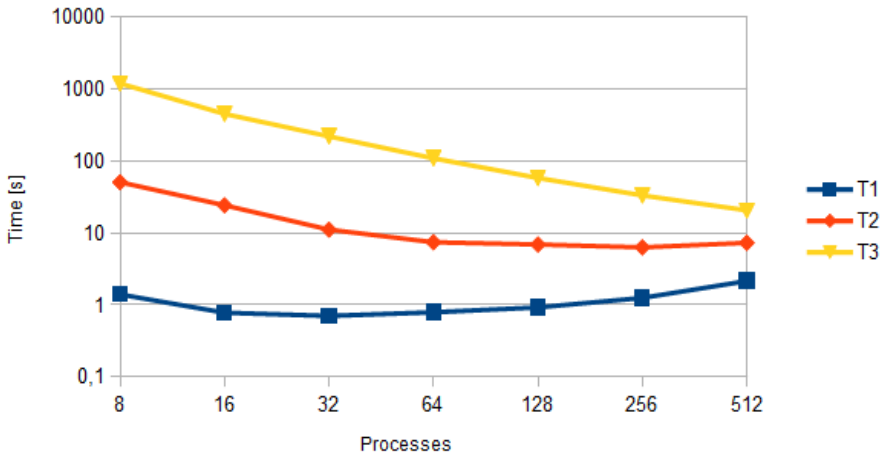


Fig. 2. Scalability of the BB algorithm

5 Conclusion

The main objective was to experimentally verify scalability of the BB algorithm. Based on the measured data, it is possible to assume that BB algorithm can scale

Table 1. Computing time

Problem	Graph nodes	Variables	Processes	Time [s]	t_s/t_i	Visited nodes	Max visited level of tree
T1	12	143	8	1.403	1.0	275	19
T1	12	143	16	0.782	1.8	261	19
T1	12	143	32	0.705	2.0	275	19
T1	12	143	64	0.793	1.8	299	19
T1	12	143	128	0.925	1.5	299	19
T1	12	143	256	1.255	1.1	293	19
T1	12	143	512	2.153	0.7	297	19
T2	14	195	8	50.778	1.0	3953	61
T2	14	195	16	24.336	2.1	3999	57
T2	14	195	32	11.153	4.6	3593	57
T2	14	195	64	7.436	6.8	3243	57
T2	14	195	128	6.916	7.3	3225	57
T2	14	195	256	6.324	8.0	3233	57
T2	14	195	512	7.298	7.0	3237	57
T3	16	255	8	1193.950	1.0	43267	92
T3	16	255	16	451.259	2.6	35819	66
T3	16	255	32	219.854	5.4	36197	66
T3	16	255	64	109.129	10.9	35429	66
T3	16	255	128	58.168	20.5	35477	66
T3	16	255	256	33.309	35.8	35559	66
T3	16	255	512	20.637	57.9	35599	66

well if it has enough branching of the tree, which depends on the given number of nodes. More branching means more opportunities for parallel computing.

The implemented algorithm for solving MTSP is not fast enough to solve large problems. It could be further optimized. For example by using interior point method instead of the simplex algorithm. BB algorithm also could be optimized by changing the order of computed nodes.

References

1. DIABY, Moustapha. *Linear Programming Formulation of the Traveling Salesman Problem*. Connecticut, 2006. University of Connecticut.
2. LIŠKA, Miroslav. *Metody operačního výzkumu*. Ostrava, 2002. University of Ostrava.
3. MILLER, C. E.; TUCKER, A. W.; ZEMLIN, R. A. *Integer programming formulations and traveling salesman problems*. Journal of the ACM, 1960, pages 326–329. New York.
4. ŠŮCHA, Přemysl. *Celočíselné lineární programování* [online]. Praha, 2004. ČVUT. Available at: http://support.dce.felk.cvut.cz/pub/hanzalek/_private/ref/sucha_ilp.pdf

Finding an optimal path for collecting items in the store

Adam Silber

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
adam.silber@vsb.cz

Abstract. Imagine a large warehouse with many aisles in several halls with corridors between them. Suppose we have to pick up several different items for a customer in this warehouse, so the total time of collection depends mainly in which order we choose to pick up the items.

So how to choose the path that passes through all the points of interest? Unfortunately, this is not a computationally simple task. The problem is close to the famous travelling salesman problem. The travelling salesman problem is one of the so-called NP-hard problems. Therefore, it is not known how, in general, to find an exact solution in a reasonable time. For industrial solution it is not possible, when finding the optimal way, to wait for several hours, the result should be available in a matter of seconds in the worst case.

We show the idea of an algorithm, based on the Christofides algorithm, which suits this problem and can generate very good results quickly.

Keywords: store optimization, travelling salesman problem, Christofides algorithm, approximation algorithm.

1 Introduction

For the process, of collecting items from the store, it is often important what path is chosen, because the longer the path is, the more time it takes. For every single order it is necessary to collect several items in the store, and we will try to find the optimal route to make it as fast as possible. Actually we will try to find one of the shortest possible routes or at least something close to it.

Imagine the situation in figure 1. You want to pick up 10 items from the shelves (red numbers) in a large warehouse with many aisles in separate halls. If you choose to pick them up in the given order $1, 2, \dots, 10$ it takes much longer than if you pick up the odd ones first and then the even ones $1, 3, \dots, 9, 2, 4, \dots, 10$.

And it's easy to see, that reordering $1, 3, 7, 5, 9, 10, 8, 4, 2, 6$ would be even better. So, if we want to pick it up as fast as possible, the order in which we pick up items does matter.

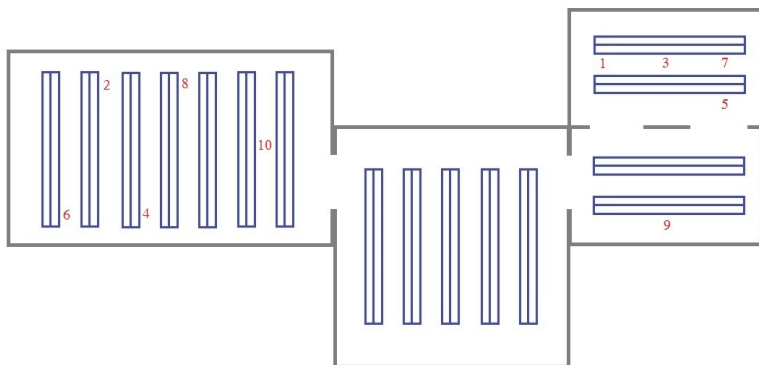


Fig. 1. example of the store with items to pick.

2 Travelling salesman problem

This problem is very close to travelling salesman problem (TSP for further), the famous problem of a salesman who want to visit by shortest possible route specified set of cities, each exactly once. The condition, of visiting each point of interest exactly once, is not so crucial for the store optimization, although in the optimal solution will be met.

TSP is an NP-hard problem, so we don't expect¹ that there exists a polynomial algorithm for an exact solution.

In Graph theory this is a classical optimization task of finding Hamiltonian cycle with the lowest sum of edge labels in a edge-labelled complete graph, where the labels correspondent to distances [3]. We consider only the symmetric variant of the problem, assuming the distance from vertex A to vertex B is equal to the distance from B to A.

3 Approaching the solution

A specific case of the travelling salesman problem is Metric TSP (also known as *delta*-TSP) where distances between vertices (cities) satisfy the triangle inequality, which become a very natural condition for most real life problems. In this case distances must satisfy the conditions for a metric

$$\rho : \mathcal{V}(\mathcal{G}) \times \mathcal{V}(\mathcal{G}) \rightarrow \mathbb{R}$$

¹ one of the Millennium Prize Problems

1. Create a minimum spanning tree T of G (e.g. *Kruskals* algorithm) – step 2 Figure 3.
2. Let \mathcal{O} be the set of vertices with odd degree in T (full vertices) and find a perfect matching M with minimum weight in the complete graph over the vertices from \mathcal{O} - step 3 Figure 3.
3. Combine the edges of M and T together to form a new multigraph H – step 3 Figure 3.
4. Find an Eulerian circuit in H (H is Eulerian multigraph because it is connected and all vertices are of even-degree).
5. The final Hamiltonian circuit is made by Eulerian multigraph H by skipping previously visited nodes (shortcutting²) – step 4,5 Figure 3.

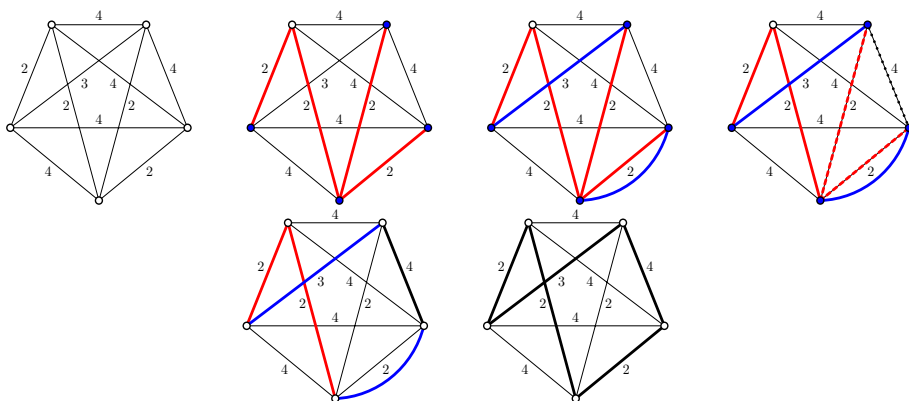


Fig. 3. Steps of The Christofides algorithm.

3.2 Generalization of the problem

We want to represent this problem in the language of graph theory. We will transform items to collect to vertices and every possible route between them in the store to the appropriate edge. That we obtain a graph giving us the first insight into the problem.

If the shelves of the store have more then one level, we merge them into one layer. You always should collect all items from same column at once.

Then, if we are finding the shortest route to collect specified stuff in the warehouse, we reduce the numbers of vertices, there can be thousands of them, will be the vertices represent items we want to collect at this route. So we form an induced subgraph (clique).

² Shortcutting is always possible because the graph G is complete. Because the triangle inequality holds, it never makes the route longer.

Moreover, we use the observation: at the aisle with the dead end there is always need to go for the item, which is from the entrance to the aisle furthest, and on the way back gradually pick up the remaining items.



Fig. 4. The aisle with a dead end.

This means that all items in aisle with the dead end can be represented by single vertex. A similar reduction can be applied to any aisle with two entries. This reduction, properly argued, reduced the size of the problem dramatically from thousands to dozens and it was the crucial contribution in our solution.

Next we describe the algorithm to perform on this reduced graph.

3.3 Algorithm

We will introduce the algorithm for finding an optimal route in the store, which is based on the Christofides algorithm.

input Algorithm has two input parameters: simplified graph G represent given store and set of vertices U represent items to collect, $u \in V(G)$.

1. Assemble the matrix D of the shortest paths between vertices $i, j \in V(G)$.
 $D_{i,j}$ = length of shortest (i, j) -path in G .
2. Generate complete graph K_U (with U as the set of vertices).
3. Label each edge of K_U by $f(\{i, j\}) = D_{i,j}$, $\forall \{i, j\} \in E(K_U)$.
4. Find a minimum spanning tree of K_U .
5. Find a tree T in graph G corresponding to the minimum spanning tree of K_U (Notice - if G is not complete, edges of K_U corresponding to paths in G) so that each edge of G is in $E(T)$ at most once.
6. By W denote the odd-degree vertices of T , $W \subseteq V(G)$. Remark $|W|$ is even by parity principle.
7. Generate complete graph K_W (with W as the set of vertices).
8. Label each edge of K_W by $f(\{i, j\}) = D_{i,j}$, $\forall \{i, j\} \in E(K_W)$.
9. Find minimum cost perfect matching M in K_W .
10. Unite the edges in $E(T)$ and in the matching M to form a (connected) Eulerian multigraph $H = (V(T), E(T) \cup M)$.
11. Find an Eulerian cycle \mathcal{E} in H .

12. Shortcutting, if there is a vertex $x \in \mathcal{E}$ which appears more than once, then its second and any further appearance is erased.

This is how we can find an optimal route for collecting items. Sometimes one vertex represent several items and the order of collecting them depends on the previous simplification, in our case from one side of the aisle to another or from the top downwards.

We have implemented the algorithm in MATLAB and tested it with real data from a real store, where we had to optimize the process of collecting items. We also compared the algorithm with exact TSP algorithm implemented in MATLAB graphs toolbox. The computation time of both algorithms was for a small number of collecting items of the same order. But for orders where have to be visited twelve or more different places in the store the proposed algorithm has found the solution in matter of seconds, while TSP algorithm found the solution of the same length in considerably longer time, it took several minutes.

4 Conclusion

This paper gives a brief introduction to the problem of store optimization, and algorithm to solve it in reasonable time. We proposed some generalizations and we proposed a several ways how to simplify or reduce the theoretical model so it contains only crucial information. Our method was in no case worse than the theoretical minimum.

References

1. D. L. APPLGATE, R. E. BIXBY, V. CHVTAL, W. J. COOK, The Traveling Salesman Problem: A Computational Study, *Princeton University Press*, 606pp, 2006.
2. W. J. COOK, In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation, *Princeton University Press*, 2012.
3. P. KOVÁŘ, Teorie grafů, Graph Theory script [online], VŠB-TU Ostrava, 2012 [cit. 2014-08-06], available at http://homel.usb.cz/~kov16/files/skriptum_teorie_grafu_rozsirene.pdf.
4. J. B. KRUSKAL, JR., On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*, 1956.
5. N. CHRISTOFIDES, Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem. *GSIA-Carnegie Mellon University*, 1976, Report 388.

Boundary Element Method for Mathematical Homogenization of Composite Materials

Matyáš Theuer

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`matyas.theuer@vsb.cz`

Abstract. Numerical realization of mathematical homogenization of elliptic equations is based on homogenization theorem which is unfriendly to boundary element methods. Homogenized coefficients of a periodic material are given by an integral formula that includes derivation of a function - the solution of an auxiliary periodic equation on a unit cell. It turns out that for some composite materials the volume variational formulation of the auxiliary problems can be reduced to the interface between the parts of the composite material and periodic boundary by means of Steklov-Poincaré operator. Furthermore the homogenized coefficient matrix can be computed directly using the auxiliary function values only on the interface.

Keywords: homogenization, boundary element method, composite material

1 Introduction to homogenization

Strictly speaking there are no homogeneous materials in this world. Every material is heterogeneous at some scale. In this paper we deal with composite materials with periodic inhomogeneities on length scales that are much larger than atomic scale but which are essentially homogeneous at macroscopic scale so they seem like homogeneous. Many of their macroscopic properties are dependent on their microscopic structure.

The basic task of homogenization is to determine properties of those seemingly homogeneous materials by studying its microscopic structure. Let's consider simple Dirichlet equation with zero boundary condition that describes stationary heating distribution problem in composite material in domain $\Omega \subset \mathbb{R}^N$ with given inner sources $f : \Omega \rightarrow \mathbb{R}$

$$-\operatorname{div} \left(A \left(\frac{x}{\epsilon} \right) \nabla u^\epsilon(x) \right) = f(x),$$

where A is material function which is rapidly oscillating. We study the sequence of solutions $\{u^\epsilon\}$ for $\epsilon \rightarrow 0^+$. Expectedly under some reasonable conditions this sequence has a limit u^0 that is a solution of a homogenized problem

$$-\operatorname{div} (A^0 \nabla u^0) = f.$$

This corresponds with an idea that heterogeneous material becomes seemingly homogeneous (see fig.1) and it can be represented by constant material function $A^0 \in \mathbb{R}^{N \times N}$. Main problem of mathematical homogenization is to find the

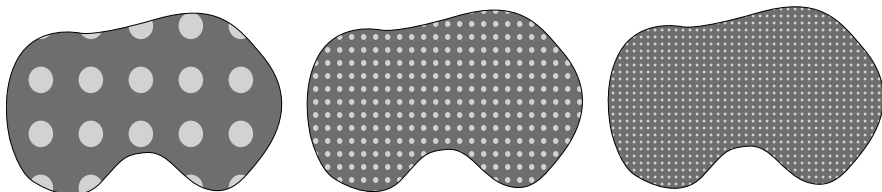


Fig. 1: Changing scale of heterogeneities in composite material with $\epsilon \rightarrow 0^+$.

coefficient matrix A^0 . There are various approaches for various types of materials [7]. For periodic materials there exists an integral formula in which appears a derivation of a function that is a solution to an auxiliary periodic problem on a unit cell that represents the given material [1].

In our work we consider only periodic two-part compound materials with piecewise constant material function. We present a boundary integral formulation of the auxiliary periodic problem so it can be solved by boundary element method. Furthermore this solution can be effectively used for computing the homogenized coefficients since we present adapted formula that needs the auxiliary function values only on the interface between the two parts of the material.

2 Boundary formulation of auxiliary homogenization problem

Let $\Omega \subset \mathbb{R}^N$, $N \in \{2, 3\}$, be a Lipschitz domain and $Y = \langle 0, 1 \rangle^N$ be a unit cube in \mathbb{R}^N . We denote $Y_1 \subsetneq \text{int}Y$ simply connected Lipschitz subdomain of Y and $Y_2 = Y \setminus \overline{Y_1}$ (fig. 2). Let's consider a material coefficient function

$$A(y) = \begin{pmatrix} a(y) & 0 \\ 0 & a(y) \end{pmatrix}, \tag{1}$$

where $a(y)$ is periodic extension of

$$a_b(y) = \begin{cases} a_1, & y \in Y_1, \\ a_2, & y \in Y_2 = Y \setminus Y_1, \end{cases}$$

$$a_1, a_2 \in \mathbb{R}.$$

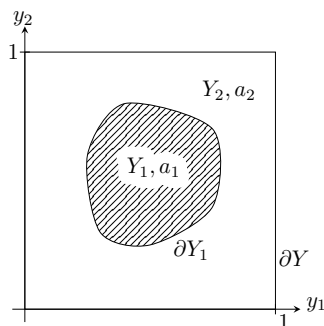


Fig. 2: Unit cell of composite material

Theorem 1 (Homogenization theorem). *Let $\alpha, \beta \in \mathbb{R}, 0 < \alpha \leq \beta$, $A \in M_{per}(\alpha, \beta, \Omega)$ and $f \in L^2(\Omega)$. We denote $\epsilon > 0$ $A^\epsilon(x) = A(\frac{x}{\epsilon})$ and $u^\epsilon \in H_0^1(\Omega)$*

weak solution of

$$\begin{cases} -\operatorname{div}(A^\epsilon \nabla u^\epsilon) = f & \text{in } \Omega, \\ u^\epsilon = 0 & \text{on } \partial\Omega. \end{cases} \tag{2}$$

Then for every sequence (ϵ_n) such that $\epsilon_n \rightarrow 0^+$:

$$u_{\epsilon_n} \rightharpoonup u^0 \text{ v } H_0^1(\Omega),$$

where $u^0 \in H_0^1(\Omega)$ is defined as weak solution of homogenized problem

$$\begin{cases} -\operatorname{div}(A^0 \nabla u^0) = f & \text{in } \Omega, \\ u^0 = 0 & \text{on } \partial\Omega, \end{cases} \tag{3}$$

where

$$A^0 = \int_Y \left(A(y) - A(y) (\nabla_y \chi(y))^T \right) dy$$

and function $\chi(y)$ is solution of

$$\begin{cases} \operatorname{div}(\nabla_y \chi(y) (A(y))^T) = \operatorname{div}_y (A(y))^T & \text{in } Y, \\ \chi(y) \text{ is } Y\text{-periodic.} \end{cases} \tag{4}$$

Proof. See [5], [1] or [8].

According to the homogenization theorem 1 the elements of the coefficient matrix A^0 are given by

$$a_{i,k}^0 = \int_Y \left(a_{i,k}(y) - \sum_{j=1}^2 a_{i,j}(y) \frac{\partial \chi^{(k)}}{\partial y_j}(y) \right) dy.$$

Using Green’s formula, periodicity of function $\chi^{(k)}$ and considering our function A defined in (1), we get alternate formula

$$a_{i,k}^0 = \int_Y a_{i,k}(y) dy + (a_1 - a_2) \int_{\partial Y_1} \chi^{(k)}(y) (n_{Y_1})_i ds_y. \tag{5}$$

We can see from (5) that entries of the homogenized coefficient matrix depends on values of function χ only on the interface between domains Y_1 and Y_2 .

The Auxiliary periodic problem (4) reads in weak sense as follows:

$$\left\{ \begin{array}{l} \text{Find } \chi^k \in H_{\text{per}}^1(Y) : \\ \int_Y a(y) \nabla \chi^k(y) \nabla v(y) dy = \int_Y a(y) \frac{\partial v}{\partial y_k}(y) dy \quad \forall v \in H_{\text{per}}^1(Y), \end{array} \right. \tag{6}$$

for $k = 1, \dots, N$. This problem has unique solution up to a constant and it can be equivalently formulated as follows

$$\left\{ \begin{array}{l} \text{Find } \chi_1^{(k)} \in H^1(Y_1), \chi_2^{(k)} \in H^1(Y_2) \\ -\Delta \chi_1^{(k)} = 0 \quad \text{in } Y_1, \\ -\Delta \chi_2^{(k)} = 0 \quad \text{in } Y_2, \\ \chi_1^{(k)} = \chi_2^{(k)} \quad \text{on } \partial Y_1, \\ a_1 \left\langle \frac{d\chi_1^{(k)}}{dn_{Y_1}}, T_1(v) \right\rangle + a_2 \left\langle \frac{d\chi_2^{(k)}}{dn_{Y_2}}, T_2(v) \right\rangle = \\ \qquad \qquad \qquad = \int_{\partial Y_1} (a_1 - a_2) (n_{Y_1})_k T_1(v) \, ds_y \quad \forall v \in H_{per}^1(Y), \\ \chi_2^{(k)} \text{ satisfies periodic boundary condition} \quad \text{on } \partial Y, \end{array} \right. \tag{7}$$

where functions $\chi_1^{(k)}$ and $\chi_2^{(k)}$ are restrictions of function $\chi^{(k)}$ to domain Y_1 and Y_2 respectively. For proof of equivalence see [3]. Problem (7) can be understood in terms of Steklov-Poincaré boundary integral equations

$$\left\{ \begin{array}{l} \text{Find } \chi_1^{(k)} \in H^{1/2}(\partial Y_1), \chi_2^{(k)} \in H_{per}^{1/2}(\partial Y_2) \\ a_1 \left\langle S_1(\chi_1^{(k)}), v|_{\partial Y_1} \right\rangle + a_2 \left\langle S_2(\chi_2^{(k)}), v \right\rangle = \\ \qquad \qquad \qquad = \int_{\partial Y_1} (a_1 - a_2) (n_{Y_1})_k T_1(v) \, ds_y \quad \forall v \in H^{1/2}(\partial Y_2) \\ \chi_1^{(k)} = \chi_2^{(k)} \quad \text{on } \partial Y_1. \end{array} \right. \tag{8}$$

We use standard Galerkin discretization of the single-layer, double-layer, and hypersingular boundary integral operators on ∂Y_1 and ∂Y_2 respectively. Finally we get system of equations

$$\begin{pmatrix} a_1 \mathbf{S}^{(1)} + a_2 \mathbf{S}_{1,1}^{(2)} & a_2 \mathbf{S}_{1,2}^{(2)} \mathbf{P} \\ a_1 \mathbf{P}^T \mathbf{S}_{2,1}^{(2)} & a_2 \mathbf{P}^T \mathbf{S}_{2,2}^{(2)} \mathbf{P} \end{pmatrix} \begin{pmatrix} \chi_1^{(k)} \\ \chi_2^{(k)} \end{pmatrix} = \begin{pmatrix} (a_1 - a_2) \mathbf{n}^{(k)} \\ \mathbf{0} \end{pmatrix}$$

where \mathbf{S} represents the matrix of discretized Steklov-Poincaré operator and \mathbf{P} is permutation matrix ensuring periodic boundary conditions. The discretization error estimate is as follows

$$\left| a_{i,k}^0 - a_{i,k}^{0,h} \right| \leq |a_1 - a_2| c(Y_1, \chi^{(k)}) h. \tag{9}$$

Numerical results presented in next section shows that for some inclusion shapes the convergence rate can be even better than h .

3 Numerical results

At first we compare classical FEM and our BEM solutions u^ϵ of (2) with solution u^0 of (3) for various epsilons. We measure $e_s = \|u^\epsilon - u^0\|_{sup}$. As a testing inclusion we used a square shape and Ω is also a square domain. As we can see from tab. 1 the difference between solutions of equations with rapidly oscillating

coefficients and the homogenized solution is getting smaller as $\epsilon \rightarrow 0^+$. We can also observe that there is no difference between FEM and BEM results. This is caused by fact that the homogenized matrices A^0 obtained by FEM and BEM are nearly the same which shows that both methods are equivalent in means of correctness.

ϵ	e_{s-FEM}	e_{s-BEM}
1	$1,27 \cdot 10^{-2}$	$1,27 \cdot 10^{-2}$
0,5	$2,92 \cdot 10^{-3}$	$2,92 \cdot 10^{-3}$
0,25	$1,68 \cdot 10^{-3}$	$1,68 \cdot 10^{-3}$
0,1	$8,10 \cdot 10^{-4}$	$8,10 \cdot 10^{-4}$
0,05	$4,08 \cdot 10^{-4}$	$4,08 \cdot 10^{-4}$

h	$e_{max-\square}$	e_{max-U}	e_{max-rU}
$2,5 \cdot 10^{-2}$	$1,24 \cdot 10^{-2}$	$1,54 \cdot 10^{-2}$	$2,04 \cdot 10^{-2}$
$1,25 \cdot 10^{-2}$	$4,52 \cdot 10^{-3}$	$6,06 \cdot 10^{-3}$	$6,72 \cdot 10^{-3}$
$6,25 \cdot 10^{-3}$	$1,60 \cdot 10^{-3}$	$2,15 \cdot 10^{-3}$	$2,30 \cdot 10^{-3}$
$3,13 \cdot 10^{-3}$	$5,28 \cdot 10^{-4}$	$7,19 \cdot 10^{-4}$	$7,54 \cdot 10^{-4}$
$1,56 \cdot 10^{-3}$	$1,41 \cdot 10^{-4}$	$1,92 \cdot 10^{-4}$	$2,00 \cdot 10^{-4}$

Table 1: Difference between u^ϵ and u^0 .

Table 2: Discretization errors.

Let's now concentrate on convergence of approximate coefficients $a_{i,k}^{0,h}$ to $a_{i,k}^0$ as $h \rightarrow 0^+$. We use three testing shapes of inclusions (fig. 3) for comparing discretization error $e_{max} = \|A_{min}^0 - A_h^0\|_{max}$. Results are presented in tab. 2 and visualization of errors (fig. 4b) in logarithmic scale shows that rate of convergence is between h and h^2 .

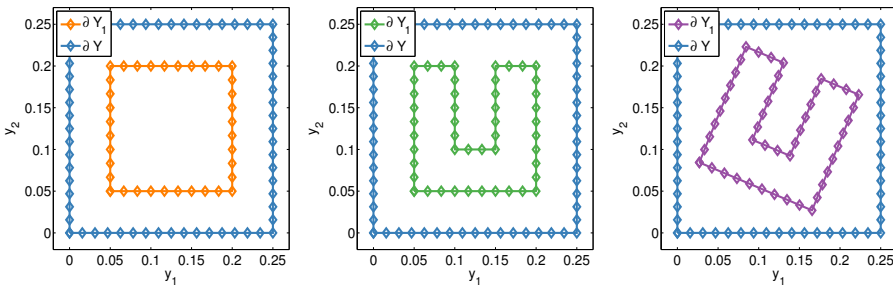


Fig. 3: Testing inclusions: square, u-shape and rotated u-shape.

4 Conclusion and outlook

In this paper we briefly presented derivation and numerical results of boundary element method for solving auxiliary periodic problem that arises in homogenization of two-compound composite materials. The results are comparable with classical FEM approach. This topic is more extensively studied in [3]. There we present for example proof of equivalence of problem (6) and (7) and rate of convergence for Galerkin discretization.

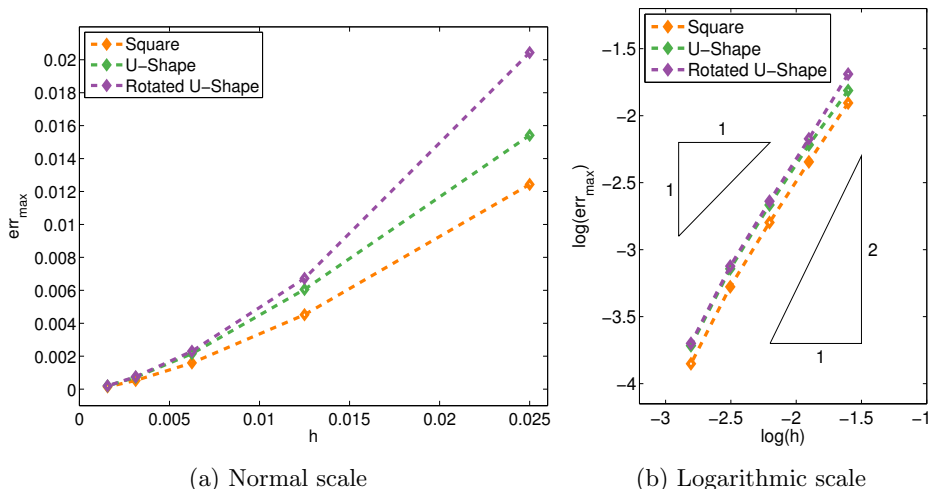


Fig. 4: Discretization errors for testing inclusions.

Acknowledgements

This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme. The work was also supported by VŠB—Technical University of Ostrava under the grant SGS SP2013/191.

References

1. Allaire, G.: Homogenization and two-scale convergence. *SIAM J. Math. Anal.* 1992, 23, 6, pp. 1482-1518.
2. Allaire, G.: *School on Homogenization, Two-Scale Convergence and Homogenization of Periodic Structures*, pp. 1-24. ICTP, Trieste, 1993.
3. Bouchala, J. - Lukáš, D. - Theuer, M. Boundary Element Method for Homogenization *Mathematics and Computers in Simulation* (in preparation).
4. Chechkin, G. A. - Piatniski, A. L. - Shamaev, A. S. *Homogenization: Methods and Applications*. USA : AMS, 2007. ISBN 0-8218-3873-3.
5. Cioranescu, D. - Donato, P. *Introduction to Homogenization*. New York : Oxford University Press Inc., 1999. ISBN: 0-19-856554-2.
6. Kouznetsova, V. G. *Computational homogenization for the multi-scale analysis of multi-scale materials*. PhD thesis, Technische Universiteit Eindhoven, 2002. ISBN 908-386-2734-3.
7. Pavliotis, G. A. - Stuart, A. M. *Multiscale methods - Averaging and Homogenization*. New York : Springer, 2008. ISBN 978-0-387-73828-4.
8. Theuer, M. *Homogenization of Elliptic Equations* Master's thesis, VŠB - Technical University of Ostrava, 2013.

Graph Partitioning using Spectral Methods

Radek Tomis

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

Abstract. This article deals with graph partitioning based on spectral bisection. Graph partitioning can be used in many technical applications. Our main criterion for graph partitioning is to create components with a minimum number of interconnections. Although the spectral bisection is a well known and frequently used partitioning algorithm, it is not optimal in terms of our criteria. There are many approaches for choosing the best edge separator to produce spectral bisection. Some of these approaches are tested and compared at the end of this article. In addition, a comparison of spectral bisection with a graph partitioning algorithm from METIS software package is made.

Keywords: graph partitioning, spectral bisection, road network

1 Introduction

Various technical problems can be represented as a graph and solved by graph algorithms. If these problems are computationally intensive, it might be beneficial to split the graph into multiple smaller parts. Graph algorithms can usually work much faster on smaller graphs, because they can operate in parallel. Therefore, many computational problems can be significantly simplified with graph partitioning.

For testing purposes the graph of a road network from the open source project OpenStreetMap¹ was used. Road networks are specific with their high sparsity levels. Partitioning the graph into smaller subgraphs can be a solution for many problems. For example, the common shortest path problem can be speeded up substantially [1] by creating partitions with minimum interconnections (edges connecting different partitions). Therefore, in our case the minimum number of interconnections is our main criteria for graph partitioning.

Graph partitioning also provides interesting results in other applications. For example in conjunction with social networks, which by their nature are very similar to road networks. Social networks are very sparse and contain clusters, which can be used for communities identification. Graph partitioning of the social network graph can naturally partition the graph by clusters, resulting in easy communities identification.

¹ <http://www.openstreetmap.org/>

The main focus of this article are two methods for solving graph partitioning. The first method is spectral bisection [6] and the second is multilevel k-way from a METIS software package [2]. In this work, we will focus more on spectral bisection. Theory about the spectral bisection will be described in section 2. The algorithm was implemented and tested on Anselm supercomputer. The graph algorithm from METIS software package was used to compare the quality of spectral bisection partitioning. Some results of spectral bisection testing and comparison with METIS will be presented in section 3.

2 Theory

2.1 Basic Concepts

Graph partitioning algorithm should partition given undirected² weighted³ graph $G = (V, E)$ into two partitions in such a way, that its vertex set V should be divided into two disjoint sets such that $V = V_1 \cup V_2$. These sets should be approximately the same size. In other words, we try to find an edge separator E_s as the subset of E such that removing E_s from E divides G into two disconnected subgraphs G_1 (containing V_1) and G_2 (containing V_2).

There are various criteria for graph partitioning. For example in road networks the criteria might be to create partitions with minimum interconnections. In other words we need to minimize the number of edges connecting partitions, i.e. minimize the number of edges in E_s . Other possible partitioning criteria might be to minimize the value of edge cut, i.e. minimize the sum of weights of edges in E_s .

2.2 Spectral Bisection

We decided to use spectral bisection because it is known that it can partition graph by our criteria. In brief, for creation of spectral bisection we first need to rewrite given graph in the form of Laplacian matrix [7] described below in section 2.3. Then, we calculate the smallest nonzero eigenvalue of the Laplacian matrix and the eigenvector belonging to this eigenvalue, called Fiedler vector [3]. The elements of this vector determine the partitioning of the graph.

2.3 Laplacian Matrix

Let $A = A(G)$ be $n \times n$ adjacency matrix. $A_{i,j} = w(v_i, v_j)$, where $w(v_i, v_j)$ is the weight of the edge between vertices v_i and v_j . If there is no edge between v_i and v_j then $A_{i,j} = 0$.

Let $D = D(G)$ be $n \times n$ degree matrix. $D_{i,i} = d(v_i)$, where $d(v_i) = \sum_{v_j} w(v_i, v_j)$.

² Graph can actually be directed, because the direction of edges has no impact on resulting partitions, therefore we can easily convert directed graph to undirected.

³ We can use edge weights of all ones in case of unweighted graph

Laplacian matrix [7] $L = L(G)$ is $n \times n$ symmetric matrix defined as $L = D - A$. We can also define Laplacian matrix as follows: $L = l(i, j)$,

$$l(i, j) = \begin{cases} d(v_i) & \text{if } i = j \\ -w(v_i, v_j) & \text{if } i \neq j \text{ and edge between } v_i \text{ a } v_j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

Laplacian matrix represents graph connectivity. All eigenvalues of the Laplacian matrix are real, because L is a symmetric matrix. We also know that the eigenvalues are nonzero from Gershgorin circle theorem. Thus L is positive semidefinite. Number of zero eigenvalues L corresponds to the number of continuous components of G . Therefore L has always at least one zero eigenvalue. Eigenvalues are also naturally sorted by value [3], i.e. $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. The smallest non-zero eigenvalue (λ_1 for continuous G) defines an algebraic connectivity of G .

2.4 Fiedler Vector

Eigenvector belonging to the smallest nonzero eigenvalue is called Fiedler vector [3]. Fiedler vector has specific features. The most important feature is that value n at position i in this vector corresponds to the connectivity of vertex v_i . We choose separator value s and create partitions of G easily by comparing n to s . If $n > s$, v_i is put to G_1 , otherwise G_2 .

Separator value s can be chosen in various ways. Common values for s are for example zero or a median of Fiedler vector values. However, these values doesn't have to partition graph with minimal interconnections or minimal value of edge cut. Therefore we can use different values for s with the goal to minimize interconnections or edge cut value.

3 Testing Results

Spectral bisection was used for graph partitioning of OpenStreetMap road network. Algorithm was implemented in *C++* with libraries PETSc [4] and SLEPc [5]. These libraries are using MPI for paralelization. Our implementation can be used recursively to create more than two partitions, however the number of partitions has to be always the power of two. METIS⁴ graph partitioning package was used for comparison of partitioning quality.

The Anselm supercomputer with Intel Sandy Bridge E5-2665 2.4GHz processors was used. The application was running with the specified number of computational cores.

In the spectral bisection the most computationally expensive operation is to calculate the smallest non-zero eigenvalue and its eigenvector (Fiedler vector). To calculate the eigenvalues and eigenvectors the Krylov-Schur [8] method was used.

⁴ Parallel version of METIS called ParMETIS wasn't used because it doesn't guarantee to create contiguous partitions.

The precision of eigenvalues is very important, because even a small change of eigenvalue (such as rounding error) can have an essential influence on the related eigenvector. SLEPc allows to set the relative error tolerance of the solution. For testing the default value of 10^{-8} was used.

Testing was conducted on various sizes of OpenStreetMap road network graphs. Table 1 lists tested graphs. Table 2 shows computing time related to the number of computing cores used. Figure 1 shows the scalability of our spectral bisection implementation.

Table 1. Sizes of test graphs.

Graph	Vertex count	Edge count
Eastern Moravia	117497	291142
Czech republic	570368	1447868

Table 2. Computation times for our implementation of spectral bisection.

Graph	Computing time [s] for given number of cores							
	1 core	2 cores	4 cores	8 cores	16 cores	32 cores	64 cores	128 cores
Eastern Moravia	590	290	157	82	47	62	89	194
Czech republic	16270	7373	3987	2211	1824	1646	1655	2368

The partitioning test results of Eastern Moravia points out that the fastest calculation occurred with 16 cores. The reason for this is that there are exactly 16 cores within one computing node, and therefore, the application uses the local communication between processes. Larger number of cores start to slow down the computation.

The situation is different for the graph of Czech Republic. There is an observable acceleration of calculation beyond the borders of one computing node, specifically 2 (32 cores) and 4 (64 cores). Therefore we can conclude that a much larger graphs could make better use of higher number of computing nodes. However, we should note that PETSc and SLEPc does not scale too well in our case.

We have to note that METIS software used for partitioning comparison ran only on one computation core and its computing time was in all scenarios below 5 seconds, which is negligible compared to spectral bisection.

The comparison of partitioning quality is in Table 3. The lower number of interconnections is better. Explanation for separator value s is in section 2.4. Separator value named "best" is the value within the range of zero and median with best partitioning result, i.e. the value with minimal interconnections. Spectral bisection was worse than METIS in both test instances.

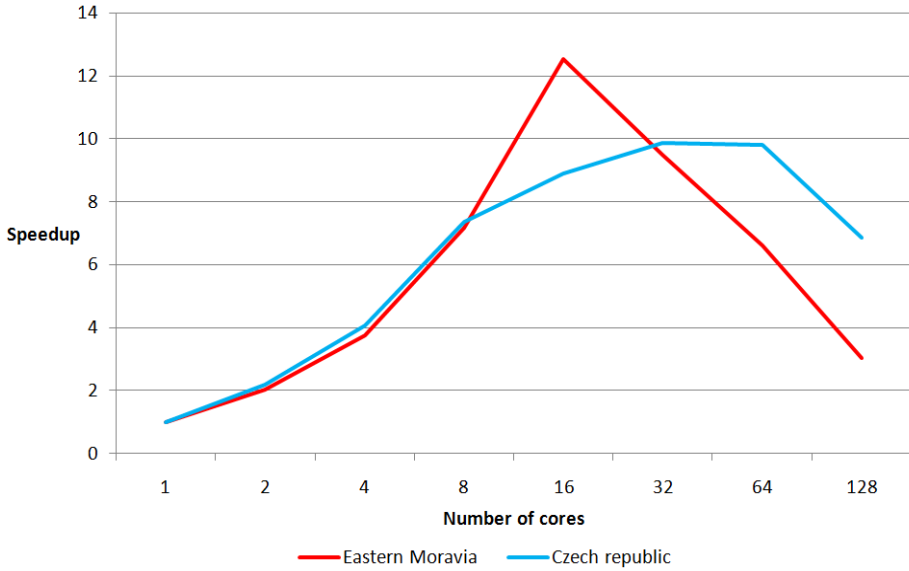


Fig. 1. Scalability of our implementation of spectral bisection.

Table 3. Number of interconnections for different separator values compared with METIS.

Graph	Number of interconnections			
	Spectral bisection			METIS
	$s = 0$	$s = median$	$s = best$	
Eastern Moravia	160	164	134	74
Czech republic	376	334	190	188

4 Conclusion

Although spectral bisection is very interesting in terms of numerical methods, the time required for its calculation is significantly larger in comparison with METIS. In addition, the quality of partitioning using spectral bisection method does not improve over partitioning obtained by METIS. Therefore METIS is far better option for graph partitioning.

References

1. Flinsenber, Ingrid, et al. Creating graph partitions for fast optimum route planning. *WSEAS Transactions on Computers* 3.3 (2004): 569-574.
2. George Karypis and Vipin Kumar. A Fast and Highly Quality Multilevel Scheme for Partitioning Irregular Graphs. *SIAM Journal on Scientific Computing*, Vol. 20, No. 1 (1999), pp. 359-392.
3. Fiedler, Miroslav. "Algebraic connectivity of graphs." *Czechoslovak Mathematical Journal* 23.2 (1973): 298-305.
4. Satish Balay, William D. Gropp, Lois Curfman McInnes, and Barry F. Smith. 1997. Efficient management of parallelism in object-oriented numerical software libraries. In *Modern software tools for scientific computing*, Erlend Arge, Are Magnus Bruaset, and Hans Petter Langtangen (Eds.). Birkhauser Boston Inc., Cambridge, MA, USA 163-202.
5. Vicente Hernandez, Jose E. Roman, and Vicente Vidal. 2005. SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans. Math. Softw.* 31, 3 (September 2005), 351-362.
6. Pothen, Alex, Horst D. Simon, and Kang-Pu Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications* 11.3 (1990): 430-452.
7. Fan R. K. Chung. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics, 92 (1997).
8. G. W. Stewart. A Krylov-Schur Algorithm for Large Eigenproblems. *SIAM J. Matrix Anal. Appl.* 23, 3 (March 2001), 601-614.

Elastic image registration based on domain decomposition with mesh adaptation

Alena Vasatova and Ales Ronovsky

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
alena.vasatova@vsb.cz

Abstract. Medical images are increasingly being used within healthcare for diagnosis, planning treatment, and monitoring disease progression. These images are acquired at different circumstances. Thus it often yields that additional clinical information that is not revealed in the separate images. In our contribution, we use elastic registration which supposes that the images are two different observations of an elastic body which is discretized by the finite element method. Because medical images usually contain large area of background and even only small area of changes, regular discretization results in waste of computational resources due to the fine refinement of the space outside the region of interest (especially in 3D). To avoid this, we use coarser grid with local refinement that takes into account specific features of the images and their differences. The related elasticity problems are solved by TFETI, which is a variant of the FETI domain decomposition method for massively parallel numerical solution of elliptic PDE with optimal complexity.

Keywords: elastic image registration, domain decomposition, mesh adaptation

1 Introduction

Image registration is one of the challenging problems in image processing. Since the information gained from two (or more) images is usually of a complementary nature, proper integration of useful data obtained from the separate images is often desired. Thus there is a need to establish an exact point-to-point correspondence between the voxels in one image and those in the other image, which makes direct comparison possible. Although manual alignment of images is possible, it is time-consuming (especially in more than two dimensions) and lacks reproducibility. In practice, the specific type of the geometric transformation as well as the meaning of the correspondence depends on the specific application.

The medical image registration is a vital component of a large number of applications, see [1], [2]. Such applications include the alignment of data sets from different modalities, comparison of follow-up scans to a base-line scan, alignment of pre- and post-contrast images, updating treatment plans for radiotherapy and

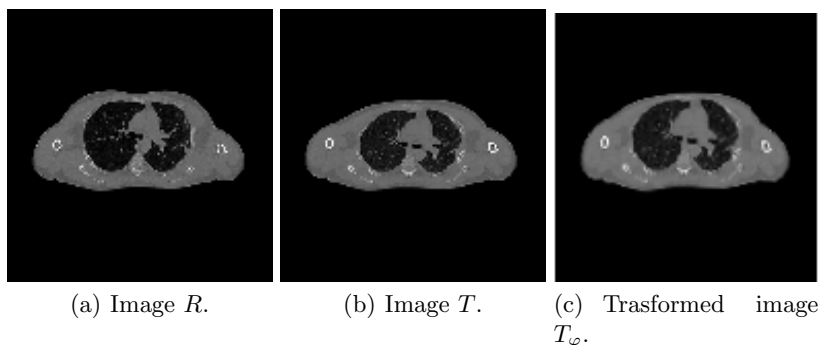


Fig. 1. Registration example. Images in this paper are from department of Oncology at the University Hospital of Ostrava. The images show the cuts in the chest inhale and exhale obtained by CT method.

surgery, monitoring of diseases, atlas-based segmentation and creating models of anatomy.

Usually, one of the images is viewed as a reference (target, fixed, baseline) R and the other one as a deformable template (source, moving, floating) T . The optimal transformation φ is estimated by minimizing the cost function, called distance measure (similarity measure or metric) \mathcal{D} , which determines how much is the image R , in a certain sense, similar to the image T . So the problem can be formulated as

$$\mathcal{D}[R, T; \varphi] := \mathcal{D}[R, T_\varphi] \xrightarrow{\varphi} \min. \quad (1)$$

Problem (1) is a challenging optimization problem, in particular in 3D, whose solution requires effective solution of elasticity problems. In our paper, we propose to use a variant of the FETI domain decomposition method, although usual applications of FETI method lie in the field of material sciences and structural mechanics.

The performance of the FETI method can be improved by effective mesh generation. In case of images of human body, e.g. CT scans, there are many heterogeneous parts of the image domain in which it is not necessary to have so fine mesh as in others. Mesh adaptation is clearly recognized as an efficient and powerful method for improving the accuracy of the solution as well as for capturing the physical phenomena behavior. Reducing the number of degrees of freedoms (DoF) allows to substantially reduce the CPU time.

The adaptive meshing in the field of elastic image registration was used for example in [3], but their approach requires special and careful discretization of the variational form on adaptive quad/octree grids.

2 Elastic image registration

The elastic potential of deformation in connection with the image registration has been introduced by Broit [4]. 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 $u : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. As the regularizer we use linearized elastic potential, where λ and μ are the so-called Lamé constants. We choose $\lambda = 0$, thus the body shows no contraction under deformation, and to choose μ we use heuristic approach. 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 replaced by a viscous fluid model.

The optimal transformation is estimated by minimizing the cost function, called the distance measure \mathcal{D} , which determines how much is the image R , in a certain sense, similar to the 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)

$$\mathcal{D}[R, T; u] := \frac{1}{2} \|T_u - R\|_{L_2(\Omega)}^2, \quad (2)$$

where $T_u(x) := T(x - u(x))$. The forces $f : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ are derived from its Gâteaux-derivative. 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 $x \notin \Omega$.

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 by the solution of partial differential equation

$$\mu \Delta u + (\lambda + \mu) \nabla \operatorname{div} u = -f. \quad (3)$$

Modersitzki [2] solves this equation by finite difference approximation with periodic boundary conditions, the resulting matrix is highly structured. This structure allows him to use Fast Fourier transformation to diagonalize and invert matrix. Thus he gains very effective solution with complexity $O(N \log N)$. But this approach is sensitive to other constraints, because the structure of the matrix could be destroyed.

We discretize the problem using a finite element method with piece-wise affine basis functions on triangular elements. To approximate the image gradient we use a convolution with an appropriate kernel of the Sobel operator.

We would like to mention some examples for future use of elastic registration. Non-rigid image registration was already used for automatically quantify small changes in volume of anatomical structures in brain over time by means of segmentation propagation, in [5] for dementia progression, in [6] monitoring of the response of the brain to drug therapy. Another example can be modeling

organs motion during the respiratory cycle in radiotherapy to allow for increased radiation dose to the tumor while minimizing the dose to healthy tissue, in [7] modele liver motion, in [8] study thorax and lung deformation to developpe tools to evaluate for reproducibility of ABC (Active Breath Control).

3 Adaptive mesh generation

Different methods exist for adaptive mesh generation, see e.g. [9] or [10]. We used a method that locally increase resolution depending on the domain. As a refinement criteria could be used manual selection of the area of interest, but we are more focused on automatisation of the process so we used a priory estimate based on geometry of the differences of given images. The geometry closely correspond to areas where the image is transforming so the solver needs higher resolution there to hold required accuracy.

Full refinement algorithm starts with initial mesh that is generated uniformly by subdividing domain of the image. Then the algorithm works in number of levels given by user of by resolution of given image to get highest mesh resolution. The refinement process marks all elements from processed level as selected for refinement, transition elements and convenient ones. For refining we use two types of refinement patterns. Pattern for full refinement of selected elements which uses mid-edge points, and transition pattern which ensure that the resulted adaptive mesh will be still conformal. Transition template uses bisection of an element. Mesh for the next level of processing is then build from convenient elements and all elements created by refining.

To keep that the quality of the resulted mesh is not decreasing much we mark the elements that share node with originally selected elements as selected elements too. That moves transition little bit further from the border of area of interest and that ensures that each element is refined by transition pattern only once. Because standard pattern for full refinement keeps quality of the mesh, and transition pattern is used on each elements at most once, the worst angle form the initial mesh is halved in worst case.

4 Problem Solution

To solve the partial differential equation TFETI (proposed in [11]), a variant of FETI (Finite Element Tearing and Interconnecting), was used. It is one of the most successful methods for parallel solution of elliptic partial differential equations. The basic idea of TFETI is to simplify the inversion of stiffness matrices of subdomains by using Lagrange multipliers not only for gluing of the subdomains along the auxiliary interfaces, but also to 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.

5 Mesh partitioning

Using local mesh refinement we end up with irregular mesh, which need to be divided into equal-sized pieces with few connecting nodes because of domain decomposition method. For this purpose we use Matlab Mesh Partitioning and Graph Separator Toolbox. This toolbox contains Matlab code for several graph and mesh partitioning methods. Specifically, we used quite simple, but also quite efficient geometric method [12].

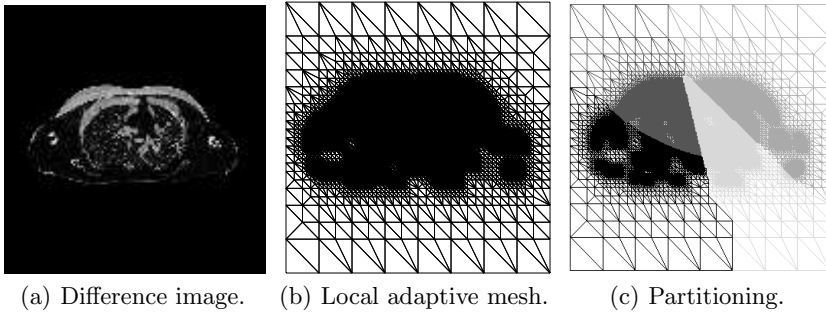


Fig. 2. Example of image differences, adaptive mesh and mesh partitioning.

6 Numerical experiments

Algorithm for creating adaptive image-based mesh described in this paper and assembling matrix for TFETI was implemented in MATLAB. TFETI method itself, was performed by FLLOP (FETI Light Layer On top of PETSc), the package for constrained quadratic programming and FETI domain decomposition, see [13].

The numerical experiments were, for now, performed on personal computer with just two processors. In the next stage we will be testing on supercomputer cluster Anselm.

7 Conclusion

In our work we enhanced capability of image registration algorithm solved by TFETI method by connecting with local adaptive mesh creation. Because body scans, medical images or similar models usually contain huge areas where is no need for computation with higher accuracy, using local adaptive mesh allows us to obtain higher precision of an solution with adequate number of DoF with the comparable computational cost. Navier–Lamé equation are solved by TFETI, a domain decomposition method that is a variant of the FETI method, which

is very suitable for parallel numerical solution of an elliptic PDE. Using local adaptive mesh combining with domain decomposition method could provide strategy for future computing where very huge meshes will be required for solving 3D medical images.

8 Acknowledgment

This publication was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and the project of major infrastructures for research, development and innovation of Ministry of Education, Youth and Sports with reg. num. LM2011033.

References

1. J. B. Antoine Maintz, M. A. Viergever, A survey of medical image registration *Medical Image Analysis*, Vol. 2, No. 1, 1998, pp. 1–36.
2. J. Modersitzki *Numerical Methods for Image Registration*, Oxford : Oxford University Press, 2004.
3. J. Modersitzki Adaptive mesh refinement for nonparametric image registration, *SIAM Journal on Scientific Computing*, Vol. 30, No. 6, pp. 3012-3027
4. C. Broit, *Optimal Registration of Deformed Images*, PhD thesis, University Pennsylvania, 1981.
5. W. R. Crum, R. I. Scahill, N. C. Fox, Automated Hippocampal Segmentation by Regional Fluid Registration of Serial MRI: Validation and Application in Alzheimers Disease, *NeuroImage*, Vol. 13, No. 5, 2001, pp. 847-855.
6. M. Holden, J. A. Schnabel, D. L. G. Hill, Quantification of Small Cerebral Ventricular Volume Changes in Treated Growth Hormone Patients Using Nonrigid Registration, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, Vol. 21, No. 10, 2002, pp. 1292-1301.
7. T. Rohlfing, C. R. Maurer, Jr., W. G. O'Dell, J. Zhong, Modeling liver motion and deformation during the respiratory cycle using intensity-based free-form registration of gated MR images, *SPIE Proceedings Medical Imaging 2001: Visualization, Display, and Image-Guided Procedures*, Vol. 4319, pp. 337, 2001.
8. V. Boldea, D. Sarrut, S. Clippe, Lung Deformation Estimation with Non-Rigid Registration for Radiotherapy Treatment, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, Vol. 2878, 2003, pp. 770–777, 2003.
9. P.J. Frey, P.L. George, Mesh Generation. Applications to Finite Element Methods, London, ISTE, 2008.
10. M.C. Rivara, Local modification of meshes for adaptive and/or multigrid finite element methods, *J. Comput. Appl. Math.* 36 (1991) 79–89.
11. Z. Dostal, D. Horak, R. Kucera, 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.
12. J. R. Gilbert, G. L. Miller, S. Teng, Geometric mesh partitioning: Implementation and experiments, *SIAM J. Scientific Computing*, Vol. 19, No. 6, 1998, pp. 2091–2110.
13. D. Hork, V. Hapla, TFETI coarse problem massively parallel implementation, *EC-COMAS 2012 - European Congress on Computational Methods in Applied Sciences and Engineering*, 2012, pp. 8260-8267.

Shape Optimization for Free Surface Problems Using Subdivision Techniques

Jan Zapletal

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.zapletal@vsb.cz

Abstract. We present a shape optimization workflow based on subdivision techniques widely used in computer graphics. This procedure approximates the optimal shape on a hierarchy of meshes with different resolution and understands every node of a coarse control mesh as a design parameter. To solve the underlying state and adjoint boundary value problems we use the boundary element method, which reduces the problem from the domain to its boundary and thus avoids meshing of the volume. This is a great advantage over the finite element method, where the volume mesh can deteriorate with shape perturbations and a costly remeshing is usually necessary.

1 Introduction

We aim to solve the exterior Bernoulli free surface problem by tracking the Neumann data [4], i.e., we look for an optimal domain $\Omega^* \in \mathbb{R}^3$, such that

$$J(\Omega^*) \leq J(\Omega) \text{ for all } \Omega \in \mathcal{O},$$

where \mathcal{O} denotes the set of admissible domains,

$$J(\Omega) := \frac{1}{2} \left\| \frac{\partial u}{\partial \mathbf{n}} - Q \right\|_{L^2(\Gamma_f)}^2 = \frac{1}{2} \int_{\Gamma_f} \left(\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) - Q \right)^2 d\mathbf{s}_{\mathbf{x}} \quad (1.1)$$

with the fitting constant $Q \leq 0$, Γ_f denoting the free surface, and u solving the Dirichlet boundary value problem for the Laplace equation

$$\begin{cases} -\Delta u = 0 & \text{in } \Omega, \\ u = 1 & \text{on } \Gamma_0, \\ u = 0 & \text{on } \Gamma_f. \end{cases} \quad (1.2)$$

In (1.2), Γ_0 denotes the fixed component of the boundary $\partial\Omega = \Gamma_f \cup \Gamma_0$.

2 Shape Calculus

To find the optimal shape we use the ‘first-optimize-then-discretize’ technique based on the first-order shape calculus [3, 11]. Geometry updates are described

by the perturbation of identity mapping $\mathbf{T}(t, \mathbf{x}) := \mathbf{x} + t\mathbf{V}(\mathbf{x})$. Starting from an initial domain Ω , we are looking for a suitable speed field \mathbf{V} , such that the updated domain $\Omega_t := \mathbf{T}(t, \Omega)$ with t small enough reduces the cost (1.1). The whole process runs iteratively and is performed by a first-order minimization algorithm. The gradient information is provided by the shape derivative of J in the direction \mathbf{V}

$$J'(\Omega)(\mathbf{V}) := \lim_{t \rightarrow 0^+} \frac{J(\Omega_t) - J(\Omega)}{t}.$$

In the case of (1.1) the shape derivative reads [3–5]

$$J'(\Omega)(\mathbf{V}) = \int_{\Gamma_f} -\frac{\partial p}{\partial \mathbf{n}} \frac{\partial u}{\partial \mathbf{n}} - \frac{H}{2} \left(\left(\frac{\partial u}{\partial \mathbf{n}} \right)^2 - Q^2 \right) \langle \mathbf{V}, \mathbf{n} \rangle \, ds \quad (2.1)$$

with H denoting the additive curvature and the adjoint state p solving

$$\begin{cases} -\Delta p = 0 & \text{in } \Omega, \\ p = 0 & \text{on } \Gamma_0, \\ p = \frac{\partial u}{\partial \mathbf{n}} - Q & \text{on } \Gamma_f. \end{cases} \quad (2.2)$$

The Hadamard-Zolésio form of the shape derivative (2.1) allows us only to define \mathbf{V} on Γ_f . Using the representation $\mathbf{V} := \sum_i \alpha_i \varphi_i$ with coefficients α_i and ansatz functions φ_i defined as unit normal perturbations of mesh nodes \mathbf{x}_i , the i -th component of the gradient is given by

$$g_i := J'(\Omega)(\varphi_i). \quad (2.3)$$

Thus, the coefficients $\alpha_i := -g_i$ define a decrease direction (see Figure 5.1a for an example of such a speed field).

3 Subdivision Surfaces

To describe the free surface we use subdivision techniques known from computer graphics. The free surface is represented by a coarse mesh, whose repeated refinement leads to a smooth limit surface. Since every refinement level represents the same limit surface, this approach allows us to use different meshes for the optimization and analysis stages. While BEM requires a fine enough mesh to obtain solutions to (1.2), (2.2), and to calculate the shape derivative (2.1) and gradient (2.3), it is advantageous to use a coarser mesh to perform shape perturbations.

In the work presented we use the Loop subdivision scheme [8], which leads to quartic box splines on regular triangular meshes, i.e., meshes where every node is incident to six edges. Although it is not possible to construct a triangular mesh of a closed surface in \mathbb{R}^3 only with regular vertices, it can be shown that for an arbitrary mesh the limit surface is C^1 continuous around nonregular vertices and C^2 continuous everywhere else.

The subdivision step consists of two stages. Firstly, every triangle of the mesh is refined by quadrisection by introducing new vertices at the midpoint of every edge (see Figure 3.1a). In the second step the positions of both old and new vertices are updated (see Figures 3.1b, 3.1c).

A vertex \mathbf{x} that had already been present in the mesh on the previous refinement level is updated to the position $\tilde{\mathbf{x}}$ as

$$\tilde{\mathbf{x}} := (1 - a)\mathbf{x} + \frac{a}{v} \sum_{i=1}^v \mathbf{x}^i, \tag{3.1}$$

with

$$a := \frac{1}{64} \left[40 - \left(3 + 2 \cos \frac{2\pi}{v} \right)^2 \right],$$

v denoting the number of edges adjacent to \mathbf{x} , and \mathbf{x}^i denoting the neighbouring vertices.

The position of a newly inserted vertex is computed as

$$\tilde{\mathbf{x}} := \frac{3}{8}(\mathbf{x}^\ell + \mathbf{x}^r) + \frac{1}{8}(\mathbf{x}^u + \mathbf{x}^d), \tag{3.2}$$

where \mathbf{x}^ℓ and \mathbf{x}^r denote the vertices sharing the current edge and \mathbf{x}^u and \mathbf{x}^d stand for the remaining vertices of adjacent faces.

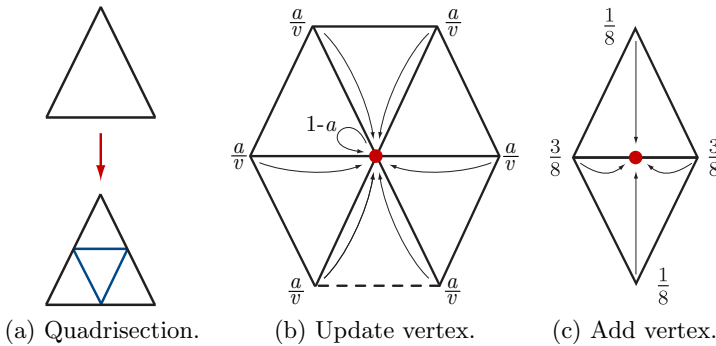


Fig. 3.1. The Loop subdivision scheme.

4 Multiresolution Optimization

In this section we summarize the algorithm for the multiresolution shape optimization. The input for the algorithm consists of the constant Q , an analysis ready (i.e., fine enough) triangular mesh representing the fixed surface, and a coarse representation of the to-be-optimized free surface. The user also has to

supply the number of refinements necessary to create an analysis ready free mesh, we denote this constant by n . The algorithm itself is iterative and is described below.

1. Subdivide the optimization mesh at level ℓ_o (initially set to 0) until the analysis level $\ell_a = n$ is reached. Append the already fine enough fixed mesh.
2. Solve the primal and adjoint boundary value problems (1.2) and (2.2), respectively. These problems are solved using the single-layer boundary integral equation [12, 13].
3. Evaluate the cost functional (1.1) and the shape gradient (2.3) for every node of the free mesh on the level ℓ_o . To perform this step we need to transfer the normal perturbation φ_i of the coarse mesh to the fine level $\tilde{\varphi}_i$. This can be done using the vertex update rules (3.1), (3.2). In this way, applying φ_i to the coarse mesh is equivalent to perturbing the fine mesh by $\tilde{\varphi}_i$.
4. If the decrease in the cost is not sufficient, increase the optimization level ℓ_o , if the maximum level n is reached, terminate the algorithm.
5. Perturb the optimization mesh using a gradient based algorithm.
6. Go to step 1.

5 Numerical Examples

In the last section we present a numerical example validating the procedure described above. We consider the Bernoulli problem with Γ_0 represented by an L-shaped domain and $Q := -3.0$. A regular icosahedron is chosen as the initial coarse mesh representing the free mesh Γ_f (see the wireframe in Figure 5.1b), after $n := 4$ subdivision steps the surface resembles a sphere (see the smooth surface in Figure 5.1b). The starting value of the cost functional is $J = 9.59 \cdot 10^1$. Figure 5.2 depicts optimal shapes on optimization levels $\ell_o \in \{1, 2, 3, 4\}$ and the corresponding values of cost functional. Again, the wireframe represents the current optimization mesh, the smooth surface enclosing the L-shape is the analysis mesh.

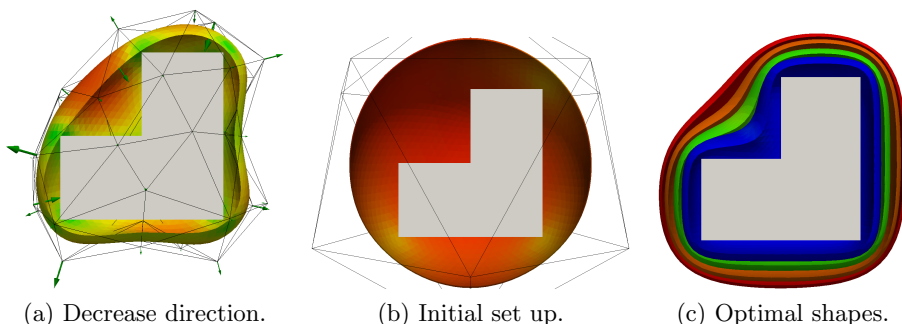


Fig. 5.1. Bernoulli free-surface problem.

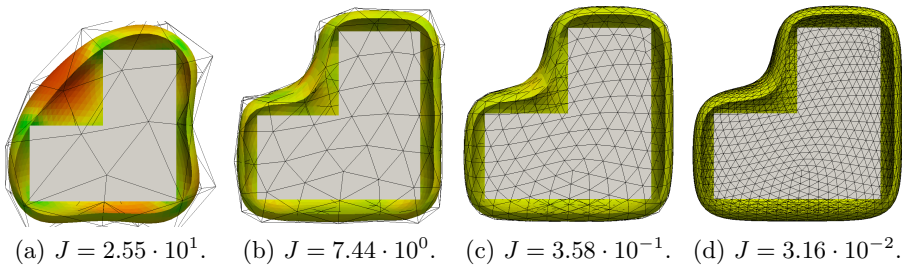


Fig. 5.2. Optimal surfaces on optimization levels $\ell_o \in \{1, 2, 3, 4\}$.

Furthermore, in Figure 5.1c we present a series of optimal shapes for different values of Q , namely $Q \in \{-1.2, -1.4, -2.0, -3.0\}$. The bigger the absolute value of Q , the closer the surface gets to the fixed surface.

6 BEM4I

The above described approach to shape optimization has been implemented in a newly developed library BEM4I based on the boundary element method [10, 12, 13, 15]. The problems (1.2), (2.2) were solved by direct BEM using the single-layer boundary integral equation. The optimization module uses OpenMesh [2] for the internal representation of the control mesh. For the gradient based optimization we use the MMA algorithm [14] implemented in the nlopt package [6].

The project BEM4I itself aims at HPC environment and utilizes modern programming techniques including in-core vectorization by Vc [7] and hybrid OpenMP/MPI parallelization. As of today, the package implements boundary element solvers for the Laplace, Helmholtz or time domain wave equations. Moreover, the assembly of the Laplace system matrices is sped up by the fast multipole method (FMM) [9]. Lately, support for sparsification by the adaptive cross approximation (ACA) has been added both for the Laplace and Helmholtz solvers allowing to model large sound scattering problems.

Acknowledgements

The work on this topic was initiated during the author's stay at the Institute of Computational Mathematics at TU Graz from September 2012 to March 2013 within the FP7 Marie Curie IAPP Project CASOPT. Together with the colleagues from TU Graz (G. Of, O. Steinbach) and the University of Cambridge (K. Bandara, F. Cirak) a joint paper [1] summarizing our results has been submitted to the impacted Journal of Computational Physics and is awaiting a review. In addition to the above presented results the paper also includes 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.

Talks related to the software package BEM4I have been presented at

- the SPOMECH workshop held in Ostrava on 13/11/2013,
- TU Graz within the Workshop on Industrial Optimization on 16/4/2014,
- the Modelling conference held in Rožnov pod Radnoštěm on 6/6/2014.

Jointly with Michal Merta, a Ph.D. student at the Dept. of Applied Mathematics and a Research Assistant at IT4Innovations, we plan to submit several papers related to the software and its applications including sound scattering and shape optimization problems.

References

1. Bandara, K., Cirak, F., Of, G., Steinbach, O., Zapletal, J.: “Boundary element based multiresolution shape optimisation in electrostatics” Submitted to *J. Comp. Phys.*
2. Botsch, M., Steinberg, S., Bischoff, S., Kobbelt, L.: “OpenMesh - a generic and efficient polygon mesh data structure”, 1st OpenSG Symposium, 2002.
3. Delfour, M. C., Zolésio, J.-P.: “Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization”, SIAM, 2011, 621 p.
4. Eppler, K., Harbrecht, H.: “Tracking Neumann Data for Stationary Free Boundary Problems”, *SIAM J. Control Optim.* 48/5, 2009, p. 2901–2916.
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. Johnson, S. G.: “The NLOpt nonlinear-optimization package”, available online at <http://ab-initio.mit.edu/nlopt>.
7. Kretz, L., Lindenstruth, V.: “Vc: A C++ library for explicit vectorization”, *Software: Practice and Experience* 42/11, 2012, p. 1409–1430.
8. Loop, C. T.: “Smooth Subdivision Surfaces Based on Triangles”, MSc. thesis, Department of Mathematics, University of Utah, 1987. 60 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, p. 272–296.
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. Rjasanow, S., Steinbach, O.: “The Fast Solution of Boundary Integral Equations”, New York: Springer–Verlag, 2008, 386 p.
13. Steinbach, O.: “Numerical Approximation Methods for Elliptic Boundary Value Problems: Finite and Boundary Elements”, New York: Springer–Verlag, 2007, 279 p.
14. Svanberg, K.: “The method of moving asymptotes – a new method for structural optimization”, *Int. J. Numer. Meth. Eng.* 24, 1987, p. 359–373.
15. 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 Influence of Gamma Irradiation on the Optical Coupler

Lukas Bednarek, Petr Winkelhofer, and Martin Papes

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{lukas.bednarek, petr.winkelhofer, martin.papes}@vsb.cz

Abstract. The article is dealing with ageing of optical couplers. Traffic increase in optical networks makes downgrading of all operational parameters faster. The article mentions most important power and polarization parameters of coupler with splitting ratio 1:2 and power splitting 30/70. Particularly power parameters of individual branches with wavelength 1310 nm and 1510 nm and polarization parameters with wavelength 1510, 1550 and 1612 nm. The findings were obtained by measuring of new coupler's parameters and comparison of matching results after first and second irradiation.

1 Introduction

Experience gained over the years showed, that optical elements in passive optical network are ageing faster than expected. It was caused mainly by increase of operational power and higher quantity of operated channels. This ageing makes operational parameters of individual components worse. Hastening of the ageing process was accomplished by gamma radiation. The article is dealing with optical coupler with splitting ratio 1:2 where are in first branch 30% of power and in second branch 70% of power. Observed are power and polarization parameters. Observed power parameters are insertion loss of individual branches, exact distribution of power among the branches, total loss, homogenization of loss and directionality. Analyzed polarization parameters are azimuth, ellipticity, power splitting ratio, phase difference and level of polarization. This article shows results of power parameters with wavelengths of 1310 nm and 1550 nm and polarization parameters with wavelengths of 1510, 1550 and 1612 nm after two irradiations. The findings were obtained by measuring of new coupler's parameters and comparison of matching results after first and second irradiation.

2 Parameter's description

Optical coupler has many different parameters. Power and polarization parameters are among fundamental parameters.

Total losses, by another name residual attenuation, are all losses of the optical coupler. Ideally they should be zero, by all means they are in the interval up to 1 dB usually.

Insertion loss or insertion attenuation are losses in individual route, accordingly from input branch to the particular output branches. They are different for each output branch and the values are round about 3 dB, in case of a coupler 1:2, with splitting ratio 50/50.

Splitting ratio is percentage of input power allocation among particular output branches. Splitting ratio is important for couplers 1:2, since several variants of power splitting are existing, from 1/99 to 50/50. Couplers with more branches usually have constant splitting ratio.

Homogenization of loss is difference between maximal insertion loss in any branch and minimal insertion loss in any branch. Homogenization should be minimal, for a coupler 1:2 in decimal to centesimal dB. The calculation is simple subtract of two above mentioned insertion losses.

Polarization depending losses, or PDL, are losses caused by different light polarization during the measuring. They are usually mentioned in dB and they are deviations of insertion loss while changed polarization of. They aren't calculated but determined from the deviation.

Directivity or crosstalk attenuation is loss between 2 adjacent inputs or outputs. It is in dB and it's needed as large as possible, because with low crosstalk attenuation would be coupler not working properly, instead of splitting of power from the input to the output side, would be power partially radiated on the input side of the coupler, and therefore to increase of residual loss and insertion loss.

Reflection return loss is last of couplers' fundamental power parameters. It determined what power will be reflected back from the coupler to the source. This is unwanted and that's why should be this loss maximized. It is in dB again.

Polarization says how much is the direction of light's vector oscillation (electromagnetic wave) directionally limited. The main value is called degree of polarization, varies from 0.0 to 1.0 and it is completely unpolarized light (normal light, standard light bulb), the more is the value closer to 1, this more is the light polarized to a direction, if the value is 1, the light is completely polarized (i. e. good/ideal laser). Degree of polarization is calculated value, it is determined either by Stokes' parameters or the respective angles. Stokes' parameters are 4, S_0 represents the overall beam intensity, S_1 is difference between horizontal direction and vertical direction intensity, S_2 is difference between intensities of diagonal directions (R+45P minus R-45P). The last parameter S_3 is difference between intensities in the right circularly polarized direction against the left circularly polarized direction. The hardest part is usually determining of S_3 (depending on type of measuring). Obviously, Stokes' parameters could be determined also with knowledge of angles between the oriented beam and the axis, therefore these angles also could be used for calculation of polarization [1,2,3].

3 Evaluation of the measured data

Following part summarizes and evaluates reached results. The data are not showed here, because they are huge in amounts. That's why only analyzed and evaluated results are here.

First part is focused on the power parameters, while the second part summarizes results of polarization parameters.

3.1 Evaluation of power parameters

Among power parameters are insertion loss, residual loss, splitting ratio, loss homogenization and directionality. These parameters are calculated according the equations and from measured input power and output powers after passing through the coupler. Measuring of these powers was performed 20x in total for each branch separately, for average and more accurate the results.

First thing was the measuring of a new coupler and comparison of measured values with catalogue sheet of the coupler. Insertion loss for the first branch is 8,19 dB on average 7,39 dB respectively for both above mentioned wavelengths. This result is similar to the values in the catalogue sheet from the manufacturer. Also the insertion loss of the second branch has values similar to values declared by the manufacturer - 2,76 dB and 2,7 dB, for the given wavelengths. Splitting ratios for the first branch were 22,25 % and 25,38 % and in case of the second branch 77,75 % and 74,62 %. These values are different from the values declared by the manufacturer, because of different measuring method. Total residual loss is round about 1,5 dB for both wavelengths and is higher than the one measured by the manufacturer, possible reason of different measuring method again. Loss homogenization is in this case 5,43 dB for wavelength 1310 nm and 4,69 dB for wavelength 1550 nm. The last analyzed parameter is directionality, which is for both wavelengths higher than 60 dB, therefore is this parameter of the coupler within permitted limits.

The first dose of radiation had power of 5 kGy. One would expect, that the parameters will get worse, but reality is different, see bellow. Only by seeing on irradiated coupler was found out, that radiation slightly damaged ferrule and the entire input connector. Average insertion loss of the first branch were 6,83 dB and 6,28 dB. For the second branch was the insertion loss 1,76 dB and 1,77 dB. Therefore parameters were improved for approximately 1 dB in both cases. Splitting ratios were changed against measuring before irradiation only in tenths of percents, therefore we can say, that the dose of irradiation didn't have influence on power splitting among particular branches of the coupler. Residual loss was reduced to 0,59 dB and 0,45 dB for wavelengths 1310 and 1550 nm. Therefore it's clear, that the irradiation had quite the opposite effect than was expected. Loss homogenization has changed for both wavelengths round about 0,1 dB, this parameter wasn't influenced much by the irradiation. Directionality was practically unchanged for wavelength 1310 nm, for wavelength 1550 nm was the directionality round about 3 dB higher, than for the measuring before irradiation.

Power of the second dose of irradiation isn't known yet, because the evaluation of the dose is still in progress at a research institute Řež. Coupler's ferrule was damaged more than after the first dose of irradiation, it was black, there were scratches visible by a microscope. That's why increase of insertion losses was expected, however insertion loss of the first branch was increased by 0,2 dB for both wavelengths, when compared with the first exposure. That is negligible. Also in case of the second branch, the insertion loss was increased by round about 0,2 dB for both wavelengths. Splitting ratio was almost identical as results after first irradiation, this dose of irradiation didn't influenced split of the power between particular branches. Coupler's residual loss was also increased by 0,2 dB for both wavelengths, when compared with the first exposure. This increase is caused by same increase of both the insertion losses, it is usually in correlation, with increase of insertion loss is usually connected also total loss. Homogenization of loss almost has not changed for wavelength 1310 nm, but for the wavelength 1550 nm has decreased by 0,3 dB. Directionality for wavelength 1310 nm was improved to 64 dB, for the second wavelength it was deteriorated to 58 dB. It means, that this parameter have been deteriorated, by the radiation, but it is within tolerated limits.

3.2 Evaluation of polarization parameters

These were a very time-consuming measurements, we were measuring each coupler's transfer for three wavelengths, measuring period is 30,72s (1024 samples) and each measuring is performed 5x for better relevance.

Polarization measurement isn't trivial as well as measuring instrument for that. Polarimeter is measuring many values, which are calculated via Fourier transformation, then is each of the calculated values measured on the basis of the calculations. The resulting values are usually mathematically related among themselves. Unfortunately transitions among them (deviation) are sometimes bigger than measuring accuracy of the value [4,5], after consultation with the manufacturer of the device we decided to analyze only azimuth angles and ellipticity. For values, which are describing polarization, we are favoring values in watts instead of decibels.

Regarding the analysis, we are looking for characteristic value changes of azimuth and ellipticity during time, which are some kind of instabilities and which could help us to understand, why are fibers practically ageing more quickly than expected. We are not seeing anything special during laic processing so far, i.e. if we make an average of 1024 samples of a value, each particular sample isn't as far from the average as the accuracy of the value. Of course we will also have more sophisticated analysis, but these are based on statistics themes, which we will go through yet. Only really interesting measured phenomenon are the loss changes of the routes in the lateral wavelengths 1510 and 1612 nm so far.

Another interesting problem of the measuring is a fact, that the manufacturer guarantees accuracy of parameters' measuring only within certain range of ellipticity value and up to certain power level on the output of the Polarimeter. Unfortunately, when measuring of a coupler with 8 outputs we have, for example, 4 results within limits and 4 results out of limits and manufacturer isn't able to guarantee the accuracy at all.

Next, we are considering other type of measurement, more samples in longer time-frame, but these considerations are under negotiation only so far. Representatives of our department have developed new software for that purpose.

Finally, the measuring is quite a bit time-consuming, and it will be under development (other settings), interpretation of results is still difficult, given the characteristics of the polarimeter and our knowledge of the mathematical analysis. Considering the minimal global knowledge and publications in this sector (ageing of optical fiber), we don't know much what to expect, we also have weak theoretical background. We are not publicizing any tables of results demonstrating the measuring, because with the amount of the values the table would take more than 5 sheets with minimum of meaningful values. Only thing we could say and prove for now, is the fact that the loss is significantly changing in lateral wavelengths (1510 nm and 1612 nm), improving and deteriorating, change of 6 dB is quite normal.

4 Conclusion

The article is summarizing results acquired by measuring of a coupler irradiated by gamma radiation. It is not presenting all measured data, because all of them aren't significant, presented are only averages and meaningful data. All measurements were very time - consuming and sensitive to change of conditions, temperature was most influential in case of power parameters, fiber has to be kept in stabile state in case of polarization parameters, to avoid changes of polarization during the measurement.

First part theoretically describes examined coupler's parameters. Second part is dedicated to evaluation of measured figures.

Power parameters of examined new coupler were compared with figures measured by the manufacturer. It was detected, that these data are almost matching. Comparison of data measured before and after first irradiation showed, that irradiation improved parameters, insertion loss, for example. This fact wasn't expected before irradiation. Second irradiation had only minor impact on the coupler and measured values were just slightly different from the first irradiation's measured values.

There haven't been approved any significant changes in scope of polarization properties yet. Only different loss levels at side wavelengths (1510 nm and 1612 nm) before and after irradiation.

5 References

1. COLLETT, Edward. *Polarized light in fiber optics* [online]. Lincroft, N.J.: PolaWave Group, c2003, ix, 540 p. [cit. 2014-07-30]. ISBN 09-677-1671-3. Available from: http://books.google.cz/books/about/Polarized_Light_in_Fiber_Optics.html?id=I3lBS7RBX3oC&redir_esc=y
2. BŘEZINA, VOJTĚCH. POLARIZAČNÍ VLASTNOSTI OPTICKÉHO SVAZKU. In: [online]. 2012. vyd. [cit. 2014-07-30]. Available from: <https://dspace.vutbr.cz/bitstream/handle/11012/10049/bakal%C3%A1%C5%99sk%C3%A11%20pr%C3%A1ce%20-%20Vojt%C4%9Bch%20B%C5%99ezina.pdf?sequence=2>

3. PAVEZA, VÁCLAV. Konstrukce polarimetru. In: [online]. 2013. vyd. [cit. 2014-07-30]. Available from:
https://theses.cz/id/1h2dc2/Konstrukce_Polarimetru_Vaclav_Paveza_2013.pdf
4. PAX5710 / PAX5720 Series Polarimeters. In: *Www.thorlabs.com* [online]. [cit. 2014-07-30]. Available from:
<http://www.thorlabs.com/thorcat/12900/PAX5710IR3-T-SpecSheet.pdf>
5. Operation Manual. In: *Www.thorlabs.com* [online]. 2008. vyd. [cit. 2014-07-30]. Available from:
<http://www.thorlabs.com/thorcat/12900/PAX5710IR3-T-Manual.pdf>

Detection of the Network Anomalies by Processing of NetFlow with Machine Learning Systems

Miroslav Bures, Pavel Nevlud, and Jaroslav Zdralek

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{miroslav.bures, pavel.nevlud, jaroslav.zdralek}@vsb.cz

Abstract. The article deals with NetFlow and its processing with machine learning systems for detection of network anomalies. The types of the data collected and the collection methodology are the first step in detecting the network anomalies. The second important step is processing of the acquired data as well as of the methods of evaluation of this data. Machine learning systems can be considered as a support or a limited type of artificial intelligence for processing of the network data. There are several machine learning techniques available. We tested the decision tree learning and Bayesian networks. The open source data-mining framework Weka was the tool that was used for testing the classification and for visualization of the gained results.

Keywords: Attack, bayesian networks, decision tree, NetFlow, network anomaly, Weka.

1 Introduction

Nowadays, computer network is a frequent target of attacks in order to obtain confidential data, or unavailability of network services. To detect and prevent these attacks, there are a large number of software or hardware solutions such as IDS (Intrusion Detection Systems), firewalls and monitoring systems.

These attacks increased normal network traffic that appears as something undesirable, what would not occur in the network. Such deviations from normal operation are called as network anomalies. Between network anomalies include everything that is quite different from the normal operation of the network [1].

Anomalies are values in a statistical sample which does not fit a pattern that describes most other data points. Figure 1 illustrates anomalies in a simple 2-dimensional data set. The data has one normal regions, since most observations lie in this region. Three points that are sufficiently far away from the regions are anomalies. One of these points is border point that can be detected as anomaly.

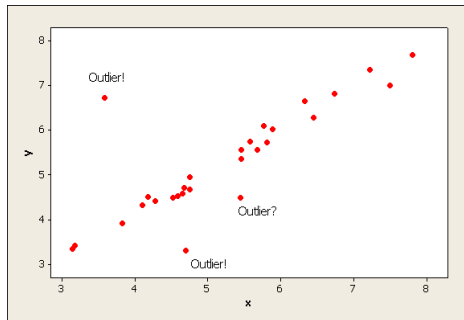


Fig. 1. A simple example of anomalies

2 Detection of Network Anomalies

Network anomalies can be detected in several ways. Each method has its advantages and disadvantages, but in practice there are three commonly used methods. Them together they can develop systems such as IDS software [2].

2.1 Comparing signatures

The principle of this method is the comparison of network data with a database of signatures. Signature database contains patterns of data anomalies. Data anomaly pattern is actually a description of a typical data sequence that characterizes the anomaly. The principle can be seen in Figure 2. It used the same principle as in the anti-virus programs.

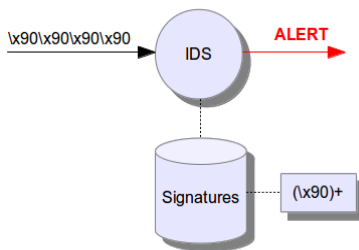


Fig. 2. Comparing signatures

The effectiveness of anomaly detection using signature recognition is highly dependent on the quality of the database of signatures. The big disadvantage is almost no detection of new types of attacks called Zero day attack, because it is not in the database signature pattern for this type of anomaly [3].

2.2 Stateful protocol analysis

Stateful protocol analysis assumes that each protocol used for network communication is specified, such as RFC. Thanks to precise specifications, all connections using protocols defined state. Each event must occur at the right moment, the state. This makes it possible to describe the protocol as a state machine.

The advantage of this method is less frequent updates. The stateful analysis needs update only after the change of protocol or the installation of a new one.[4].

2.3 Behavioral analysis

The method of behavioral analysis is based on the assumption that the emergence of anomalies can be detected by the deviation from the normal or expected network behavior. Model of normally or anticipated behavior of the network is created based on network monitoring and collecting reference information.

The reference information is compiled model normal behavior and network traffic is subsequently compared with this model. Any deviation from such a learned model is automatically considered an anomaly. The principle can be seen in Figure 3. For behavioral analysis and create network's model can be used MLS (Machine Learning Systems).

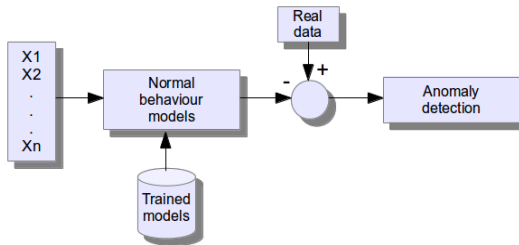


Fig. 3. Behavioral analysis

The disadvantage of this method is precisely the fine detection. Any deviation from the normal model is detected even though it is not an attack or threat. It is due to the fact that the creation of the model can not capture all types of network traffic and user activity on the network. This model is created to some extent distorted.

On the other hand, behavioral analysis provides an advantage in terms of detection of completely new types of threats, for example, by comparing detection signatures did not react at all.

3 NetFlow

NetFlow is currently the most widely used industry standard for measuring and monitoring of computer networks based on IP flows. Flow is defined as a

sequence of packets with the same group of five data: target/source IP address, destination/source port number and protocol. For each flow is recorded time of its origin, its duration, the number of transmitted packets and bytes, and more data. NetFlow statistics generated over IP service offers information about who communicates with whom, when, how long, how often, how much data was transferred data [5], [6]. In Figure 4 u can see principle of exporting and collecting of NetFlow in computer network.

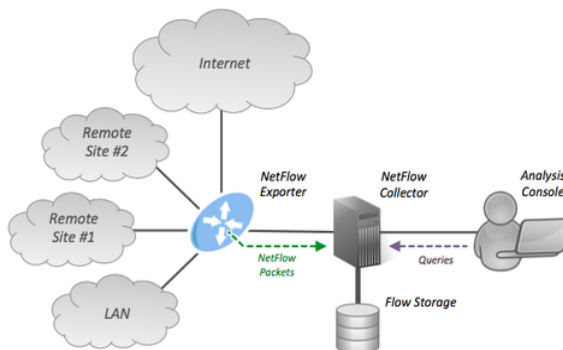


Fig. 4. Architecture of NetFlow

3.1 Internet Protocol Flow Information Export

Internet Protocol Flow Information Export (IPFIX) is a relatively new protocol for describing the transfer of information about IP flows from exporter to collector. IPFIX was created by the IETF working group. Previous versions of NetFlow has always been the proprietary Cisco Systems. IPFIX is often called as NetFlow of version 10.

4 Machine Learning Systems

If we want to be able to solve the computer problem, some intelligence is needed. Machine learning can be considered as a support or a limited type of artificial intelligence. Algorithms MLS can move on with the development of computers. This means that computers are no longer just a database comparing sets of data.

A machine learning system usually starts with some knowledge and a corresponding knowledge organization so that it can interpret, analyze, and test the knowledge acquired. The principle can be seen in Figure 5.

Training is the process of making the system able to learn. It may consist of randomly selected examples that include a variety of facts and details including irrelevant data. The learning techniques can be characterized as a search through

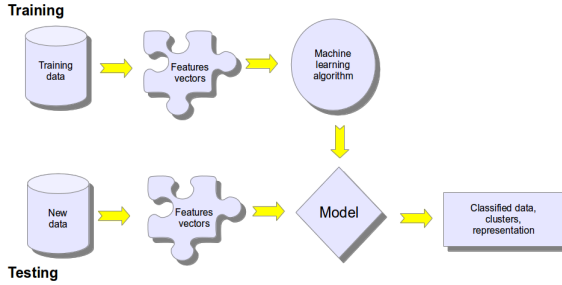


Fig. 5. Principle of machine learning system

a space of possible hypotheses or solutions. Background knowledge can be used to make learning more efficient by reducing the search space.

The success of machine learning system also depends on the algorithms. These algorithms control the search to find and build the knowledge structures. The algorithms should extract useful information from training examples. There are several machine learning techniques available [7].

5 Decision tree learning

Decision tree learning is ‘a method for approximating discrete valued functions that is robust to noisy data and capable of learning disjunctive expressions’ according to [8].

Decision tree induction takes a set of known data and induces a decision tree from that data. The tree can then be used as a rule set for predicting the outcome from known attributes. The initial data set from which the tree is induced is known as the training set. The decision tree takes the top-down form. At the top is the first attribute and its values, from this next branch leads to either an attribute or an outcome. Every possible leaf of the tree eventually leads to an outcome.

5.1 Decision Trees – C4.5

C4.5 is an algorithm developed by Ross Quinlan that generates Decision Trees (DT), which can be used for classification problems. It improves (extends) the ID3 algorithm by dealing with both continuous and discrete attributes, missing values and pruning trees after construction. Its commercial successor is C5.0/See5, a lot faster than C4.5, more memory efficient and used for building smaller decision trees. J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool.

The generic description of how C4.5 works is shown in Algorithm 1. A decision tree is built top-down from a root node and involves partitioning the data into

Algorithm 1 C4.5(D)

```

Require: an attribute-valued dataset  $D$ 
1:  $Tree = \{\}$ 
2: if  $D$  is "pure" OR other stopping criteria met then
3:   terminate
4: end if
5: for all attribute  $a \in D$  do
6:   Compute information-theoretic criteria if we split  $a$ 
7: end for
8:  $a_{best} =$  Best attribute according to above computed criteria
9:  $Tree =$  Create a decision node that tests  $a_{best}$  in the root
10:  $D_v =$  Induced sub-datasets from  $D$  based on  $a_{best}$ 
11: for all  $D_v$  do
12:    $Tree_v = C4.5(D_v)$ 
13:   Attache  $Tree_v$  to the corresponding branch of  $Tree$ 
14: end for
15: return  $Tree$ 

```

subsets that contain instances with similar values (homogenous). Decision tree algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one. The entropy of class random variable that takes on c values with probabilities p_1, p_2, \dots, p_c is given by:

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i. \quad (1)$$

Figure 6 shows the form of the entropy function relative to a binary classification.

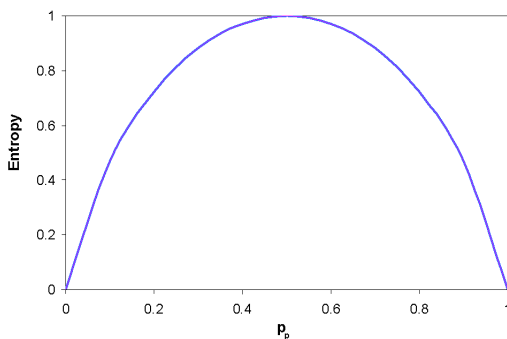


Fig. 6. Entropy function

The estimation criterion in the decision tree algorithm is the selection of an attribute to test at each decision node in the tree. The goal is to select the attribute that is most useful for classifying examples. A good quantitative measure of the worth of an attribute is a statistical property called information gain that measures how well a given attribute separates the training examples

according to their target classification. This measure is used to select among the candidate attributes at each step while growing the tree.

The information gain is based on the decrease in entropy after a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain (i.e., the most homogeneous branches).

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v), \quad (2)$$

where $Values(A)$ is the set of all possible values for attribute A , and S_v is the subset of S for which attribute A has value v (i.e., $S_v = \{s \in S | A(s) = v\}$).

6 Bayesian networks

Bayesian networks are graphical representation of the relationship between variables. Graphical representation of Bayesian networks are directed acyclic graphs with nodes and edges. Nodes represent variables, parameters or hypotheses and edges represent conditional dependencies.

6.1 Algorithm of Naive Bayesian

The Naive Bayesian classifier is based on Bayes' theorem with independence assumptions between predictors. Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}, \quad (3)$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c), \quad (4)$$

where $P(c|x)$ is the posterior probability of class (target) given predictor (attribute), $P(c)$ is the prior probability of class, $P(x|c)$ is the likelihood which is the probability of predictor given class and $P(x)$ is the prior probability of predictor.

7 Experimental results

For data mining platform was chosen open source project Weka [9]. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from one's own Java code.

For the experiment, the data obtained from the real network traffic was used. The volume of the processed data was approximately 200,000 records. The NetFlow records were being collected for about half a year in laboratory of

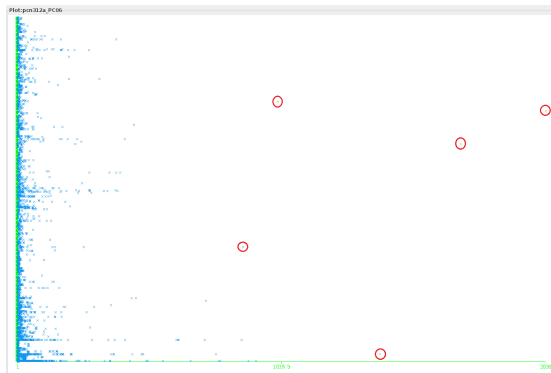


Fig. 7. The dependence of the number of packets to source address for J48 classifier

computer networks. NetFlow was then converted into an external csv file with the script using Nfdump [10]. Weka accepts *.csv files, *.arff files, or it was connected to the database. For this research, the converted csv files opened in Weka were used.

Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Figures 7 and 8 show the results in Weka of the J48 and the Bayes classifier training run on NetFlow. These figures show the dependence of the number of packets on the source address for different types of classifiers.

The marked outliers outside the main area indicates of samples with a high probability of containing anomalous behavior, i.e. some type of a network attack. From many attributes in the NetFlow records only ten of them were chosen to be

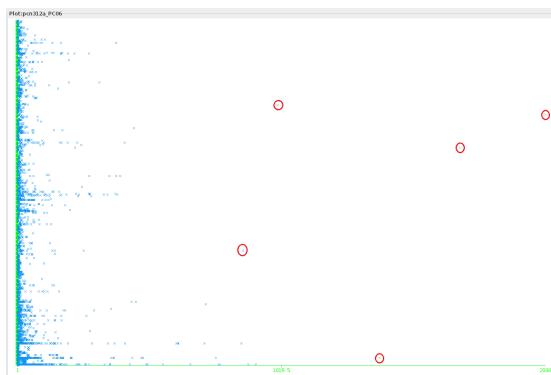


Fig. 8. The dependence of the number of packets to source address for Bayes classifier

processed in Weka. For visualization was choosed theree main attributes: source

IP address, number of packets and number of transfer bytes in one IP flow. From these few parameters it is possible to create a model of normal network behavior. A more accurate estimate is possible from sufficient number of records.

Figures 9 and 10 show the results in Weka of the J48 and the Bayes classifier training run on NetFlow again. In these figures the dependence of the number of packets on the number of bytes for different types of classifiers can be seen. The marked outliers are again the samples with a high probability of containing

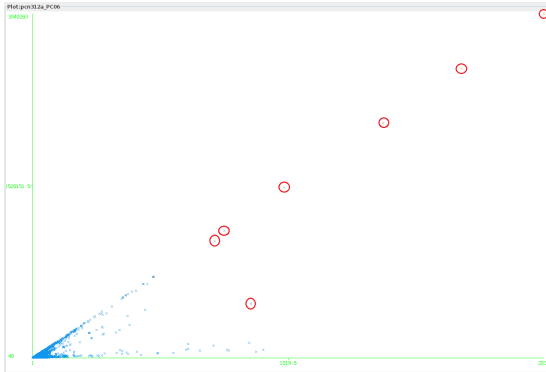


Fig. 9. The dependence of the number of packets to transfer bytes for J48 classifier

anomalous behavior. Anomalous behavior can be detected by monitoring of traffic trends. Comparison of the number of packets, depending on the size of bytes is a very good way to detect anomalies in IP flows. In this type of detection, a large volume of the processed data network is a major advantage. NetFlow is possible statistical process and the new data can be compared with the median of the previous ones. If the data exceeds the threshold in any of the set values, the sample can be described as an anomaly in network traffic, i.e. a potential attack.

8 Conclusion and future work

This article introduces an experiment processing NetFlow with a certain type of artificial intelligence. Machine learning systems, namely the Bayes classifier and J48, were used as artificial intelligence. A machine learning system usually starts with some knowledge and during the rounds its knowledge can be improved.

The information about IP flows from real network traffic was tested. The volume of the processed records was approximately 200,000. Data processing was carried out in the Weka open-source program. NetFlow was converted into an external csv file using the Nfdump program. Finally, the classified data was visualized by the Weka software.

In their further work, the research team are going to focus on the use of other classification methods and other data mining algorithms. Furthermore,

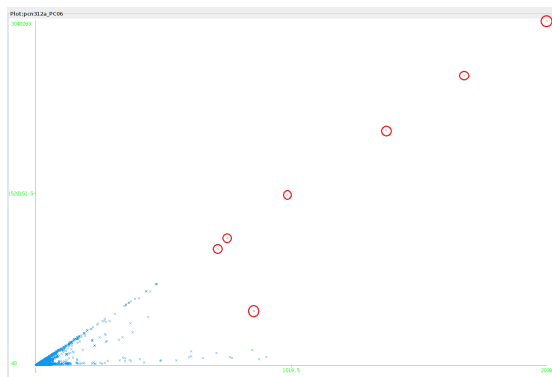


Fig. 10. The dependence of the number of packets to transfer bytes for Bayes classifier

they are going to carry out some research on the correlation NetFlow and status information, and their mutual processing by machine learning systems.

References

1. Fowler, Ch. A., Hammell II, R. J. *Building Baseline Preprocessed Common Data Sets for Multiple Follow-on Data Mining Algorithms* In Proceedings of the Conference on Information Systems Applied Research. New Orleans: 2012. ISSN: 2167-1508.
2. Nevlud, P., M. Bures, L. Kapicak and J. Zdralek. Anomaly-based network intrusion detection methods. *Advances in Electrical and Electronic Engineering*. 2013, vol. 11, iss. 6, pp. 468–474. ISSN: 1336-1376.
3. Richard, M. *Intrusion Detection FAQ: Are there limitations of Intrusion Signatures?*. [online 8.2.2014]. <http://www.sans.org/resources/idfaq/limitations.php>.
4. Das, K. *Protocol Anomaly Detection for Network-based Intrusion Detection*. [online 8.2.2014] http://www.sans.org/reading_room/whitepapers/detection/protocol_anomaly_detection_for_networkbased_intrusion_detection_349?show=349.php&cat=detection.
5. Lucas, M. W. *Network Flow Analysis*, San Francisco: 2010, ISBN: 978-1593272036.
6. Patterson, M. A. *Unleashing The Power Of NetFlow and IPFIX*, Sanford: 2012, ISBN: 978-0-9860665-1-1.
7. Farraposo1, F., P. Owezarski and E. Monteiro. *NADA – Network Anomaly Detection Algorithm* In Managing Virtualization of Networks and Services, Lecture Notes in Computer Science. San José: Springer Berlin Heidelberg, 2007. vol. 4785, pp. 191–194. ISBN: 978-3-540-75694-1. DOI 10.1007/978-3-540-75694-1_18.
8. Mitchell, T. M. *Machine Learning*, New York: 1997, ISBN: 978-0070428072.
9. Weka 3: Data Mining Software in Java [online 8.2.2014] <http://www.cs.waikato.ac.nz/ml/weka/>.
10. Nfdump [online 8.2.2014] <http://nfdump.sourceforge.net/>.

Optical Vibration Sensor based on Michelson Interferometer Arrangement with Polarization-maintaining Fibers

Jakub Cubik

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

Abstract. Today, the interferometric sensors belong to the one of the most accurate, thanks to its great sensitivity. With their help we are able to measure temperature, strain, level, flow, vibration, stress, etc. For its operation the Michelson interferometer consist of the two arms terminated by mirrors, by which is possible to measure generated phase shift in the individual arms. Furthermore, were used polarization maintaining fibers. With this setup we will examine the effects of vibration and also how is this sensor influenced by the different setup arrangement and how it will manifest its frequency response. It is important to isolate the reference arm to increase the sensitivity of the measurement and the subsequent effect on the maximum phase shift and maximum frequencies response. In this work, we should describe various combinations of the arrangement of the measuring and reference arm and their effect on the sensitivity of different measured phenomena. Subsequently obtained frequency bands are evaluated for these various configurations and materials.

Keywords: Michelson, Interferometer, Fibers, Polarization-maintaining, Vibration

1 Introduction

The possibility to guide light through reflection, first demonstrated in 1840, resulted in much progress in the field of fiber optics. Nowadays, it is common the use of light for high-speed data transmission in backbone networks and slowly approaching a state where everyone will have optical fiber reaching up to the living space (so called FTTH fiber-to-the-home). With the development of fiber optics for data transmission another field of interest was developing simultaneously. The field of fiber optic sensors did not experience such a rapid rise, but now there is a great interest in this area. Interest in optical sensors is primarily due to the possibility to measure with high sensitivity over a long distance and also the fact that the optical fiber is a dielectric.

Optical fiber sensors are generally divided into two groups. The intensity sensor evaluates the change in light intensity depending on the measured variable, while the phase sensor evaluates the phase change of the light beam. The intensity sensors are

usually used for evaluating displacements or other physical phenomena affecting the fiber. Phase sensors also called interferometers are using phase change of light in one part of the measuring system (measure arm), compared with unchanged phase in reference part (reference arm). By comparing these phase differences, variable acting on the fiber can be measured with extreme precision. These sensors are more accurate than intensity sensors, but at the cost of a higher price and therefore are mainly used for complex applications requiring accuracy and dynamic range [1].

Many different types of interferometers, including Fabry-Perot, Sagnac, Michelson, Mach-Zehnder has been introduced. This article uses a Michelson-type interferometer.

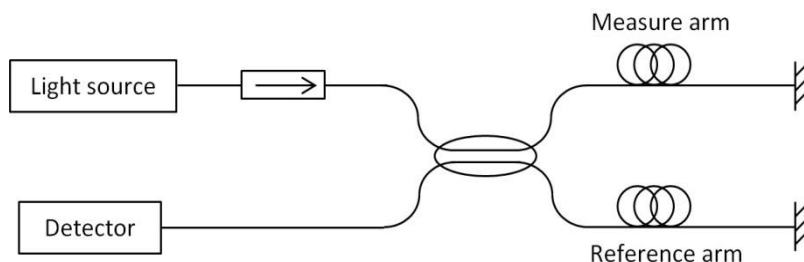


Fig. 1. Basic scheme of Michelson interferometer

In the case of Michelson interferometer the light source, which is protected against back reflection by using isolator, is divided between two arms terminated by mirrors. One arm acts as a measure arm and is directly exposed to the measurand, while the second arm act as a reference experiencing constant environment without being affected by the measured value. Division of the light and its merging occurs at the same coupler. The signal is further propagated to the optical detector, where it is converted into electric current corresponding to the measurand. Visibility of interference depends on relative intensity of measuring and reference beam, their relative state of polarization and their mutual coherence. In optimal case, relative intensity and state of polarization are equal and the difference of optical path length between measure and reference arms is lower than coherence length of source light [2].

This publication relies on previous research performed by our research team [3,4,5] and on the article [6] Interferometric sensor based on the polarization-maintaining fibers in particular.

Other possible applications that use interferometry include following articles [7,8,9,10]. These experiments, however, are based on different principles, use other components or focus on the measurement of other phenomena. For example, in the article In-Fiber Quasi-Michelson Interferometer for Liquid Level Measurement with a Core-Cladding Modes-Fiber End-Face Mirror was a quasi-Michelson interferometer used to measure liquid level. The article A Single-Mode Fiber Sensor Based on Core-Offset Inter-Modal Interferometer discusses the use of conventional single-mode fiber and a very small portion of the multi-mode fiber.

In the reference [11] Fiber Bragg Grating Vibration Sensor with DFB laser diode, bragg grating is used as a sensing element. This fiber optical sensor is a possible al-

ternative to the interference sensor, but exhibits lower sensitivity to the measured phenomenon.

2 Experimental setup

As mentioned above, the article deals with the use of fiber optic Michelson interferometer as a sensor of the impact of different objects on different surfaces. The entire diagram is composed of polarization-maintaining components to achieve higher sensitivity. As the radiation source DFB laser with a wavelength of 1550nm and a spectral line width of 0.03 nm was used. The operating point of the laser was kept by Thorlabs temperature and current controller. The diode is pigtailed with PM fiber terminated by inappropriately keyed FC/APC connector. Inappropriately keyed connector is compensated by the polarizer at the output of the laser. The suitability of this solution was verified by measuring the polarization and visualization on the Poincaré sphere. Due to the absence of the polarization-maintaining coupler 2x2 was replaced by circulator, which also serves as isolator. Further, the light is brought into the 1x2 coupler, which distributes the light beam into two arms equally. The arms were formed by 2 m long polarization-maintaining fibers terminated with mirrors. This fiber optic mirror reflects incoming light directly back, with very low optical loss. Reflected light recombines in the coupler gets on InGaAs photodetector through the circulator, where it is converted into the electrical signal. The electrical signal is further processed as shown in Figure 3. The picture 2 depicts the final measurement setup for this experiment.

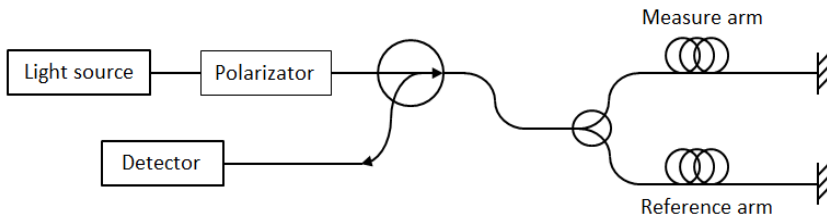


Fig. 2. Real configuration of Michelson interferometer

An electrical signal was fed to the LC high-pass filter with a cutoff frequency at 8 Hz via 50 Ω coaxial cable in order to suppress DC component. After that was fed into a measuring card NI USB-6210 with sampling rate up to 250 kS/s. The card was capturing voltage on analogue input using the application written in LabView development suite. The application also performs a discrete Fourier transformation so the voltage was transferred into the frequency domain. Hanning window function was used in our case.

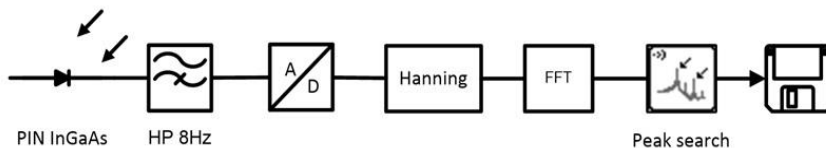


Fig. 3. Schematic diagram of electrical signal processing.

The reference arm was placed in the polystyrene box during the whole measurement. The measure arm was directly exposed to the measured phenomenon and was placed on the different surfaces, where selected objects had fallen. The first case was a concrete floor covered with linoleum the second was a pane of glass was inserted in a metal frame, which was shielded from external influences by means of rubber. The last measured surface was wooden board placed on a metal frame, from which was also shielded using the rubber.

The test measurement was performed with loudspeaker located one meter above the measure arm and it was powered by signal generator Hameg HMF2550. The rectangular signal with frequency of 200 Hz and with 50% duty cycle was used. Frequency spectrum from this measurement can be found in Figure 4.

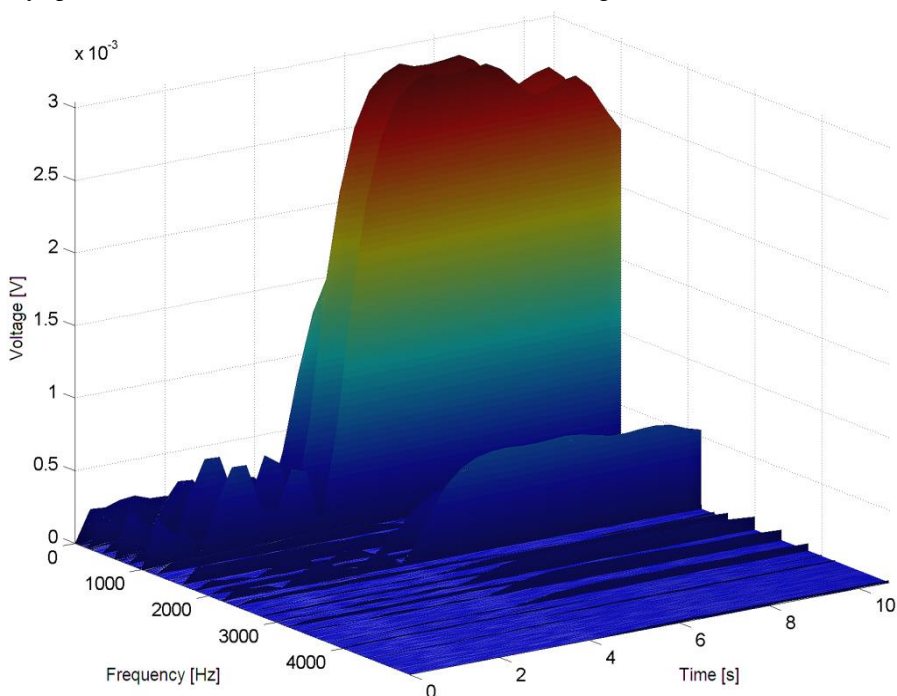


Fig. 4. Reference measurement on 200Hz

The measuring arm on all three surfaces was influenced with three selected objects by impingement on the surface from a height of one meter. These four objects were a tennis ball weighing 56g and diameter of 63mm, plasticine ball with a diameter of 3 cm, table-tennis ball weighing 2g and diameter of 37,5mm and weight of 20 grams and the last object was a steel ball with a diameter of 8 mm and weight of 2g. Each measurement was repeated 10 times for statistical reasons. Results of measurements and evaluation are shown below.

3 Results and discussion

As mentioned above, the first measurement was verification of test setup by an acoustic signal from the speaker. This measurement showed that the measuring set was able to measure a large frequency range. Thanks to the low-pass filter frequencies below 10Hz were excluded from the measurement. Figure 5 shows the spectrum of the 10 attempts in case of impact of steel ball on the pane of glass. The x-axis is frequency in Hertz and the y-axis is plotted the voltage in volts. The x-axis represents the frequency in Hertz and the y-axis voltage in Volts. For better interpretation of the data are then converted to voltage level in decibel Volts. In tables 1-3 medians of voltage level and frequency for 10 repeated measurements can be found. Values were transferred to the tables to 3 decimal places. For each case the standard deviation is further calculated. Tables with results are situated in annex, in case of compliance with requirement.

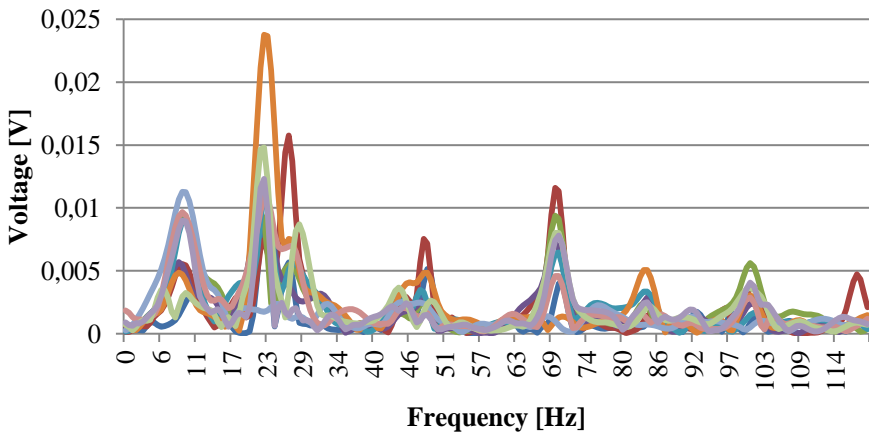


Fig. 2. The measured spectra at the time of impact of steel ball on the pane of glass.

4 Conclusion

Measurements published in this article points out the dependence of frequency spectra of various objects impacting on surface. It can be stated that interferometric measurements, fiber optic Michelson interferometer in particular, are very sensitive

and has wide frequency range of the measurement. The advantage was scheme with polarization-maintaining fibers. This improvement over the use of conventional fibers had a valuable contribution to the sensitivity of the apparatus. This measurement should prove the recognisability of individual objects, through comparing their frequency spectra in time. In the near future it will be necessary to better stabilize the amplitude of each frequency leading to a lower standard deviation. The aim is to build a measurement set, which would be able to measure virtually constant amplitude and frequency spectrum on repeated measurements. Due to this fact, a variety of objects could be detected using a software application.

References

1. Krohn, D., [Fiber optic sensors: fundamentals and applications], Instrument Society of America, Research Triangle Park, NC (1992).
2. López-Higuera, J.M. [Handbook of Optical Fibre Sensing Technology], J.Wiley and Sons, New York, 227-245 (2002).
3. Cubik, J., Kepak, S., Doricak, J., Vasinek, V., Jaros, J., Liner, A., Papes, M., and Fajkus, M., "The Usability Analysis of Different Standard Single-Mode Optical Fibers and its Installation Methods for the Interferometric Measurements." *Advances in Electrical and Electronic Engineering* 11 (6): 535-542 (2013).
4. Vasinek, V., Cubik, J., Kepak, S., Doricak, J., Latal, J., and Koudelka, P., "Mach-zehnder interferometer for movement monitoring," Paper presented at the Proceedings of SPIE - the International Society for Optical Engineering, Vol. 8370, 83700R (2012).
5. Kepak, S., Cubik, J., Doricak, J., Vasinek, V., Siska, P., Liner, A., and Papes, M., "The arms arrangement influence on the sensitivity of Mach-Zehnder fiber optic interferometer," *Proceedings of SPIE - The International Society for Optical Engineering* Vol. 8697, 877415-877415-8. DOI: 10.1117/12.2017305.
6. Cubik, J., Kepak, S., Doricak, J., Vasinek, V., Liner, A., and Papes, M., "Interferometric sensor based on the polarization-maintaining fibers," Paper presented at the Proceedings of Czech-Polish-Slovak Optical Conference, (2012).
7. Tian, Z., S. S. -H Yam, and H. -P Loock, "Refractive Index Sensor Based on an Abrupt Taper Michelson Interferometer in a Single-Mode Fiber." *Optics Letters* 33 (10): 1105-1107., (2008).
8. Rong, Q., X. Qiao, Y. Du, H. Sun, D. Feng, R. Wang, M. Hu, and Z. Feng. 2014. "In-Fiber Quasi-Michelson Interferometer for Liquid Level Measurement with a Core-Cladding-Modes Fiber End-Face Mirror." *Optics and Lasers in Engineering* 57: 53-57.
9. Fan, J., Zhang, J., Lu, P., Tian, M., Xu, J., and Liu, D., "A Single-Mode Fiber Sensor Based on Core-Offset Inter-Modal Interferometer." *Optics Communications* 320: 33-37, (2014).
10. Yu, H., Xiong, L., Chen, Z., Li, Q., Yi, X., Ding, Y., Wang, F., Lv, H. and Ding, Y., "Ultra-compact and High Sensitive Refractive Index Sensor Based on Mach-Zehnder Interferometer." *Optics and Lasers in Engineering* 56: 50-53. (2014).
11. Siska, P., Brozovic, M., Cubik, J., Kepak, S., Vitasek, J., Koudelka, P., Latal, J., and Vasinek, V., "Fiber Bragg Grating vibration sensor with DFB laser diode," *Proceedings of SPIE - The International Society for Optical Engineering* Vol. 8697, 86971P-86971P-9. DOI: 10.1117/12.2010467

Speech Quality Measurement of GSM infrastructure built on USRP N210 and OpenBTS project

Marcel Fajkus and Martin Mikulec

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{marcel.fajkus, martin.mikulec}@vsb.cz

Abstract. The paper deals with the methodology for speech quality measuring in GSM networks using Perceptual Evaluation of Speech Quality (PESQ). The paper brings results of practical measurement of own GSM network build on the Universal Software Radio Peripheral (USRP N210) hardware and OpenBTS software. This OpenBTS station was installed in open terrain and the speech quality was measured from different distances from transmitter. The limit parameters of OpenBTS station with USRP N210 were obtained.

Keywords: PESQ, OpenBTS, USRP, GSM, Asterisk

1 Introduction

GSM and UMTS technologies have become common part of almost every aspect of a human activity during its expansion of last decade. Nowadays people consider using mobile phones for calling, texting and browsing the Internet as natural service like electricity or drinkable water at home. All these services require reliable infrastructure and control mechanism to ensure quality of the provided service. This paper deals with one aspect of control mechanism, measuring and controlling the quality of speech. In GSM/UMTS environment we can take advantage of algorithms that have been invented for IP based networks. There are two basic evaluations of speech quality, objective and subjective. We have aimed at the objective method of measuring because of the zero influence of human factor and because of possibility to repeat the same methodology for measuring speech quality in every GSM or UMTS infrastructure. The mobile terminals were connected to GSM infrastructure provided by testing OpenBTS station; build on Universal Software Radio Peripheral. We have aimed at measurement of speech quality in open terrain according to the distance of mobile terminal from the base transceiver station and according to the number of simultaneous calls.

2 Testing platforms

We have designed the measuring platform, which is able to generate GSM calls automatically in regular intervals and analyze the voice sample according to the P.862 specification. The results were logged for future purposes.

As the measured platform, we have used own GSM infrastructure provided by testing OpenBTS station build on Universal Software Radio Peripheral used in our laboratory.

2.1 Measuring platform details

The measuring platform was implemented on the low-energy consumption embedded device. According to this fact, we were able to use power supply from car and get high mobility of end mobile devices. The low-cost Huawei K3765 modems were used as end mobile devices for originating and receiving testing calls. These modems were connected directly with embedded device by USB ports and controlled by Asterisk PBX system installed inside the embedded device.

The calls with original speech sample were generated automatically by Asterisk over SIP (Session Initiation Protocol) and RTP (Real-time Transport Protocol). The modems were providing SIP/GSM translation of the outgoing and incoming calls. The Asterisk PBX was recording the incoming voice data to a separate WAV file with adequate time stamp of the measurement. After the end of the call, the PESQ algorithm was applied to the original and recorded degraded signal. The result, speech quality in MOS, was stored in the database for future evaluation.

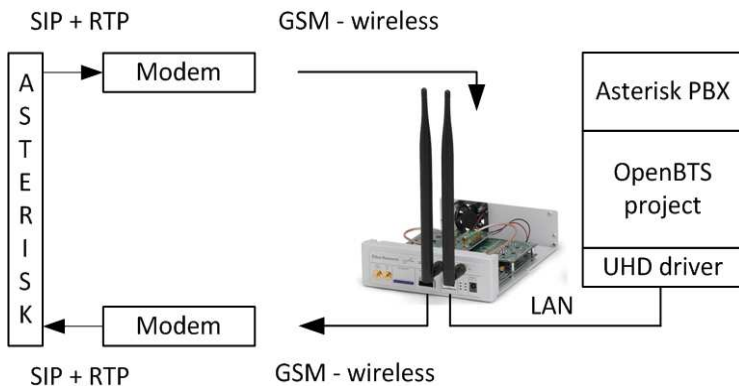


Fig. 1. Scheme of measuring – BESIP + USRP N210 connected with host OpenBTS server

The tested GSM infrastructure was built on Universal Software Radio Peripheral (USRP) hardware in connection with OpenBTS (Open Base Transceiver Station) software, which is opensource linux-based application, that provides management of USRP to create wireless GSM interface. We have used USRP N210 from Ettus Research™ in combination with Daughterboard WBX. [6]

There is UHD (USRP Hardware Driver), which provides API and drivers to the host computer for communication with USRP device. OpenBTS software is written in the C++ programming language and ensures the following functions of Um radio interface, which provides radio interface for GSM standard. [7,8]

- GMSK modulation with 13/48 MHz modulation rate and 200 kHz distance – supports GSM850, PGSM900, DCS1800 and PCS1900.
- Multiplexing and coding
- Management of network resources

Asterisk PBX is implemented inside the OpenBTS project and provides mobility management, authentication of the user and routing of the calls between registered users. Registered modems are identified by IMSI number of SIM card presented in the modems. The main advantage of this solution is that the calls can't be affected by commercial mobile operator; whole traffic is routed through our low-cost infrastructure.

We have installed USRP N210 with OpenBTS server statically outside the laboratory in an open terrain to ensure minimal influence of surrounding condition on a radio signal spreading.

The measuring platform, represented by embedded BESIP with modems, was installed in a car and placed nearby our OpenBTS station.

When the modems were registered to our OpenBTS station, there were 3 calls originated on that place. The reason of the three calls was to get statistically reasonable results of speech quality and strength of the incoming signal to the modems. Than the measurement was repeated with increasing distance from the OpenBTS station.

We were interested in a few measuring parameters:

- The maximum number of simultaneous call, that is our OpenBTS able to provide
- Dependence of receive signal strength on the distance from our OpenBTS
- Dependence of a speech quality on the distance from our OpenBTS

3 Results of measurement

After many repeated measurements in increasing distance, we have reached sufficient amount of data to statistically evaluate limit parameters of our OpenBTS system. The results were following.

3.1 Maximum number of simultaneous calls

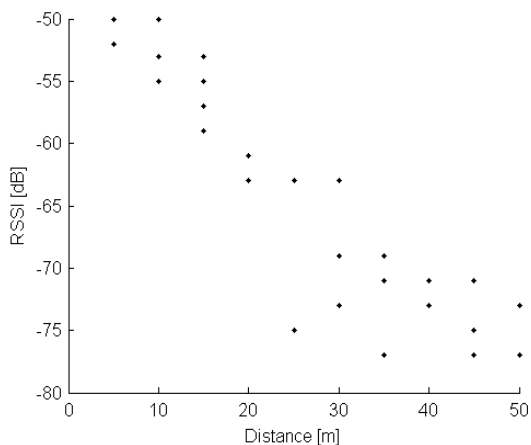
OpenBTS in basic, non-commercial version use one logical channel. GSM technologies use time multiplexing, one channel can be separated to 8 single time channels. We have found out, that in our configuration the maximum number of simultaneous calls were 3 simultaneous calls. The probability of successful call establishing was only 12% (4 successful calls from 31 tries)

Table 1. Success rate of established calls

Number of simultaneous calls	Number of originated calls	Number of successfully answered calls	Probability of successful answered calls [%]
1	35	31	88,6
2	34	23	67,6
3	31	4	12,9

3.2 Dependence of receive signal strength on the distance

There was a received signal strength (RSSI) measured during every active call. Measured and average values are depicted on the Fig.2. As the picture shows, the RSSI decreases -1,95 dB to 1 meter in average

**Fig. 2.** RSSI of GSM signal in different distance from OpenBTS

3.3 Dependence of a speech quality on the distance from OpenBTS

The measured MOS values of speech quality are depicted on the Fig.3. We have removed values smaller than 2.2 MOS, because of the poor quality according to the ITU-T P.800 recommendation.

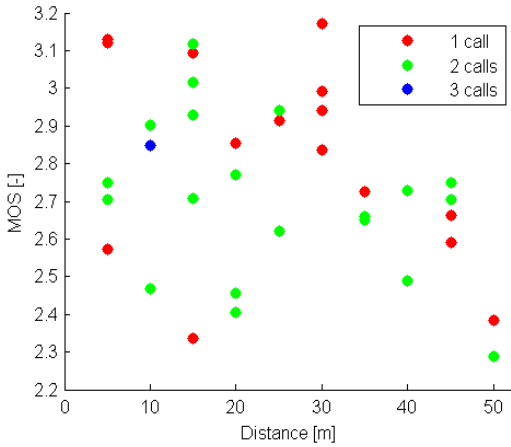


Fig. 3. Speech quality in MOS in different distance from OpenBTS

We have made regression analysis from the measured data. The results are depicted on the Fig.4. According to the low count of the successful calls for 3 simultaneous calls, we have added the MOS values < 2.2 to show regression progress for 3 simultaneous calls in Fig.4.

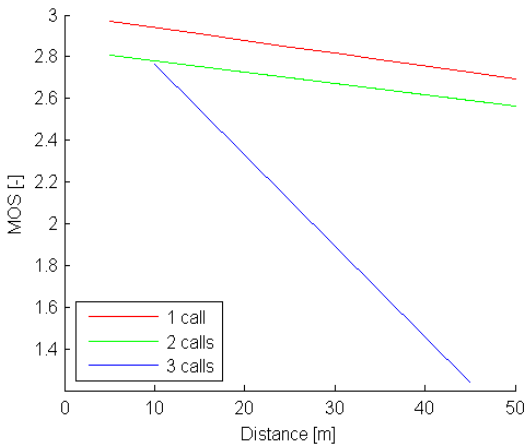


Fig. 4. Linear regression MOS value for all calls

4 Conclusion

The results of the measurement showed, that our OpenBTS station built on the USRP N210 is not suitable for practical application, only for laboratory purposes. The maximum of simultaneous calls were 3 calls, optimal load of the OpenBTS station is 1 – 2 simultaneous calls from 5 to 50 meters from the transmitter. The poor speech quality

and connection failures are detected out of the measured limits. The measurement approved theoretical prerequisite that the speech quality decreases with increasing distance from the OpenBTS station and number of simultaneous calls.

The main benefit of the paper is the methodology for automatic speech quality measurement of a general GSM infrastructure. The methodology uses low-cost embedded hardware, modems and opensource software tools.

Our next step will be to install new OpenBTS station built on more powerful USRP hardware from National Instruments, repeat the same measurement on it and compare the results.

Possibility of Computer Simulation of Modulation Formats for Real Environment Conditions of Free Space Optics

Lukas Hajek, Ales Vanderka, Tomas Hlavinka, and Jan Latal

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ales.vanderka@vsb.cz

Abstract – This article is aimed at evaluation measurement of parameters such as relative air humidity, wind speed, temperature, pressure and solar radiation received from the weather station. Random changing of these parameters creates atmospheric turbulences, absorption and dispersion centers. It is necessary to specify the value for structural parameter of refractive index C_n^2 because the need of the influence of atmosphere on Free Space Optics (FSO). The first part of this article includes the theoretical calculation for C_n^2 , there are used two models PAMELA and Macroscale-Meteorological model. In the next part there is realized a simulation and comparing value of received signal strength indication (RSSI) for the real and simulating case of Free Space Optics. The last part is aimed at testing FSO with modulation formats OOK-RZ, OOK-NRZ and BPSK.

Keywords: structural parameter of refractive index, turbulence, scintillation, modulation, BPSK, OOK, RSSI, PAMELA, MOS, FSO

1 Introduction

The Free Space Optics (FSO) are similar to Wi-Fi technology, which is used in radio communication systems. These FSO are advantageously used in areas, where is very difficult built an optical or electrical data sites. A high bit rate is uncontested advantage of FSO and it is closely related with using modulation format. Another advantage of FSO is simple and fast installation, lower operational requirements, released the ISM band for unlicensed use and for very difficult bugging (security). Disadvantages of FSO are dependence of weather and influence of atmospheric phenomena.

An atmospheric environment is stochastic transmission medium, where the refractive index of air is fluctuating during the day, especially at the turn of the day and night. Furthermore, the fluctuation of refractive index is related on wind velocity, roughness of the earth's surface, amount of solar radiation, rainfall, air pressure, density and composition of the air. The intensity of these fluctuations is described by structural parameter of the refractive index C_n^2 . For the theoretical calculation of structural parameter of refractive index basic mathematical models can be used, obtained from empirical study of the influence of the atmosphere on the structural para-

meter of the refractive index. These models are known as horizontal models and include the PAMELA model based on the Monin-Obhukov similarity theory (MOS) and Macroscale-Meteorological model based on global meteorological data released by the U.S.Army Night Vision Laboratory [2]. By comparing the two models, PAMELA model is far more difficult to re-computing performance and includes more parameters. Macroscale-Meteorological model is based especially on special weight function relative of day.

Often changing atmospheric parameters limits FSO also in the possibility of using sophisticated modulation formats in terms of achieving higher bit rates. The most commonly used modulation format is therefore modulation format OOK (On Off Keying) variants RZ (Return to Zero) and NRZ (Non-Return to Zero).

2 Description of the measured area, the equipment used and the measured values

The data was analyzed during the month of September 2012. This month was chosen due to the completeness of the data and also due to the development of RSSI value in each day, because the RSSI value fluctuated approximately within the same range. It was considered that in this month there were very few dispersion centers, which would be appreciably showed on the RSSI diagram.

2.1 Description and measurement of the FSO

Measurement of the received signal strength indication (RSSI) was realized on the installed FSO in the premises of the VŠB – Technical University of Ostrava. The total length of the optical path was 1,47 km and as a sources of light radiation three infrared lasers were used which works at $\lambda = 830 - 860$ nm, beam divergence was 2 mrad, total power of the optical head was 40 mW (2x40 mW and 1x60 mW). Transmitting lens diameter was 6 cm and receiving lens diameter was 20 cm. For detection of an optical signal has been used APD (Avalanche Photodiode) and the maximum bit rate of the FSO was $1,25 \text{ Gbit}\cdot\text{s}^{-1}$. The basic circuit diagram FSO is shown in Figure1.

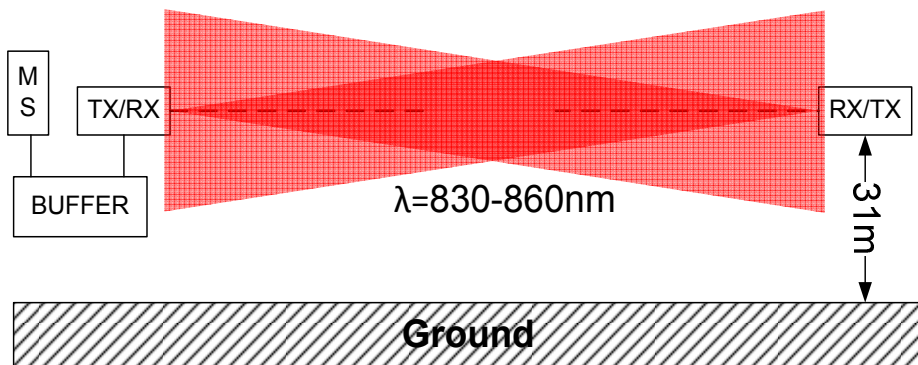


Fig. 1. Basic scheme of FSO system, MS (Meteorological station).

Optical head for FSO link transmits periodically sequence of bits and measures value of the received signal strength indication RSSI. The RSSI was recorded in one minute intervals and a total result represents 21 508 values. Therefore, it appeared preferable to perform at hourly arithmetic means of measured RSSI values. In Figure 5 is shown the continuance of hourly arithmetic means of measured RSSI values.

2.2 Meteorological data

For measurement of meteorological data weather station Davis Vintage Pro2 was used. This weather station is installed in close proximity to the optical head for FSO on the roof of the main building of VŠB – Technical University of Ostrava. In addition to the values of wind speed, temperature, pressure, relative humidity and solar radiation station the weather station is also capable of measuring wind direction and rainfall. There were made hourly averages from measured values as is shown in Figures below.

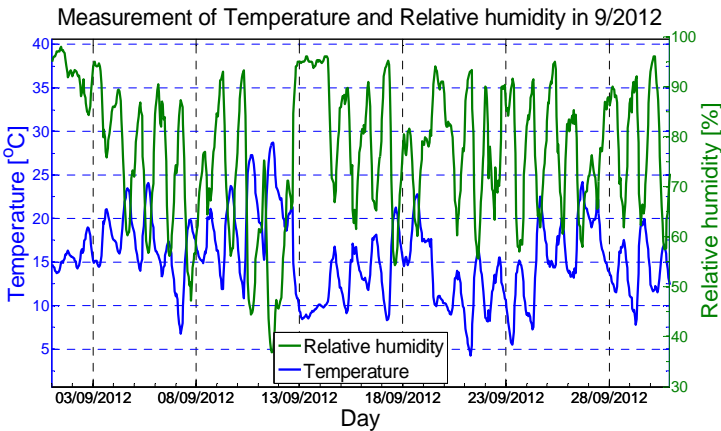


Fig. 2. Measurement of temperature and humidity in 9/2012.

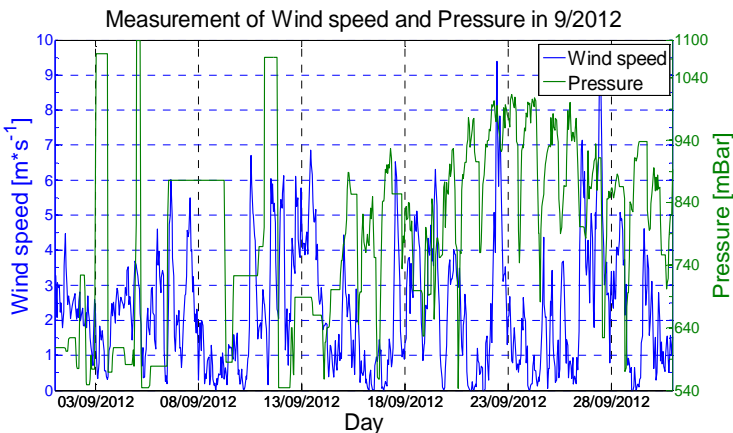


Fig. 3. Measurement of wind speed and pressure in 9/2012.

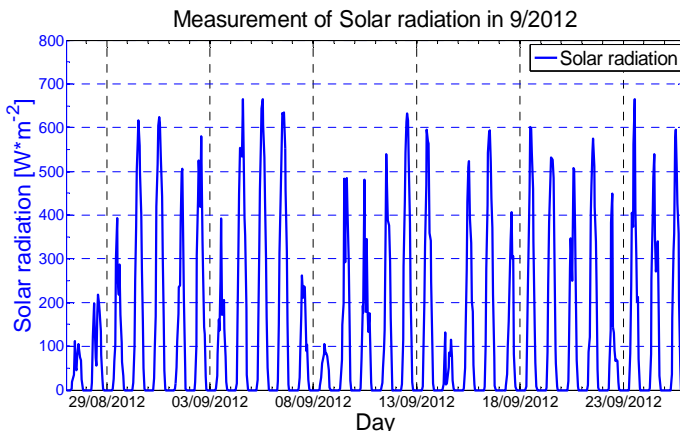


Fig. 4. Measurement of solar radiation in 9/2012.

3 Modeling of structural parameter of the refractive index C_n^2

For calculation of the structural parameter of the refractive index C_n^2 two mathematic models have been used. The first has been PAMELA model which provides estimates of C_n^2 within the surface boundary layer. The required inputs are latitude, longitude, date, time of day, percent cloud cover and terrain type, as well as single measurement of atmospheric temperature, pressure and wind speed. With the derivation of the basic equation for calculating the refractive index (equation 1) can be obtained equation to the calculate the structural parameter of the refractive index C_n^2 . [1]

$$\Delta n = \frac{77.6 \cdot 10^{-6} \cdot Pa}{T} * \left(1 + \frac{7.52 \cdot 10^{-3}}{\lambda^2} \right) \tag{1}$$

By modifying this formula, where λ represents the wavelength of the radiation light source we obtain these formula.

$$C_n^2 = \frac{b \cdot K_h}{\varepsilon^{1/3}} \left(\frac{\partial n}{\partial h} \right)^2 \tag{2}$$

, where b is a constant commonly approximated by 2,8. By ignoring the small contribution to the total differential from fluctuations in atmospheric pressure, we can differentiate (equation 1) with respect to the potential temperature θ and by using next equation from (more in [1] p.25) with ignoring small contribution due to wavelength it follows that.

$$\frac{\partial n}{\partial h} = \frac{\partial n}{\partial \theta} \frac{\partial \theta}{\partial h} = \frac{(-77,6 \cdot 10^{-6} Pa) T_* \Phi_h \left(\frac{h}{L} \right)}{k_v h T^2} \tag{3}$$

, where k_v is von Karman's constant taken to be 0,4, T_* is characteristic or scaling temperature, Φ_h is dimensionless temperature gradient, L is Monin-Obukhov length and u_* is friction velocity.

Next model is based on Macroscale-Meteorological model and carried out by the U.S. Army Night Vision Laboratory. This model is based on standard meteorological parameters measured the world over. It is also based on the concept of temporal hours or relative part of day. Relations between C_n^2 and th (temporal hour) parameter led to construction of weight function. Values of th are obtained in following way [2]:

1. one th is obtained by subtracting the hour of sunrise from that sunset and dividing by 12.
2. the current th is obtained by subtracting the hour of sunrise from current hour and dividing by the value of 1 th.

$$C_n^2 = 3,8 * 10^{-14}W + 2 * 10^{-15}T - 2,8 * 10^{-15}RH + 2,9 * 10^{-17}RH^2 - 1,1 * 10^{-19}RH^3 - 2,5 * 10^{-15}WS + 1,2 * 10^{-15}WS^2 - 8,5 * 10^{-17}WS^3 - 5,3 * 10^{-13} \quad (4)$$

,where W is temporal hour weight, T is air temperature in degrees of Kelvin, RH is relative humidity in (%) and WS is wind speed (m/s) [2].

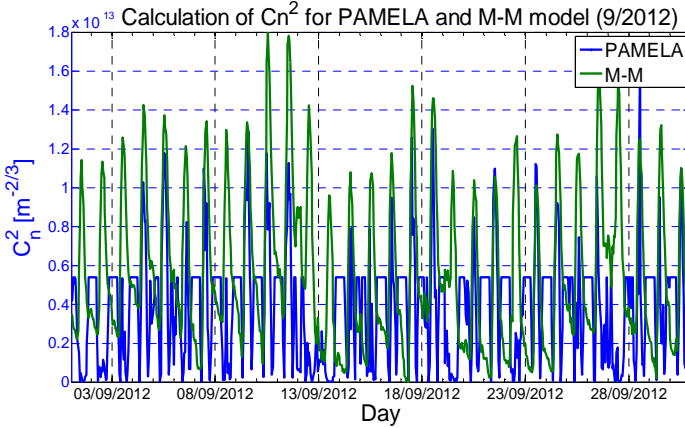


Fig. 5. Graph for PAMELA and M-Mmodel of C_n^2 .

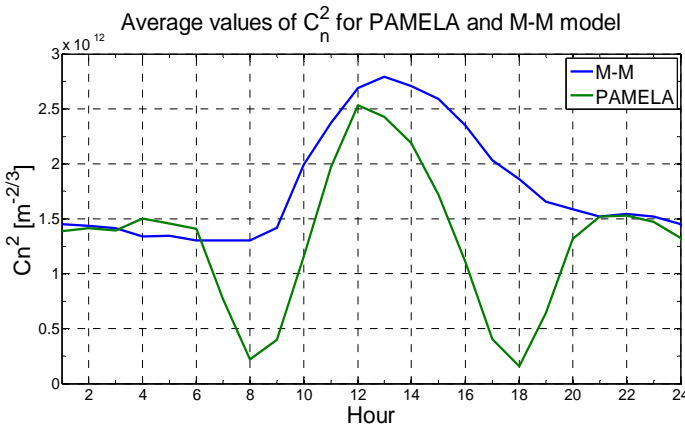


Fig. 6. Average values of C_n^2 for PAMELA and M-M model.

4 Simulation of modulation formats

4.1 OOK

OOK (On-Off-Keying) modulation format is the most commonly used in commercial terrestrial FSO communications. Advantage of this format is resistance to nonlinearities of the laser and an external modulator. However, this format is much more sensitive to turbulence and other disturbances which leads to fluctuation in the received optical power. We can reduce this fluctuation by replacing the decision-making level with adaptive decision-making level. The OOK modulation format can be used in non-return to zero (NRZ) or return to zero (RZ) format.

$$BER_{OOK-NRZ} = \frac{1}{2} \operatorname{erfc} \left(\frac{1}{2\sqrt{2}} \sqrt{SNR} \right) \quad (5)$$

$$BER_{OOK-RZ} = \frac{1}{2} \operatorname{erfc} \left(\frac{1}{2} \sqrt{SNR} \right) \quad (6)$$

4.2 BPSK

For OOK-RZ format the pulse duration is shortens at “1” thereby increasing the energy efficiency, but to the detriment of larger bandwidth requirements as compared to non-return-to-zero format. To achieve the same bit-error-rate (BER) value a half signal-to-noise ratio (SNR) is required then regular OOK-RZ format.

$$BER_{BPSK} = \frac{1}{2} \operatorname{erfc} \left(\frac{\sqrt{SNR}}{\sqrt{2}} \right) \quad (7)$$

4.3 Simulation of modulation formats in OptiSystem

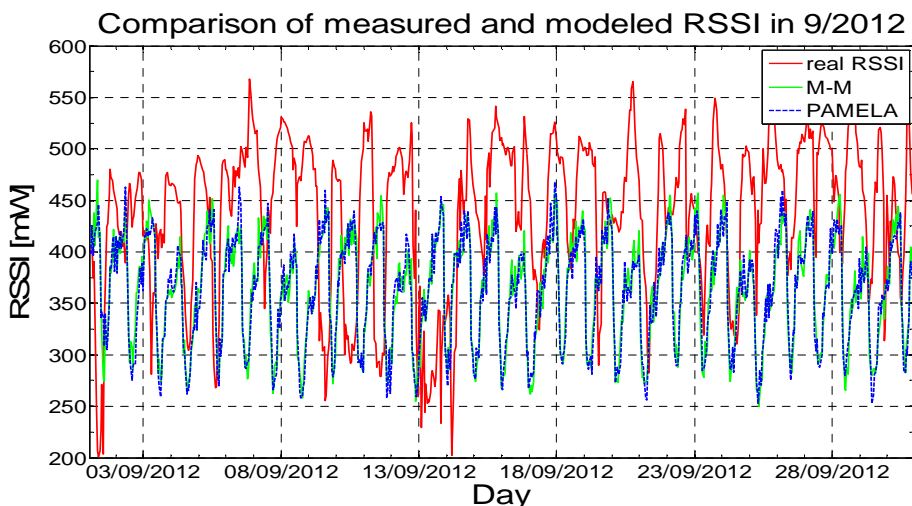


Fig. 7. Comparison of measured and modeled RSSI in 9/2012.

In Figure 7 is shown comparison of progress RSSI signal between the real measure and the both simulation models PAMELA and Macroscale-Meteorological model.

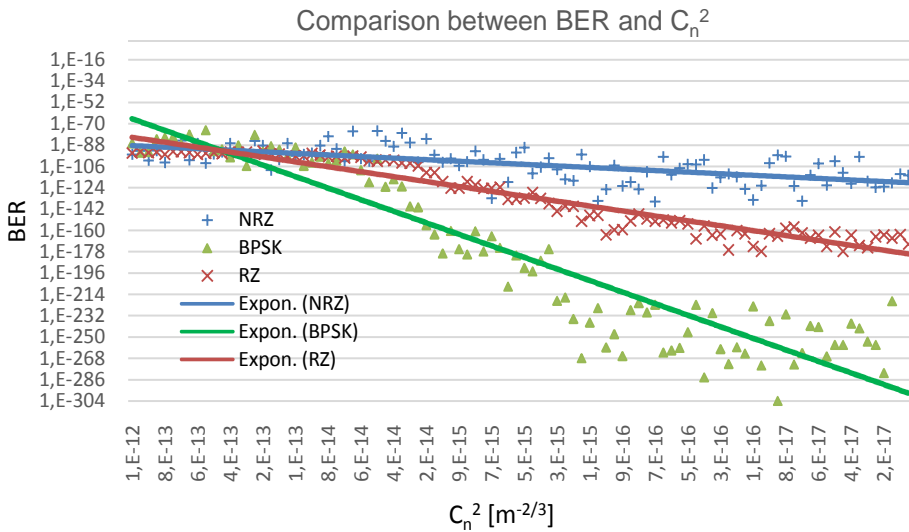


Fig. 8. Comparison between BER and C_n^2 for OOK-RZ, OOK-NRZ and BPSK.

In Figure 8 is shown the dependence between the parameter of bit error rate BER and structural parameter of refractive index C_n^2 . The values of the C_n^2 was assigned in the range $10^{-12} - 10^{-17} m^{-2/3}$.

5 Conclusion

In this article, the author’s team tried to determine the behavior, respectively the strength of turbulence in the atmosphere by using mathematical models for calculation structural parameter of refractive index (C_n^2). The terrestrial mathematical models concrete PAMELA and Macroscale-Meteorological model were used to calculate C_n^2 . Both of these models use different principle for the calculation C_n^2 when is sunrise or sunset which is reflected especially in the result. The PAMELA model shows the lowest level of turbulence, approximately from 10^{-17} to the value of high turbulence 10^{-13} in the result (Figure 7). The both models show a similar trend of progress when the average values of C_n^2 are stable at around 10^{-14} value at the night (Figure 6). The largest differences were occurred during the days between the models when the values of the solar radiation gradually decreased. The PAMELA model counts with this parameter, but Macroscale-Meteorological model calculates the weight function of the day.

The comparison of simulated and of real measured values of RSSI is good to see a similarity both calculated models PAMELA and M-M (Figure 7). The real measured values of RSSI showed to large inaccuracies which were caused an automatic adjust-

ment of optical power emitted from the transmitter, as compared with the models. These values showed some similarity around 400mW and in area where the RSSI value increased and decreased. The last part of simulation was aimed to comparison of modulation formats for FSO, concrete OOK-RZ, OOK-NRZ and BPSK format. These modulation formats have been tested on the parameters of the real Free Space Optics with the calculated values of the structural parameter of the refractive index $Cn2$ from the both models. The result of this comparison was characteristic depending of bit error rate BER on $Cn2$. Therefore implies that decreasing intensity of turbulence the BER value was most decreased with used BPSK modulation format. Vice versa the minimum decrease of value was noticed with the OOK-NRZ modulation format.

Acknowledgements

The research described in this article could be carried out thanks to the active support of the Ministry of Education of the Czech Republic within the projects no. SP2014/77 and SP2014/147. This article was supported by project VG20102015053 and Technology Agency of the Czech Republic TA03020439, TA04021263. The research has been partially supported by the project no. CZ.1.07/2.3.00/20.0217 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.

6 References

1. MAJUMDAR, Arun K and Jennifer Crider RICKLIN. Free-space laser communications: principles and advances. New York, NY: Springer, c2008, x, 417 p. ISBN 9780387286778.
2. KOPEIKA, Norman S and Jennifer Crider RICKLIN. A system engineering approach to imaging: principles and advances. Bellingham, Wash.: SPIE Optical Engineering Press, c1998, xxi, 679 p. ISBN 08-194-2377-7.
3. BOUCHET, Olivier a Jennifer Crider RICKLIN. Wireless optical telecommunications: principles and advances. Hoboken, NJ: John Wiley, 2012, xxiv, 284 p. ISBN 978-184-8213-166.
4. BENDERSKY, S., N. S. Kopeika, and N. Blaunstein, "Atmospheric optical turbulence over land in middle east coastal environments: prediction modeling and measurements," *Appl. Opt.* 43, 4070-4079 (2004)
5. DOSS-HAMMEL, Stephen, Eun OH, Jennifer C. RICKLIN, Frank D. EATON, G. Charmaine GILBREATH, Dimitri TSINTIKIDIS, Jennifer C. RICKLIN and David G. VOELZ. "A comparison of optical turbulence models." p. 236-246. DOI: 10.1117/12.563746.
<http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1281341>
6. HRANILOVIC, Steve. *Wireless optical communication systems*. 1. vyd. New York: Springer, c2005, xii, 196 p. ISBN 03-872-2785-7.
7. RAMIREZ-INIGUEZ, Roberto, Sevia M IDRUS and Ziran SUN. *Optical wireless communications: IR for wireless connectivity*. 1. vyd. Boca Raton: CRC Press, c2008, xxxi, 344 p. ISBN 08-493-7209-7.

Basic Evaluation and Testing of V2I Communication System

Stanislav Hejduk

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

Abstract.

In this paper, basic information about the proposed V2I communication system is presented. This document contains informations about the impact of the background communication on the lighting function of the used street lights. It also describes the characteristics of emerging interference zones in the covered area, where individual lights are sending different data streams.

Keywords: V2I, LED, VLC, interference.

1 Introduction

LED diodes have been significant part of everyday life for many years. Nowadays they are also replacing obsolete light sources in street lights. This is an opportunity to establish communication between street lights (Infrastructure) and road users (Vehicles). So the vehicle or infrastructure can send important data directly to the active receivers in covered area. This type of communication is called V2I (Vehicle-To-Infrastructure). [1]

Another similar type of communication is direct communication between neighboring vehicles V2V (Vehicle-To-Vehicle). This is also possible due to the increasing use of LEDs in the automotive industry.

Those types of communication can be simply used for sending telemetry in well defined area. It can be used for sending informations about the near car accidents, roadblocks and other informations about emergency situations and traffic conditions.

Since the LED street lighting uses primarily white LEDs with extremely high power output, the modulation speeds are significantly reduced. Together with extreme multipath signal propagation and outdoor interference is the expected communication speed of V2I and V2V communication too low to allow some high speed communication, but at least for basic telemetry is communication fast enough. [2,3]

The following text describes measured results and theoretical calculations of basic communication parameters like color shifts during the communication, interference zones arising due to the diffuse nature of streetlights and amount of received data depending on the vehicle speed and the covered undisturbed area.

2 Color and intensity shifts during the communication

One of the main problems with visible light communication (VLC) is its observability. Communication must not be observable by the human eye and also by some optoelectronic devices e.g. security cameras. The first precondition is to maintain a stable mean value of the transmitted signal. The second thing is the dependence of the light intensity on the modulation frequency. Since the luminophor does not manage to react fast enough on high frequencies, it leads to a reduction of luminous flux which may be observable.

2.1 Color shifts

The following figures show the effect of modulation on the color coordinates of the used white LED. Even with relatively small speeds are coordinates shifted by 148K.

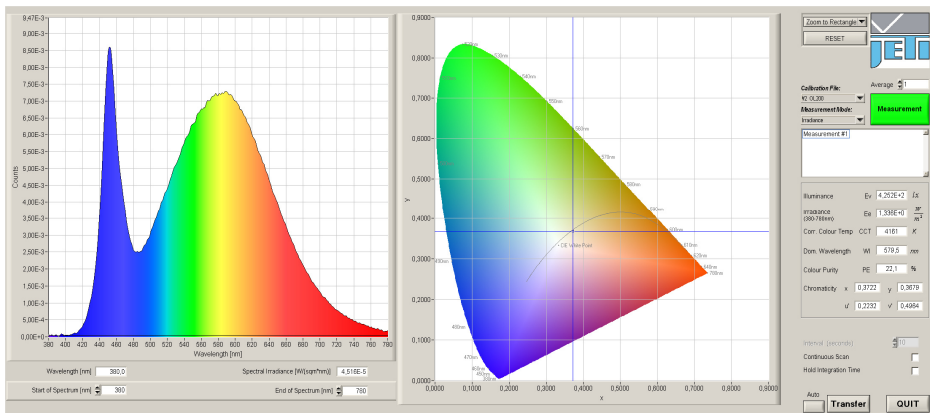


Fig. 1. White LED spectrum without communication (4161K).

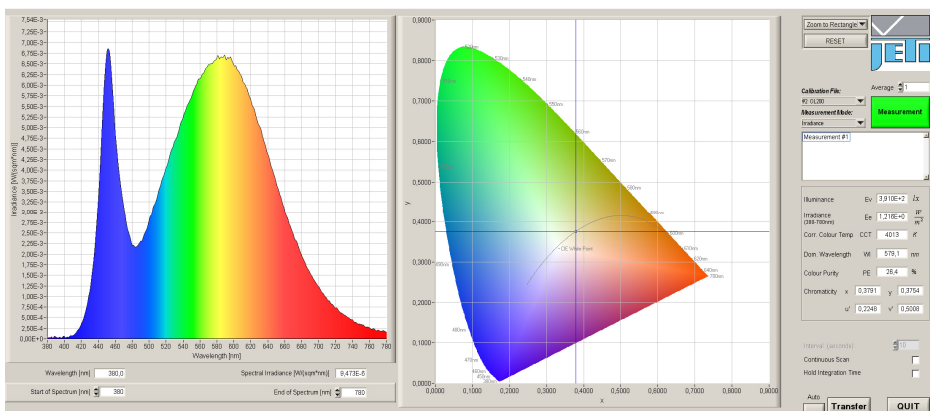


Fig. 2. White LED spectrum with 5,6kbit/s Differential Manchester communication (4013K).

2.2 Intensity vs. Frequency

High power combined with high frequency have a negative impact on light intensity. The course of change of light intensity is shown in Fig. 3.

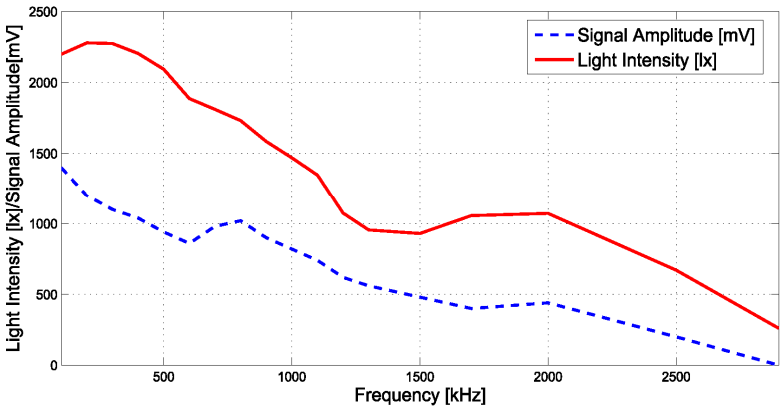


Fig. 3. Effect of frequency on light intensity and signal amplitude.[4]

3 Interference

Aim of the street lighting is to cover the biggest possible area. This means that they light is usually overlap. In case that the lights are transmitting the same data is communication perfect. However, loss of synchronization or different data resulting in formation of the interference zone, where is communication impossible.

In following text are mentioned two different scenarios which were used as an attempt to eliminate this interference.

3.1 Complete cell overlap

In this case is sensitivity of the receiver set to the maximum and individual signal covers the entire area between neighboring streetlights. This case is illustrated in Fig.4.

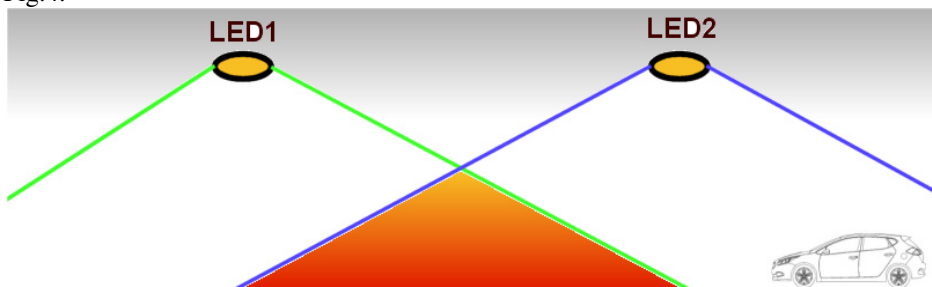


Fig. 4. Area coverage with mutual overlapping.

As a result, 30% of covered area in the middle of streetlights is jammed and unable to establish communication. In the rest of the area is the receiver able to correctly distinguish the closer signal.

3.2 No overlap

In this case were light intensities and receiver sensitivity set to no overlapping between individual lights, shown in Fig.5. However the streetlights are still overlapping, but light intensity in the middle of the lights should be to small to cause such huge interference.

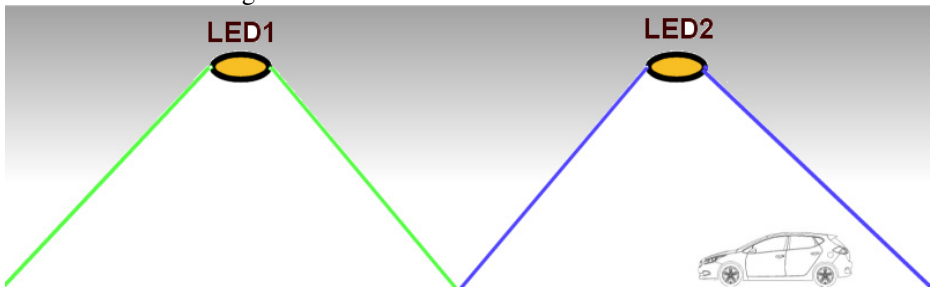


Fig. 5. Area coverage without overlapping.

Result of this experiment is same as in the previous case. Entire 30% of the covered area between lamps was full of interference and receiver was unable to decode the right signal.

This problem is caused by the diffusive nature of streetlights coverage and according to results is extremely difficult to eliminate. Only possibility is the use of the two independent photodetectors which will be aimed to different directions, so the signals will be divided by the receiving angle of photodetectors itself.

4 Amount of received data vs. Speed

Because the car is moving with different speeds in different places, the time that it spends in the covered area will be also different. Question is how much data will be car able to receive at a given speed. This case can be described by the following equation.

$$n_{RB} = \frac{s_{com} \cdot 3,6}{v \cdot t_B} \tag{1}$$

Where n_{RB} is the number of received data bursts, s_{com} is the covered area in [m], v is speed of the vehicle in [km/h] and t_B is the duration of the one data burst [s] which has to be completely received for correct decoding.

Fig. 6 shows the graph of the theoretical number of received data bursts for $s_{com}=4m$, $t_B=1ms$, together with the results of the real measurements.

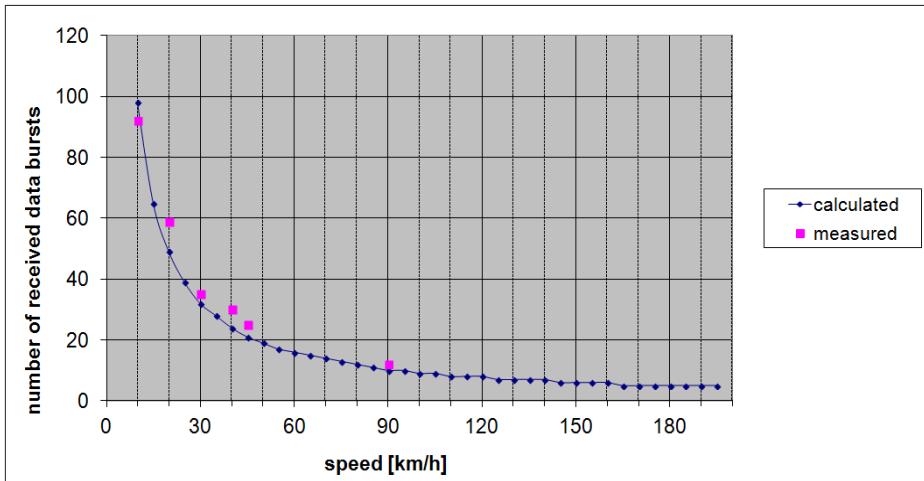


Fig. 6. Graph of measured and calculated number of received data bursts for $s_{com}=4m$, $t_B=11ms$.

5 Conclusion

As a result it can be seen, that V2I communication is possible and because of huge deployment of LEDs to automotive industry and infrastructure, it is also simple to establish this communication with already existing technology.

Unfortunately, this kind of communication is more suitable for telemetry sending instead of high speed internet and reliable data transfer. Higher speeds are also little bit difficult, because of color and light intensity shifts.

In terms of traffic safety will be this technology undoubtedly very beneficial.

6 Acknowledgement

This article was created with the active support of the Ministry of Education of the Czech Republic within the projects no. SP2014/147, SP2014/77 of VSB-Technical University of Ostrava. This article was supported by the projects no. VG20102015053 and TA03020439.

7 References

1. Biagi, M., Rinauro, S., Colonnese, S., Scarano, G., & Cusani, R. (2013). Short paper: Dynamic network selection in V2I systems. Paper presented at the *IEEE Vehicular Networking Conference, VNC*, 190-193.
2. Godoy, J., Milanés, V., Pérez, J., Villagrà, J., & Onieva, E. (2013). An auxiliary V2I network for road transport and dynamic environments. *Transportation Research Part C: Emerging Technologies*, 37, 145-156.

3. Huang, W., Yu, Z., Zhu, F., Yang, L., & Wang, F. -. (2013). Applicability of short range wireless networks in V2I applications. Paper presented at the *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 231-236.
4. Hejduk, S., Witas, K., Latal, J., Vitasek, J., Bocheza, J., Vasinek, V. Simple and Universal Current Modulator Circuit for Indoor Mobile Free-Space-Optical Communications Testing. *Advances in Electrical and Electronic Engineering, 2014. Vol. 12, No. 1.* pp. 66-74. ISSN 1804-3119.

Mach-Zehnder interferometer for movement monitoring

Stanislav Kepák

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
stanislav.kepak@vsb.cz

Abstract. Article summarizes the progress of the third year of the author's study of the field of Communication technology. The topic of dissertation thesis is the Mach-Zehnder interferometer for movement monitoring. Article summarizes activities at the Department of Telecommunications during the academic year 2013/2014, research and development activities related to the theses and research projects and a list of publications in which the author participated.

Keywords: interferometer, fiber-optic sensor, communication technology, thesis

1 Introduction

I am studying full-time doctoral study at the Faculty of Electrical Engineering and Computer Science at the Technical University of Ostrava. My study programme is called Computer Science, Communication technology and Applied Mathematics and field of study is Communication Technology. The topic of my dissertation thesis is Mach-Zehnder interferometer for movement monitoring and my tutor is Professor Vašínek. This article summarizes the third year of study.

2 Activities related to the dissertation thesis

Given the dissertation topic, my goal is to create a fully functional fiber-optic sensor that can analyze movement. During the academic year, I worked mainly on increasing the stability of the optical fiber Mach-Zehnder and Michelson interferometer. I tried various detectors of optical radiation and their influence on the measured signal. I also evaluated the long-term stability of the frequency, amplitude and polarization state of the radiation source which was DFB laser diode.

As the main component of the optical interferometer is beam splitter and coupler, respectively, I focused on the effect of temperature on the parameters of the couplers. In the case of Michelson interferometer, I studied the same for fiber mirrors. Over the year, I have performed experiments with optical connectors, as it turned out that the quality of the connection depends on the quality of connectors and especially to precision mating sleeve.

In terms of high sensitivity of the entire fiber-optic sensor polarization-maintaining fibers have proven to be advantageous. This type of fiber is minimizing the phenomenon of polarization fading. However, the use of polarization-maintaining fibers brings additional demands on quality of fiber-optic components of the entire sensor. Light from the laser must be perfectly linearly polarized, which requires high quality laser, potential improvements can be achieved by fiber polarizers.

Our department is newly equipped with a polarization state analyzer that allows us to analyze the polarization state of a wide range of optical and fiber-optic components. Therefore I also devoted to measuring the polarization state of lasers, fiber, couplers, mirrors, isolators and polarizers during the academic year. I evaluated connection quality of polarization-maintaining joints both with mating sleeves and splices. Since the original application did not allow recording more than 1024 readings, new application in LabVIEW which removed this restriction was developed along with colleagues. Figure 1 shows a comparison of both graphical interfaces.

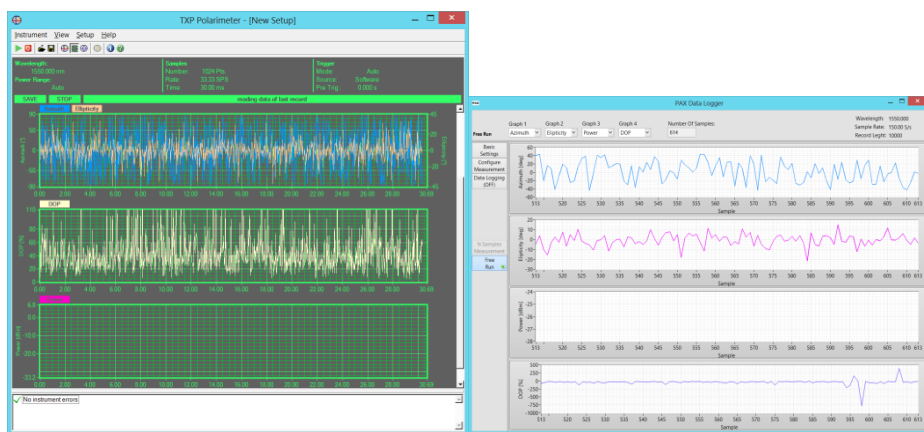


Fig. 1. GUI comparison of the original and the custom LabVIEW application

Enhancements to the functionality of the script for the visualization of measured data written in MATLAB [1] were made. The new look of the GUI is shown in Figure 2. The new functions include the ability to set parameters of FFT, data interpolation, or faster FFT calculation algorithm.

3 Activities at the Department of Telecommunications

I run practical tutorials in radio networks and telecommunications networks for the third year. I participated in the creation of template for writing bachelor and diploma theses in Microsoft Word format. At 22nd International Trade Fair of Electrotechnics, Electronics, Automation, Communication, Lighting and Security Technologies I represented the department and the faculty. I also participated in the open days in Uherské Hradiště and Ostrava.

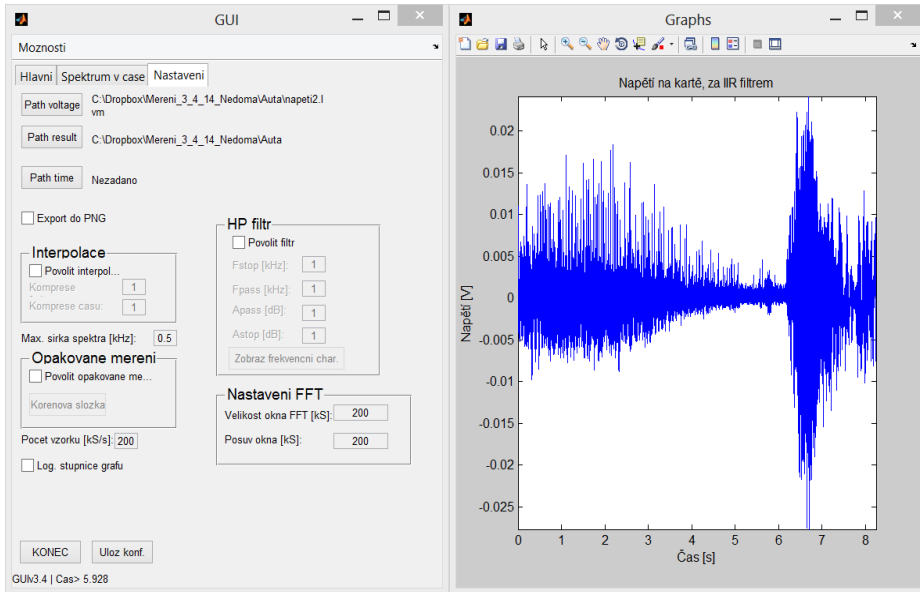


Fig. 2. Improved script for data visualization

I am working on several projects. These include Safety of optical transport network and development of optical components for energy grid and energy transport systems under the auspices of the Technology Agency of the Czech Republic, The modern structure of photonic sensors and new innovative principles for intrusion detection systems, integrity and protection of critical infrastructure – GUARDSENSE under the auspices of the Ministry of Interior, The Development of Excellence of the Telecommunication Research Team in Relation to International Cooperation under the auspices of the European Union and Student Grant Competition Projects under the auspices of Grant Agency of the VSB-Technical university of Ostrava.

4 Publications

As co-author, I participated in the following publications in last year. The usability analysis of different standard single-mode optical fibers and its installation methods for the interferometric measurements [2], Implementation of optical meanders in the temperature measurement of the extermination of basidiomycete *serpula lacrymans* using microwave heating [3], Optical vibration sensor based on Michelson Interferometer arrangement with polarization-maintaining fibers [4] and Measurement of the Microwave Emitter's Inhomogeneity Using Optical Fiber DTS [5].

References

1. Dorcak, J., Kepak, S., Cubik, J., 2012 "Software for visualization of measured data from Mach-Zehnder interferometer", Wofex conference, vol. 10, pp. 464-469. ISBN 978-80-248-2769-8
2. Cubik, J., Kepak, S., Dorcak, J., Vasinek, V., Jaros, J., Liner, A., Papes, M. & Fajkus, M. 2013, "The usability analysis of different standard single-mode optical fibers and its installation methods for the interferometric measurements", *Advances in Electrical and Electronic Engineering*, vol. 11, no. 6, pp. 535-542.
3. Liner, A., Papes, M., Jaros, J., Cubik, J., Kepak, S., Smira, P., Nasswetrova, A. & Gabriel, J. 2013, "Implementation of optical meanders in the temperature measurement of the extermination of basidiomycete *Serpula lacrymans* using microwave heating", *Advances in Electrical and Electronic Engineering*, vol. 11, no. 6, pp. 528-534.
4. Cubik, J., Kepak, S., Dorcak, J., Kajnar, T., Zboril, O., Jaros, J., Vasinek, V., Liner, A., & Papes, M., 2014, " Optical vibration sensor based on Michelson interferometer arrangement with polarization-maintaining fibers ", *Proc. SPIE 9098, Fiber Optic Sensors and Applications XI*, 90980I; doi:10.1117/12.2050528; <http://dx.doi.org/10.1117/12.2050528>
5. Jaros, J., Papes, M., Liner, A., Vasinek V., Smira, P., J., Cubik, Nasswetrova, A. & Kepak, S. 2014, "Measurement of the microwave emitter's inhomogeneity using optical fiber DTS", *Proc. SPIE 9098, Fiber Optic Sensors and Applications XI*, 909815; doi:10.1117/12.2050278; <http://dx.doi.org/10.1117/12.2050278>

The Temperature Measurement of the Extermination of Dry Rot Fungus Using Microwave Heating

Andrej Liner

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
andrej.liner@vsb.cz

Abstract. The dry rot basidiomycete *Serpula lacrymans* is the most common and destructive wood decay fungus, which attacks and damages houses and other wooden construction worldwide [1], [2]. Effective chemicals have been developed for remediation and treatment of dry rot outbreaks and for wood preservation against dry rot, but in most cases, control is most economically achieved by environmental management to avoid creating favourable growth conditions for the fungus [3]. Thermal treatment using microwaves represents one of possible approaches in fungal growth control and refurbishment of damaged wooden constructions. One of the possibilities, how to monitor this whole process seems to be the use of Optical fiber DTS (Distribution Temperature Systems). The Optical fiber DTS are unique distributed temperature systems using optical fiber as a sensor. Due to the electromagnetic resistance is this system suitable for the monitoring of these processes. This article deals with application of optical meanders in the temperature measurement during the extermination of basidiomycete *Serpula lacrymans* using microwave heating. Because of the adverse effect of microwave radiation on all other types of temperature sensors.

Keywords: Microwave emitter, microwave heating, optical fiber DTS, optical meander, *Serpula Lacrymans*.

1 Introduction

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 [4], [5]. 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 operates 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 arises 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 spectrum [6].

Spatial resolution of the DTS system is standardly about 1 m with accuracy of ± 1 °C, at a resolution of 0,01 °C.

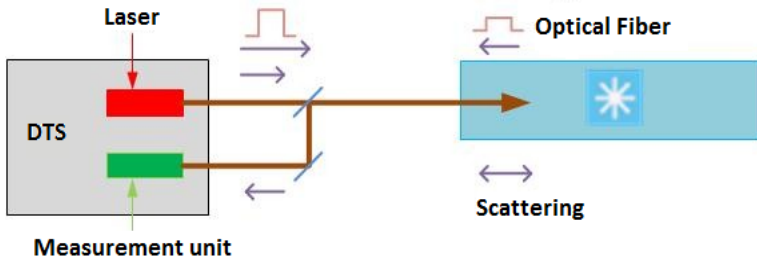


Fig. 1. Block diagram of the DTS system.

Spatial information about the temperature distribution along the optical fiber is achieved using a technique called Optical Time Domain Reflectometry (OTDR) which is nowadays mostly used for optical testing line [7] (seen in Fig. 1).

2 Extermination of *Serpula Lacrymans* Using Microwave Heating

Oat flakes were mixed with water (6:4 w/w) and autoclaved. Bricks ($10 \times 10 \times 1,5$ cm) were prepared from the material in plastic bags in aluminium form and autoclaved again after 3 days. Bricks were then inoculated with agar plugs from ME agar on Petri dishes (malt extract 7 g/L, agar 20 g/L, pH 7,0) with *S. lacrymans* (CCBAS110) and cultivated at 25 °C in darkness for 14 days. After microwave heating, samples from all bags were taken and ME agars were re-inoculated. Mycelial growth was checked daily during the next 10 days.

The measurement was carried out in the Department of Theoretical and Experimental Electrical Engineering laboratory again. In each measurement were used two samples and between these two samples was placed meander composed of measuring optical fiber (Fig. 2). The aim of this measurement was to determine the required time for the operation of the microwave emitter for the consumption of dry-rot fungus. The temperature of the sample had to be maintained over 90 °C, because this is the temperature limit for destruction of the tested *Serpula Lacrymans*. However, the temperature couldn't significantly exceed the value of 100 °C, because the measured sample could ignite. The microwave emitter was operating in sample's temperature range between 90



Fig. 2. The temperature measurement for one optical meander.

to 100 °C so that was switched off at 100 °C and switched on if sample's temperature dropped at 90 °C as can be seen in graphs.

The time intervals were chosen at 5, 15, 30, 120 and 240 minutes. This means that the sample had to be exposed to temperatures over 90 °C during this time intervals.

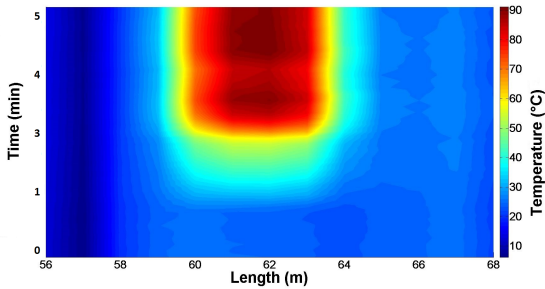


Fig. 3. The temperature measurement for one optical meander for a 5 min. X-axis represents the total length of fiber, which is zooming in to the measured optical meander. Measured optical meander is located in the section between 60 m to 63 m.

Measurements were divided into two parts. In the first part were first three time intervals (5, 15 and 30 min.) measured and in each time interval was used only one sample. The time course of the first measurement, 5 min. time interval is shown in Fig. 3.

After 5 minutes, the samples were changed and the measurement in 15 min. time interval was initiated. The time course of this measurement is shown in Fig. 4.

After 15 min. were the samples changed again and the measurement in 30 min. was initiated. The time course of this measurement is shown in Fig. 5.

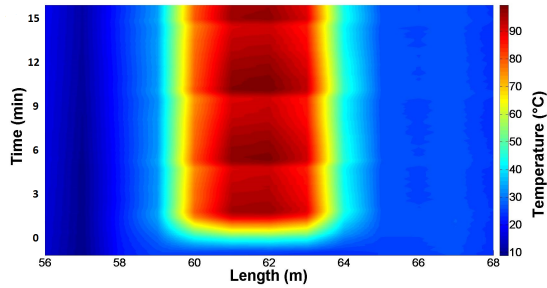


Fig. 4. The temperature measurement for one optical meander for a 15 min. X-axis represents the total length of fiber, which is zooming in to the measured optical meander. Measured optical meander is located in the section between 60 m to 63 m.

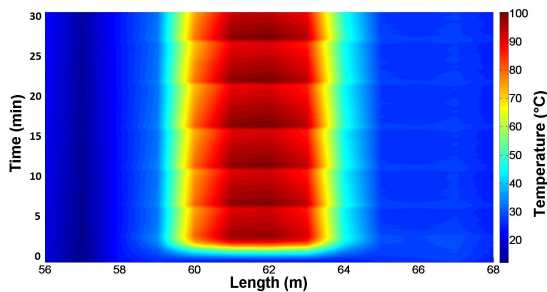


Fig. 5. The temperature measurement for one optical meander for a 30 min. X-axis represents the total length of fiber, which is zooming in to the measured optical meander. Measured optical meander is located in the section between 60 m to 63 m.

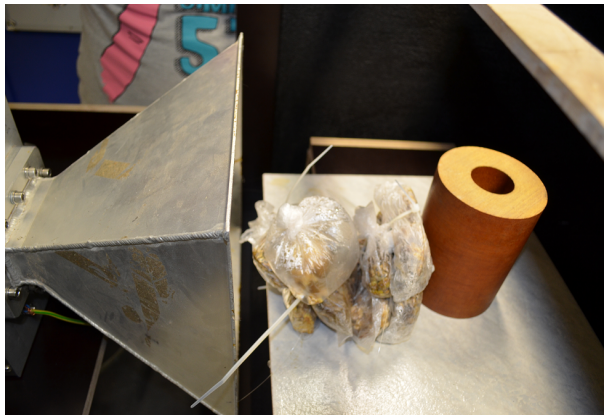


Fig. 6. The temperature measurement for two optical meanders.

Due to the time demands were the last two measurements (120 and 240 min. time intervals) merged. In this measurement were used two optical meanders. Samples were

placed one after another and heated together (Fig. 6). During measurement the first tested sample has been losing its original shape and was then less heated. Therefore, the operation of microwave emitter had to be longer. However, the second sample was heated much more. The temperature of second sample was reaching the values up to 115 °C and the sample had to be constantly monitored. After 120 min. was one sample taken and measurement continued with only the second sample (Fig. 7).

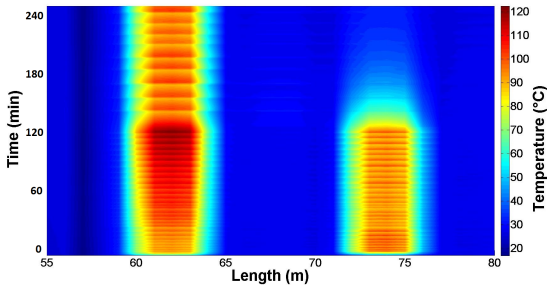


Fig. 7. The temperature measurement for two meanders. X-axis represents the total length of fiber, which is zooming in to the measured optical meanders. Measured optical meanders are located in the section: first between 71,5 m to 74,5 m and second between 60 m to 63 m.

After the measurement were the samples received by the researchers after from The Institute of Microbiology, ASCR, v. v. i. for the next research and measurements, which had to confirm or contradict the extermination of the *Serpula Lacrymans* in the different time interval measurements.

The effect of high temperature on the fungal viability was tested by re-inoculation of treated samples on agar plates. No growth of the fungus was observed even after 5 min of microwave treatment. The experiments described here demonstrated the ability of microwaves to kill efficiently mycelial cultures of *Serpula Lacrymans*. However, the results obtained with oat flakes can not be generalized and applied to other materials such is wood or timber.

3 Conclusion

The aim of these measurements was to determine the functionality of using optical meanders to monitoring the temperature inside the tested samples during extermination of *Serpula Lacrymans* with microwave radiation. After successfully completing these experiments was ascertained the unique and reliable utilization of optical meanders and DTS in processes of extermination of *Serpula Lacrymans* by use of microwave radiation and that because of measured high accuracy of temperature and resistance to electromagnetic radiation. Given the current range of materials used in the sensors this method appears as the most reliable and most accurate.

Additional measurements and applications that we can use DTS resistance consisting in application of optical meanders on a wooden beam. These beams will be heated

up in two ways. In the first case it will be heated by hot air in the reconstruction process of old buildings. In the second case it will be heated by microwave emitter in the reconstruction process of old beams in laboratory. Our research team also done another measurements such as in terrain and the also at the research workplace in Zvolen.

References

1. MAURICE, S., L. COROLLER, S. DEBAETS, V. VASSEUR, G. LE FLOCH a G. BARBIER. Modelling the effect of temperature, water activity and pH on the growth of *Serpula lacrymans*. *Journal of Applied Microbiology*. 2011, vol. 111, iss. 6, pp. 1436–1446. DOI: 10.1111/j.1365-2672.2011.05161.x.
2. SCHMIDT, Olaf. Indoor wood-decay basidiomycetes: damage, causal fungi, physiology, identification and characterization, prevention and control. *Mycological Progress*. 2007, vol. 6, iss. 4, pp. 261–279. DOI: 10.1007/s11557-007-0534-0.
3. WATKINSON S.C. and D.C. EASTWOOD. 2012. *Serpula lacrymans*, wood and buildings. *Advances in Applied Microbiology*. 2012, vol. 78, pp. 121–149. ISBN 978-0-12-394805-2.
4. KOUDELKA P., B. PETRUJOVA, J. LATAL, F. HANACEK, P. SISKÁ, J. SKAPA and V. VASINEK. Optical fiber distributed sensing system applied in cement concrete comixture research. *Radioengineering*. 2010, vol. 19, no. 1, pp. 172–177. ISSN 1210-2512.
5. LONG D. *The Raman Effect: A Unifield Treatment of the Theory Raman Scattering by Molecules*. New York: Wiley, 2002, pp. 598. ISBN 978-0-471-49028-9.
6. BALL D.W. Theory of Raman Spectroscopy. *Spectroscopy*. 2001, vol. 16, iss. 11. ISSN 0887-6703. Available at: <http://spectroscopyonline.findanalytichem.com/spectroscopy/data/articlestandard/spectroscopy/442001/836/article.pdf>.
7. RODERS A. Distributed optical-fibre sensing. *Measurement Science and Technology*. 1999, vol. 10, iss. 8., ISSN 1361-6501. DOI: 10.1088/0957-0233/10/8/201.
8. KOUDELKA P., J. LATAL, J. VITASEK, J. HURTA, P. SISKÁ, A. LINER and M. PAPES. Implementation of Optical Meanders of the Optical-fiber DTS System Based on Raman Stimulated Scattering into the Building Processes. *Advances in Electrical and Electronic Engineering*. 2012, vol. 10, no. 3, pp. 187–194. ISSN 1336-1376.
9. KOUDELKA P., A. LINER, M. PAPES, J. LATAL, V. VASINEK, J. HURTA, T. VINKLER and P. SISKÁ. New Sophisticated Analysis Method of Crystallizer Temperature Profile Utilizing Optical Fiber DTS Based on the Stimulated Raman Scattering. *Advances in Electrical and Electronic Engineering*. 2012, vol. 10, no. 2, pp. 106–114. ISSN 1336-1376.
10. DEDEK L. and J. DEDEKOVA. *Elektromagnetismus*. 2nd ed. Ed. Brno: VUTIUM, 2000, pp. 232. ISBN 80-214-1548-7.
11. HALLIDAY D., R. RESNICK and J. WALKER. *Fyzika: vysokoskolska ucebnice obecne fyziky*. 1st ed. Ed. Praha: Prometheus, 2000, xxiv, pp. 1198. ISBN 80-214-1869-9.

SOM Classifier for Speech Stress Detection

Pavol Partila, Jaromir Tovarek, and Miroslav Voznak

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{pavol.partila, jaromir.tovarek, miroslav.voznak}@vsb.cz

Abstract. This article deals with a speech emotion recognition system. We discuss the usage of a neural network as the final classifier for human speech emotional state. The introduction explains why the extraction of the emotional state of a person is an important topic in the field of speech processing. We carried our research on a database of records of both genders and various emotional states. In the preprocessing and speech processing phase, we focused our intent on parameters dependent on the emotional state. The output of this work is a system for classifying the emotional state of a man's voice, which is based on a neural network classifier. The statistical analysis focused mainly on the usability of the calculation emotional state was used to verify the parameters significance. The article compares the accuracy of the classifier, using different sets of parameters to train and test the classifier.

Keywords: Fundamental frequency, Central Clipping, SOM, Stress, Speech.

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.

2 Pre-Processing

Speech signal is stochastic by nature. However, a speech signal has a number of characteristics that may be unwanted during processing. This chapter describes a process called pre-processing that is an important part of the digital speech signal processing.. These few steps prepare the signal for subsequent extraction of signal parameters. The values of these parameters could be wrong without the pre-processing process. Figure 1 shows this pre-processing.

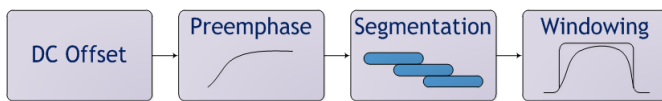


Fig. 1. Pre-Processing Process

Train respectively testing is the key step of the system. The quality of the input data, the audio signal in this case, has a direct impact on the classification accuracy. For this reason, it is used the Berlin database containing over 500 recordings of actors consists men and women. The database contains 10 sentences in the seven emotional states. This corpus of recordings is considered as a high-quality database, because it was created by professional actors in the sound studio. As mentioned, the speech signal has to be modified by routine operations such as removing the DC component, pre-emphasis and segmentation stochastic signal into quasi-periodic frames.

3 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. Human speech consists of voiced and unvoiced speech sounds. 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 F_0 can be calculated only from the voiced parts. Each method to calculate F_0 has advantages and disadvantages. This article compares four methods and their use in a self-organizing feature map classifier.

3.1 Auto-correlation function

This simple method can be used where we have a full signal. The auto-correlation function is defined in Equation 7.

$$R(m) = \sum_{n=m}^{N-1} s(n)s(n-m) \quad (1)$$

Lag k is determined by the position of maxima. The fundamental frequency of vocal cords is calculated using Equation 12.

$$k = \operatorname{argmax} R(m) \quad (2)$$

$$F_0 = \frac{F_{\text{sampling}}}{k} \quad (3)$$

3.2 Normalized cross-correlation function

The difference in energy shifted and original framework is settled by normalization:

$$NCCF(m) = \frac{\sum_{n=m}^{N-1} s(n)s(n-m)}{\sqrt{E_1 E_2}} \quad (4)$$

Where E_1 is energy of original frame and E_2 is energy of shifted frame. Lag and F_0 are calculated. F_0 extraction is performed just as in the ACF (Eqs. 11 and 12) [5].

3.3 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.

$$P(i) = \alpha \min(\text{MAX}_{i-1}, \text{MAX}_{i+1}) \quad (5)$$

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.

3.4 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 . The sum of subharmonic amplitude is described in Equation 13. Subharmonic-to-harmonic ratio can be obtained by dividing SS and SH .

$$SS = \sum_{n=1}^N A\left[\left(n - \frac{1}{2}\right)f_0\right] \quad (7)$$

4 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. There are generally two-dimensional maps of neurons. 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.

$$w_{ij}(t+1) = w_{ij}(t) + \gamma(x_j(t) - w_{ij}(t)) \quad (8)$$

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 hard-ware. This should be taken into account as these days real time speech processing is sought after in the entire telecommunications services sector.

5 Results

The system which was designed in software Matlab finally consists of three main blocks. These have been described above. Figure 5 shows the block diagram. The in-

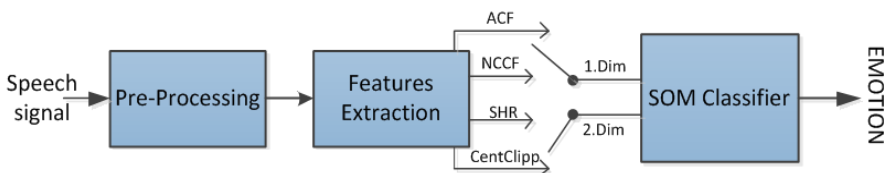


Fig. 2. Speech emotion recognition system

put 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. Pairs of SOM Classifier Inputs

Input	ACF-NCCF	ACF-SHR	ACF-CentClip	NCCF-SHR	NCCF-CentClip	SHR-CentClip
Resolution	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 below.

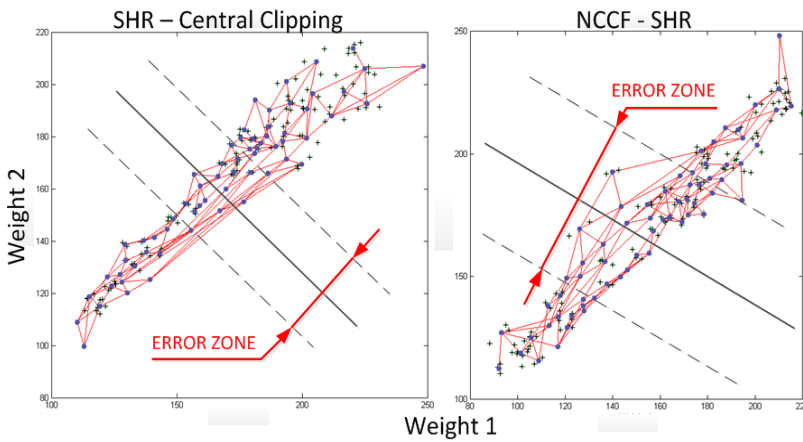


Fig. 3. Trained self-organizing maps with error zones.

6 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. The problem of delay is unacceptable for telecommunications services and applications that require real-time access. The block algorithm for pitch extraction is designed to calculate F0 using four methods. 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. Voznak, M., Rezac, F., Rozhon, J. Speech Quality Monitoring in Czech National Research Network. *Advances in Electrical and Electronic Engineering*. 2010, vol. 8, no. 5, pp. 114-117. ISSN 1804-3119
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*. 2012, vol. 10, no. 4, pp. 270-275. ISSN 1804-3119.

OFDM and QAM Modulation Under Different Weather Conditions

Frantisek Perecar, Jakub Jaros, Andrej Liner, and Vladimir Vasinek

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{frantisek.perecar, jakub.jaros}@vsb.cz

Abstract. The FSO (Free Space Optics) communication uses the visible or infrared light for transmission. As well as cable optics FSO also uses laser for the data transmission, but the data flow is not transmitted in the fiber but in the air. Nowadays, the most used modulations are QAM (Quadrature Amplitude Modulation) and OFDM (Orthogonal Frequency Division Modulation) [1]. OFDM belongs to the border group of modulations with more carrier waves, where the informations are transmitted via subcarrier waves with lower data flow and baud rate. It is mainly used in broadband wire and wireless communications. This paper deals with the modulations used in FSO. Most used modulation in FSO is OOK (On-Off Keying), but modulations OFDM and QAM are (modulation) ways of the future. The main task was to determine how much is the reach of modulations changing with the changes of density (visibility) of fog and the set transceiver power. As software environment for simulations has been used OptiSystem program. For the simulation of the atmosphere the FSO component has been used. In this component were simulated attenuations, which are responding to varying densities of fog [2]. At different intensities of fog were changing the received power.

Keywords: Free space optics, fog, modulation, OFDM, QAM, simulation.

1 Introduction

After age of revitalization of wireless networks and with advancing 4G networks, fibrous optics seems like fine competitor to frequency radio links. This method brings new opportunities for unlicensed use of spectrum without electromagnetic interference and increases the safety and speed of transmission. FSO is a innovation in optical communication, because signal spreads faster in the air than in glass, so it use the full speed of light [3]. It uses full duplex and throughput of Gigabit Ethernet. Is it enables to send 1.25 Gbps of data, voice and video simultaneously through the air. In the future, specialist count with a transfer rate of 10 Gbps. FSO can be placed behind windows, which eliminates need for costly rent of roof space. FSO technology is most suitable form of building-to-building communication. FSO is highly depending on the atmospheric effects such as clouds, snow, rain, because such condition increase the loss of power during transmission. Fog is causing biggest loss from all atmospheric effects. It is because elements of fog are extremely small compared with the wavelength and its

causing Rayleigh scattering. Fog is even worse than rain because elements of rain are larger than the wavelength [4], [5].

2 OFDM Modulation

FSO-OFDM systems support high bit rate departments of large data stream from the small number of data streams and transmit it through many narrowband subcarriers waves. Narrowband subcarrier data reach lower distortion than high-speed data and they do not need any settlement. The required signal processing is mostly done in RF field. This is beneficial because microwave ovens are much more developed than the optical frequency and because of microwave filters selectivity, and frequency stability of the microwave oscillator is substantially better than that of the optic. In addition, the level of phase noise microwave oscillators is significantly lower than that of the DFB laser diode, which means that the RF coherent detection is simpler to implement than coherent optical detection. By this we can apply the most advanced coherent modulation formats already developed for cable-free communication [7].

10 Gb/s data stream is demultiplexed into four 2.5 Gb/s streams, and each is encoded by the same LDPC encoder. Transmitted OFDM signal can be written in the following equation:

$$s(t) = S_{OFDM} + D, \tag{1}$$

$$S_{OFDM} = Re \left\{ \sum_{k=-\infty}^{\infty} w(t - kT) \sum_{i=-N_{FFT}/2}^{N_{FFT}/2} X_{i,k} \cdot e^{j2\pi \frac{i}{T_{FFT}}(t - kT)} \cdot e^{j2\pi f_{RF}t} \right\}, \tag{2}$$

is defined for: $kT - T_G/2 - T_{win} \leq t \leq kT + T_{FFT} + T_G/2 + T_{win}$.

In the above expressions $X_{i,k}$ denotes the k -th OFDM symbol in the i -th sub-carrier frequency, $w(t)$ is the window function and f_{RF} is the RF carrier frequency. The duration of an OFDM symbol is referred to as T , whereas T_{FFT} is the sequence FFT duration, T_G is the guard interval duration (duration cyclic elongation), and T_{win} is the length of the window interval. Details of the OFDM symbol are shown in Fig. 1 and Fig. 2. These symbols are made up as follows: $N_{QAM}(=k)$ consecutive input QAM symbols are "zero-padded" gain $N_{FFT}(=2^m, m > 1)$ input samples for inverse FFT, then the N_G samples is created and guard interval T_G finally OFDM symbol is multiplied by the window function [6].

3 QAM Modulation

QAM modulation is a compound modulation which uses symbols to create a combination of (ASK) amplitude and phase shift keying (PSK). Each state is represented by a certain value of amplitude and phase. It is actually a multistate modulation which is able to transmit n bits using the symbols m , it means that at any time more bits are transferred at once. Quadratic amplitude modulation is mainly used to transfer data with a narrow transmission spectrum. QAM modulation saves bandwidth, or vice versa, with the same

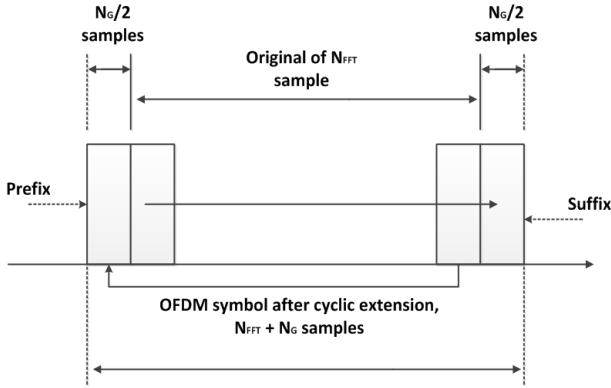


Fig. 1. FSO-OFDM system: cyclic extension of the OFDM symbol.

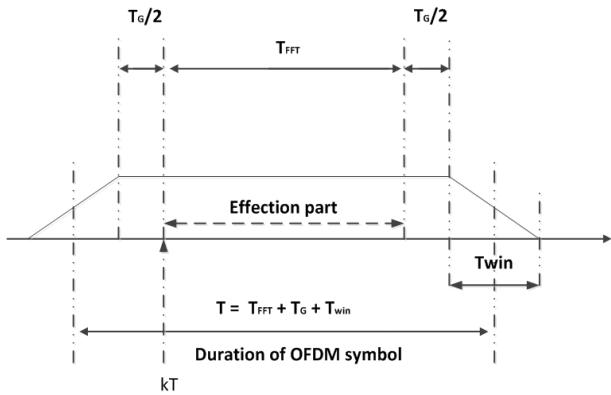


Fig. 2. FSO-OFDM system: OFDM symbol after the window (windowing).

bandwidth can increase transfer rate. Using more states, signals are more prone to signal interference and it makes it more difficult to identify the symbol [8], [9].

Relationship between states and symbols are defined as follows:

$$n = \log_2 m. \tag{3}$$

To transmit one bit it is needed either two states $\log 0$ or $\log 1$. If you want to transfer more bits per symbol you need more states. For example, to transfer 4 bits we need $2^4 = 16$ states (16-QAM). The following formula applies to the n bits we need m of modulation states according to the formula:

$$m = 2^n. \tag{4}$$

4 Simulation and Result

Simulation of fog environment was implemented in software OptiSystem 11.0, which allows users to plan, test and simulate optical coupling and transmission systems of

modern optical networks. On the FSO channel was set aperture diameter telescope transmission with 5 cm diameter aperture receiving telescope with 30 cm and wavelength of 1550 nm. Attenuation was gradually changed to kilometer, depending on the type of fog and distance of the transmitter from the receiver in the range of 50 m to 2 km. The measurements were repeated to set the power level of the CW lasers to 5, 10, 15, 30 and 40 dBm. Measurements were made for OFDM modulation with 512 subcarriers waves and 16-QAM modulation.

4.1 Dense Fog, Visibility 50 m, Attenuation 315 dB·km⁻¹

As indicated above, then the power level of the CW laser is set to 5, 10, 15, 30 and 40 dBm. The biggest setback occurred in dense fog which has a depression 315 dB·km⁻¹. Attenuation values used in the simulation were based on the model of Kim, created by Mr. Isaac I. Kim. To a distance of 50 m, a signal was usable for transmission of information. After 200 m it does not recognize the symbols OFDM and 16-QAM modulation. To achieve greater distance to transfer information, large transmission power had to be set. The fair value of the transmitted power in practice is in the values of 8–20 dBm.

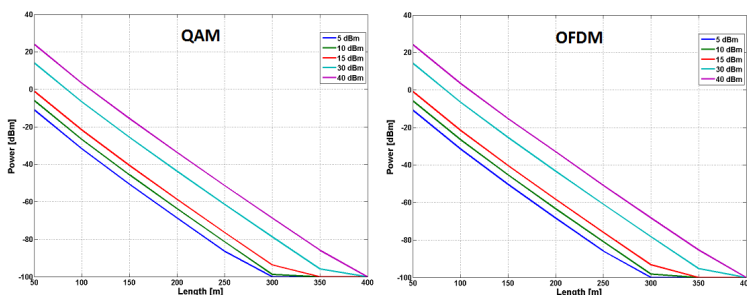


Fig. 3. Graphs of relativity of decrease in the level of the signal power at the distance and the type of modulation in dense fog.

Graphs (Fig. 3), the bottom in performance of OFDM and QAM modulation was similar. Performance can be regarded as a more usable in about -20 to -25 dBm range. Performance level after decay by fog decreased almost linearly. At power level of 40 dBm, the maximum distance where signal could be measured was 400 m, but at 5 dBm it was only 300 m. In simulations of dense fog, it was found that the type of used modulation has minimum affects on the quality of the transmission.

4.2 Medium Fog, Visibility 500 m, Attenuation 28.9 dB·km⁻¹

As noted above, the power level on CW laser was set to at 5, 10, 15, 30 and 40 dBm. In an environment with medium fog decrease of 28.9 dB·km⁻¹ signal reached a greater distance. Attenuation in medium fog in comparison to a dense fog decreased. With power level set at 40 dBm, signal coverage reached up to 1500 m.

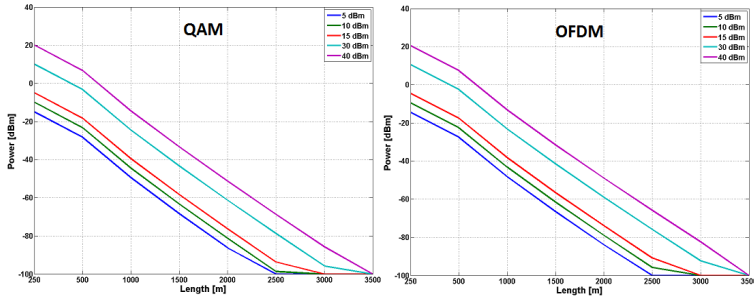


Fig. 4. Graphs dependent decrease in the level of the signal power at the distance of P and the type of modulation at moderate fog.

Graphs (Fig. 4) shows that even at medium fog decrease in the level of performance OFDM and QAM modulation is very similar. Level of performance for both types of modulations at a distance of 2000 m was very subdued and the signal will be unusable for the transmission of information. Decrease the power level was proportional to the distance.

4.3 Light Fog, Visibility 1000 m, Attenuation 14 dB·km⁻¹

As noted above, the power level at CW laser was set to 5, 10, 15, 30 and 40 dBm. In light fog, we have experienced the smallest loss at 14 dB·km⁻¹. At preset power level of 5 dBm signal could be received even 1000 m away, but was too weak -33.349 dBm and the symbols were no longer recognized. At the level of 40 dBm symbols were recognizable within 2800 m. Environment with light fog poses a problem for free-space optical network for the receiver at a distance of 2800 m from the transmitter.

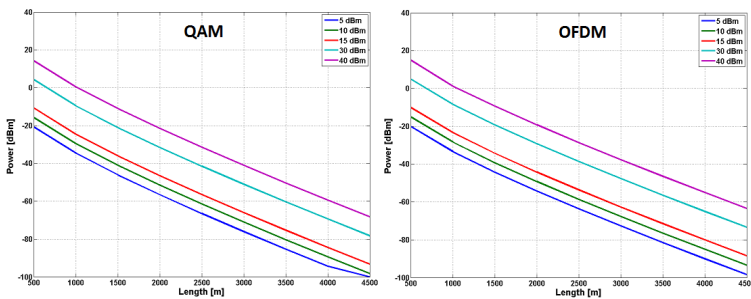


Fig. 5. Graph of the decrease in the level of the signal power at the distance of P and the type of modulation in weak fog.

No significant decrease due to the use of signal modulation occurred in light fog. Graphs indicates (Fig. 5) that the power levels could be measured at 4 km but power has been unusable for transferring information.

5 Conclusion

The aim of the team of authors of this article was to describe and implement simulation modulations and determine their impact on free-space atmospheric link in OptiSystem 11.0. Designed circuit for OFDM modulation with 512 subcarriers waves and 16-QAM modulation was set in this program. FSO component used for simulations, was set to wavelength of 1550 nm, corresponding attenuation environment under fog, distance from the transmitting after receiving part, diameter telescope transmission of 5 cm and diameter receiving telescope of 30 cm. CW laser was set to the power level of 5, 10, 15, 30 and 40 dBm. Decrease was measured in the level of performance for each power level after a signal passed certain distance through environment with fog. Optical power meter was placed just behind the FSO component. Level of fog affected signal quality, transmission speed and range. Comparison of OFDM and 16-QAM modulation simulations we have reached conclusion that the decrease of signal by fog intensity is minimally affected by the type of modulation, which is demonstrated charts and graphs. OFDM modulation suppresses multipath signal propagation better than modulation 16-QAM, while multistate QAM modulation allows to save bandwidth, or vice versa at the same bandwidth to increase bit rate.

References

1. Khvorostyanov, V. I. and Curry, J. A., "Fall Velocities of Hydrometeors in the Atmosphere: Refinements to a Continuous Analytical Power Law," *Journal of the Atmospheric Sciences* 62(12), 4343–4357 (2005).
2. Al Naboulsi, M., "Fog attenuation prediction for optical and infrared waves," *Optical Engineering* 43(2), 319–329 (2004).
3. Deng, P., Yuan, X. H. and Huang, D., "Scintillation of a laser beam propagation through non-Kolmogorov strong turbulence," *Optics Communications* 285(6), 880–887 (2012).
4. Borik, S. and Cap, I., "Measurement and Analysis Possibilities of Pulse Wave Signals," *Advances in Electrical and Electronic Engineering* 11(6), 514–521 (2013).
5. Hejduk, S., Witas, K., Latal, J., Vitasek, J., Bocheza, J. and Vasinek, V., "Simple and Universal Current Modulator Circuit for Indoor Mobile Free-Space-Optical Communications Testing," *Advances in Electrical and Electronic Engineering* 12(1), 66–74 (2014).
6. Latal, J., Bojko, M., Vitasek, J., Koudelka, P., Siska, P., Liner, A., Papes, M. and Vasinek, V., "Influence of thermal turbulences on the optical beam of atmospheric optical link." *Proc. 36th International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 138–143 (2013).
7. Zhou, G., Cai, Y. and Chu, X., "Propagation of a partially coherent hollow vortex Gaussian beam through a paraxial ABCD optical system in turbulent atmosphere," *Optics Express* 20(9), 9897–9910 (2012).
8. David, T., Latal, J., Hanacek, F., Koudelka, P., Vitasek, J., Siska, P., Skapa J. and Vasinek V., "Cross-sectional measuring of optical beam," *Advances in Electrical and Electronic Engineering* 9(4), 179–186 (2011).
9. Vitasek, J., Latal, J., Vasinek, V., Hejduk, S., Liner, A., Papes, M., Koudelka, P., and Ganiyev, A., "The fog influence on bit error ratio," *Proc. 18th Czech-Polish-Slovak Optical Conference on Wave and Quantum Aspects of Contemporary Optics* art. no. 86970L (2012).

Measuring Optimal Length of the Amplifying Fiber in Different Working Conditions of the Amplifier

Radek Poboril

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
radek.poboril@gmail.com

Abstract. The aim of this article is to highlight possible unwanted behaviour of EDFA optical amplifiers at temperature changes that act on them. At first it was necessary to build an EDFA amplifier. For this purpose we used a standard EDFA construction and the IsoGain I-6 as the amplifying optical fiber. The amplifier construction is described in the first part of the article. In pursuance of testing the amplifier and temperature instability, we focused on the temperature impact on the ideal length of the Erbium-Doped Fiber. This measurement is discussed in the second part of the article.

Keywords: EDFA, erbium, optical amplifier

1 Introduction

The crucial matter in optical telecommunication systems is the need to get over long distances between the endpoints of a topology. The optical signal from a transmitter must pass through an optical fiber and then must be detected by a receiver. However, we can encounter certain problems, the biggest one of which is the attenuation.

The attenuation arises in optical fibers as well as in other optical components such as couplers, splitters, AWGs, etc. It makes a signal very weak for its detection, that is why, the main objective in the optics in the last decade has been to find a method to regenerate the signal so it can be easily detectable. Devices performing such regeneration are regenerators, and also optical amplifiers [1].

Regenerators only receive the optical signal, convert it to the electric one, repair it (i.e. amplify it) and then convert it back. As for the latter, i.e. the amplifiers, they do the same, plus their main advantage, compared to regenerators, is their versatility. Optical amplifiers do not need know the network settings, modulation, bitrate etc. The systems which are using optical amplifiers can be easily upgraded [2].

Over the last few years, optical fiber amplifiers have experienced a huge development, especially as far as their properties are concerned. These are influenced not only by the material they are made of, but also by the way this material is used.

2 Concept of Erbium Doped Fiber Amplifier

In this chapter, we will focus on the construction of an amplifier and properties of the individual components.

As a model, we used a conventional erbium doped fiber amplifier schematic diagram [3] This schematic can be seen in Fig. 1.

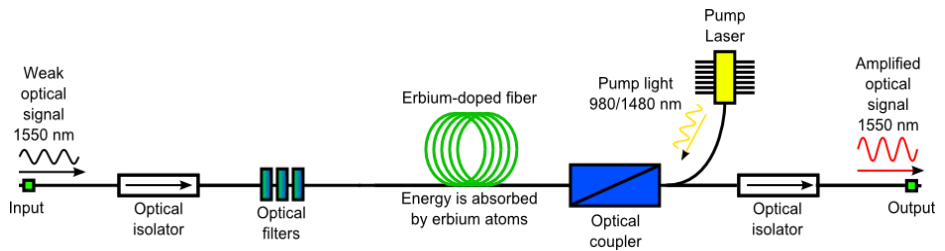


Fig. 1. Circuit of the constructed amplifier.

Instead of components used in standard amplifiers much bigger components were used, so it was impossible to achieve at least partial integration. For that reason, the entire amplifier was placed on a mounting plate covered by a soft foam protecting individual components and optical fibers.

Since it was expected from the beginning that it would be necessary to move the amplifier from one place to another, we built a two-storey construction that is easy to manipulate and at the same time "component-friendly". The construction can be seen in Fig 2.

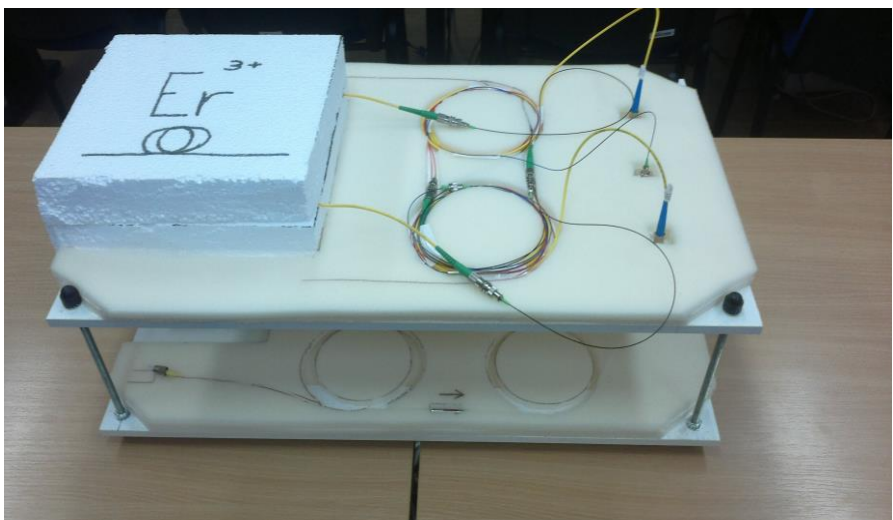


Fig. 2. Construction of the erbium doped fiber amplifier.

3 Measuring Optimal Length of the Amplifying Fiber in different Working Conditions of the Amplifier

For experiments were selected three basic temperature conditions for the Fiber – 1°C, 25°C and 50°C - because we wanted to achieve three basic states: hypothermia, standard state and hyperthermia. According to the maker's datasheet, the ideal length of fiber is 13m. Since we had 20m at our disposal, was decided to start measuring on 20m and then successively cut off 2m long pieces and, thus, measure with 20m, 18m, 16m, 14m, etc. of the fiber. Another step consisted in splicing pigtailed to both ends of the fiber. Particular emphasis was placed on maximum precision of the work so that each weld had the lowest possible attenuation. Due to this accuracy, the highest attenuation achieved was 0.01dB. The spliced fiber was then put in a plastic foil and carefully sealed in order to prevent penetration of water to the fiber when it was placed in a temperature bath. Along with the fiber there was also a fiber thermometer PalmSence FTC-PALM-ST. Figure 3 illustrate placing the fiber in the thermal or snow bath. The disadvantage of this solution was condensation of water inside the plastic foil. For that reason the measurement had to be performed quickly.



Fig. 3. Testing erbium doped fiber IsoGain I-6 at 1°C temperature and 50°C

The following charts (Fig: 4,5,6) show the ideal fiber length at the pumping power of 10mW, 50mW and 125mW. As for the amplified signal was chosen the signal with the 1553nm wavelength and 1mW power. It is obvious from the chart that the power for sufficient excitation of erbium ions is up to 125mW. At lower powers the fiber was attenuated or amplified only partially. The optimal fiber length at all three temperatures was 16 m. The conclusion resulting from the measurements is that the change in temperature of the fiber did have an impact on the power of the amplifier. The lowest power of the amplifier could be observed at 50 °C. The change in temperature causes a change in the stimulated emission cross section: with the temperature increase the cross section of erbium decreases, which affects the overall gain of the amplifier. The measurement also shows that the temperature stress has not much influence on the change of the spectrum of the output signal of the amplifier as such. This fact can be caused by low resolution and spectral range of the used spectrometer. In fact, power fluctuations can also be caused by moving the energy among parts of the spectrum that the spectrometer possibly have not measured at all [4].

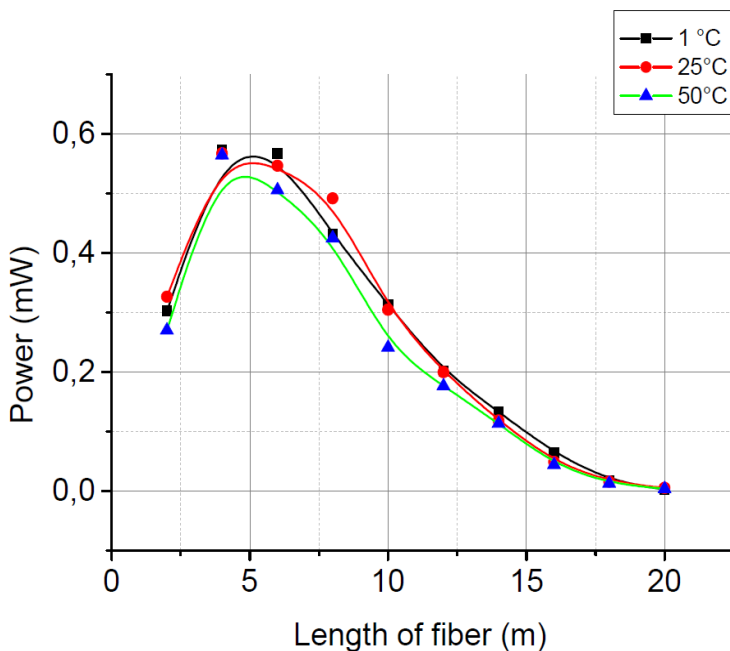


Fig. 4. Measuring different lengths of doped fiber at 10 mW pump power and dissimilar temperatures

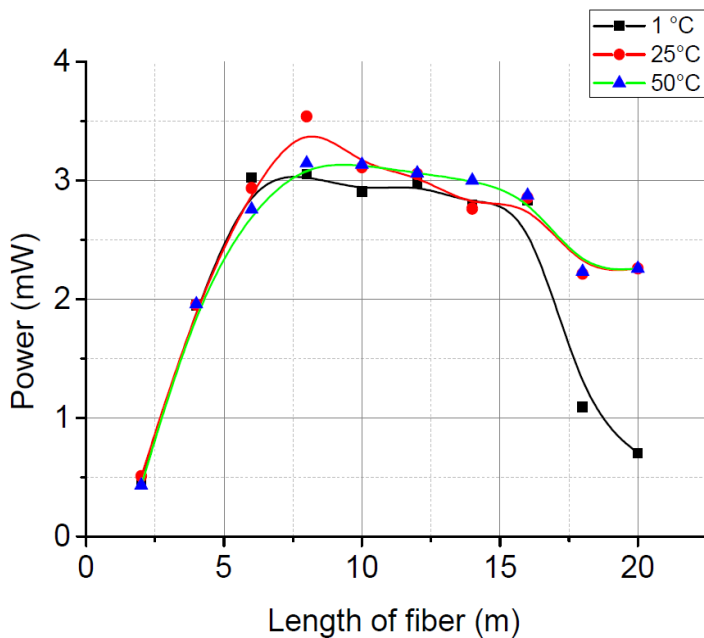


Fig. 5. Measuring different lengths of doped fiber at 50 mW pump power and dissimilar temperatures

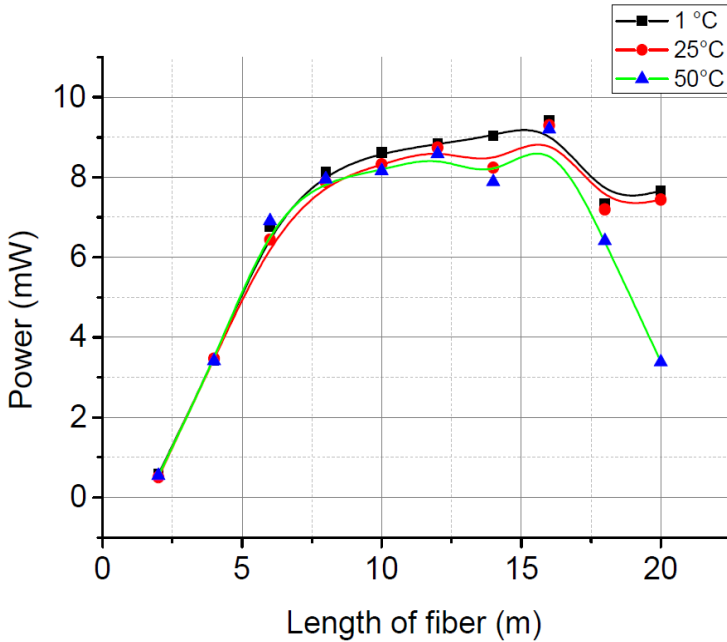


Fig. 6. Measuring different lengths of doped fiber at 125 mW pump power and dissimilar temperatures.

4 Conclusion

The aim of this article was to highlight negative impacts on erbium doped fiber amplifiers. Since the EDFA amplifiers represent one of the most important components used in optical WDM networks, it is necessary to keep improving them constantly. The main thing to test was an ideal length of the fiber in different temperature conditions. This measurement has not confirmed that the temperature fluctuation has any significant effect on the ideal fiber length. However, the measured results clearly say that temperature influences on the output power of the amplifier which is directly dependent on the state of the pumping laser. What can be read from the charts is e.g. amplifying fiber characteristics in various power levels of the pumping source. The last chart concerning the 125 mW pumping power indicates the influence of temperature on the power of the amplifier depending on the amplifying fiber length. At all the testing temperatures, the optimum fiber length was 16 m, which corresponds to higher pumping powers and partly also to theoretical assumptions based on the datasheet provided by the fiber manufacturer.

References

1. KOZAK, M., B. JAUMARD and L. BOHAC. On Regenerator Placement in Loss-less Optical Burst Switching Networks. In:36th International Conference on Telecommunications and Signal Processing. Piscataway: IEEE, 2013, pp. 311-315. ISBN 978-1-4799-0403-7.
2. RAMASWAMI, Rajiv; N SIVARAJAN, Kumar and Galen H SASAKI. Optical networks: a practical perspective. 3rd ed. Amsterdam: Morgan Kaufmann, 2009, xxxiv, Morgan Kaufmann series in networking. ISBN 978-0-12-374092-2.
3. BECKER, P., OLSSON, N and SIMPSON, J. Erbium-doped fiber amplifiers: fundamentals and technology. San Diego: Academic Press, c1999, xv, ISBN 01-208-4590-3.
4. KOROSHETZ J., B. BRICKEEN, S. FAHR, A. RAPAPORT and M. BASS, Athermal emission in Yb,Er:glass. In:Opt. Express 15 pp. 11530-11535 ,2007.

Monitoring of Speech Quality in full-mesh Networks

Jan Rozhon

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.rozhon@vsb.cz

Abstract. The quality of speech in network environment is a growing concern of all the companies and institutions that migrated or are planning to migrate their voice communications to IP based technologies. This trend together with the increasing amount of Internet traffic in general poses a great challenge for network administration to be able to configure and manage the network so it can carry the multimedia traffic without the excessive delays, which degrade the quality of speech as it is perceived by the end users. Continuous monitoring of key nodes of internal infrastructure as well as external interconnections is the possible way to increase the quality of multimedia transmissions for most users, because it allows the network administrators to be informed as soon as the problem rises and consequently to change the network routing and queuing policies to bypass the connection experiencing the quality issues. The tool for this type of monitoring and its architecture are described in detail in this paper.

Keywords: Network Administration, Network Policing, PESQ, Speech Quality Monitoring

1 Introduction

Past decade was marked by two important factors from the multimedia communication point of view. The first one, a rapid development of the Voice over IP (VoIP) technology in the business communications was an aspect of the long-term trend of voice and data networks convergence. This trend was driven by the both economic and network management reasons. VoIP technologies, no matter whether commercial or open-source, have been a mean to achieve higher efficiency of intra-organization collaboration by establishing a platform for in given time enterprise level services like conference, joint calendar events, video telephony, call transfer functions, etc. From the economic perspective, the single infrastructure advantage and inherent VoIP features such as performing a least cost routing (LCR) or callback function catalyzed the ongoing process even more, which led to a broad VoIP adoption across all the levels of private and public institutions.

Alongside this fundamental transition between technologies, the Internet multimedia content began to form the most of the Internet content in general. This resulted in

heavy burden laid on the back of the network infrastructure, which had to be capable of delivering the content not only reliably and securely, but with given quality parameters as well. Through systematic analysis and upgrades the infrastructure was in many cases made robust enough to handle this load. However, under some special circumstances (device malfunction, traffic peak, etc.) the transmission parameters can temporarily deteriorate. These special occasions can significantly decrease the multimedia service rating by the customers, but can be countered easily as well. To counter the effects of these anomalies many routing and queuing policies can be employed, but to be able to do that efficiently the anomalies need to be identified, which in terms of speech quality (or multimedia quality in general) cannot be achieved directly on network elements, such as routers or switches.

One-time or irregular measurements cannot address this problem, but the continuous monitoring of speech quality can achieve the benefits being sought like already mentioned quick or even immediate notification about the problem or in some cases even the speech quality prediction.

This paper deals with the tool for the aforesaid periodic quality measurements in the full-mesh IP networks, which is designed to use the industry standard for speech quality calculation – PESQ (Perceptual Evaluation of Speech Quality) as well as the state-of-art monitoring and call generation tools available. In this paper the basic idea, functionality and architecture of this system will be presented and the planned future development of the project outlined.

2 State of the Art

The introductory part of this paper mentioned one of the most used algorithms for evaluation of speech quality - PESQ, but this is just one representative of the extensive group of methods to measure and evaluate speech quality. These methods can be subdivided into two groups according to the approach applied, conversational and listening.

Conversational tests are based on mutual interactive communication between two subjects through the whole transmission chain of the tested communication system. These tests provide the most realistic testing environment but they are very time consuming. Listening tests do not provide such plausibility as conversation tests but are recommended more frequently. According to methods of assessment, speech quality evaluation methodologies can be subdivided as subjective methods and objective methods. To evaluate speech quality, MOS (Mean Opinion Score) scale as defined by the ITU-T recommendation P.800 is applied [1]. In order to avoid misunderstanding and incorrect interpretation of MOS values, ITU-T published ITU-T recommendation P.800.1 [2]. This recommendation defines scales for both subjective and aforementioned objective methods as well as for aforementioned individual conversational and listening tests.

The subjective evaluation methods are based on evaluation by human beings (listeners), i.e. subjects, which makes them unsuitable for real-time applications.

The objective evaluation methods substitute the necessity to involve humans in the testing by mathematical computational models or algorithms. Their output is again a MOS value or, depending on the algorithm applied, a different value which can be transferred to a MOS value using a suitable mapping function. The aim of objective methods is to estimate, as precisely as possible, the MOS value which would be obtained by a subjective evaluation involving sufficient number of evaluating subjects. Objective testing's exactness and efficiency is therefore a correlation of results from both subjective and objective measurements. Objective methods can be subdivided into two groups, Intrusive and Non-intrusive.

The core of intrusive (also referred to as input-to-output) measurements is the comparison of the original sample before releasing it into a transmission chain of a communication system with the output sample, transmitted through the system (degraded).

This type of testing includes, among other, the following methods: PSQM (Perceptual Speech Quality Measurement), PAMS (Perceptual Analysis Measurement System) developed by British Telecommunications, P.OLQA (Perceptual Objective Listening Quality Assessment) and PESQ (Perceptual Evaluation of Speech Quality) [3]. P.OLQA may become the successor of PESQ because avoids weaknesses of the current P.862 model and has an extension towards higher bandwidth audio signals. Yet, PESQ is the most common objective intrusive method. Computational technique applied by this method combines PAMS' robust temporal alignment techniques and the PSQM' exact sensual perception model. Its final version is contained in ITU-T recommendation P.862 [3]. As was stated above, the principle of this intrusive test is based on comparison of original and degraded signals, their mathematical analysis using FFT and interpretation in the cognitive model.

The non-intrusive methods do not require the original sample. This is why they are more suitable to be applied in real time. On the other hand, however, much less accurate results can be obtained by using these methods, which is why this paper focuses on the implementation of the intrusive methods.

3 System Architecture

From the previous chapters it is clear why and what algorithm was chosen for the speech quality calculation. However, the algorithm itself is just a small piece in a greater picture of the system architecture necessary to provide the full set of functionalities needed for continuous monitoring. To fully understand the system architecture, the necessary background has to be provided.

In extensive networks, the individual elements are monitored on the network level, but in some cases, they can influence the quality of speech as it is perceived by end users even without the sign of a problem on the network level. This is due to the specific features and requirements of the speech (or multimedia in general) traffic. Among others the low and stable latencies and small number of lost packets can be named. From the aforementioned two necessary features of the monitoring system are implied. First, to be able to discover and identify the problem on the application level,

the application level measurements need to be implemented. Second, the numerous network probes need to be deployed, so that all the key multimedia interconnections can be observed.

The term probe itself is very important from the perspective of the system architecture. Due to the fact, that it is expected for the probe to be connected with many other probes in many different locations, the number of function the probe performs must be limited. The reason for this is economic, if the functionalities of the probes were vast, the system resources, it would require, would render the whole system economically inefficient. To counter this, the probes come in form of low-cost hardware computer and the sole purpose of theirs is to initiate, terminate and record the call. When deployed the probe is tapped to the network of the hosting institution. The probe then communicates with other partner probes in the monitored network using standard VoIP protocols, namely SIP for signaling and RTP for media.

Two pieces of software were used for the probe functionalities. The Asterisk PBX for call generation and recording and OpenSSH/Rsync for the security and data transfer. Both programs are open-source and are time-proven with large community of users, which makes them excellent for deployment from the economic and future development point of view.

To make the system fulfill its role, the central element is also required to supplement the probes. This element - a server - provides following features to the system:

- The interface for the user to manage the probes and inspect the results
- The interface for the probe to register automatically,
- System access control and encryption of data communication,
- PESQ calculation,

The server communicates with the probes using a secure connection based on SSH tunnels. Within these tunnels the information about all the probes, the measuring intervals and probe registration credentials are exchanged. The probe cannot access the server without the SSH tunnel, which can only be established with correct private keys. This mechanism allows control of access for the individual probes and secures the connections between probes and the server. The second role of the server is to calculate the MOS score using the PESQ algorithm. To be able to handle the number of PESQ calculations, the server needs to be of high performance and should therefore be implemented as the virtual computer with high-end hardware resource parameters. Virtualization offers two key aspects - the existing infrastructure can be used and the server backup can easily be created and deployed in case of hardware malfunction. The server also provides the means for the result visualization and probe management to the system administrator.

For the server to be able to perform the abovementioned tasks, the well-tested and stable open-source software was used. Upon Debian linux, OpenSSH and Zabbix are deployed. While the former provides the security layer, the latter forms the user-system interface to present the results to the administrator and to allow him to manage the probes and to control the monitoring process. Both the programs share the active community of users and developers, which makes them safe for long term deployment even in critical systems.

The entire system architecture in the example deployment is depicted in the Fig.1.

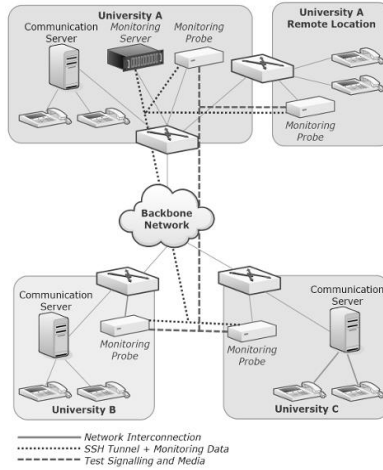


Fig. 1. Overview of a system architecture.

4 Conclusion

In the limited space of this paper the idea and proposition of the system for continuous monitoring of speech quality was presented. This monitoring system introduces application layer monitoring of the speech-oriented multimedia infrastructure, which makes it possible to detect the irregularities and anomalies of the network on the application layer, thus allowing for faster and more appropriate response to the situation in the form of routing and queuing policy changes. This system is based on PESQ algorithm to provide the high accuracy in comparison to subjective methods for speech quality measurement. This algorithm works in conjunction with broadly used open-source applications such as Asterisk PBX and Zabbix.”

The system is intended for the non-commercial operation in the network of the academic institutions, but thanks to its robust design it can easily be used in any environment, for example to allow the private organizations to monitor the stability of their VPN connections. The full-mesh operation of the system is a possibility but not the necessity, because the probes can be instructed to test only certain interconnections. This is done by splitting the probes into the separate groups in the monitoring system, which results in a subset of the probes to be propagated from the monitoring system to the individual probes, thus allowing to monitor only certain interconnections.

The contribution of the paper is an integration of speech quality evaluation into common monitoring system. Our design of monitoring probes was implemented and verified in testbed. This testbed was created in the experimental environment with the custom defined traffic and distortion generation. During the one month period of testing the visible correlation between network impairments and resulting speech quality was observed, thus confirming the expected behavior of the system.

The future development will be marked with an increased effort for transition from the usage of PESQ algorithm to the neural network based model for speech quality prediction to further decrease the resources needed by the system and to increase the responsiveness of the system as the whole.

Acknowledgements

This work was supported by project of Moravian-Silesian Region No. 02540/2013/RRC and by internal grant of Faculty of electrical Engineering and Computer Science SGS No. SP2014/72.

References

1. ITU-T P.800 – Recommendation P.800 of the International Telecommunication Union: *Methods for subjective determination of transmission quality*, ITU-T, 1996.
2. ITU-T P.800.1 - Recommendation P.800.1 of the International Telecommunication Union: *Mean Opinion Score (MOS) terminology*, ITU-T, July 2006.
3. ITU-T P.862 – Recommendation P.862 of the International Telecommunication Union: *Perceptual evaluation of speech quality (PESQ) – An objective Method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, ITU-T, February 2001.

Previous work of the Author

This paper is a shortened version of author's contribution to the CN2014 conference. The individual parts of the author's work were presented on several occasions. Following list provides an overview of author's publication activities in this topic.

1. J. Rozhon, F. Rezac, J. Slachta, M. Voznak, *Monitoring of Speech Quality in Full-Mesh Networks*, In SPRINGER Communications in Computer and Information Science CCIS, Volume 431, 2014, pp. 157-166. DOI 10.1007/978-3-319-07941-7_16.
2. F. Rezac, J. Rozhon, M. Voznak, J. Slachta, J. Safarik, *Multi-agent System for Monitoring the Quality of Speech in Computer Networks*, In Proc. 37th International Conference on Telecommunication and Signal Processing, Berlin, July 1-3, 2014, ISBN 978-80-214-4983-1, ISSN 1805-5435, pp. 150-154.
3. J. Frnda, M. Voznak, P. Fazio, J. Rozhon, *Network Performance QoS Estimation*, In Proc. 37th International Conference on Telecommunication and Signal Processing, Berlin, July 1-3, 2014, ISBN 978-80-214-4983-1, ISSN 1805-5435, pp. 287-291.
4. M. Voznak, J. Rozhon, P. Partila, J. Safarik, M. Mikulec, M. Mehic, *Predictive model for determining the quality of a call*, In Proc. SPIE. 9118, Independent Component Analyses, Compressive Sampling, Wavelets, Neural Net, Biosystems, and Nanoengineering XII, 91180Y. (Baltimore, Maryland, USA, May 22, 2014) doi: 10.1117/12.2050661.
5. J. Frnda, M. Voznak, J. Rozhon, M. Mehic, *Prediction Model of QoS for Triple Play Services*, 21st Telecommunications Forum Telfor, TELFOR 2013, 2013, Article number 6716334, pp. 733-736.

Classification of VoIP Honeypot Data

Jakub Safarik

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

Abstract. Information about attacks in computer networks is gathered via various methods. One of the technique uses monitoring server masquerading as a production one, called honeypot. These honeypots behave like production system, typically with some security flaw, misconfiguration or in unsecured parts of the network. All interactions with the honeypot is monitored and provides valuable information about methods used by malicious intruder. Using a distributed network of honeypots brings unwanted overhead for data analysis and attack evaluation, but provide more valuable and independent result, so an automatic classification mechanism is needed. This article describes a multilayer perceptron neural network for classification of attacks in voice over IP telephony infrastructure on a centralized server. With a centralized server for aggregation and data processing is possible to detect attacks from honeypots in different networks, take a precaution step against them or harden existing production IP telephony infrastructure. Crucial part of the proposed solution is a proper attack classification by neural network.

Keywords: Attack classification, multilayer perceptron network, neural networks, VoIP attacks

1 Introduction

The SIP (Session Initiation Protocol) is a popular design for handling IP telecommunications services. SIP is an open-source protocol and is defined in RFC 3261 [1]. With the growing popularity of SIP protocol, the number of implementations and various extensions raise. Juniper Research forecast 640 million VoIP users by 2016 [2]. The IP telephony infrastructure based on SIP is very fragile to different types of attacks because of its similarity with HTTP and SMTP protocols. This can lead to loss of money, trust and other unpleasant consequences [3].

This situation could be solved partially with strict security rules, encryption and properly set VoIP servers. Some of mechanisms can disrupt or mitigate certain types of attacks, but there are still remaining attacks, which can impact VoIP servers.

Detection of SIP infrastructure attack is solved with different ways in a range of studies and papers. This paper covers a SIP attack classification with MLP (multilayer perceptron) neural network from honeypot application Dionaea. The information about SIP attacks from a honeypot application brings valuable source of network attacks.

With an automatic classification system is possible to automatically detect attacks on IP telephony from various set of honeypots. This honeypot network concept is shown on Fig. 1.

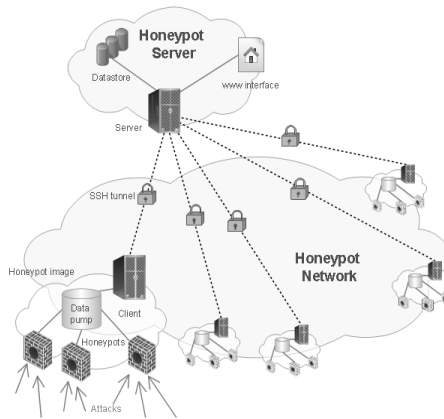


Fig. 1. Honeypot network concept

Neural network algorithm described in this paper is used in honeypot network hierarchy as a module on the centralized server for classification of SIP based attacks. More information about distributed honeypot network covers separate article [7].

2 MLP neural network

Neural networks try to model information processing capabilities of the nervous system of mammals, which is composed of millions of interconnected cells in a complex arrangement. The artificial neural network tries to model this design. The function of a single neuron is well known and serves as a model for an artificial one.

Even that we do not completely understand the complexity and massive hierarchical networking of the brain, with its incredible processing rate. The artificial neural network handle complex problems by using different topologies, each of them has its pros and cons.

The feed-forward MLP neural network was used for VoIP attack classifications. It consists of several layers, each containing the specific number of neurons called perceptron. These perceptrons in one layer are interconnected to each other in the following layer (this connection could be also called a synapse) [8].

The figure 2 shows the inner structure of used MLP network. The MLP network has two hidden layers, with one input and one output layer. Each neuron in the input layer has a value based on input parameters. This layer has the same number of neurons as there are parameters in the input set. After the input layer continues two hidden layers and output layer. The output layer has the same number of neurons as the number of attack classes, so each neuron is then a single class of learned attack. Number of neurons inside hidden layers depends on neural network configuration and are typically higher than the number of neurons in input or output layers.

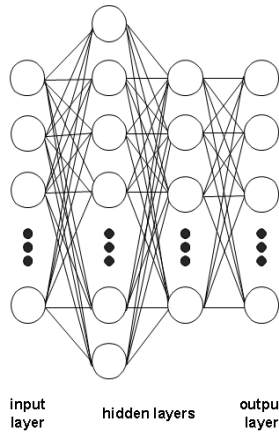


Fig. 2. MLP neural network topology

2.1 Perceptron

The perceptron is a more general computational model than McCulloch-Pitts units. The main innovation is in including numerical weights for connections and a special interconnection pattern. The activation function for neuron – sigmoid impacts the final potential of a neuron. This result potential is then transmitted through connections to neurons in the next layer while afflicted by each connection weight. These weights serve as a memory for neural network. Inputs for the activation function are real inputs x_1, x_2, \dots, x_n from the previous layer, with the associated connection weights w_1, w_2, \dots, w_n .

The output of a neuron is between 0 and 1, where 0 means inhibition and 1 excitation. The final value on the output of neuron (y) depends on its activation function. As was mentioned before, this function is a real sigmoid function (equations 1 and 2).

$$S_c: \mathcal{R} \rightarrow (0,1) \tag{1}$$

$$y = S_c(z) = \frac{1}{1+e^{-cz}} \tag{2}$$

$$z = \sum_{i=1}^n w_i x_i \tag{3}$$

The equation 3 shows parameter z , which is the sum of the output from previous layer neuron x and multiplies by corresponding connection weight w . Parameter c represents a skewness of the sigmoid function (typically it is 1.0). Higher values of parameter c bring the skewness of a sigmoid function closer to the step function [8, 9].

3 Backpropagation algorithm

The memory of neural network is saved in connection weights and backpropagation is used to acquire these values. Backpropagation works as a reverse mechanism to feed-

forward, with the specific set of data called training set. Training set has the same format as attack inputs for neural networks but contains also the final result of classification (or the class of the specific attack).

The core of a backpropagation algorithm and the neural network learning is a process of weight adaptation. It is done on the training set of inputs with known outputs. The solution of learning problem is a combination of weights with the minimal error function. Learn rate parameter (η) affects connection weight correction, used to lower the value of the error function (equation 5).

The equation 4 shows computation the of backpropagation error (δ) for connection weight in one layer (indexed as j). It is counted as a multiplication of higher layer (indexed as k) backpropagation error, actual output, expected output and actual weight of the connection. Parameter y represents the output of neuron, x always its inputs. Parameter c is the expected output and w the connection weight. The backpropagation error is then used in weight adaptation equations 5 and 6.

$$\delta_j = \sum_{k=1}^n \delta_k y_k (1 - y_k) c w_{jk} \quad (4)$$

$$\Delta w_{ij} = \eta \delta_j y_i \quad (5)$$

$$w_{ij} = w_{ij} + \Delta w_{ij} \quad (6)$$

The learn rate parameter (η) serves to set a proper step of correction in one backpropagation iteration [8, 9]. One iteration of backpropagation learning uses all records from the training set. The last parameter w_{ij} is the connection weight from the previous layer (i) to the actual layer (j) as shown figure 3.

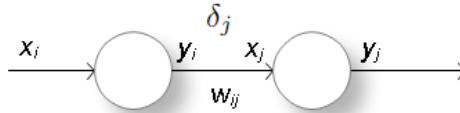


Fig. 3. Indexing between layers

4 Practical implementation

4.1 MLP neural network configuration

In previous research covered in [10] was used MLP neural network for detection of six basic SIP attacks. Inner structure of MLP network was improved for higher attack detection accuracy. The final structure of new neural network contains 10 input layer neurons, 30 and 24 neurons in hidden layers and 8 output neurons. With the change to eight classes, there is more robust and accurate detection of attacks.

Both generations of MLP neural networks are learned, when the successfulness on training set is lower than 5%. This ensures statistically significant classification capability on the training set. Both generations use the same configuration of skewness (1.0) and learn momentum (0.8).

4.2 Data source for classification & input vector parameters

All attack information is gathered through honeypots running Dionaea. It emulates and monitors traffic of a SIP PBX and save specific set of information about malicious traffic to the internal database. All running honeypots are accessible through internet on public IP addresses (IPv4). No legitimate calls or devices connect to this honeypot, so only malicious traffic or misconfigured devices communicate with it.

All attack data saves Dionaea to a sqlite database. The database contains several tables for specific protocols and functions, so information regarding SIP protocol is gathered from specific tables and aggregated. Selecting only single lines from these tables for classification is valueless, partly because of request/response behavior of SIP protocol. Final data is an array with 10 attributes.

These individual 10 attributes then serve as an attack vector (or neural network input). The aggregation depends on attack origin and also time of last message occurrence (there is 5 minute sliding window). Attributes are in the following order: attack time duration; connection count; count of various SIP messages, connection rate. The connection count attribute holds the number of connection from a single source on honeypot. The connection rate is the ratio of all received SIP messages to connection count.

4.3 Backpropagation configuration

Before backpropagation starts, all connection weights are set randomly from range $(-1,1)$. After first initialization of weights starts backpropagation iteration. Each iteration cycle uses all items from the training set and after specific number of iteration cycles is automatically checked successfulness of classification. If the successfulness on the training set is lower than specified threshold, backpropagation runs another iteration cycle. When the successfulness is higher than 95%, the neural network learning is done. To avoid the possibility of stuck in local extreme, system automatically re-initialize all connection weights after 2 500 000 backpropagation cycles.

As a source for the training set was used real attack traffic. This traffic is aggregated and classified by hand. From this classified set is prepared training set, which consists of 104 items, 13 items for each attack group. These classes are options tests; options scanning; call testing; unknown protocol; register and call; registration test, registration flooding; register attempt.

Reference set.

As a reference set serves an aggregated set of attack vectors detected on various honeypots. All attacks detected on these honeypots were classified by hand, so we cannot eliminate human factor error.

Tests of new MLP neural network on three reference data set bring exciting information. These set do not contain data from the training set. Result of analyses with MLP networks has following successfulness: 94.94%; 79.85% and 97.54%. Totally were detected 1631 attack groups and 57752 SIP messages (data for three months period). The lowest classification precision 79.85% was caused by new call attack, which was not included in the training set.

5 Conclusion

The proposed distributed honeypot network in combination with neural network classifiers serves as another security level. But the potential lies not only in detection capability and attack research. With the possibility to change firewall rules or network routing, whole system can prepare precaution mechanisms against attacks even when it do not influence the target network. The proposal of such monitoring system is distributed honeypot network.

Classification by human is very precise, but time consuming and expensive. It is typically conducted after the damage is done. Automatic classification mechanism brings a solution for VoIP classification and simplifies the analysis of attacks. The biggest disadvantage of this solution is its strong bindings on the training set. The MLP neural network cannot adapt to new attack classes or scenarios.

Test of the new MLP classification network on reference sets prove its quality but shows its limits and new ways for improvements. The future plans for neural network classification cover improving the accuracy of existing solutions and implementation of other evolutionary and statistical algorithms for attack classification.

References

1. Rosenberg, J., Schulzrinne, H., Camarillo G., Johnston, A., Peterson, J., Sparks, R., Handley M., Schooler, E., "SIP: Session Initiation Protocol" June 2002
2. "Mobile voice & video calling: strategic opportunities & business models 2011-2016", Tech. rep. Juniper research, 2011
3. Rezac, F., Voznak, M., Ruzicka, J., "Security Risks in IP Telephony", Cesnet Conference 2008-Security, 31-38, 2008
4. Gomez, J., Gil, C., Padilla, N., Banos, R., Jimenez, C., "Design of a Snort-Based Hybrid Intrusion Detection System", IWANN 2009, Springer-Verlag Berlin Heidelberg 2009, 515-522, 2009
5. Kang, H.J., Zhang, Z., Ranjan, S., Nucci, A., "Sip-based VoIP Traffic Behavior Profiling and Its Applications", MineNet'07, 2007
6. Sengar, H., Wang, H., Wijesekera, D., Jajodia, S., "Detecting VoIP Floods Using the Hellinger Distance", IEEE transactions on parallel and distributed systems, Vol. 19, No. 6, June 2008
7. Safarik, J., Voznak, M., Rezac, F., Partila, P., Tomala, K., "Automatic Analysis of Attack Data from Distributed Honey-pot Network", Mobile Multimedia/Image Processing, Security, and Applications 2013, 875512 (May 28, 2013)
8. Rojas, R., "Neural Networks", Springer-Verlag, 1996, ISBN 978-3540605058
9. Heaton, J., "Introduction to Neural Networks for JAVA, 2nd Edition", Heaton Research, 2008, ISBN 978-1604390087
10. Safarik, J., Partila, P., Rezac, F., Macura, L., Voznak, M., "Automatic Classification of Attacks on IP Telephony", Advances in Electrical and Electronic Engineering, Vol. 11, Issue 6, 2013, pp. 481-486, ISSN 1336-1376

Radio Jamming

Martin Tomis

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

Abstract. This article describes the basic theory of radio jamming in radio networks, systems and devices. Simple, but ineffective methods of radio jamming are described following with a slightly more complex but much more efficient jamming methods.

Keywords: radio jamming, radio networks

1 Introduction

With the development of wireless radio communication and radiolocation it was sometimes necessary to interrupt the communication of someone else or to disable the function of radar. In World War II interference of the enemy radio communications or radar meant a strategic advantage. Radio jammers [1] are now frequently used in prisons and offices of high-ranking officials to prevent the use of mobile phones and bugs. In this case it is necessary to disrupt at least the GSM frequency bands. For example, in military areas and airports the GPS signals are intentionally jammed, making it impossible to use GPS-guided missiles against those objects. Cell phone jammers [2] can be also found in theaters to prevent anyone to unnecessarily disturb the performance.

In general, to disrupt any radio technology it is always possible to use a sufficiently powerful source of electromagnetic radiation with sufficiently large width of the disrupted frequency band. It is then usually necessary to use a significantly higher power of disrupting signal than the transmitting power of the disrupted transmitter. According to this basic theory of radio jamming, to disrupt the GSM mobile networks, where the transmitter power is in hundreds of watts, radio jammer transmission power must be at least in kilowatts, covering frequencies from 800 MHz to 2.2 GHz. Very high transmission power at these frequencies reliably disables the GSM network, but it also has a negative impact on the environment by a dangerous exceeding of electromagnetic frequency exposure limits. It can cause diseases such as cancer, has large dimensions, extremely high power consumption, and last but not least it cancels a large number of other radio services, such as GPS, TV transmissions, police radios etc.

For effective jamming of the specific radio equipment or radio networks the significantly more energy efficient devices which cancel only the specific radio traffic and only to the desired distance are commonly used. This is achieved by well-known properties of a particular radio system. Similarly as in other attempts to attack the system the weakest link is chosen from the system characteristics. The attack on the weakest link can simply and effectively disable the entire system, or a substantial part thereof.

2 Radio Jamming Methods

For different types of radio systems there are different ways of jamming. The success of jamming is determined by several parameters:

- The number and width of the frequency bands available in the radio system
- The number and width of the frequency bands disrupted by a jammer
- Modulation types available in the radio system
- Interfering signal types used by a jammer
- The length of the radio link, transmitting powers and antenna gains
- Obstacles in the path of the radio link (path loss)
- The quality of the radio link receivers (filters, decoders)
- Transmitting power of radio jammer and the type of used antenna [3]
- Distance of the jammer from the interfered radio link receiver and obstacles in the interference path

2.1 Weakest link

Effective, small, energy-efficient jamming device must effectively disrupt the radio link in its weakest link. An example might be the GSM interference.

Mobile phone with a normal, almost omni-directional antenna transmits power up to 33dBm (2W) and has a normal antenna with approximately +2dBi of gain, which gives 35dBm EIRP. GSM base station transmitter power is for example 44dBm (25W) and also it has an antenna with +15dBi of gain, which gives 59dBm EIRP. In the case that it is necessary to prohibit the use of mobile phones, the simplest way is to use a very powerful wideband jammer that disrupts all of the base stations transmitted signals. However, if we look into the principle of GSM networks, we find that two-way communication is necessary for a phone call to be made. In this regard, instead of jamming signal transmitted from the base station, simply signals received by the base station are being jammed. As the power density of the transmitted signal decreases with the square of the distance, to jam a signal received from mobile phone in the nearby base station receiver the significantly less transmission power of radio jammer is sufficient. Theoretically it is necessary to have a jammer with transmitting power higher than 33dBm (2W), which is the transmit power of a mobile phone that it is necessary to disrupt. However, taking into consideration the great distance of mobile phones from a base station, which is essential for jamming, simply by putting

a jammer on high ground near a base station, and it may have a transmitter power of 33dBm or less for successful disruption of the whole base station. Since the receiving antenna and receiver of the base station are in this case the weakest link.

More often it is necessary to prohibit the use of mobile phones in buildings, such as prisons. Effective jamming in this case is also based on the weakest link. Instead of jamming of the transmitted signal from the mobile phone, the received signal in mobile phone is being jammed. This method does not restrict the functionality of other mobile phones in the entire area, and it is very effective. Pocket radio jammer with transmitting power of 30dBm (1W) reliably prevents the use of mobile phones in an ordinary room, and in the case of a remote base station also the surrounding rooms are being jammed.

2.2 Jamming signal types

Another important parameter for successful radio jamming is the type of the jamming signal. From the basic theory the sufficiently strong sinusoidal signal from radio jammer is considered enough to disrupt radio system. In some cases, a single sinusoidal signal of a single frequency is not sufficient for successful disruption of the radio communication system. Mobile phone networks have several different frequency bands with hundreds of radio channels. Radio jammer with a sinusoidal signal can usually disrupt only one radio channel. In the case of very high transmit power also several neighboring channels can be disrupted. Mobile phone uses a different radio channel or a different frequency band to communicate, and whole radio jammer with sinusoidal signal is meaningless. For this purpose the multiband radio jammers with varying signal is used. Common multi-band radio jammer contains multiple similar jammers for different frequency bands in a single device. Varying signal means for example sweeping or step-changing frequency of the transmitter in the specific frequency range. The mentioned jammer many times per second disrupts every radio channel for a certain period of time, and at all frequency bands used. This method of jamming is very effective against GSM networks.

The problem occurs with the radio networks with broadband transmission such as spread spectrum or OFDM. Information in such network is transmitted in parallel across the used frequency range. Interference signal with sweeping or step-changing frequency may not be sufficient to disrupt broadband radio system in order to break-up the connection. This simple interference in broadband radio system occurs only in terms of increased bit error rate. Error correction techniques of broadband radio systems, however, simply correct the errors in transmission and the connection is kept active even at the cost of transmission bitrate reduction. For effective jamming of broadband radio systems it is necessary to use jammer with a very similar jamming signal, ideally identical, or at least a sinusoidal signal with a very fast step-changing frequency, where the rate of frequency change approximately equals to the speed of step-changes in the signal of the radio system.

3 Conclusion

Powerful sine wave transmitter is not always the best solution to jam a radio system. The above described methods of interference to radio systems were tested on different radio networks and terminal equipment. Some systems it is easy to jam, others are more resistant. Some jamming methods are more effective than others, but it all depends on the particular circumstances of a particular system. The next step, and at the same time the main goal of this research, is the feedback jammer, which disrupts receiver of the radio system by the signal transmitted from the system itself, suitably or randomly phase-shifted and amplified. Feedback radio jammer receives signal from a radio system, and immediately returns the modified and amplified signal back to the system, which should cause very effective interference.

4 Projects and Publications

Martin Tomis is a member of team solving the grant of SGS, Czech Republic, No. SP2014/72 and grant of TAČR, Czech Republic, No. TA03020439.

Martin Tomis has also submitted a total of eight patent applications to Czech Industrial Property Office, expecting their granting.

References

1. Patel, I; Kulkarni, R.; Khan, J.A, "Intelligent FM signal jamming system," *Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on* , vol., no., pp.1,6, 26-28 July 2012
2. Shah, S.W.; Babar, M.I; Arbab, M. N.; Yahya, K. M.; Ahmad, G.; Adnan, T.; Masood, A, "Cell phone jammer," *Multitopic Conference, 2008. INMIC 2008. IEEE International* , vol., no., pp.579,580, 23-24 Dec. 2008
3. Sambhe, V.K.; Kale, D.S.; Wasule, A; Shikha, N., "Antenna for Mobile Phone Jammer," *Emerging Trends in Engineering and Technology, 2008. ICETET '08. First International Conference on* , vol., no., pp.856,859, 16-18 July 2008

Precision of Classification Methods for Berlin Database of Emotional Speech

Jaromir Tovarek, Pavol Partila, and Miroslav Voznak

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jaromir.tovarek, pavol.partila, miroslav.voznak}@vsb.cz

Abstract. Emotional state classification of human speech and recognition accuracy of the classifiers is disclosed in this paper. Recent developments in speech recognition places more emphasis on the extraction of information about the speech source. This means obtain information about who and how it was said. This article describes research which seeks to recognize the information from speaking, emotional state in particular. Emotional state is recognized by using different classifiers and features of speech by nowadays known systems. Berlin database of emotional recordings was used to train and test the system. Mel-frequency spectral coefficients and dynamic coefficients were extracted from the audio signal of the database. For classification were used Gaussian Mixture Model, k-Nearest Neighbours and Artificial Neural Networks methods. The main effort of this research is to examine the accuracy and usability of classifying methods for detection of human stress status from his speech.

Keywords: Emotional state, MFCC, Speech, Stress

1 Introduction

As the title suggests, this article describes a system for classifying emotional state of human speech. Emotion is one of the characteristics of human which describes his mental condition affecting physiological changes in the human body. These changes are also reflected in the human speech. Information about the emotional state is requested in many fields. Statistical evaluation of customer satisfaction, his interest in the products is evaluated by affected emotional state. This information is a direct response to any stimulus. Information about the emotional state can be used for call centers and the military, police and fire dispatching.

These are just the first examples of utilizations for speech emotion recognition systems. It is obvious that the system will have great application in human-machine interaction. Therefore it is appropriate to identify a classification ability of different classifiers for different emotional states. [1]

2 Speech Emotion Recognition System

System design consists of several blocks, which it distributed to major functions. Input values for the training and testing to create database of audio signals. Block diagram of the system is shown in Figure 1

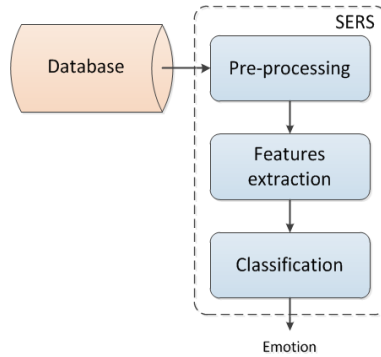


Fig. 1. Block diagram of Speech Emotion Recognition System (SERS). The system consists of a database that is used for training and testing and other blocks that describe the function of the algorithm.

Train respectively testing is the key step of the system. The quality of the input data, the audio signal in this case, has a direct impact on the classification accuracy. For this reason, it is used the Berlin database containing over 500 recordings of actors consists men and women. The database contains 10 sentences in the seven emotional states. This corpus of recordings is considered as a high-quality database, because it was created by professional actors in the sound studio. As mentioned, the speech signal has to be modified by routine operations such as removing the DC component, pre-emphasis and segmentation stochastic signal into quasi-periodic frames.

Speech recognition system is context-independent, that meaning take into account only signal parameters, not content informations. These parameters are the training and testing vectors for classifiers. The calculation parameters are represented by the Features Extraction block that extracts the Mel-Frequency Cepstral Coefficients (MFCC) and dynamic parameters (first and second derivative of MFCC). [2]

3 Classifiers

Individual research shows that cannot be said which classifier for emotion recognition is the best. Each classifier or their combination achieved some results accuracy, which depends on several factors. The success of classifier is directly dependent on the data. This is derived from the fact that the accuracy varies with the data character such as the quantity, density distribution of each class (emotions) and the language also. One classifier has different results with acted database, where the density of each emotion are equitable and different with real data from call centre where normal (calm) emotion state occupies 85 to 95 percent of all data. Appropriate choice of parameters has a considerable effect on the accuracy of these classifiers. The following subsections describe the used classification methods.

3.1 Gaussian Mixture Model

A Gaussian Mixture Model is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in biometric system, such as vocal tract, in speaker recognition systems as well. Probability distribution of the parameter vectors derived from human speech can be described using GMM.

$$p(o|\lambda^s) = \sum_{i=1}^{M^s} w_i^s p_i^s(o) \quad (1)$$

Where M is number of components for s class, w_i , $i=1, \dots, M$ are weights of components complying condition that sum of all weights is 1, p means the probability density of the components represented by the mean value and covariance matrix C_i . Speaker model can be described mentioned mixture characterized by the equation below.

$$\lambda^s = \{w_i^s, \mu_i^s, C_i^s\}, \quad i, \dots, M^s \quad (2)$$

The criterion of maximum likelihood is found lambda parameters with maximum p probability density based on sequence parameters $O = o_1, o_2, \dots, o_n$ obtained from speech, seen below. [3]

$$\lambda^s = \operatorname{argmax} p(o|\lambda^s) \quad (3)$$

3.2 K-Nearest Neighbour

In pattern recognition, the k-Nearest Neighbour algorithm (KNN) is a method for classifying objects based on the closest training examples in the feature space. kNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The kNN algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbour. The various distances between the vector x_s and x_t .

$$d_{st}^2 = (x_s - x_t)(x_s - x_t)' \quad (4)$$

The neighbourhood distance is calculated through Euclidean metric. Given an m -by- n data matrix X , which is treated as m (1-by- n) row vectors x_1, x_2, \dots, x_m .

3.3 Artificial Neural Network

Our emotional state classification problem with high number of parameters can be considered as a pattern-recognition problem. In this case, it can be used two-layer feed-forward network. A two-layer feed-forward network, with sigmoid hidden and output neurons, can classify vectors arbitrarily well, given enough neurons in its hidden layer. The network is trained with scaled conjugate gradient (SCG) backpropagation.

Let's say that, MFCC and dynamic coefficients in this case, are the input vectors x_i where $i = 1, \dots, d$. The first layer of network forms M linear combinations of these inputs to give a set of intermediate activation variables $a_j^{(1)}$

$$a_j^{(1)} = \sum_{i=1}^d w_{ij}^{(1)} x_i + b_j^{(1)} \quad j = 1, \dots, M, \quad (5)$$

with one variable $a_j^{(1)}$ associated with each hidden unit. Here $w_{ij}^{(1)}$ represents the elements of first-layer weight matrix and $b_j^{(1)}$ are the *bias* parameters associated with the hidden units.

SCG training implement mean squared error $E(w)$ associated with gradient ∇E and avoids the line-search per learning iteration by using Levenberg-Marquardt approach in order to scale the step size. A weights in te network will be expressed in vector notation. [4]

$$w = \left(\dots, w_{ij}^{(1)}, w_{i+1j}^{(1)}, \dots, w_{N_{1j}}^{(1)} \theta_j^{(l+1)}, w_{ij+1}^{(1)}, w_{i+1j+1}^{(1)}, \dots \right) \quad (6)$$

4 Results and Discussion

The following settings and features were used in the experiment:

- Input samples - Berlin database of emotional utterances.
 - 10 different sentences recorded by 10 different actors (both genders).
 - Over 530 samples consisting with 7 emotions: *anger, bored, disgust, fear, happiness, sadness, neutral*.
- Feature Extraction - computing of input vectors (speech parameters).
 - 13 Mel-frequency Cepstral Coefficients $c_m = [c_m(0), \dots, c_m(12)]$.
 - Dynamic coefficients of MFCC - Δc_m and $\Delta^2 c_m$ (acceleration coefficients).
- Emotion Classification.
 - GMM - 64 mixture components.
 - K - Nearest Neighbours (set up 5 neighbours.)
 - Artificial Neural Network - Feed Forward Backpropagation.

4.1 Conclusion

This research was aimed at recognizing the stressed out person. It means recognizing deviations from the neutral state. This state is not defined in the Berlin database. Therefore, it was necessary to assemble a set of data, so-called "stress cocktail" from defined emotional states. The stress of a person can be assembled from emotional states, other than neutral. Emotions Anger and Fear were used to compile the stress data set that because these emotional states are reflected most often when a person is exposed to stressful situations.

As before, the GMM, k-NN and the ANN were used to classify the stress versus neutral. Results for all three classifiers are shown in Fig. 2. The Receiver Operating

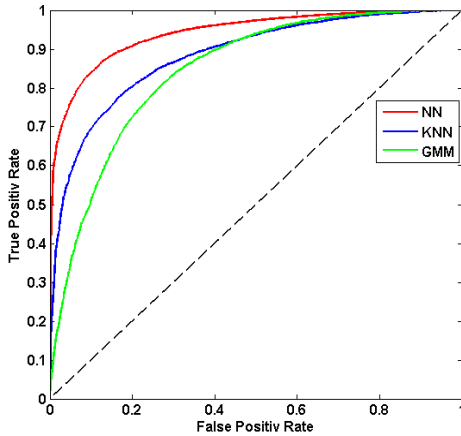


Fig. 2. Receiver Operating Characteristic of GMM, k-NN and ANN classifier for Neutral vs. Stress Recognizing.

Characteristic (ROC) is applied for better understanding the system. ROC curve is a tool for the evaluation and optimization of binary classification system (test), which shows the relationship between the sensitivity and specificity of the test or the detector for all possible threshold values.

The graph shows that the classification accuracy for the detection of stress is comparable to the classification of different emotional states. ANN achieved the best results in this experiment. Confusion Matrix for all classifiers is shown in Table 2. Explanation of cells for this matrix is described in Table 1. The results of Table 2 are presented for clearer representation of ROC curves.

Table 1. Confusion Matrix - description of cells.

		Classifier		
Output classes		True Positive	False Positive	Positive Predictive Value
		False Negativ	True Negativ	Negative Predictive Value
		Senzitivity	Specificity	Precision
	Target classes			

This experiment shows that these classification methods can be used on the recognition of emotional state. At the same time, the question arises, what emotional states will characterize stress. The answer will probably depend on which system would be applied.

Table 2. Confusion Matrix for ANN, k-NN and GMM classifiers and results for neutral vs. stress recognition.

	ANN			k-NN			GMM		
Neutral	7445 22.1%	1901 5.7%	79.7%	4919 14.6%	2209 6.6%	69.0%	6305 18.8%	3507 10.4%	63.3%
Stress	1684 5.0%	22587 67.2%	93.1%	4210 12.5%	22279 66.3%	84.1%	2824 8.4%	20981 62.4%	88.1%
	81.6%	92.2%	89.3%	53.9%	91.0%	80.9%	69.1%	85.7%	81.2%
	Neutral	Stress		Neutral	Stress		Neutral	Stress	

Acknowledgement

The research leading to these results received funding from the SGS project no. SP2013/94 and was partially supported by the project OPVK No. CZ.1.07/2.2.00/28.0062, Joint activities of BUT and TUO while creating the content of accredited technical courses in ICT.

References

1. Zarkowski, M. Identification-driven emotion recognition system for a social robot. In: *Methods and Models in Automation and Robotics (MMAR), 2013, 18th International Conference on*, pp.138–143, 26–29 Aug. 2013
2. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B. A Database of German Emotional Speech. In: *Interspeech 2005 Eurospeech, 9th European Conference on Speech Communication and Technology*. September 4–8, 2003, Lisbon, Portugal, Proceedings, pp. 1517–1520, 2005
3. Knox, M., Mirghafori, N. Automatic laughter detection using neural networks. In: *Interspeech. 2007*. pp. 2973–2976.
4. Vlasenko, B., Schuller, B., Wendemuth, A., Rigoll, G. Combining frame and turn-level information for robust recognition of emotions within speech. In: *Interspeech. 2007*. pp. 2249–2252.

List of the Publications

References

1. Partila, P., Tovarek, J., Voznak, M., Safarik, J. Classification Methods Accuracy for Speech Emotion Recognition System. In: *Springer Advances in Intelligent Systems and Computing AISC, Vol. 289, 2014*, pp. 439–447, DOI: 10.1007/978-3-319-07401-6_44.
2. Partila, P., Tovarek, J., Frnda, J., Voznak, M., Penhaker, M., Peterek, T. Emotional Impact on Neurological Characteristics and Human Speech. In: *Springer Advances in Intelligent Systems and Computing AISC, Volume 298, 2014*, pp. 527–533, DOI: 10.1007/978-3-319-07773-4_52

Thermal inducted change of beam profile in bend fibers

Petr Závodný

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.zavodny.st@vsb.cz

Abstract. This article examines effect of the temperature on the transmission characteristics of telecommunication fibers. Presented results show the behavior of variously curved optical fibers exposed to the temperatures from 30°C to 90°C. Temperature effect was studied from the viewpoint of the beam profile change and variation of the fiber attenuation. There was observed slight dependence on the temperature on measured fibers, which caused increase in numerical aperture and reduction in fiber's attenuation.

1 Introduction

The bent losses are very important parameters of the optical fibers that have influence on the transmission parameters. Because the attenuation of the bent fibers is wavelength depended, it must be monitored in the WDM system to reach equal power in all channels. Thermal dependence of the fiber is also important, because some parts of the optical lines can be exposed to the very large temperature differences from very low, during winter, to very high, when closed parts are exposed to the direct sunshine and the thermo optical effect can cause drift in the wavelength powers.

In this experiment was examined both bend and thermal properties of the single mode fibers. We examined the change of the FWHM (full width at half maxima) of the beam (measured by the beam profiler) and the change in the fiber attenuation.

2 Theoretical impact of the fiber bending and heating

2.1 Temperature

The temperature change causes several changes in the properties of the optical fiber. The main influence is variation of the refraction index and the length. The change of optical fiber length is small and results in very small differences in the optical properties of the fiber. On the other hand, thermally inducted change in the refraction index, caused by thermo optic effect, has significant impact on the properties of the fiber. The typical refraction index change is from 10^{-4} to 10^{-6} °C⁻¹ [1] and its influence to the optical fiber properties can be demonstrated by beam optic using the Snell's law:

$$n_1 \sin \phi_1 = n_2 \sin \phi_2 \quad (1)$$

Where n_1 and n_2 are refraction indices of the core and the cladding, ϕ_1 and ϕ_2 are the angles of the incidence and reflection. In the case of the total reflection (which is case of the optical fibers) equation can be expressed by critical angle ϕ_c :

$$\sin\phi_c = \frac{n_2}{n_1} \quad (2)$$

If we think about the change of the refraction index as function of the temperature T , the previous equation can be rewritten in the following form:

$$\sin\phi_c(T) = \frac{n_2(T)}{n_1(T)} \quad (3)$$

The equation shows that the variation of the temperature induces change in the magnitude of the critical angle. This affects wave guiding properties of the optical fiber and may cause change of the attenuation, especially when the optical fiber is bent.

When we look on the fiber from outside environment, we can use numerical aperture to describe its coupling abilities:

$$NA = \sqrt{n_1^2 - n_2^2} \quad (4)$$

In this case we can also express the refraction index as function of temperature.

$$NA(T) = \sqrt{n_1^2(T) - n_2^2(T)} \quad (5)$$

The change of numerical aperture causes change in the size of emitting/receiving cone of the light beam i.e. the beam profile. Because thermo optic coefficient is positive and $n_1 < n_2$, we can suppose that increasing temperature also increases numerical aperture of the fiber and FWHM of the beam profile.

2.2 Fiber bend

The bending of a fiber causes coupling light from the fiber core to the cladding. In simple models with infinite cladding this results only in attenuation, because energy is lost in the cladding. But the real fiber cladding is finite and on the end of the cladding is interface between cladding and (usually) coating which is also able to guide the light. Several experiments showed that the real attenuation of the fiber is not continuously increasing with bent radius, but there are attenuation peaks and valleys causing ripple in the attenuation curve [2, 3]. These oscillations of the attenuation are caused by interference of the cladding modes (also known as Whispering Gallery modes) with core mode [6].

Other experiments showed that attenuation peak and valleys are also depended on the temperature and causes oscillating of the optical fiber attenuation in order of tenths dB of magnitude [4].

3 Measurement and results

The examined fibers were conventional single mode telecommunication fibers according to ITU-T G.652 with coating and protective jacket. The fibers were bent in diameter 15 mm, 20 mm, 25 mm, and 35 mm in multiple turns (1 to 4). Bent fibers were heated from 30 to 90 °C in water in heater with automatic regulation. Two attributes of fiber were measured: an output power and a beam profile. The beam profile was measured in two axes by rotating slit beam profiler BP-109. The output power (attenuation) was measured by standalone power meter. Measuring was performed on the wavelengths 1310 nm and 1550 nm. The optical fibers was reel in on the pipes with given diameter and fixed with adhesive tape.



Fig. 1. The beam profile and the output power measurement diagram.

The measurements results (Fig. 2 and Fig. 3) shows the beam profile of the single mode fiber at 1550 nm at x and y axis. We can see different behavior between axis x and y. At x axis change in the beam profiles is minimal, but in y axis the beam profiles amplitude and diameter increases together with temperature. Increase of the beam diameter indicates that numerical aperture of the fiber has also increased. Different results in x and y axis can be result of the polarization of light in the bend deformed fiber.

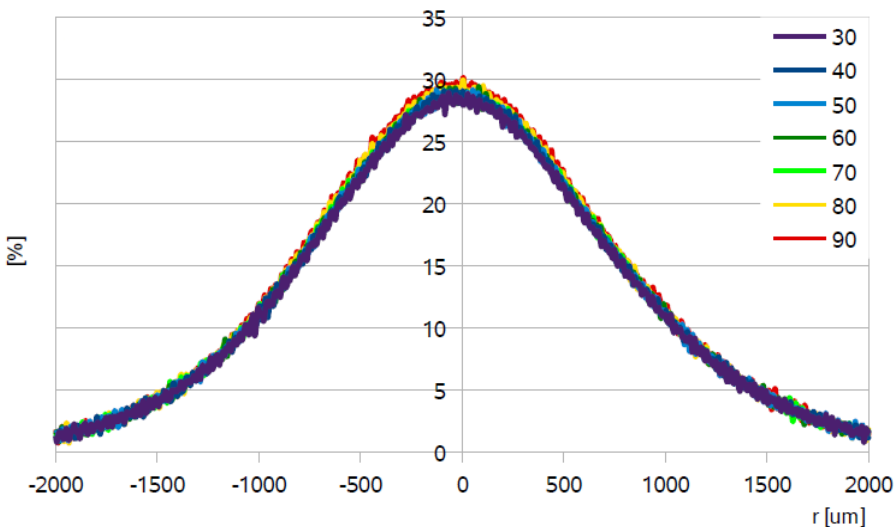


Fig. 2. The beam profiles in y axis at 1550 nm, 15 mm bend one turn.

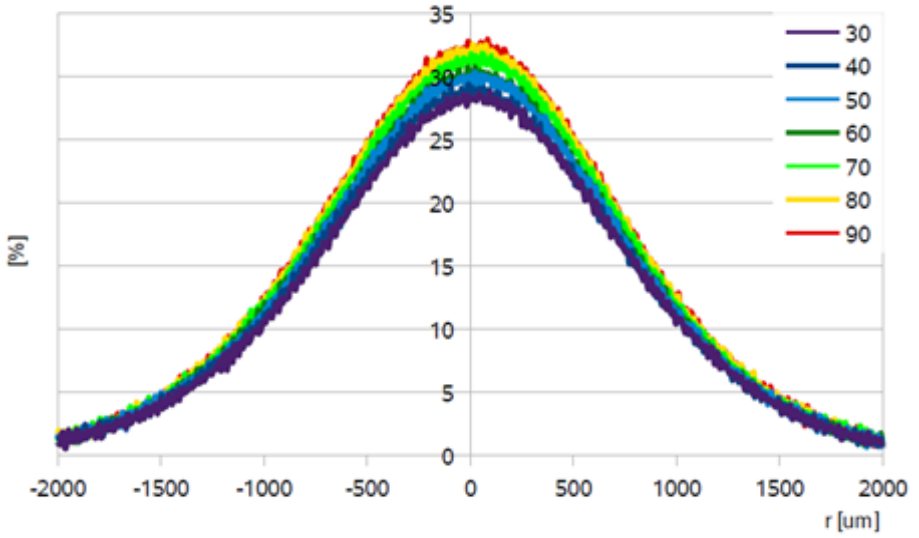


Fig. 3. The beam profiles in y axis at 1550 nm, 15 mm bend one turn.

Similar result of the beam profiles has been measured on different combination of wavelength, bend diameter and turns count. The change in profiles was always most visible only in one of the axis (x or y). The most noticeable change in the beam profile was at small diameters. Here is summary of the FWHM for the 15 mm bend diameter for both 1550 nm and 1310 nm wavelengths, from one to four fiber turns.

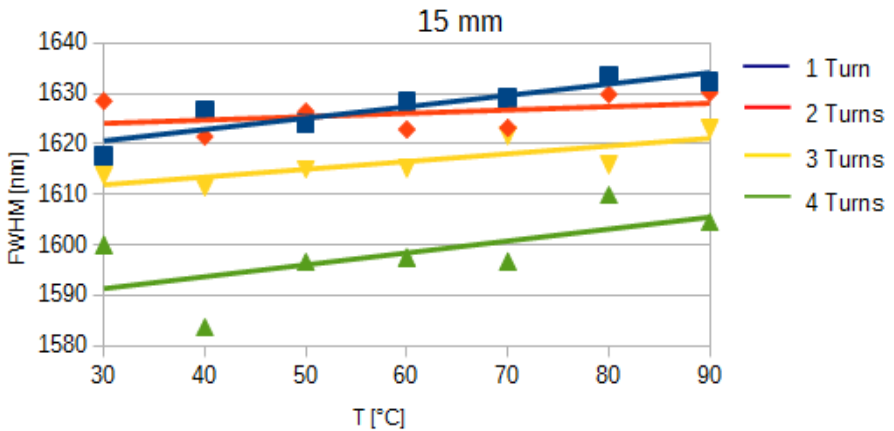


Fig. 4. The dependence of the FWHM on the temperature at 1550 nm, 15 mm bend diameter.

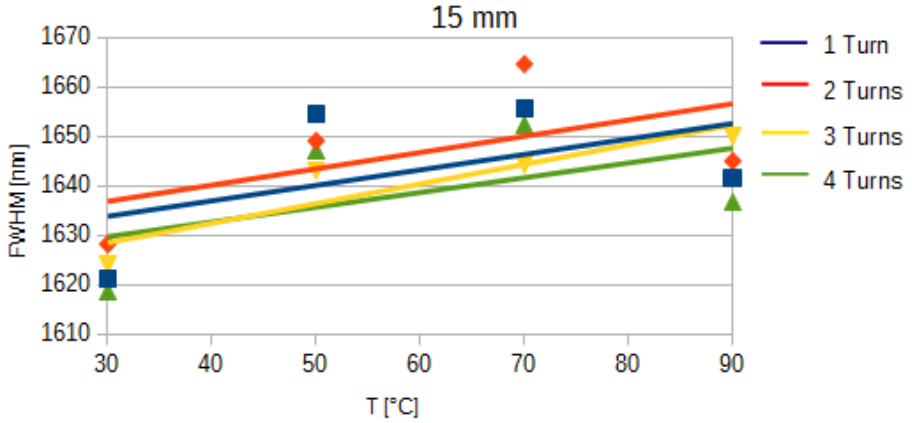


Fig. 5. The dependence of the FWHM on the temperature at 1310 nm, 15 mm bend diameter.

Graphs on the figures **Fig. 6** and **Fig. 7** show the output powers that correspond to the previous beam profile measurement results. There is slight increase in the output power with temperature acknowledging increase of the numerical aperture and magnitude of the beam profile. No oscillation of the attenuation peaks was observed mainly due to limited available resolution of the power meter.

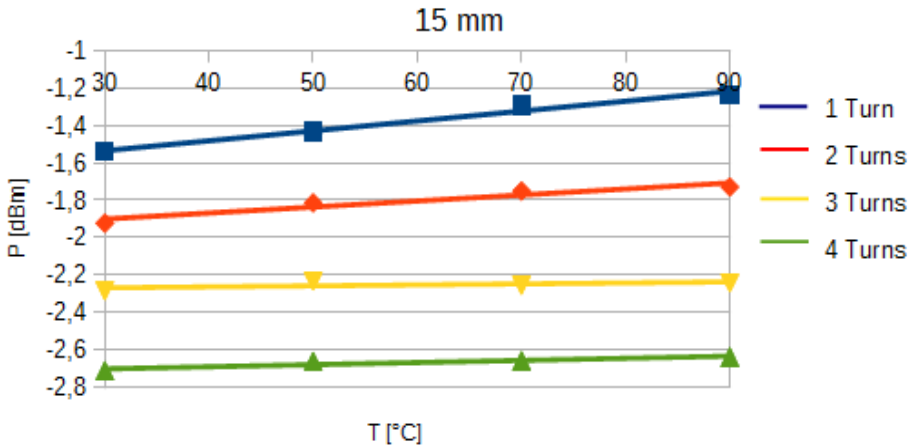


Fig. 6. The dependence of the output power on the temperature 1550 nm

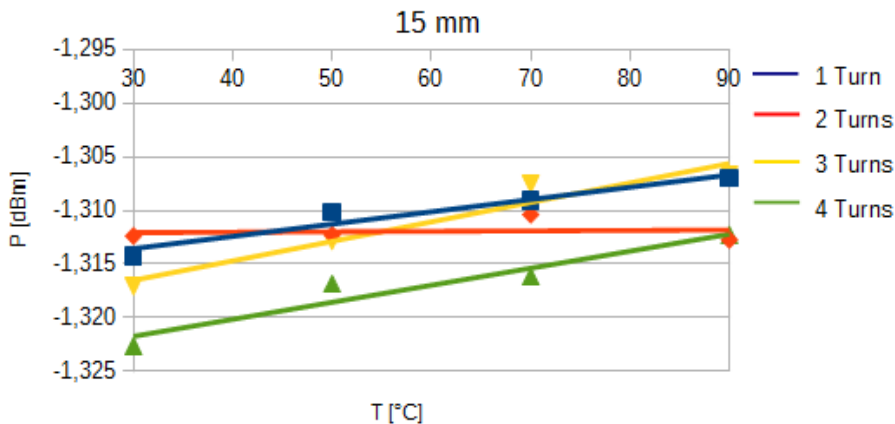


Fig. 7. The dependence of the output power on the temperature 1310 nm

4 Conclusion

The beam profile measurement results show increase of the FWHM and numerical aperture of the single mode optical fiber with increasing temperature. Increase of the FWHM was between 10 μm and 20 μm with the temperature change from 30 $^{\circ}\text{C}$ to 90 $^{\circ}\text{C}$. The optical fiber was most sensitive for small diameters between 15 mm and 20 mm. Stronger dependence can be probably reached at smaller diameters. Increasing number of turns only increased the attenuation, but didn't change results of FWHM.

References

1. YIN, Shizhuo, Paul B RUFFIN a Francis T YU. Fiber optic sensors. 2nd ed. Boca Raton, FL: CRC Press, c2008, 477 s. Optical science and engineering (Boca Raton, Fla.), 132. ISBN 14-200-5365-5.
2. RENNER, H.; , "Bending losses of coated single-mode fibers: a simple approach," Lightwave Technology, Journal of, vol.10, no.5, pp.544-551, May 1992doi: 10.1109/50.136086
3. FAUSTINI, L.; MARTINI, G.; , "Bend loss in single-mode fibers," Lightwave Technology, Journal of, vol.15, no.4, pp.671-679, Apr 1997 doi: 10.1109/50.566689
4. ANANDKUMAR S. LALASANGI, VIJAY K. KULKARNI, I. I. PATTANSHETTI, U. S. RAIKAR, Bend induced loss in single mode fiber for designing simple interferometric temperature sensor, Journal Of Optoelectronics And Advanced Materials Vol. 8, No. 5, October 2006, p. 1901 – 1904
5. A. ZENDEHNAM, M. MIRZAEI, A. FARASHIANI, L. HORABADI FARAHANI, Investigation of bending loss in a single-mode optical fiber, Indian Academy of Sciences, Vol. 74, No. 4 April 2010 pp. 591-603

Conical Termination of Optical Fibers for Sensor Applications

Ondřej Zbořil, Tomáš Kajnar, and David Hrubý

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ondrej.zboril@vsb.cz

Abstract. This article deals with the implementation of the conical ends of optical fibers and a description of their shapes. The reader is introduced to the problems of optical fibers and their potential use for sensor applications.

Keywords: Optical Fiber, Splicing, Sensor, Tapering, Conical Termination

1 Introduction

First of all optical fibers has earned a place the most important telecommunication medium for high-speed transmission, with a number of advantages over metallic connections, whether it's a much higher transmission rate, immunity to electromagnetic interference and eavesdropping, and last but not least, a much smaller space requirements. In recent years, optical fibers are increasingly used in the fields of sensor technology.

Today there is a huge variety of optical fiber sensors that can measure a large portion of physical quantities. These sensors use fiber for its activities of various external influences acting on the fiber and thereby affect the modulation of optical radiation, which is then evaluated in an appropriate manner.

2 Splicing of optical fibers

We must follow certain guidelines to proper connecting optical fibers. The ends of the optical fibers must be prepared for such an application. When there is imprecise connecting or connecting fibers of different cross-sections, it leads to coupling losses. Optical fibers can be connected with different methods. The choice of method to bond the fibers depends on how we use them. [1]

The methods can be divided as follows:

- Detachable connections
 - the mechanical splicing
 - the optical connector

- Permanent connection
 - Fusion splicing

For this case was used fusion splicing of optical fibers. This is an undetachable method of connecting fibers. For splicing of optical fibers is used a special arc fusion splicer. This is the most reliable method for connecting the ends of the fibers. Attenuation of classical splice is in the order of hundredths of a dB. Very low attenuation values of this method predispose it to connecting long distance paths.

2.1 Tapering

Tapered fibers have narrowed part of the profile. This is done by drawing the optical fiber, for example, in the hot zone of the drawing furnace. In this case was forming by dragging two opposite ends of optical fibers from each other during fusion splicing. Example of tapered structure is shown in Figure 1. [2]

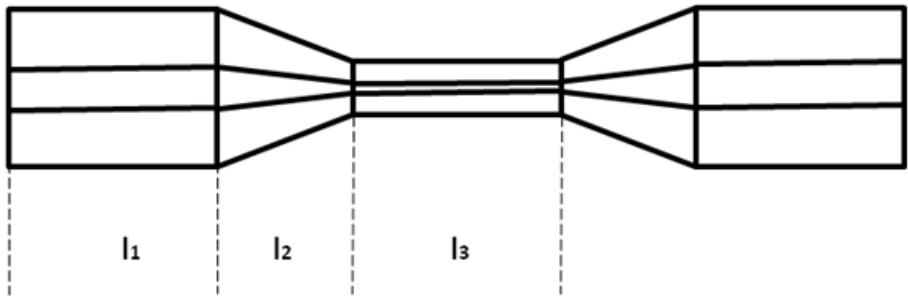


Fig. 1. Profile of tapered optical fiber. [3]

The tapered fiber could be divided into three parts:

- l1 - the first part is a classic fiber, usually with a diameter of 125 micron
- l2 - section with varying diameter fibers called transition
- l3 - the central part of the tapered fiber is called waist

The tapered part of the optical fiber is interesting for fiber sensor technology, where can be experimented with the application of such shapes. Above all it comes to fiber sensors with evanescent wave eventually sensors working with the decoupling of the optical beam. The more waists are thinner, the better is approach of evanescent waves to the analyte.

Fiber sensors have recently become increasingly popular and their use is still expanding. The fiber sensor can detect physical and chemical variables (temperature, pressure, vibration, presence and concentration of certain chemical components, etc..)

and their changes. This method is widely used due to its size, weight, low energy requirements and resistance to electric and magnetic fields. Fiber sensor applications found use in non-traditional fields such as in engineering, in wastewater treatment plants or in the biochemical industry and often replaced traditional methods of measurement of the variables. [2]

The real profile of created tapered fiber is shown in Figure 2.

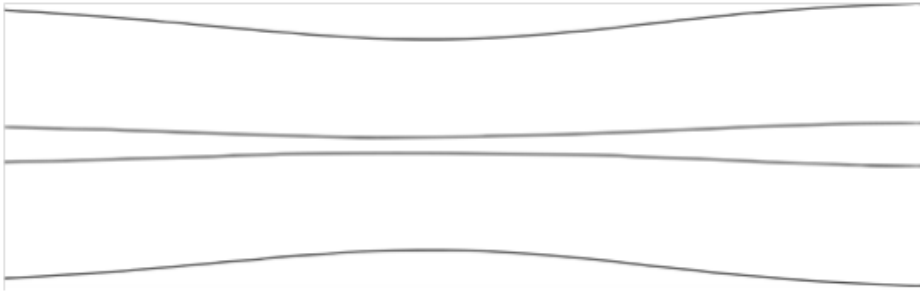


Fig. 2. Real profile of tapered fiber.

3 Optical Fiber Sensors

The sensor is responsive to external physical or chemical stimuli, which then converts the physical signal, which is processed by the device. The fiber sensor thus responds, for example, to temperature, vibrations or chemicals, and it will appear on the signal which passes by the measuring fiber. There are used primarily changes the amplitude, phase, and wavelength of the optical signal, which are caused due to the measurement.

Fiber sensors are suitable due to their small size and light weight. Such measurements have low power consumption, and although they are sensitive in a wide range. Their resistance to external electric and magnetic fields can be in many cases also useful. The different types of measurements can be used and current portion of optical telecommunication routes. Use of such sensors can be found for example in the production of mechanical or chemical, grocery, pharmacy and a large number of other industries. [4]

Separation of sensors according to the measured magnitudes:

- Mechanical - vibration, pressure, displacement
- Temperature - the temperature around the fiber
- Electromagnetic - electric and magnetic fields
- Radiation - reaction using photonic materials
- Chemical - the presence and concentration of the substances
- The flow of a fluid vorticity - flow measurement control points
- Biomedical quantities - monitor temperature, pressure and chemicals

3.1 Sensor with evanescent wave

These sensors are based on the interaction of the detected quantity of evanescent wave, which is part of the guided wave, but it is not the wave at all. The phase is constant in the propagation direction, since the phase constant is zero, and the amplitude is exponentially decreasing as the attenuation constant is nonzero. It forms at the interface of two media with different optical properties and spread along it, the most famous case is a total internal reflection in the fiber. In the direction perpendicular to the interface amplitude decreases exponentially and strikes only tens of nm behind the interface between two media, where it originated. Optical properties of the cladding react to changes due to the detected value. These changes in the optical properties affecting the transmitted optical beam and are subsequently detected. [5]

The intensity of the evanescent wave decreases exponentially with increasing distance from the interface of the core and cladding. The more evanescent wave extends into the analyte, the detection sensitivity. [6]

4 Conclusion

The aim of this article was to realize tapered end on different types of optical fibers. Tapering was predominantly on conventional single mode optical fibers with dimensions of 9/125 micron. It was also created to compare several samples of tapered splices on multi mode fibers with dimensions of 50/125 micron.

Profile of tapered fibers was changed as is shown in Figure 2. For single mode fibers have taper waist dimensions 6.2/101.5 microns. Multi mode fibers have a size in the waist 35.1/102.3 microns. Geometry of individual tapered fiber samples was very similar.

Attenuation of the tapered single mode fibers was around 0.328 dB for wavelength of 1310 nm and 0.296 dB to 1550 nm. This value has exceeded expectations, taking into account the deformation profile of the fiber. Tapered joint does not behave like attenuator, which is good.

Tapered multi mode fibers have attenuation value around 0.738 dB for wavelength of 850 nm and 1300 nm was 0.473 dB attenuation. These values are still acceptable, but are not so favorable as attenuation values for single mode fibers.

For better access evanescent waves to the analyte would certainly be interesting to taper microstructure optical fibers (MOF). Holes in cladding would increase the detection sensitivity of the conically tapered optical fiber.

We are currently working with a material suitable for the production of optical elements. Once we master the processing of this material, we can form additive layers on tapered fibers. These additional layers in combination with the profile of tapered fibers would be suitable for the specific sensoric applications. Another research is in patent proceedings and for this reason is forbidden by prof. Vašínek to publish this nowadays.

The use of fiber optics in sensors will definitely help the further development of optoelectronic technology and the gradual replacement of conventional sensors and measuring instruments, which are often large, heavy and sometimes energy intensive.

5 Other Activities

- Co-author on the article for the SPIE conference in Baltimore.
 - “Optical vibration sensor based on Michelson Interferometer arrangement with polarization-maintaining fibers”
- Participation in the project TAČR: TA03020439
 - Proposal arrangement of project laboratories for science and research and project documentation.
 - Research and development of fiber optic networks and systems, work according to specifications and schedule of the project manager.
- Participation in the project SGS: SP2014/147
 - Studies on the mechanisms of aging elements of fiber optic networks.

6 References

1. Novotný K.: *Optická komunikační technika*. Vydání třetí. Praha : ČVUT, 2007. 148 s.
2. Matějec, V., Podrazký, O., Mrázek, J., Peterka, P., Hayer, M.: "*Tažení optických vláken v Ústavu fotoniky a elektroniky*", *Jemná mechanika a optika*, Roč. 55, č.4 (2010), s. 107-110. Praha: Fyzikální ústav Akademie věd ČR, 2010. ISSN 0447-6441
3. Martan, T.: "*Zúžená optická vlákna*", *Jemná mechanika a optika*, Roč. 55, č.4 (2010), s. 111-114. Praha: Fyzikální ústav Akademie věd ČR, 2010. ISSN 0447-6441
4. Yin,S., Ruffin,P.B., Yu,F.T.S.: *Fiber Optic Sensors*. CRC Press, FL, USA, 2008, ISBN 978-1-4200-5365-4
5. Martan T., Kaňka J., Podrazký O., Kašík I., Matějec V.: Kónicky zúžená optická vlákna. *Optické komunikace OK 2007*, 89-97 (2007).
6. Kulhánek, P.: Hyperčočky. *ALDEBARAN BULLETIN* [online]. 2009, roč. 7, č. 20, s. 1-5. Internet: http://www.aldebaran.cz/bulletin/2009_20_met.php. ISSN 1214-1674.

An analysis of the geometry of a power line span on the basis of high-resolution images

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. In paper the author proposes that the measurements of the location of power lines should be taken by means of a hybrid method which the author calls the method of virtual images. This method consists in comparing factual photographs which have undergone transformation for the purpose of highlighting analysed fragments with a virtual photograph generated in the CAD environment on the basis of a three-dimensional model of a studied object.

1 Introduction

Any material cable similar to a power line, assumed to be perfectly flaccid, with a constant specific weight, suspended in two points, adopts, in the state of equilibrium and in the vertical plane, a shape describable by means of the catenary equation. A freely sagging line is always deflected from the chord connecting its suspension points, and the measure of this deflection is sag f (Fig. 1), which constitutes the maximum deflection of the line from the chord. Only when the vertical coordinates of the cable suspension points are the same, the sag is determined in the middle of its length. The shift between the centre of the cable's cross-section and the location of the maximum deflection between the chord and the cable is called deviation, which is measured horizontally.

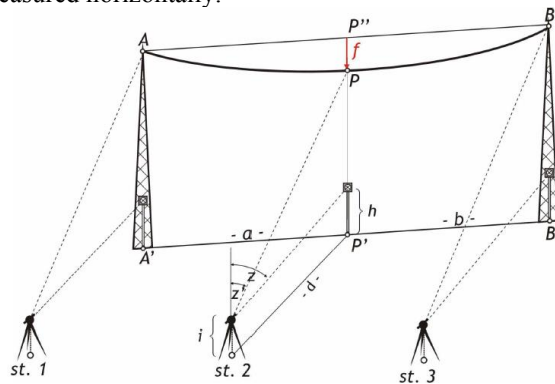


Fig. 1. Measurement of sag wire position using measurements of angles and distances

The measurement of the location of power lines along a span is difficult to carry out. Currently the most frequently measured value is the distance between the lines and the ground in selected points. It is measured geodetically by means of tacheometers or dedicated ultrasonic gauges. It is also possible to measure distances by means of a laser distance measurer. However, such measurements are difficult to take because of a small laser beam diameter and power lines' diameter of a few centimetres. The aforementioned measuring methods provide information on the coordinates of individual points of an analysed span. If the surveying methods are used, measurements are taken from a certain distance from the axis of the span, while, if a distance measurer is used, the instrument has to be located precisely over the lines. The position of the whole length of the line can be measured by means of a laser scanner, but in view of the cost of measuring equipment, this method is quite expensive. The knowledge of the so-called sag is necessary in the engineering practice first of all in the places where overhead power lines cross with road, rail or waterway transport lines.

2 The measurements with used the method of virtual images

The method which can be used successfully for this purpose is photogrammetry or the taking of measurements on the basis of photographs. Thanks to such advanced functions as DSM (Dense Surface Modeling), the software creates a cloud of points (similarly to what happens in a laser scanner) which subsequently create nodes of a network mapping the studied geometry. Results acquired from the Photomodeler Scanner software for preliminarily processed photographs are presented in Figure 2. Despite the fact that the photographs present just one power line, the matching of the acquired points is far from ideal.

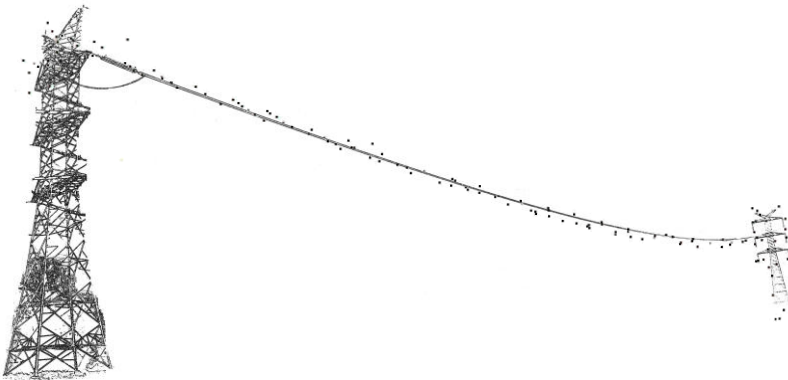


Fig. 2. Examples of the processed image with the background removed and with superimposed points obtained from the photogrammetric measurement

Measurement results are obtained similarly to measurements taken by means of a laser scanner. In order to carry out a complete survey of an object by means of both a scanner and the photogrammetric method, it is necessary to take measurements from at least a few to more than ten locations around an object. An important feature

of this method is a possibility of returning to source data at any time and obtaining required information thanks to a survey of surroundings at a particular point in time.

The author proposes that the measurements of the location of power lines should be taken by means of a hybrid method which the author calls the method of virtual photographs [2][3]. This method consists in comparing factual photographs which have undergone transformation for the purpose of highlighting analysed fragments with a virtual photograph generated in the CAD environment on the basis of a three-dimensional model of a studied object. A virtual photograph can be the result of rendering, with particular colours allocated to particular elements of a model or with a perspective view in the shading mode. In either case, the knowledge of such parameters of the real camera as the focal length as well as the horizontal and vertical viewing angles is of primary importance. The consistency of selected elements of analysed objects recorded in a photograph with their models rendered in a virtual equivalent of a photograph allows the identification of their characteristics or parameters.

In order to carry out the task successfully it is necessary to determine the position of the camera and the vector specifying the direction of the optical axis in a coordinate system aligned with the analysed object. Data concerning the position of the camera can be acquired on the basis of measurements of distances and/or angles from selected characteristic points, on the basis of a photogrammetric backward intersection or on the basis of known distances between natural characteristic points or artificial markers mounted especially for this purpose, for example on a power line's supporting structure. In the analysed case, the adopted points of reference are the elements of the supporting structure marked in Figure 3 with red dots.



Fig. 3. The reference points

In connection with the fact that the required value is the line's sag f , the model includes assumptions concerning the line's specific weight g , the initial value of the line's stress σ , the span length a , and for the value of the coordinate x , we calculate the distance y between the line and the chord. For a flat span, we can use equation 1, which approximates a catenary to a parabola.

$$y = \frac{g}{2\sigma} x(a - x) \quad (1)$$

After the virtual photograph is generated, it is compared with the factual photograph, and the result of such comparison is an index determining a degree to which a selected line matches its model. The index is based on the definition of the total error mean-square, which is the sum of the squared differences between the curves'

distances, with the use of the weighting factors w_i related to the distance of the analysed point from the site of measurement taking, in order to allow for uncertainty in the determination of the line's position, which grows together with distance (equation 2). The span is divided into N identical sections.

$$s = \sqrt{\sum_{i=0}^N w_i \cdot d^2} \quad (2)$$

An alternative index, which is easy to implement in the graphic environment, can be the surface area between the analysed lines. However, in the case of the analysed data, the foreshortening effect will result in a situation when, during the course of matching, the line located closer to the camera is taken into consideration to a larger extent.

A model case would be to take a photograph comprising the whole span (representing its longitudinal cross section) in such a manner that the camera's optical axis is perpendicular to the line's axis and the photograph's plane is vertical (the single picture method). However, because of limited space around power lines, the taking of such pictures is frequently impossible. Consequently, it is proposed that photographs taken at a large angle in relation to a line's axis should be used. In such a case, the calculation of coordinates should include transformations based on both transformations by rotation and in the further course of processing on projective transformations [1].

The measurements were taken by means of a camera equipped with a 21 MPx full-frame CMOS sensor, thanks to which a single line has from 2 to 9 pixels on photographs, depending on the length of a span. A specimen photograph is presented in Figure 4.



Fig. 4. The analyzed image

In order to carry out a further part of the analysis, it was necessary to determine the location of the camera and the selected distances of the points of reference, i.e. the elements of the supporting structure. The positions of the elements of the supporting structure and camera at the time of taking measurements were determined by means of a GPS receiver and a Disto D8 laser distance measurer; it was also determined on the basis of a photogrammetric backward intersection by means of the PhotoModeler Scanner programme [4]. The same programme was used to determine the distances from the points of reference. The acquired distances between the camera and the

selected points of reference located on the supporting structures as measured by means of the laser distance measurer and calculated in PhotoModeler are presented in Table 1, the differences do not exceed 2%.

After the determination of the position of the camera, the further analysis required information about the location of the point at which the line was suspended on an insulator. This position can be calculated by means of the PhotoModeler software or acquired from the technical documentation. Furthermore, it is very important to know the difference in the elevations of the supporting structures. Its approximation can be acquired from detailed topographic maps; it can be also measured by means of GPS receivers. In the analysed case, the trusses were located at the same level.

Joining the lines' suspension points with a chord (see Figure 5), it is possible to read the distance between the analysed line and the chord in pixels and subsequently convert the result, taking into consideration the angles and the perspective view.

Tab. 1. Distance of the camera from the characteristic points of poles

Point	Pole I			Pole II		
	calculated distance	measured distance	relative difference	calculated distance	measured distance	relative difference
	m	m	%	m	m	%
1	51	51,38	0,74	380	379,3	0,18
2	46	45,20	1,8	375	373,2	0,48
3	52	51,45	1,1	380	379,4	0,16
4	45	45,71	1,5	-	-	-
5	52	52,15	0,28	382	380,0	0,53
6	51	52,12	2,1	-	-	-

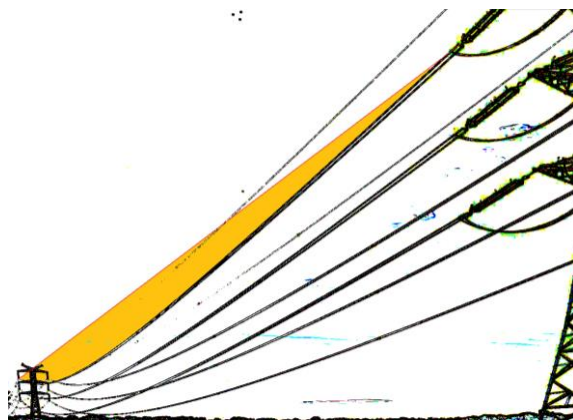


Fig. 5. The schematic of the analyzed position of a single cable

The proposed method provides for the creation of a simplified, three-dimensional model of a line in which vertical lines represent the suspension heights of the lower line. The line's sag was calculated after the adoption of stress characteristic for the calculating conditions. The use of the "Camera" function, e.g. in the AutoCAD environment, and the parameters of the camera's setting and position allows the acquisition of a perspective view. A vector picture of a virtual photograph is subsequently

scaled to a binary picture of a transformed factual photograph in such a manner that the line suspension points on both pictures overlap.

A line in a virtual photograph can be located above or below a line in a factual photograph. In the former case, it is necessary to reduce stress; in the latter, it is necessary to increase the adopted value (step) and calculate the catenary again. The procedure of creating a virtual photograph as well as comparing and calculating the index should be repeated iteratively. Every time the mutual position of the lines changes (in the calculations, the mark of the coordinate X of the distance vector), the value of the stress step was reduced by half. The iteration is stopped when the adopted criterion is fulfilled. It can be, e.g. a difference between the indexes of less than 5% between successive iterations or a maximum distance between the lines of, e.g. 1 cm.

4 Conclusions

Currently there is no measuring device which could be used for a relatively simple, cheap and quick method of determining the position of power transmission lines. The application of the known solutions can be limited by high costs, required qualifications and unfavourable terrain conditions, which has resulted in the commencement of work on new methods. The method proposed here allows a relatively quick and inexpensive measurement of the position of power lines by way of analysing a series of photographs taken by means of a high resolution digital camera. The paper presents an algorithm allowing the user to perform analyses to a considerable extent automatically. The method presented by the author yields good results, but it requires some further development work.

References

1. Bernasik J., Mikrut S.: Fotogrametria inżynierska, Wyd. Akademii Górniczo-Hutniczej, Kraków 2007
2. Zator S., Gasz R.: Identyfikacja elementów linii elektroenergetycznych z wykorzystaniem zdjęć wirtualnych, Zator S., Tomaszewski M. (ed.): Wybrane zagadnienia gospodarki remontowej energetyki, Wyd. Nowa Energia, Racibórz 2012, pp. 94-112
3. Gasz R., Zator S.: Evaluation of selected elements of a power line with using CAD environment, [w:] Tomczuk B., Waindok A., Zimon J., Wajnert D. (ed.): Electrodynamic and Mechatronic Systems SELM 2013, pp. 47-48
4. Cost Effective 3D Scanning, <http://www.photomodeler.com/products/pm-scanner.htm>
5. Alkan R.M., Karsidag G.: Analysis of the accuracy of terrestrial laser scanning measurements. FIG Working Week 2012, Knowing to Manage the Territory, Protect the Environment, Evaluate the Cultural Heritage. Rome, Italy, 6-10 May 2012. Electronic Journal of Map Technologies. Vol. 4, no: 2, 2012. pp. 1-10

Diagnostics of technical condition of boiler tubes

Mirosław Lasar

Opole University of Technology, Faculty of Electrical Engineering, Automatic Control
and

Computer Science, Department of Power Plants and Measuring Systems,
ul. Prószkowska 76, 45-758 Opole, Poland
m.lasar@po.opole.pl

Abstract. This paper presents the construction of a power boiler plant screens and diagnostic methods measure the wall thickness of boiler tubes, using active thermography. Presented the results of the first measurement and the results of numerical calculations in the cylindrical axial-symmetric system.

1 Introduction

Screens power plant boiler undergo constant periodic reviews diagnostic in order to observe changes in combustion processes. Early diagnosis of tubular elements of the boiler allows to prevent accidents that could hinder or completely stop the power plant unit. Diagnostics screens industrial process is, in this case the change in wall thickness which varies in the mixture of the pulverized coal combustion. Pipes screen standard with a thickness of 5 mm. Accuracy of a single field of the screen shows the table 1 in which thickness measurements are presented ultrasonic thickness gauge made Olympus DL37. The standard deviation is 0.086 mm.

Tab. 1. The results of measurements of the thickness of the screen

Measurements	Pipe 1	Pipe 2	Pipe 3	Pipe 4
1.	5,00	5,02	5,13	4,85
2.	4,96	5,05	5,12	5,04
3.	4,92	5,04	5,05	4,99
4.	4,99	4,90	4,92	4,83
5.	5,05	5,04	4,89	4,88
6.	4,96	4,99	4,92	4,87
7.	4,94	5,01	4,96	4,89
8.	4,94	5,09	4,92	4,82
9.	4,94	4,99	5,02	4,98
10.	4,90	4,99	4,91	4,95
11.	4,90	5,16	4,99	4,98
12.	5,03	4,97	4,97	4,97
13.	5,24	4,91	4,92	5,16

Figure 1 shows a screen, which consists of 15 interconnected pipes with a length of 3,8 m. Measurements were performed on every fifth tube at a distance of 30 cm.



Fig. 1. Part of the screen before installing

Before measuring the thickness of the walls were made a series of scans the surface of the screen (Figure 2.) Using a laser scanner Trimble FX.

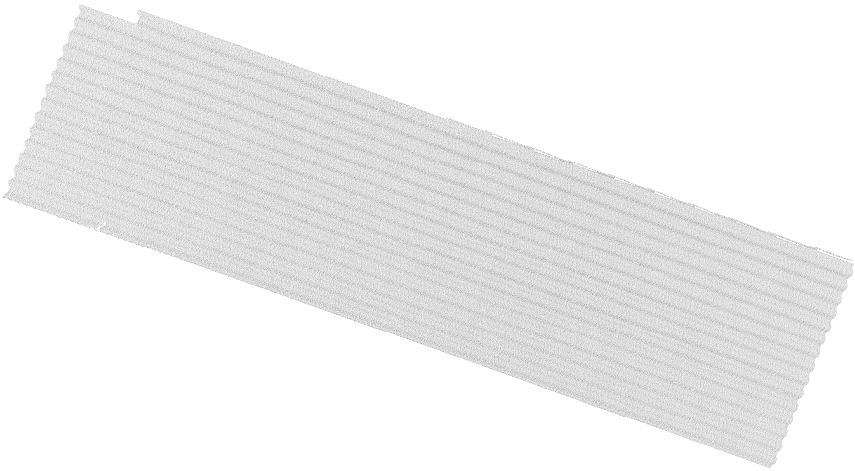


Fig. 2. Scan the boiler screen

The scanner uses a measurement of the phase shift and the accuracy of the measurement can be expressed by the equation:

$$\delta_z \approx \frac{\lambda_m}{4\pi\sqrt{SNR}} \quad (1)$$

where:

λ_m – length of the electromagnetic wave,
 SNR – signal to noise ratio.

This was done after 10 scans for three distances (5 m, 10 m, 15 m) from the scanner screen. Then scans were superimposed using crutches landmarks that were attached to the edge of the screen before scanning (Figure 3).



Fig. 3. Scanning process

During the repair of the block is also the wall thickness is measured using a thickness gauge. The measurement values are interpolated because it is not measured each pipe only every fifth or every tenth pipe at intervals of 2-3 meters vertically. The areas where the thickness is below 3 mm screens are replaced. This process is time consuming because the screen of the boiler is about 90 meters high.

Figure 4 shows a single scan line of the screen and the cross of the boiler. We can see that the geometry of the boiler screen isn't perfect.

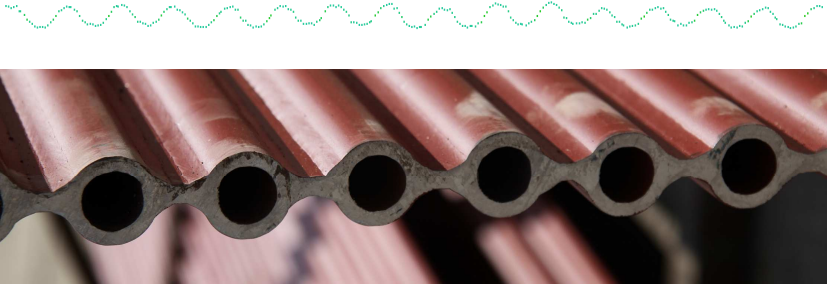


Fig. 4. Scanning process

The proposed method is dedicated to the initial diagnosis of screens involving the identification of the wall thickness of boiler pipes using active thermography. The source of initiating a change in the pipe wall temperature of the stream of water.

In the laboratory, was used the infrared camera VarioCam Head with a standard lens with a focal length $f = 25$ mm and an angle of view of $32^\circ \times 25^\circ$. For a real object (because of the considerable size of the installation) will be used a telephoto lens that has a focal length of $f = 100$ mm and an angle of view of $8^\circ \times 6^\circ$ - with it you can get a 16-fold increase in resolution small field of view. Has developed a method of the sequence of images and obtaining high-resolution thermal images [1].

Part of the flow test stand consists of a storage tank of the water heater and the installed thermostat temperature control system, from which, by means of a pump, water is pumped into two pipes of the nominal diameters 50 and 25 mm, which then returns to the tank. Currently, the main section of the measuring pipe is roughly in front of which was installed water meter with pulse output. At the beginning and at the end of nozzles installed on a Pt100 temperature sensor and a series of thermocouples. From the hardware for data acquisition module is used for measuring and control CompactDAQ 9178 C Series modules, temperature measurement with RTD's and thermocouples, and pulse input. Program execution is carried out measurements in LabVIEW application.

During registration was observed adverse event related to automatic correction of thermograms and self-calibration of the camera. Measurement corrections based on the known a priori known temperature in the area.

Figure 5 shows a pipe with three ground joints (0.05 mm 1 mm, 1.5 mm). On the thermogram these are noticeable during the pumping process has the system.

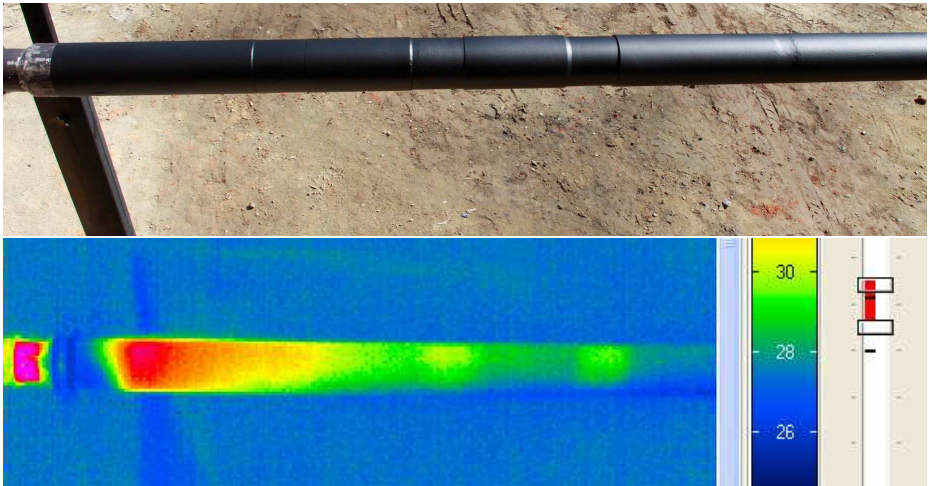


Fig. 5. Pipe with ground joints and thermogram during the experiment

There is a noticeable difference in temperature (approximately 1 K) in the steady state comparable between the measuring points. It can also be used to determine the changes in wall thickness [2].

The problem of heat exchange system shown cylindrical and both the temperature distribution in the jacket tube and in the space surrounding a well recognized due to numerous applications, including heat exchangers, in particular air heaters. Is typically determined or the temperature distribution of the heat flux, depending on the parameters of the test system. Are used for calculation of energy balance equation Fourier-Kirchhoff (2), which describes the phenomenon of heat conduction

$$k\nabla^2 T = c\rho \frac{\partial T}{\partial t} - q \quad (2)$$

where:

- k- coefficient of thermal conductivity
- c - heat capacity,
- ρ - density of the material,
- q - performance of the heat source

Due to the nonlinearity of the phenomena, the solution of the inverse problem requires numerical methods used. For this purpose, a mathematical model was built in an ANSYS'14 using FLUENT tool (Figure 6).

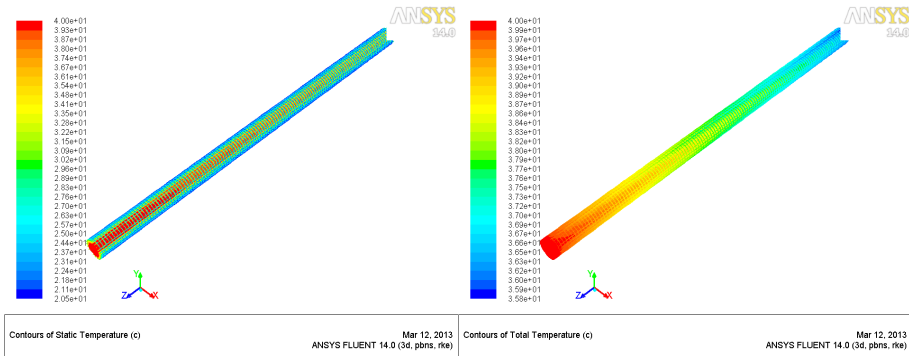


Fig. 6. Model rury w środowisku Ansys FLUENT

There is a need to model the phenomenon of fluid flow (at a given time as a function of temperature change), heat conduction and convection heat transfer. Currently, the model parameters are tuned to achieve compliance measurements and calculations.

Referenses

1. Zator S., Lasar M.: Rekonstrukcja termogramów wysokiej rozdzielczości na podstawie standardowych obrazów termowizyjnych, *Pomiary Automatyka Kontrola*, nr 11/2012, s. 965-967.
2. Laaidi N., Belattar S., Elbaloutti A.: Pipeline Corrosion, Modeling and Analysis, *Journal of Nondestructive Evaluation*, 01.2011.
3. Bose N.K., Kim H.C., Venzuela H.M.: Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes, *IEEE Conference Acoustic, Speech and Signal Processing*, Mineapolis 1993
4. Irani M., Peleg S.: Improving Resolution by Image Registration, *Graphical Models and Image Processin*, Vol. 53, 1991, s. 231-239
5. Chan R.H., Chan T.F., Shen L., Shen Z.: Wavelet algorithms for high-resolution image reconstruction, *J. Sci. Comp.* Vol 24, No 4, pp. 1408-1432



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Work co-financed by European Social Fund

Homogenization of Electrical Machines Stator Windings

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. The paper describes a method for obtaining an equivalent thermal model of a winding with impregnation resin, in order to simplify the calculation of temperatures in electrical machines.

1 Introduction

Thermal performance is one of the main factors limiting electrical machines output capability. Thermal phenomena determine volume power density and lifetime of electrical motors. Recently, in the thermal analysis of electrical machines, field methods are used [1]. One of the main problems in thermal study of electrical machines concerns modeling windings, where the temperature rises the highest value. However, use of the finite element method (FEM) to locate hot spots in slots is associated with high cost of numerical analysis. Due to heterogeneous structure of winding it is needed to apply a highly detailed and dense meshed numerical model. It is especially for windings comprising individual copper conductors, enamel, impregnation resin and air. In such cases proper modeling needs subdivision the slots regions into separate parts representing each material. The alternative approach might be a model wherein the winding is represented by the homogenous region, which allows to reduce the computation cost. The objective of the present study is to replace the winding with randomly placed conductors by the homogeneous material which results in the same thermal properties [3].

2 Equivalent thermal conductivity of the slot

Proper modeling of the windings, in particular the winding with randomly placed coils, poses much difficulty due to the high complexity of this element. In order to determine the effective thermal conductivity λ_{ei} of the winding with impregnation resin it was decided to build a set of two coils with and without impregnation resin (Fig.1a). The coils were arranged in the slot of the stator segment. In both coils there were placed five K-type thermocouples (T_1 - T_5) along the height of the slot (Fig.1b).

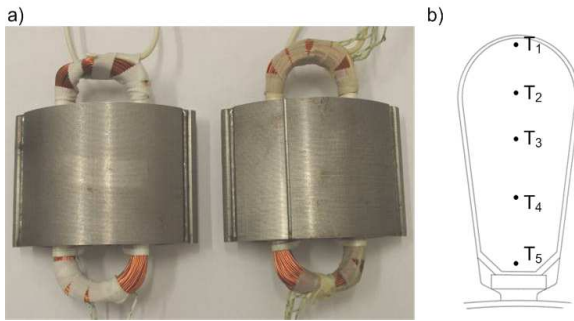


Fig. 1. Constructed coils (a) and arrangement of the thermocouples in the coil (b)

Figure 2 shows a schematic diagram of the proposed method of determining the thermal conductivity of the impregnated windings.

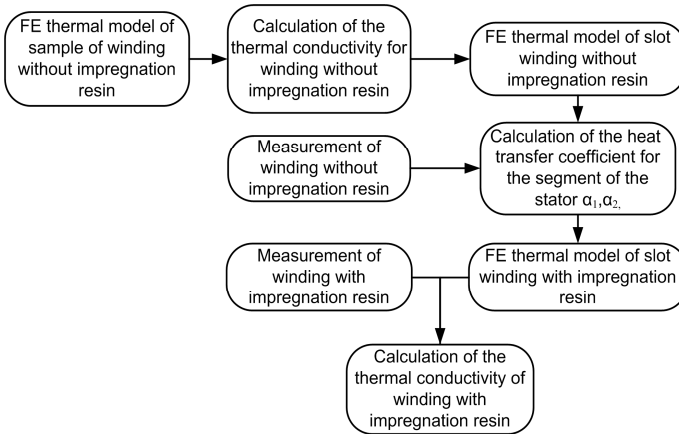


Fig. 2. The proposed method of determining the thermal conductivity of impregnated windings

The first step is the homogenization of the conductors with the enamel. In order to determine the equivalent thermal conductivity of the conductors with enamel the formula presented below was used:

$$\lambda_{ec} = \left(\frac{d' - \delta_e}{2\delta_e} \right) \lambda_e \tag{1}$$

where: d' – diameter of the conductor with enamel, δ_e – thickness of the enamel layer, λ_e – thermal conductivity of enamel.

Next, the effective thermal conductivity of the winding taking the random distribution of the conductors into account was determined. For that purpose a numerical model of the representative sample of the winding was built. Utilizing the elaborated model the effective thermal conductivity of the winding without impregna-

tion resin λ_{eni} can be determined. Imposing boundary conditions shown at Fig. 3.a, in x-axis direction for example, this quantity was obtained using following formula [2]:

$$\lambda_{eni\ x} = \frac{q_x \cdot l}{\vartheta_1 - \vartheta_2} \tag{2}$$

where: q_x – heat flux transferred across the wall, l – the thickness of the wall, ϑ_1, ϑ_2 – the temperatures of the walls, $\lambda_{eni\ x}$ – thermal conductivity of the winding in the x-axis direction.

For the model defined with the slot fill factors $f=51\%$ the coefficients of thermal conductivity determined in x- and y axis direction are: $\lambda_{eni} = 0.1 \text{ W/(mK)}$.

Knowing the power losses in the coil without impregnation and its thermal conductivity the heat transfer coefficients α form the stator segment can be estimated. In the analyzed problem the coefficient α for the outer surface of the stator segment $\alpha_1 = 10 \text{ W/(m}^2\text{K)}$, while for the remaining surface $\alpha_2 = 6 \text{ W/(m}^2\text{K)}$ (Fig. 3b).

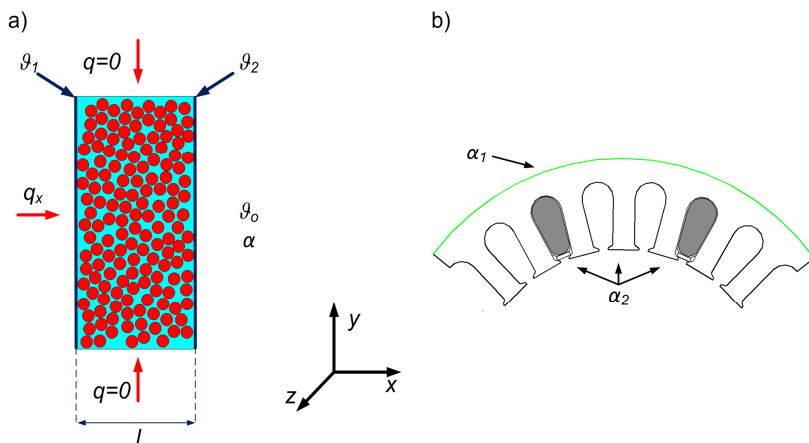


Fig. 3. FE model of the winding sample (a); The values of the coefficient of heat transfer for stator segment (b)

While maintaining the same measurement conditions for both coils it is also possible to use these heat transfer factors for coils with impregnation resin. Then, based on the impregnated coil FE thermal model and the results of measurements the coefficient of thermal conductivity of the impregnated winding λ_{ei} , were estimated. Their values are equal 0.18 W/(mK) .

Figure 4 shows the temperature distribution along the height of the stator slot obtained from the FE thermal model, which was compared with the measurement results. As it can be seen from the figure the error between the calculations and measurements results in the middle of the slot did not exceed 0.2°C . The differences in the other measurement points are probably due to the nonuniform distribution of the winding wires in the slot and the technical difficulties in direct and uniform distribution of the thermocouples along the height of the slot.

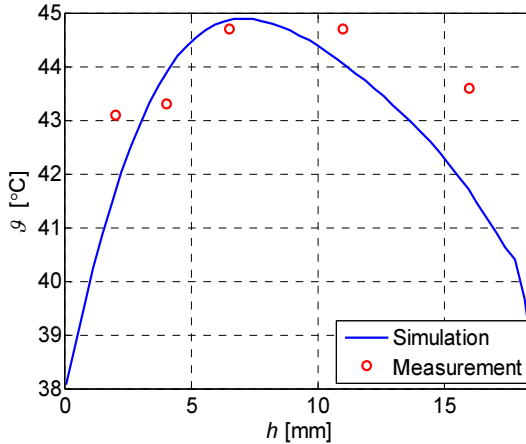


Fig. 4. Variation of the temperature along the height of the stator slot

3 Summary

The method of homogenization of the impregnated winding with randomly placed conductors was presented in the paper. The method of modeling the randomly wound coils in case of presence of slot insulation results in better representation of the temperature distribution over the stator slot. The method of analysis is characterized by relatively low computation costs.

References

1. Boglietti A., Cavagnino A., Staton D., Shanel M., Mueller M., Mejuto C.: *Evolution and Modern Approaches for Thermal Analysis of Electrical Machines*, IEEE Transactions on Industrial Electronics, vol. 56, no. 3, March 2009, pp.871-882.
2. Idoughi L., Mininger X., Bouillault F., Bernard L., Hoang E.: *Thermal Model with Winding Homogenization and FIT Discretization for Stator Slot*, IEEE Transactions on Magnetics, vol. 47, no. 12, December 2011, pp. 4822-4826.
3. Wrobel R., Mellor P. H., Holliday D.: *Thermal Modeling of a Segmented Stator Winding Design*, IEEE Transactions on Industry Applications, vol. 47, no. 5, September/October 2011, pp. 2023-2030.



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



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

StimSys: functional MRI Stimulation System

Jan Rydlo^{1,2}, Pavel Dvořák^{1,3}, and Jaroslav Tintěra¹

¹Institute for Clinical and Experimental Medicine, Vídeňská 1958/9, 140 21 Prague 4, Czech Republic

²Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Břehová 7, 115 19 Prague 1, Czech Republic

³Czech Technical University in Prague, Faculty of Electrical Engineering, Technická 2, 166 27 Prague 6, Czech Republic
{rydj, jati}@ikem.cz

Abstract. This paper describes stimulation system designed for functional magnetic resonance imaging (fMRI). In the Radiodiagnostic and Interventional Radiology Department (ZRIR) of the Institute for Clinical and Experimental Medicine (IKEM) in Prague a system for stimulating brain activity in fMRI experiments has been developed. In usual cases, such system is not a standard part of MRI scanners. Therefore, to be able to fulfil new requirements in research projects and clinical fMRI examinations a new Stimulation System (StimSys) had to be designed and implemented. The StimSys system incorporates hardware part and Java-based control software. The hardware consists of a scanner room transceiver and StimSys module equipped with transmitters and receivers for optical data transmission, and a converter (Arduino Mega 2560) for communication with dedicated computer.

Keywords: StimSys, fMRI, functional MRI, Arduino

1 Introduction

Functional MRI (fMRI) is a magnetic resonance imaging based neuroimaging technique which allows to detect the brain areas being activated during a task, a process, or an emotion by detecting associated changes in blood flow. This technique is based on the fact that cerebral blood flow and neuronal activation are coupled. When a region of the brain is activated, blood flow to that region increases (hemodynamic response). Change of oxygenated and deoxygenated haemoglobin concentration ratio causes distinguishable increase of MR signal called BOLD contrast (Blood Oxygen Level Dependent). This contrast enables mapping neural activity in the brain related to energy use by brain cells [1].

Particular fMRI images of the brain are acquired in different conditions – in a state of rest or during stimulation. Each of the condition is repeated several times. The stimulation is mostly evoked by an external stimulus, which activates certain brain regions. In fMRI measurements it is necessary to synchronize alternation of the rest states and

the stimulation with the image acquisition for later statistical evaluation. Also recording of the examined subjects responses to the stimuli with the precise time of these responses are required.

The system for automatic synchronization is implemented by StimSys system. It incorporates software and hardware part. The control software allows creating audio as well as visual sequences to be presented to the examined subject. Design of the hardware (see Fig. 1.) is influenced by possible disturbance of the scanner by external radiofrequency (RF) signals. Every conductor directly led into the magnet room transmits undesirable interferences from the outside causing artifacts and noise in the acquired MR images. The most convenient and reliable possibility to resolve this is using an optical fiber connection between the magnet room and the dedicated computer. The next possibility is the combination of a metal wire and a band-rejection filter.

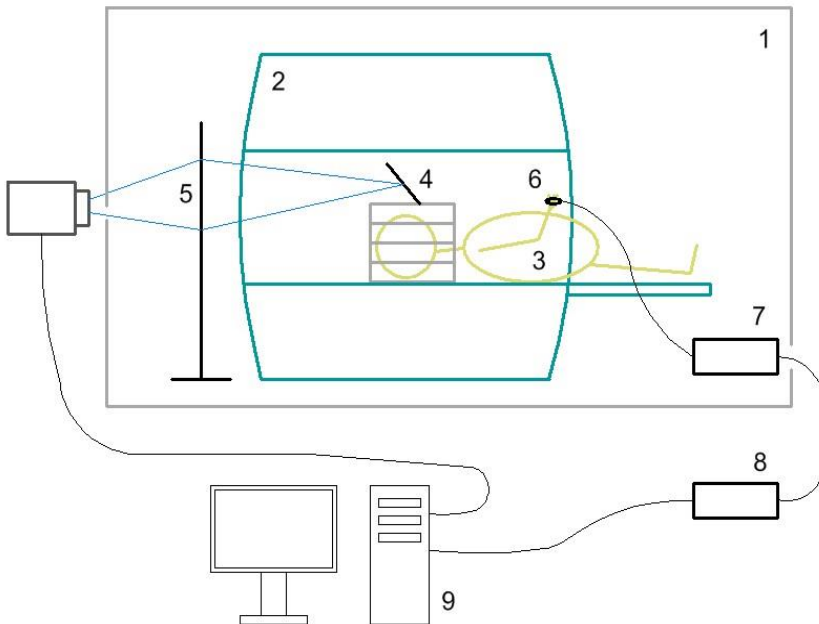


Fig. 1. Schematic organization of the fMRI experiment. 1. the shielded magnet room, 2. the MR scanner, 3. the examined subject, 4. a head coil equipped with a mirror, 5. a screen for rear projection, 6. a response pad, 7. the magnet room transceiver, 8. the StimSys module, 9. the control computer.

2 Hardware StimSys

The hardware of the StimSys system consists of two modules. The first module (magnet room transceiver) was designed to be a separate device in the magnet room. It is supplied from a rechargeable lead-acid battery to avoid use of metallic line conducted from the outside of the shielded magnet room. Fig 2. shows the block diagram of the transceiver.

The 12 V and 2 Ah battery is suitable for the power supply due to the 60 mA optical transmitter HFBR-1528 current consumption. The battery is protected from deep discharge by the battery control board. A connector for an external charger is connected to the battery. While the charger is plugged in, this connector disconnects all boards' power supply and only the battery is connected to the charger. The voltage indicator shows the battery voltage on a LED display. The transmitter board fitted with four optical transmitters HFBR-1528 is connected to D-SUB periphery connector and LED indicators. The periphery is a MR compatible device for examined subject feedback, e. g. response buttons. The receiver board fitted with four optical receivers HFBR-2522 is connected to the switch board which enables control optional external devices using relays.

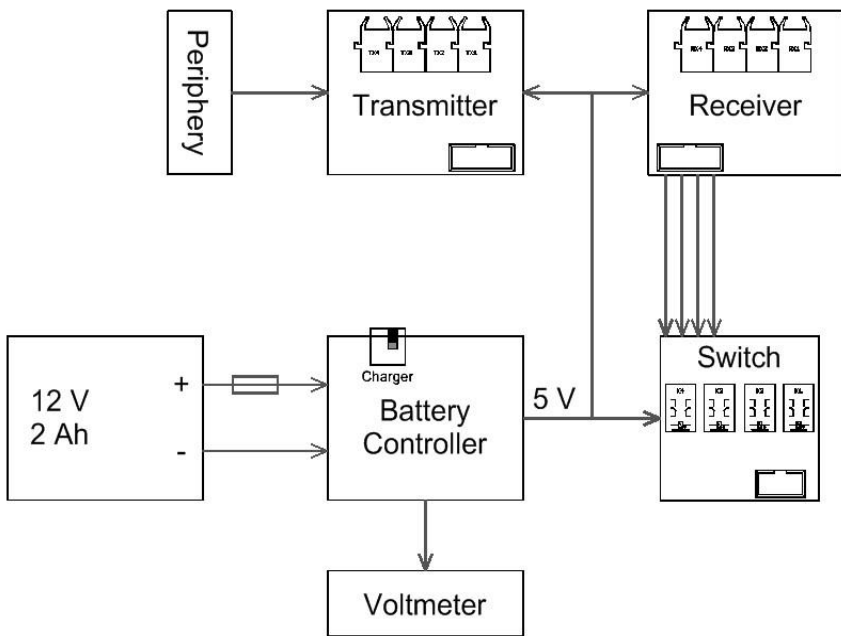


Fig. 2. Block diagram of the scanner room transceiver.

The second module (StimSys module) was designed to be a device connected to the dedicated computer. It is supplied from the computer USB interface. Fig 3. shows the block diagram of the device.

The trigger board can receive scanner synchronization pulse from optical fiber (connected Siemens MR system trigger outputs) using optical receiver HFBR-2528 or from coax cable TTL level signal using optocoupler 6N137. A monostable multivibrator prolongs the pulse to a duration of about units of milliseconds because of the sampling rate limit of Arduino converter. The receiver board fitted with four optical receivers HFBR-2522 is connected to the Arduino converter. The converter controls the switch board inside the nodule and also controls the transmitter board fitted with four optical transmitters HFBR-1528 to control the switch board in the magnet room transceiver. Two

HFBR-4505 optic connectors equipped with phototransistors are connected to Arduino converter analog input pins. These inputs are used for receiving signals (axes x and y) from MR compatible joystick which is implemented as a separate device inside the magnet room. The module incorporates also the audio board which indicates power of sound delivered to the subject by scanner audio system [2].

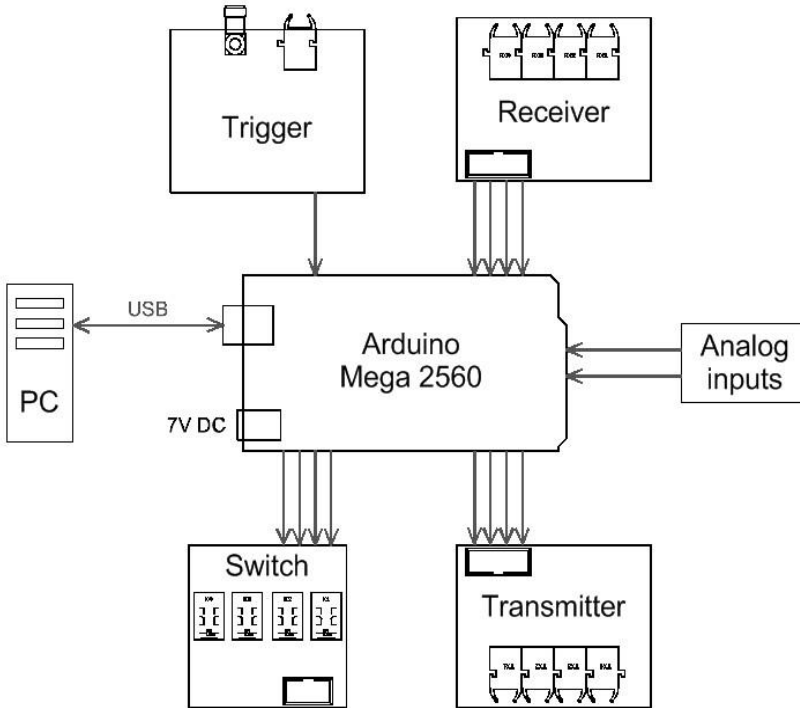


Fig. 3. Block diagram of the StimSys module.

3 Software StimSys

The StimSys software has been built to facilitate the design of paradigms for fMRI. It registers scanner synchronization pulses, performs presentation of paradigms according to previously set parameters, records digital (e. g. response buttons) and analog (e. g. joystick) signals, and drives digital outputs used for control external devices from the magnet room transceiver and from the StimSys module. The graphical user interface was designed to be user friendly according to user demands (see Fig. 4.).

An open source electronic platform Arduino was chosen to interface the application with the used hardware, specifically the Arduino Mega 2560 board. This board provides sufficient time accuracy and also meets the requirements for easy settings. The Arduino communicates with computer means of USB virtual serial port.

The StimSys software allows creating audio as well as visual (images or movies) sequences to be presented to the subject. It can also display any grid or image as a background for cursor movement controlled by the joystick. Important feature of the software is possibility to split repetition time (TR) to any number of intervals and execute more stimuli during one fMRI image acquisition. To avoid an undesirable delay of the stimulation, all data are loaded before the fMRI image acquisition starts and are only displayed on screen during the acquisition. The image presentation is delayed about units of milliseconds, the audio sequence is delayed not more than 50 ms and movies not more than 100 ms. Due to these delays the measurement is sufficiently accurate. During the image acquisition digital and analog data are registered from the inputs. These data are saved to a text. Pairs of values are saved from digital inputs (start time and duration of onset), from analog inputs any time and value are saved.

The StimSys application uses external libraries

- jSSC (Java Serial Connector) for serial communication with the Arduino
- JMF (Java Media Framework) for playing the audio files
- FOBS library for playing movies in AVI format

The application was developed in Java language and emphasis was laid on easy service and future extensions. All files are structured to packages and also Model-View-Controller pattern was used for implementing user interfaces [3].

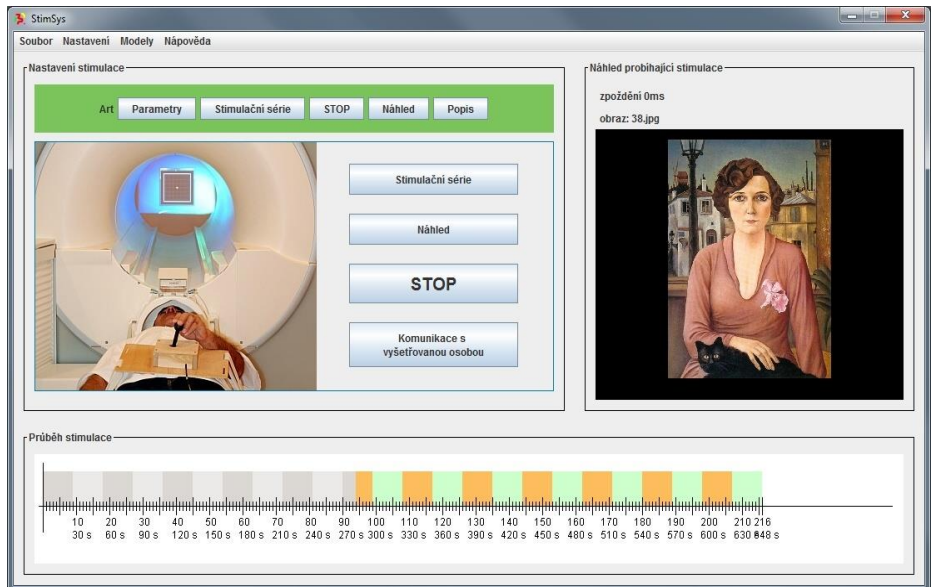


Fig. 4. Graphical user interface for control the fMRI experiment.

4 Conclusion

The StimSys hardware and software was designed, implemented and tested. After tuning the system has been commonly used for the fMRI experiments in IKEM in the

research projects and the clinical examinations. The entire system can be extended according to user requirements.

Acknowledgments

This work was supported by IGA-MZCR-NT14291, GACR P407/13-23940, and MZCR00023001IKEM.

References

1. HEUTTEL, Scott A.; SONG, Allen W.; MCCARTHY, Gregory. *Functional Magnetic Resonance Imaging*. Second Edition. Sunderland (MA.): Sinauer Associates, Inc., 2009. 542 s.
2. RYDLO, Jan. *Extension of stimulating system for fMRI*. Master degree thesis. Faculty of Electrical Engineering, CTU in Prague. Prague 2011
3. DVOŘÁK, Pavel. *Stimulation software for functional magnetic resonance imaging*. Master degree thesis. Faculty of Electrical Engineering, CTU in Prague. Prague 2014

Author Index

- Babskova, Alisa, 311
Baresova, Katerina, 138
Bednarek, Lukas, 568
Bednář, Pavel, 316
Beránek, Eduard, 1
Berek, Petr, 322
Bielik, Marian, 328
Blaheta, Radim, 484
Borski, Martin, 144
Bures, Miroslav, 574
Burianek, Tomas, 334
- Carbol, Zbyněk, 24
Cernohorsky, Jindrich, 300
Cerny, Martin, 300
Cigánek, Jiří, 6
Ćosić, Rajko, 466
Cubik, Jakub, 584
- Dočekal, Tomáš, 162, 174
Dohnálek, Pavel, 340
Drábek, Petr, 168
Drholec, Jiří, 12
Dvořák, Pavel, 687
- Đurica, Maroš, 18
- Fajkus, Marcel, 590
Farana, Radim, 434
Friedrischková, Kristýna, 162, 174, 181
Fusek, Radovan, 346, 352
- Gasz, Rafał, 671
Grunt, Ondřej, 472
- Hajek, Lukas, 596
Hajovsky, Jiri, 102
Hasal, Martin, 275, 478
Haska, Jiri, 187
Haslinger, Jaroslav, 484
Hejduk, Stanislav, 604
Helštýnová, Barbara, 24
Hlavinka, Tomas, 596
Holba, Miloslav, 203
Holis, Michal, 411
Holuša, Michael, 352
Holý, Branislav, 358
Horáček, Ondřej, 203
- Horák, Bohumil, 174, 198, 245, 263, 269, 305
Houdek, Vít, 30
Hradílek, Zdeněk, 66
Hrtus, Rostislav, 484
Hrubý, David, 666
Hrvolová, Barbora, 193
- Janoušek, Jan, 364
Janša, Jiří, 36
Jaros, Jakub, 626
Jaroš, Milan, 490
Jeżowicz, Tomáš, 370
Junwen, Shen, 181
- Kajnar, Tomáš, 666
Kalina, Jiří, 193
Kazárik, Jiří, 198, 305
Kelnar, Michal, 120, 203
Kepák, Stanislav, 610
Košíňár, Michael A., 376
Kovalev, Sergey M., 287
Kravčenko, Michal, 496
Krohova, Jana, 209, 281
Kurecka, Ales, 215
- Lasar, Miroslaw, 677
Latal, Jan, 596
Liner, Andrej, 614, 626
Lukáš, Petr, 383
- Malý, Lukáš, 502
Marcincin, Oliver, 42
Meca, Ondřej, 428
Meca, Roman, 389
Merta, Michal, 508
Michalikova, Marketa, 221
Mikolajek, Martin, 227
Mikulec, Martin, 590
Motyčková, Kristina, 514
Mozdřeň, Tomáš, 48
Mrovec, Tomas, 150
Mynarek, Piotr, 683
- Nevlud, Pavel, 574
Ney, Michal, 54
Novák, Tomáš, 24
Nowaková, Jana, 233, 275

Nudzikova, Pavlina, 239

Otevřel, Vít, 245, 263, 269
Otýpka, Jan, 108, 287

Papes, Martin, 568
Partila, Pavol, 620, 654
Perecar, Frantisek, 626
Peter, Lukas, 251
Petrtyl, Ondřej, 114
Platos, Jan, 322
Poboril, Radek, 632
Pokorný, Miroslav, 233
Pospíšil, Lukáš, 520

Rajmanová, Petra, 257
Rapant, Lukas, 526
Rawicki, Jakub, 181, 263
Režnar, Tomáš, 532
Ronovsky, Ales, 556
Rozhon, Jan, 638
Rusnok, Stanislav, 120
Rydlo, Jan, 687

Safarik, Jakub, 644
Shen, Junwen, 269
Silber, Adam, 538
Skanderova, Lenka, 395
Slabý, Roman, 275
Slíva, Ladislav, 126
Slivka, Martin, 60
Složil, Zbyněk, 401, 417
Smoček, Martin, 66
Smondrk, Maros, 209, 281
Snasel, Vaclav, 322
Sobota, Pavel, 120
Sojka, Eduard, 346
Sokanský, Karel, 24
Stachoň, Martin, 407
Stolfa, Jakub, 411
Stříbný, Daniel, 401, 417
Stuchly, Jindrich, 72

Sukhanov, Andrey V., 287

Špaček, Michal, 78
Špička, Jiří, 144
Štrba, Radoslav, 422
Šulák, Radek, 156
Šurkovský, Martin, 428

Theuer, Matyáš, 544
Tintěra, Jaroslav, 687
Tomeček, Jan, 434
Tomis, Martin, 650
Tomis, Radek, 550
Tovarek, Jaromir, 620, 654

Uher, Marián, 84
Uher, Vojtěch, 441

Vaculík, Jan, 90
Vala, David, 294
Vanderka, Ales, 596
Vantuch, Tomas, 328, 448
Vasatova, Alena, 556
Vasinek, Vladimir, 626
Vašinek, Michal, 454
Verner, Tomáš, 132
Vítek, Aleš, 407
Vozda, Michal, 300
Voznak, Miroslav, 620, 654
Vramba, Jakub, 96
Vukojević, Boris, 181, 305

Winkelhofer, Petr, 568

Yu, Tiantian, 181

Zaorálek, Lukáš, 460
Zapletal, Jan, 562
Závodný, Petr, 660
Zbořil, Ondřej, 666
Zdrálek, Jaroslav, 574
Zelinka, Ivan, 395

Editors: Michal Krátký, Jiří Dvorský, Pavel Moravec
Dept. of Computer Science
VŠB – Technical University of Ostrava

Title: WOFEX 2014

Place, year, edition: Ostrava, 2014, 1st

Page count: 711

Edit: VŠB – Technical University of Ostrava,
17. listopadu 15, 708 33 Ostrava-Poruba
Czech Republic

Impression: 200

Not for sale.

ISBN 978-80-248-3458-0