

Roman Szewczyk
Cezary Zieliński
Małgorzata Kaliczyńska *Editors*

Recent Advances in Automation, Robotics and Measuring Techniques

Advances in Intelligent Systems and Computing

Volume 267

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

For further volumes:
<http://www.springer.com/series/11156>

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc.Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

Roman Szewczyk · Cezary Zieliński
Małgorzata Kaliczyńska
Editors

Recent Advances in Automation, Robotics and Measuring Techniques

Editors

Roman Szewczyk
Industrial Research Institute for Automation
and Measurements PIAP
Warsaw
Poland

Małgorzata Kaliczyńska
Industrial Research Institute for Automation
and Measurements PIAP
Warsaw
Poland

Cezary Zieliński
Industrial Research Institute for Automation
and Measurements PIAP
Warsaw
Poland

ISSN 2194-5357
ISBN 978-3-319-05352-3
DOI 10.1007/978-3-319-05353-0
Springer Cham Heidelberg New York Dordrecht London

ISSN 2194-5365 (electronic)
ISBN 978-3-319-05353-0 (eBook)

Library of Congress Control Number: 2014932680

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Broadly perceived control, automation, robotics and measuring techniques belong to the most relevant fields of science and technology, both from the point of view of theoretical challenges and practical importance. In spite of being separate areas of research, knowledge and expertise, they are strongly related, both in terms of paradigms and tools and techniques employed, as well in terms of their industrial scope of applications. Therefore, an industrial, practice oriented perspective is an important aspects of those areas. Moreover, automation, robotics and measuring techniques have a significant innovative potential as the current industrial practice calls for a further integration of all kinds of production systems, more ecological and energy efficient solutions as well as cost and time effective production and manufacturing processes.

Among many important problems and challenges faced by automation and control, most of which have been reflected in the scope of the papers included in this volume, one can mention, for instance, discrete systems, actuators, diagnostics, and modern tools exemplified by fuzzy logic, evolutionary computation, neural networks, probabilistic approaches, etc.

In robotics, in particular in its part related to the development of mobile robots, one can quote as crucial problems and challenges various problem solving tasks related to the control of walking robots, control of manipulators, motors and drivers, mechatronic systems, and tracking control.

Measuring techniques and systems have to overcome, first of all, barriers implied by environmental conditions and limitations. They call for the development of novel sensors (also utilizing novel materials such as graphene), advanced signal processing and a more foundational development focused on the theory of metrology.

This book presents the recent advances and developments in control, automation, robotics, and measuring techniques that are trying to meet those challenges and to fulfil those technological, economic and social needs. It presents contributions of top experts in the fields, focused on both theory and industrial practice. The particular chapters present a deep analysis of a specific technical problem which is in general followed by a numerical analysis and simulation, and results of an implementation for the solution of a real world problem.

We strongly believe that the presented theoretical results, practical solutions and guidelines will be useful for both researchers working in the area of engineering sciences and for practitioners solving industrial problems.

Warsaw
January 2014

Roman Szewczyk
Cezary Zieliński
Małgorzata Kaliczyńska

Editors

Professor Roman Szewczyk received both his Ph.D. and D.Sc. in the field of mechatronics. He is specializing in the modelling of properties of magnetic materials as well as in sensors and sensor interfacing, in particular magnetic sensors for security applications. He is the leading the development of a sensing unit for a mobile robot developed for the Polish Police Central Forensic Laboratory and of methods of non-destructive testing based on the magnetoelastic effect. Professor Szewczyk was involved in over 10 European Union funded research projects within the FP6 and FP7 as well as projects financed by the European Defence Organization. Moreover, he was leading two regional and national scale technological foresight projects and was active in the organization and implementation of technological transfer between companies and research institutes. Roman Szewczyk is Secretary for Scientific Affairs in the Industrial Research Institute for Automation and Measurements (PIAP). He is also Associate Professor at the Faculty of Mechatronics, Warsaw University of Technology and a Vice-chairman of the Academy of Young Researchers of the Polish Academy of Sciences.

Professor Cezary Zieliński received his M.Sc./Eng. degree in control in 1982, Ph.D. degree in control and robotics in 1988, the D.Sc. (habilitation) degree in control and robotics in 1996, all from the Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland, and Full Professorship in 2012. Currently he is Full Professor both in the Industrial Research Institute for Automation and Measurement (PIAP) and the Warsaw University of Technology, where he is Director of the Institute of Control and Computation Engineering. Since 2007 he has been a member of the Committee for Automatic Control and Robotics, the Polish Academy of Sciences. Professor Zieliński is Head of the Robotics Group in the Institute of Control and Computation Engineering working on robot control and programming methods. His research interests focus on robotics in general and in particular include: robot programming methods, formal approach to the specification of architectures of multi-effector and multi-receptor systems, robot kinematics, robot position-force control, visual servo control, and design of digital circuits. He is the author/coauthor of over 160 conference and journal papers as well as books concerned with the above mentioned research subjects.

Dr. Małgorzata Kaliczyńska received her M.Sc./Eng. degree in cybernetics from the Faculty of Electronics, Wrocław University of Technology, and her Ph.D. degree in the field of fluid mechanics from the Faculty of Mechanical and Power Engineering in this same university. Now she is Assistant Professor in the Industrial Research Institute for Automation and Measurement (PIAP) and Editor of the scientific and technological magazine “Measurements, Automation, Robotics”. Her areas of research interest include distributed control systems, information retrieval and webometrics.

Contents

Part I: Control and Automation

Application of Artificial Neural Network for Modelling of Electrohydraulic Drive	3
<i>Mirosław Adamczyk, Andrzej Milecki</i>	
Cyclic Steady State Space Refinement	11
<i>Grzegorz Bocewicz, Zbigniew Banaszak, Paweł Pawlewski</i>	
Using Fuzzy Logic for Improving the Control Performance of Digital Servo Drives with Elastic Coupling	21
<i>Bogdan Broel-Plater</i>	
Chaos Synchronization of the Modified Van der Pol-Duffing Oscillator of Fractional Order	33
<i>Mikołaj Busłowicz, Adam Makarewicz</i>	
Stability Analysis of Descriptor Continuous-Time Two-Term Linear Systems of Fractional Orders	45
<i>Mikołaj Busłowicz</i>	
Design of Integrated Information Systems for the Security of People and Objects	55
<i>Małgorzata Cupriak, Sławomir Jasiński, Małgorzata Kaliczyńska</i>	
The Systematized Data Structures Oriented Towards Diagnosis and Prediction	63
<i>Mariusz Piotr Hetmańczyk, Jerzy Świder, Grzegorz Wszędek</i>	
The Analysis of the Registration Accuracy of Distributed Drives Parameters	73
<i>Mariusz Piotr Hetmańczyk, Jerzy Świder, Grzegorz Wszędek</i>	

CPDev Engineering Environment for Modeling, Implementation, Testing, and Visualization of Control Software	81
<i>Marcin Jamro, Dariusz Rzońca, Jan Sadolewski, Andrzej Stec, Zbigniew Świder, Bartosz Trybus, Leszek Trybus</i>	
Development and Execution of POU-Oriented Performance Tests for IEC 61131-3 Control Software	91
<i>Marcin Jamro</i>	
A New Formulation and Solution of the Minimum Energy Control Problem of Positive 2D Continuous-Discrete Linear Systems	103
<i>Tadeusz Kaczorek</i>	
The Impact of an ERP System on the Technical Preparation of Production	115
<i>Sławomir Kłos</i>	
Minimum Energy Control of Fractional Discrete-Time Linear Systems with Delays in State and Control	127
<i>Rafał Kociszewski</i>	
Efficient Mechanism of Output Constraint Handling for Analytical Predictive Controllers Based on Hammerstein Models	137
<i>Piotr M. Marusak</i>	
Piezoceramic Transformer Based Ionization-Deionization System	147
<i>Piotr Mateusiak, Roman Szewczyk</i>	
The Rapid Prototyping of Active Magnetic Bearings	155
<i>Paulina Mazurek, Maciej Henzel</i>	
Application of the MFC Method in Electrohydraulic Servo Drive with a Valve Controlled by Synchronous Motor	167
<i>Andrzej Milecki, Dominik Rybarczyk, Piotr Owczarek</i>	
Electromechanical Actuators – Selected Safety-Related Problems	175
<i>Tadeusz Missala</i>	
Immune Algorithm for Fuzzy Models Generation	187
<i>Bogumiła Mrozek</i>	
Augmented Reality of Technological Environment in Correlation with Brain Computer Interfaces for Control Processes	197
<i>Szczepan Paszkiel</i>	

Modeling and Dynamic Analysis of the Precise Electromechanical Systems Driven by the Stepping Motors	205
<i>Agnieszka Pęgowska, Tomasz Szolc, Andrzej Pochanke, Robert Konowrocki</i>	
Practical Stability and Asymptotic Stability of Interval Fractional Discrete-Time Linear State-Space System	217
<i>Andrzej Ruszewski</i>	
Interfacing Inputs and Outputs with IEC 61131-3 Control Software	229
<i>Dariusz Rzońca, Jan Sadolewski, Bartosz Trybus</i>	
Reachability of Fractional Positive Continuous-Time Linear Systems with Two Different Fractional Orders	239
<i>Łukasz Sajewski</i>	
A Hybrid Approach to the Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP)	251
<i>Paweł Sitek</i>	
Identification of Thermal Response, of Plasmatron Plasma Reactor	265
<i>Jakub Szałatkiewicz, Roman Szewczyk, Eugeniusz Budny, Tadeusz Missala, Wojciech Winiarski</i>	
Computational Problems Connected with Jiles-Atherton Model of Magnetic Hysteresis	275
<i>Roman Szewczyk</i>	
The Tester of the Actuator with ARINC 429 Data Bus	285
<i>Ewelina Szpakowska-Peas</i>	
Relay Self-tuning of Industrial PID Temperature Controller with Set-Point Weighting	295
<i>Leszek Trybus, Zbigniew Świdler, Andrzej Stec</i>	
Pointwise Completeness and Pointwise Degeneracy of Linear Continuous-Time Systems with Different Fractional Orders	307
<i>Wojciech Trzasko</i>	
Implementation Aspects of Hybrid Solution Framework	317
<i>Jarosław Wikarek</i>	
On Choice of the Sampling Period and the Horizons in Generalized Predictive Control	329
<i>Antoni Wysocki, Maciej Ławryńczuk</i>	

Part II: Robotics

A Compact Walking Robot – Flexible Research and Development Platform	343
<i>Dominik Belter, Krzysztof Walas</i>	
Towards Practical Implementation of an Artificial Force Method for Control of the Mobile Platform Rex	353
<i>Mateusz Cholewiński, Krzysztof Arent, Alicja Mazur</i>	
Static Modeling of Multisection Soft Continuum Manipulator for Stiff-Flop Project	365
<i>Jan Fraś, Jan Czarnowski, Mateusz Maciąś, Jakub Główka</i>	
Estimation of Altitude and Vertical Velocity for Multirotor Aerial Vehicle Using Kalman Filter	377
<i>Przemysław Gaśior, Stanisław Gardecki, Jarosław Gościński, Wojciech Giernacki</i>	
The Influence of the End Effector Gyroscopic Torques on a Base of the Manipulator	387
<i>Jarosław Gościński, Stanisław Gardecki, Wojciech Giernacki</i>	
A Virtual Receptor in a Robot Control Framework	399
<i>Włodzimierz Kasprzak, Tomasz Kornuta, Cezary Zieliński</i>	
Kinematic Structures of Functional Assemblies of the Table for Patients Verticalization with Lower Limbs Rehabilitation Functions	409
<i>Wojciech J. Klimasara, Dariusz Grabowski</i>	
Basic 3D Solid Recognition in RGB-D Images	421
<i>Tomasz Kornuta, Maciej Stefańczyk, Włodzimierz Kasprzak</i>	
Selected Issues of Collecting Forensic Evidence with a Mobile Robot	431
<i>Grzegorz Kowalski, Mateusz Maciąś, Adam Wołoszczuk</i>	
Lesson Learned from Eurathlon 2013 Land Robot Competition	441
<i>Karol Majek, Paweł Musialik, Piotr Kaczmarek, Janusz Będkowski</i>	
Analysis of Thrust of Underwater Vehicle with Undulating Propulsion	453
<i>Marcin Malec, Marcin Morawski</i>	
Direct Local Communication for Distributed Coordination in a Multi-robot Team	463
<i>Marta Rostkowska, Michał Topolski, Piotr Skrzypczyński</i>	

TALOS – Mobile Surveillance System for Land Borders and Large Areas	475
<i>Agnieszka Sprońska, Jakub Główka, Mateusz Maciąś, Tomasz Rokosz</i>	
Localization of Essential Door Features for Mobile Manipulation	487
<i>Maciej Stefańczyk, Michał Walecki</i>	
Motion Planning for the Mobile Platform Rex	497
<i>Krzysztof Tchoń, Krzysztof Arent, Mariusz Janiak, Łukasz Juszkiejewicz</i>	
Trajectory Tracking Control of a Four-Wheeled Mobile Robot with Yaw Rate Linear Controller	507
<i>Maciej Trojnacki, Przemysław Dąbek, Janusz Kacprzyk, Zenon Hendzel</i>	
Universal Control System for Managing Multiple Unmanned Engineering Machines	523
<i>Rafał Typiak</i>	
Motor Cascade Position Controllers for Service Oriented Manipulators	533
<i>Tomasz Winiarski, Michał Walecki</i>	
Specification of Tasks in Terms of Object-Level Relations for a Two-Handed Robot	543
<i>Cezary Zieliński, Tomasz Kornuta</i>	
Using Integrated Vision Systems: Three Gears and Leap Motion, to Control a 3-finger Dexterous Gripper	553
<i>Igor Zubrycki, Grzegorz Granosik</i>	
Part III: Measuring Techniques and Systems	
Influence of the Environment on Operation of Checkweigher in Industrial Conditions	567
<i>Piotr Bazydło, Michał Urbański, Marcin Kamiński, Roman Szewczyk</i>	
Method for Limitation of Disturbances in Measurement Data in 3D Laser Profilometry	579
<i>Piotr Czajka, Wojciech Mizak, Jacek Galas, Adam Czyżewski, Maciej Kochanowski, Dariusz Litwin, Maciej Socjusz</i>	
Preisach Based Model for Predicting of Functional Characteristic of Fluxgate Sensors and Inductive Components	591
<i>Piotr Frydrych, Roman Szewczyk</i>	
Automated Measurement Systems for Meters of Heat	597
<i>Tadeusz Goszczyński</i>	

Influence of Stresses on Magnetic B-H Characteristics of X30Cr13 Corrosion Resisting Martensitic Steel	607
<i>Dorota Jackiewicz, Roman Szewczyk, Jacek Salach, Adam Bieńkowski, Maciej Kachniarz</i>	
Anode Current Control in the Microwave Heating Equipment	615
<i>Marek Kuna-Broniowski, Piotr Makarski</i>	
The High-Resolution Camera in Estimation of the Position of the Hydraulic Valve Spool	623
<i>Piotr Owczarek, Dominik Rybarczyk, Jarosław Gośliński, Adam Owczarkowski</i>	
Sensitivity and Offset Voltage Testing in the Hall-Effect Sensors Made of Graphene	631
<i>Oleg Petruk, Roman Szewczyk, Tymoteusz Ciuk, Włodzimierz Strupiński, Jacek Salach, Michał Nowicki, Iwona Pasternak, Wojciech Winiarski, Krzysztof Trzcinka</i>	
Digitally Controlled Current Transformer with Hall Sensor	641
<i>Oleg Petruk, Roman Szewczyk, Jacek Salach, Michał Nowicki</i>	
Measuring Station for Testing of Graphene Flow Sensors	649
<i>Marcin Safinowski, Wojciech Winiarski, Kamil Domański, Oleg Petruk, Szymon Dąbrowski, Roman Szewczyk, Krzysztof Trzcinka</i>	
Influence of Tensile Force On Magnetic Properties of Amorphous Fe₈₀B₁₁Si₉ Alloys in Different States of Thermal Relaxation	665
<i>Jacek Salach, Dorota Jackiewicz, Adam Bieńkowski, Michał Nowicki, Magdalena Gruszecka</i>	
Application of X-ray Fluorescence to Determine Qualitative Parameters of Coal	677
<i>Waldemar Sobierajski, Marek Kryca, Artur Kozłowski</i>	
Measurement and Control System of the Plasmatron Plasma Reactor for Recovery of Metals from Printed Circuit Board Waste	687
<i>Jakub Szałatkiewicz, Roman Szewczyk, Eugeniusz Budny, Tadeusz Missala, Wojciech Winiarski</i>	
Influence of Operating Conditions on Functional Properties of High Resolution Analog to Digital Converter	697
<i>Krzysztof Trzcinka, Roman Szewczyk, Oskar Gińko</i>	
FPGA Based Processing Unit for a Checkweigher	713
<i>Robert Ugodziński, Łukasz Gosiewski, Roman Szewczyk</i>	

Improving of the Type A Uncertainty Evaluation by Refining the Measurement Data from a Priori Unknown Systematic Influences	721
<i>Zygmunt L. Warsza, Jerzy M. Korczyński</i>	
Evaluation of the Standard Deviation of the Random Component of the Measured Signal from Its Autocorrelated Observations	733
<i>Zygmunt L. Warsza</i>	
Author Index	743

Part I

Control and Automation

Application of Artificial Neural Network for Modelling of Electrohydraulic Drive

Mirosław Adamczyk and Andrzej Milecki

Poznań University of Technology
ul. Piotrowo 3, 60-965 Poznań, Poland

M.Adamczyk@iizp.uz.zgora.pl, Andrzej.milecki@put.poznan.pl

Abstract. The article describes the use of the Artificial Neural Network for modelling and simulation of electrohydraulic drive. The investigation test stand for this drive is presented and some investigation results are included. The structure of artificial neural network used for modeling is described and shortly discussed. The teaching procedure is described and some simulation results are presented. The accuracy of simulation results network are included.

Keywords: electrohydraulic drive, artificial neural network, modeling.

1 Introduction

Electrohydraulic servo drives are in their nature non-linear systems and therefore their design and control is not easy. That's why in their design and control, computer modelling methods play very fundamental role. In the literature there are several publications about electrohydraulic drives modelling [1–3]. In most cases the presented models are linear in nature, but only some non-linear blocks are added. In fact not all features of electrohydraulic drives is good enough recognized and described, thus the models and simulation results does not fits very well to characteristic obtained during laboratory investigations.

Neural networks are increasingly employed in a wide range of application such as modelling, control, classification, pattern recognition, signal processing, and may other areas [4, 5]. Feed forward ANN are based on neurons in hidden layers which compute a non-linear function of the scalar product of input vector and a weight vector. There are theoretical justifications that, if the network topology is sufficiently large with sufficient number of hidden layers and neurons, then any continuous function can be approximated by ANN by carefully choosing the parameters of the network [6]. In literature there are publications about applications of artificial neural network in modelling of devices dynamics and in their controllers [7–10]. Therefore authors decided to apply of such networks in modelling of electrohydraulic servo drives. Such models doesn't require the formulation of complicated derivative equations and the transformation them into simulation model. Also knowledge of accurate parameter values is not needed. Instead of special, rather expensive simulation software like, for example MATLAB/Simulink other, also freeware

software for neural network simulation software can be used. For preparing the neural network model, the electrohydraulic drive should be investigated at the laboratory test stand and the input and output signals should be recorded. Such investigations results will be then used as learning information for artificial neural network.

It has already been shown that neural network controllers offer advantages over the classical control methods [11–13]. After modelling, the control of the system can be, for example achieved solely by an inverse model of the process [14, 15] or by a combination of both, the inverse and forward models [16].

The most important problem which one faces during design of ANN design and application is the establishment of network structure i.e., number of layers, number of neurons in each layer and activation function of every neuron. At first, these parameters may be established basing on previous research results presented in literature and then corrected during tests. Another important parameter used in data preparation for teaching of ANN is the sampling frequency of the recorded signals. It's common practice that this parameters can be successfully established after several attempts, but one must be aware, that in cases which require high accuracy the selection of ANN parameters may be very difficult [7, 17].

2 Laboratory Tests of Electrohydraulic Drive and Signals Recording

In Figure 1 the scheme diagram and the photo of the test stand are presented. In the research electrohydraulic drive with servo valve with mechanical feedback linkage was used. The output motor was single rod hydraulic cylinder with stroke equal to 400 mm, piston diameter 100 mm and piston rod diameter 60 mm. The cylinder was controlled by servo valve type SM4-20. The drive actuated a lathe table with a mass about 150 kg (see Fig. 1). The actual position was measured with resolution 10 μm by incremental encoder type MSA 6803, which was connected to the computer by special input/output card. The same computer was acting as a servo drive controller i.e. input signal generator. The control and signal recording tasks were performed by DASYLab software, which was installed on the computer.

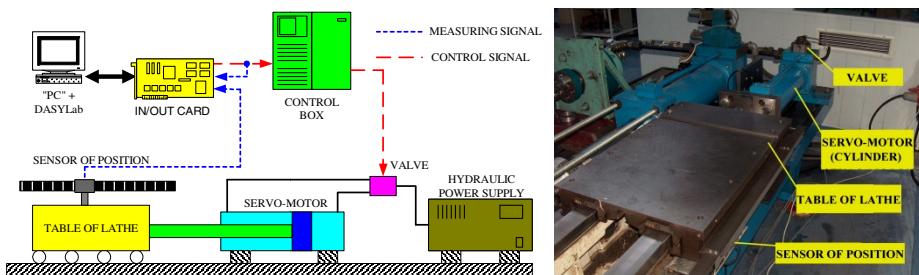


Fig. 1. Test stand: scheme block diagram and a photo

In the computer an input/output card type DaqBoard 3001 was installed. The control signal to the servo valve control card was send using 16-bit DA converter, which output voltage was in a range ± 10 V. The piston position was measured using two channel (+A, +B) encoder which was worked using following parameters: x4 count mode; latch on scan; pulse detection: edge detect – rising edge, debounce trigger – after stable, debounce time 500 ns. The use of x4 count mode enabled the increase of position measurement resolution to 2 μm . The position was measured on a distance 242 mm, so the maximum number of pulses obtained from encoder was equal to 121 000. The measuring time was equal to 20 sec. The supply pressure was equal to 10 MPa and the fluid temperature was in a range + 55 °C to + 63°C. The servo valve control and the piston position signals were recorded to a files and later transformed to the Matlab-Simulink software.

3 Structure of Artificial Neural Network Used for Modelling

During tests we tried to teach the ANN of different structures using the same training data (signals). The structure of the artificial neural network used in presented here investigations was established after making many time consuming trials. Therefore the obtained results are suitable only for specific electrohydraulic drive and for assumed measurement parameters. In the investigations we used and checkup the usefulness of different three layers ANN structures with input layer: from 4 to 50 artificial neurons with different activation functions and with hidden layer: from 4 to 50 artificial neurons with different activation functions. We tried also different teaching methods and different delays for input signal values (from 2 to 20) and for output signal values (from 2 to 20).

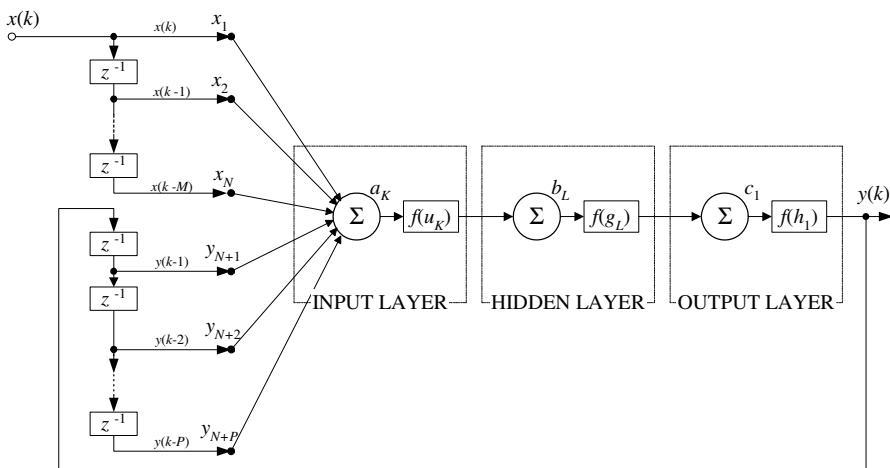


Fig. 2. Simplified scheme diagram of ANN

Finally, we find out the best solutions and decided to use the following ANN structure, which simplified scheme diagram is shown in Fig. 2:

- three layers network with a feedback taken from output,
- input layer: 25 artificial neurons, hyperbolic tangent activation function,
- hidden layer: 35 artificial neurons, hyperbolic tangent activation function,
- output layer: 1 artificial neuron, linear activation function,
- number inputs on which servo drive input signal values are given: 5;
- number inputs on which servo drive output signal values are given: 14;
- delay time for z was equal to 0.01 sec.,
- ANN process time was equal to 0.01 sec.

If we use as input signal the voltage given to servo valve control card and as output signal the piston displacement, then such a drive is modeled in the literature as forth order system connected serially with integrative element [1–3]. So, such a system is asymptotically not stable and therefore cannot be modelled by ANN. Therefore we decided to model a drive with an input signal which was voltage and output signal which was velocity. As input signal the random step-like signal was generated. The output signal was derivative and the average of the last 5 signal values (velocity) are taken for recording, which means that we measured the average velocity for every 0.01 sec. One example of collected signals is shown in Fig. 3.

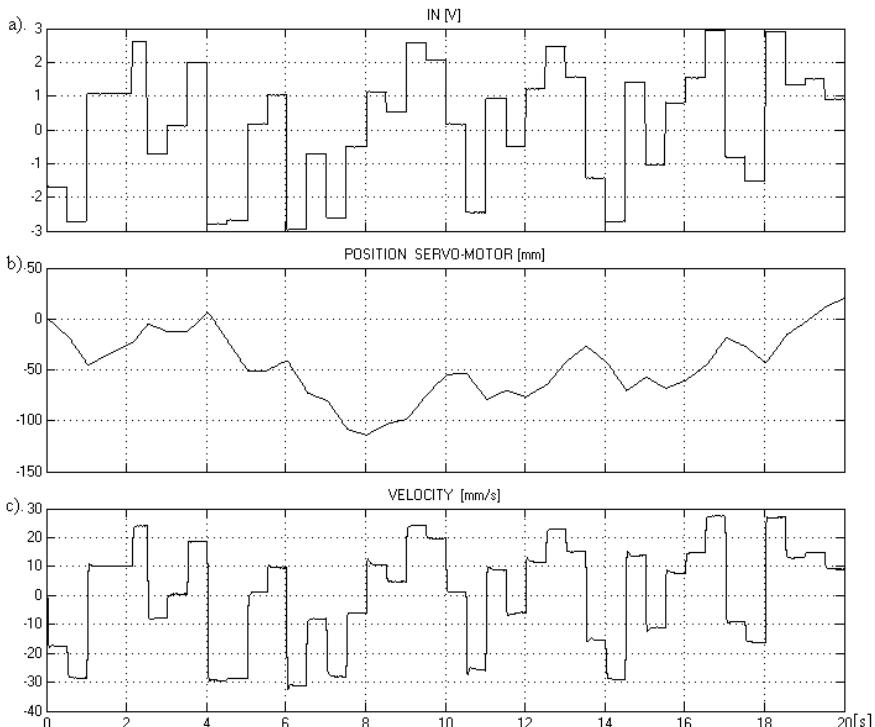


Fig. 3. Signals recorded during electrohydraulic drive investigations: a) servo valve input voltage, b) piston displacement changes, c) average piston velocity (in time 0.01 sec.)

4 Modelling Results and Tests

In teaching process on the artificial neural network inputs, the servo valve input voltage was given and as teaching signal, the piston velocity signal was used. We noticed that in the teaching process, there is optimal number of teaching iterations. In the described here case, it was equal to about 50 iterations. If there were less iterations the ANN was not trained accurately enough. However, if there are more than 50 iterations, the ANN error during teaching process decreased, but after that, when on ANN inputs unknown signal (not used in teaching mode) was given, the error of output signal drastically increased. Therefore in every teaching process we stopped it after 50 iterations.

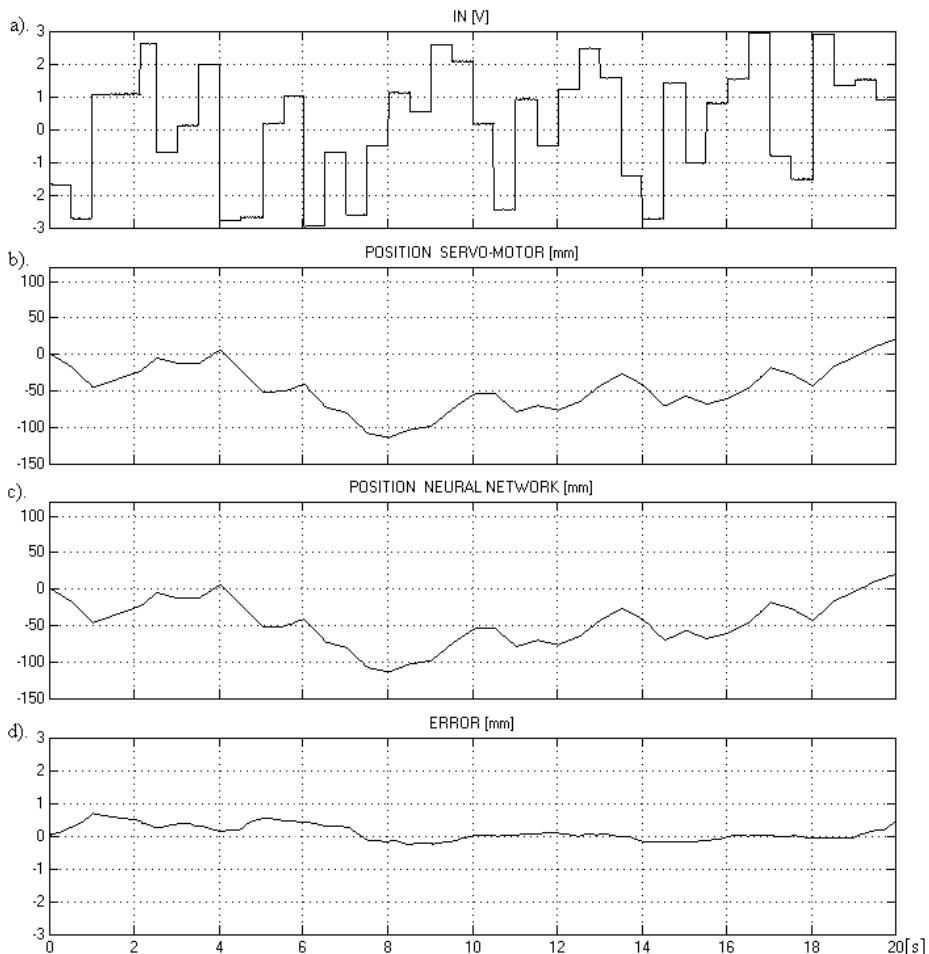


Fig. 4. Signals changes during teaching phase of ANN: a – voltage signal given on valve control card; b – piston position changes; c – ANN responses (position changes); d – ANN error

During the investigations of modelling we also noticed, that on the ANN errors, the values of input and output signals have significant influence. The best solution are obtained when both signals: input and output vary in the same range. The shown in Fig. 3 input signal vary from -3 V to $+3$ V, but in the same time the output signal vary from -30 mm/sec to $+30$ mm/sec. Therefore in teaching process, a scaling module was applied, which reduced the velocity signal to ± 3 mm/s. On the output of ANN the signal was again rescaled to proper range i.e. ± 30 mm/s. These changes are made using MATLAB-Simulink blocks. As mentioned above, the artificial neural network was used to model the velocity changes as a response of voltage on valve electronic card input. Therefore, on the output of ANN an integrator block was added in order to obtain the position on the output. Such a model may be useful in different model based positioning controllers of electrohydraulic servo drive.

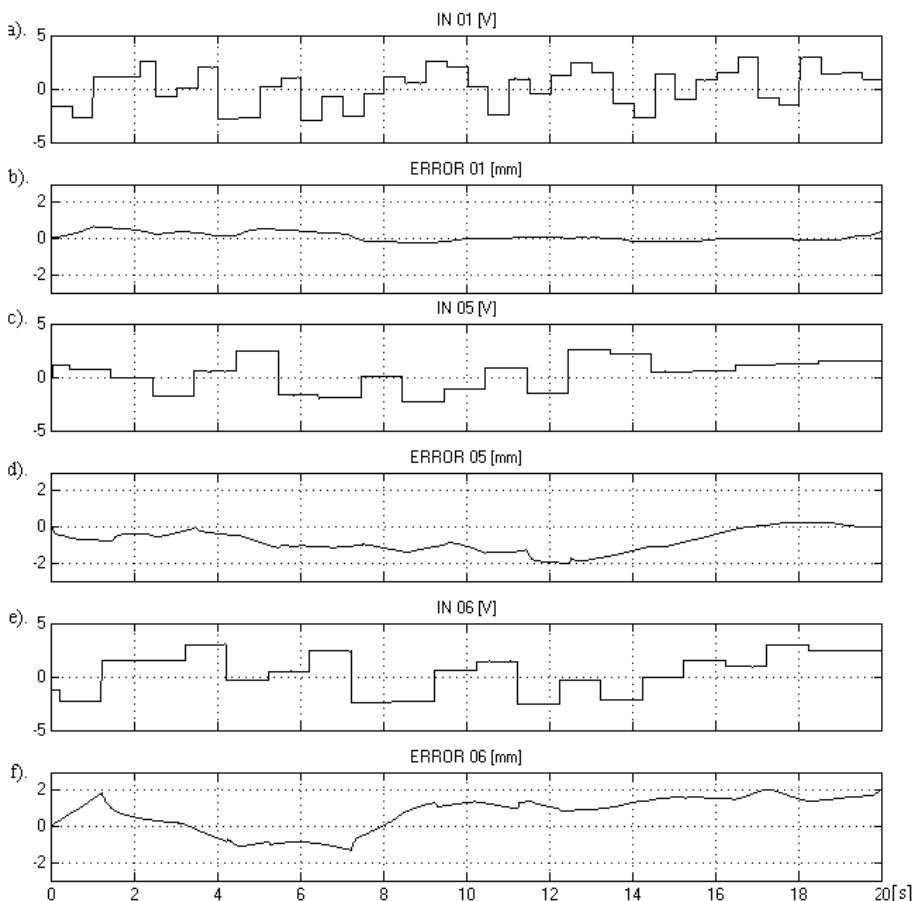


Fig. 5. Simulation results and errors; description see text

For the teaching procedure the signals shown in Fig. 3 were used. During this teaching process we looked for obtainment assumed acceptable position error, which was established in a range of ± 1 mm. For comparison, in Fig. 4 signals obtained after teaching process are shown. On the ANN model inputs the same signal as used during teaching was given. One can note, that the ANN output signal is very similar to the output signal of a real drive. Also ANN model error was in acceptable range e.g. less than 1 mm.

In Figure 5 errors of ANN model are shown. At first, the valve input signal, used in teaching is shown (Fig. 5a) and below ANN model error curve obtained during simulations is presented. In Figure 5c and Fig. 5e two other input signals, not used during teaching process, are shown and in Fig. 5d and 5f ANN model error curves obtained during simulation are respectively shown. One can note, that when on the input is the same signal as used during simulation, the modelling error is very low. Unfortunately, when the not known to the ANN input signal is used, the errors are significantly bigger, but the calculated error was always in a range of ± 2 mm, which can be regarded as acceptable.

5 Conclusion

The experiments described in the paper showed that the artificial neural networks can be used as effective simulation model of electrohydraulic drive with servo valve. The design of the neural network model does not require a complicated design procedure. Although, usually at a first look it sees that the use of ANN for modelling is very easy, in fact there are several problems to be solved and several questions to be answered. To the most difficult one can regard the choosing of the network architecture, its number of inputs, number of layers, art of activation functions and teaching algorithm. Other questions concern the learning methods and the accuracy and sampling frequency of recorded investigation results. The last factor influence really heavy on accuracy and behavior of artificial neural network model. The results obtained in simulation of electrohydraulic drive shown, that the use of ANN for modelling is possible and gave satisfactory results.

References

1. Merrit, H.E.: Hydraulic Control Systems. John Wiley & Sons, Inc. (1967)
2. Murrenhoff, H.: Servohydraulik. Verlag Meinz, Aachen (1998)
3. Milecki, A.: State space models of electrohydraulic servo. Archives of Mechanical Technology and Automation 23(2) (2003)
4. Haykin, S.: Neural Networks. Macmillan College Publishing Company (1994)
5. Pham, D., Liu, X.: Neural Networks for Identification. Prediction and Control. Springer, Heidelberg (1995)
6. Cybenko, G.: Approximation by superposition of sigmoidal function. Math, Control Signal System (1989)

7. Adamczyk, M., Bachman, P.: Modeling elektrohydraulic elements with the use of artificial neural networks and virtual reality, *Virtual design and automation: new trends in collaborative product design*. In: Weiss, Z. (ed.) Publishing House of Poznan University of Technology, pp. 485–491 (2006)
8. Ng, G.W.: *Application of Neural Networks to Adaptive Control of Nonlinear Systems*. Research Studies Press Ltd., London (1997)
9. Yamazaki, K., Chung, J.H.: Application of neural network in mechatronics control, Japan/USA Symp. Flex. Auto. ASME 1, 243–248 (1992)
10. Abdollahi, F., Talebi, H.A., Patel, R.V.: Stable Identification of Nonlinear Systems Using Neural Networks: Theory and Experiments. *IEEE/ASME Transactions on Mechatronics* 11(4) (2006)
11. Albus, J.S.: A new approach to manipulator control: Cerebellar model articulation control. *Trans. ASME, J. of Dyns. Syst., Meas. and Contr.* 97, 220–227 (1975)
12. Ito, M.: *The Cerebellar and Neural Control*. Raven Press, New York (1984)
13. Kawato, M., Uno, Y., Isobe, M., Suzuki, R.: Hierarchical neural network model for voluntary movement with application to robotics. *IEEE Contr. Sys. Mag.*, 8–15 (1988)
14. Psallis, D., Sideris, A., Yamamura, A.: A multilayered neural network controller. *IEEE Contr. Syst. Mag.*, 17–22 (April 8, 1988)
15. Miller, W.T.: Real-time application of neural networks for sensor-based control of robots with vision. *IEEE Trans. Syst. Man., Cybern.* 19(4), 825–831 (1989)
16. Werbos, P.J.: Backpropagation through time: what it does and how to do it. Proceeding of the IEEE, 1550–1560 (1990)
17. Knohl, T., Unbehauen, H.: Adaptive position control of electrohydraulic servo systems using ANN. *Mechatronics* 10, 127–143 (2000)

Cyclic Steady State Space Refinement

Grzegorz Bocewicz¹, Zbigniew Banaszak², and Paweł Pawlewski³

¹ Dept. of Computer Science and Management, Koszalin University of Technology,
Sniadeckich 2, 75-453 Koszalin, Poland
bocewicz@ie.tu.koszalin.pl

² Dept. of Business Informatics, Warsaw University of Technology,
Narbutta 85, 02-524 Warsaw, Poland
Z.Banaszak@wz.pw.edu.pl

³ Dept. of Management Engineering, Poznan University of Technology,
Strzelecka 11, 60-965 Poznan, Poland
pawel.pawlewski@put.poznan.pl

Abstract. A method aimed at refinement of the cyclic steady state space reachable in the given multimodal transportation network is proposed. The paper introduces the concept of a System of Concurrent Multimodal Cyclic Processes in which several subnetworks interact each other via distinguished subsets of common shared workstations as to provide a variety of demand-responsive work-piece transportation/handling services. Searching for the cyclic steady state behavior the following question is considered: Is the cyclic steady state space reachable in the given network structure ? The declarative approach employed makes it possible to evaluate the reachability of cyclic behaviors on a scale that reflects real practice.

Keywords: state space, cyclic steady state, cyclic scheduling.

1 Introduction

The aim of effective management (of a different nature and character) of concurrently executed flows, especially the processes implemented in transport systems, is to make sure that the system behaves in a planned manner which minimizes the operating costs. In terms of the transport system that constitutes a network of multimodal modes of material transport, the aim of logistic management is to select such a structure that guarantees the anticipated variants of traffic at various times of the day as well as during mass events and in crisis situations. It is quite obvious that various behaviors of the system and, as a result, the implementations of various variants of a journey are determined by the parameters of a transport structure, such as the number and length of particular lines, number of modes of transport as well as limitations related with their capacity and speed.

In general terms, it means that various structures of a transport system may produce different variants of its performance, especially divergent scenarios of the traffic control in transport. In the presented context, it seems necessary to evaluate and, consequently, to prepare variants of a potential multimodal structure of a transport network, e.g. in the

phase of its design or modernization. The present work focuses on this problem. The accepted model assumes there is a network of concurrently implemented cyclic processes (modes of transport) that use commonly available resources (elements of routes, stops, stations etc.). The local cyclic processes, synchronized with a protocol of mutual exclusion, make it possible to implement the so-called multimodal processes, i.e. processes understood as material flow traffic in particular directions (e.g. east- west) or between the selected points in the network. For such a model, it is crucial to find a method of prompt variant preparation for the particular structures and organization of material transport multimodal networks.

The problems of evaluating the cyclic implementation of processes are usually considered in terms of cyclic scheduling problems. The literature on the subjects provides numerous methods of solving these problems [5]. Among them there are mathematic programming methods [1, 10], max-plus algebra methods [6], constraint programming techniques [3, 4], Petri nets [9], etc. Most of them deal with seeking solutions than minimize the period of cyclic scheduling. The approaches that make it possible to evaluate cyclic behaviors and, at the same time, avoid deadlocks are quite rare. Therefore, the presented method makes it possible to determine the non-blocking behaviors (i.e. cyclic schedules) in the systems of concurrently executed multimodal processes.

Section 2 provides a description of the System of Concurrent Multimodal Cyclic Processes (SCMCP) – model of a structure and behavior of the considered class of multimodal network of material transport. Automated Guide Vehicle System (AGVS) is considered as a real-life case of the multimodal network composed of different components such as lifts, AGVs, transporters and so on. In Section 3 we formulate the problem of prompt prototyping of behaviors observed in the given structure of the multimodal network. The method of solving this problem is shown in Section 4. Then, Sections 5 and 6 describe the conducted experiments and provide conclusions resulting from them.

2 System of Concurrent Multimodal Cyclic Processes

2.1 Structure

Fig. 1 shows an example of a AGVS structure of which is a composition of numerous recurring fragments (subsystems): the material transport takes place within them. Work pieces are transported along two set routes: north-south (blue line – mP_1) and east-west (red line – mP_2). These routes, setting the courses of multimodal processes, are composed of fragments of the proper lines of transport (AGV routes). In the considered case, there are four AGV lines (P_1, P_3, P_2 and P_4) used to transport materials along the routes. A system of this kind can be modeled as the SCMCP shown in Fig. 2a). The class SCMCP is assumed to include two types of processes:

- *local processes* (representing modes of transport – P_1, P_2, P_3, P_4), whose operations are cyclically repeated along the set routes (sequences of successively used resources). For the system from Fig. 2a), the routes of local processes are defined as follows:

$$p_1 = (R_6, R_2, R_3), p_2 = (R_5, R_1, R_2), p_3 = (R_4, R_1, R_8), p_4 = (R_3, R_4, R_7),$$

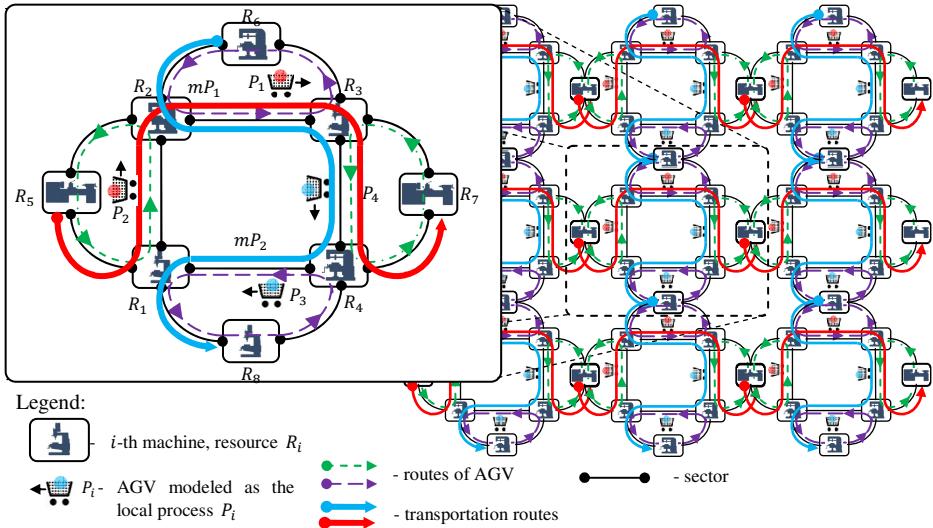


Fig. 1. Example of an AGVS

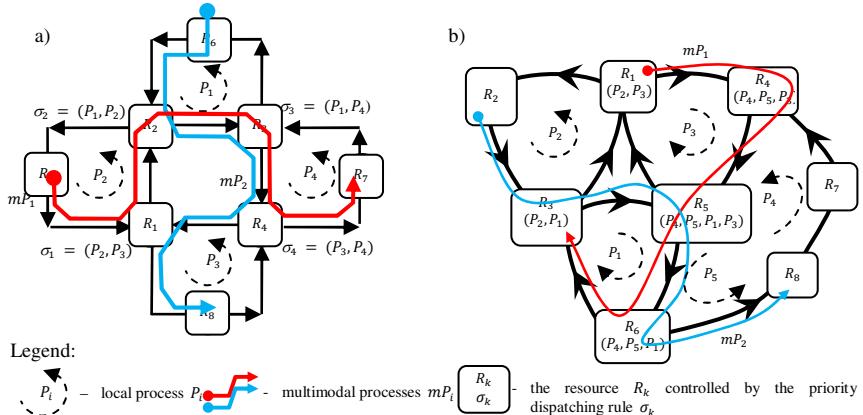


Fig. 2. SCMP model of an AGVS from: Fig 1 a), and for an example from Fig. 3 b)

- *multimodal processes* (mP_1 , mP_2) representing streams of materials. Operations of the multimodal processes are implemented cyclically along routes being compositions of fragments of routes of local processes representing resources used for transporting materials along a given route. For the system from Fig. 2a), the routes of multimodal processes are defined as follows:

$$\begin{aligned} mP_1 &= ((R_5, R_1, R_2), (R_2, R_3), (R_3, R_4, R_7)), \\ mP_2 &= ((R_6, R_2, R_3), (R_3, R_4), (R_4, R_1, R_8)). \end{aligned}$$

Process operation are implemented on two kinds of resources: own resources (each of them is used by only one process of a given kind – R_5, R_6, R_7, R_8) and shared

resources (each of them is used by more than one process of a given kind – R_1, R_2, R_3, R_4). Process uses resources that are shared in the mode of mutual exclusion, i.e. in a given moment only one process operation of a given kind can be implemented on a resource (yet, process operations of different kinds, local and multimodal, can be implemented simultaneously).

The access to shared resources is given in the sequence determined by the dispatching rules $\Theta = \{\Theta^0, \Theta^1\}$. It is assumed that $\Theta^l = \{\sigma_1^l, \dots, \sigma_k^l, \dots, \sigma_{lk}^l\}$, where σ_k^l – is the sequence whose elements determine the order in which the processes (local $l = 0$ / multimodal $l = 1$) are provided with access to the resource R_k . In case of the system from Fig. 2a), the access to shared resources is determined by the following rules

$$\begin{aligned}\sigma_1^0 &= (P_2, P_3), \sigma_2^0 = (P_1, P_2), \sigma_3^0 = (P_2, P_3), \sigma_4^0 = (P_3, P_4), \\ \sigma_1^1 &= (mP_2, mP_1), \sigma_2^1 = (mP_1, mP_2), \sigma_3^1 = (mP_1, mP_2), \sigma_4^1 = (mP_2, mP_2).\end{aligned}$$

The subsequent operation starts right after the current operation is completed, providing that the resource indispensable to its implementation is available. While waiting for the busy resource, the process does not release the resource which was assigned for implementing the previous operation. Moreover, an assumption is made that processes are of no-expropriation nature, and the times and sequence of operations performed by the processes do not depend on external interferences.

The parameters described above constitute the structure of SCMP that determines its behavior. Formally, the structure of SCMP is defined as the following tuple [3]:

$$SC = ((R, SL), SM), \quad (1)$$

where: $R = \{R_k \mid k = 1, \dots, lk\}$ – set of resources,

$SL = (P, U, O, T, \Theta^0)$ – structure of local processes, where:

$P = \{P_i \mid i = 1 \dots ln\}$ – set of local processes, P_i – i^{th} process,

$U = \{p_i = (p_{i,1}, \dots, p_{i,j}, \dots, p_{i,lr(i)}) \mid i = 1 \dots ln\}$ – set of routes of local processes, p_i – i^{th} route, $p_{i,j} \in R$ – resource required for implementing j^{th} operation of the process P_i ,

$O = \{o_i = (o_{i,1}, \dots, o_{i,j}, \dots, o_{i,lr(i)}) \mid i = 1 \dots ln\}$ – set of sequences of operations, $o_{i,j}$ – j^{th} operation of the process P_i ,

$T = \{T_i = (t_{i,1}, \dots, t_{i,j}, \dots, t_{i,lr(i)}) \mid i = 1 \dots ln\}$ – set of sequences of operation performance times, $t_{i,j}$ – time of performing an operation $o_{i,j}^h$,

$\Theta^0 = \{\sigma_k^0 = (s_{k,1}^0, \dots, s_{k,d}^0, \dots, s_{k,lh(k,0)}^0) \mid k = 1 \dots lk\}$ – set of dispatching rules, σ_k^0 – dispatching rule for the resource R_k , $s_{k,d}^0$ – local process, $lh(k,0)$ – length of the rule σ_k^0 ,

$SM = (mP, mU, mO, mT, \Theta^1)$ – structure of multimodal processes, where:

$mP = \{mP_i \mid i = 1 \dots lw\}$ – set of multimodal processes mP_i , lw – number of the processes

$mU = \{mp_i = (mp_{i,1}, \dots, mp_{i,j}, \dots, mp_{i,lm(i)}) \mid i = 1 \dots lw\}$ – set of routes of local processes, mp_i – i^{th} route,

$mO = \{mO_i^h = (mo_{i,1}, \dots, mo_{i,j}, \dots, mo_{i,lm(i)}) | i = 1 \dots lw\}$ – set of sequences of operations, $mo_{i,j}$ – j^{th} operation of the process mp_i ,
 $mT = \{mT_i = (mt_{i,1}, \dots, mt_{i,j}, \dots, mt_{i,lm(i)}) | i = 1 \dots lw\}$ – set of sequences of operation times, $mt_{i,j}$ – time of operation performance
 $mo_{i,j}$,
 $\Theta^1 = \{\sigma_k^1 = (s_{k,1}^1, \dots, s_{k,d}^1, \dots, s_{k,lh(k,1)}^1) | k = 1 \dots lk\}$ – set of dispatching rules of multimodal processes, σ_k^1 – dispatching rule for the resource R_k , $s_{k,d}^1$ – multimodal process, $lh(k, 1)$ – length of the rule σ_k^1 .

2.2 Behavior

In the systems of concurrent cyclic processes, the behavior is usually presented [5], [6], as schedules determining the moments of initiating all the operations implemented within them. Fig. 3a) provides an example of such a schedule that determines the way of implementing the processes of SC structure from Fig. 2b).

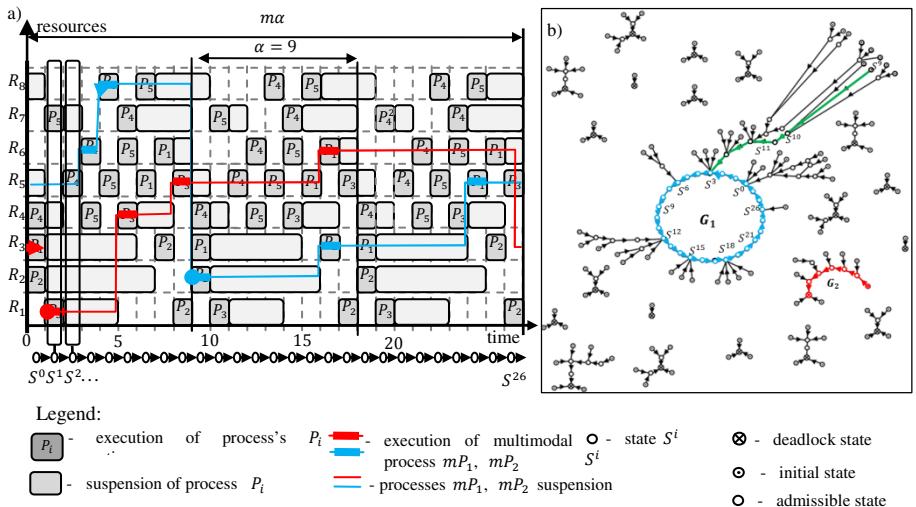


Fig. 3. Cyclic schedule for structure from Fig. 2b) a), and the corresponding states space \mathcal{P} b)

The presented schedule is an example of cyclic behavior, i.e. the successive states of the processes are reachable with the constant period (the operations of local processes are repeated with the period $\alpha = 9$ t.u. (time units) and the multimodal processes with $m\alpha = 27$ t.u.).

In this approach, each behavior can be comprehended as a sequence of successive states (subsequent allocations of processes, as well successively changing, according to the rules Θ of access rights). In case of the schedule from Fig. 3a), it is a sequence of 27 states $S^0, S^1, S^2, \dots, S^{26}$. Formally the SCMCP state is defined as follows [3]:

$$S^r = (Sl^r, mS^r), \quad (2)$$

where Sl^r means the r^{th} state of local processes:

$$Sl^r = (A^r, Z^r, Q^r),$$

$A^r = (a_1^r, a_2^r, \dots, a_k^r, \dots, a_{lk}^r)$ – **allocation** of local processes In the r^{th} state, $a_k^r \in P \cup \{\Delta\}$; $a_k^r = P_i$ – **allocation** meaning that the resource R_k occupied by the process P_i , and $a_k^r = \Delta$ – means that the resource R_k is unoccupied.

$Z^r = (z_1^r, z_2^r, \dots, z_k^r, \dots, z_m^r)$ – **sequence of semaphores** of the r^{th} state, $z_k^r \in P$ – **semaphore** determining the process (an element of rule σ_k^0), which has an access to the resource R_k next in the sequence, i.e. $z_k^r = P_i$ means that process P_i is the next to access the resource R_k .

$Q^r = (q_1^r, q_2^r, \dots, q_k^r, \dots, q_m^r)$ – **sequence of semaphore indexes** of the r^{th} state, q_k^r – **index** determining the position of the semaphore value z_k^r in the dispatching rule σ_k^0 , $z_k^r = s_{k,(q_k^r)}$, $q_k^r \in \mathbb{N}$. For example, $q_2^r = 2$ and $z_2^r = P_1$ means that the process P_1 is the second element of the dispatching rule σ_2^0 .

mS^r means the r^{th} state of multimodal processes:

$$mS^r = (mA^r, mZ^r, mQ^r),$$

$mA^r = (ma_1^r, ma_2^r, \dots, ma_k^r, \dots, ma_m^r)$ – sequence of multimodal processes allocation in the r^{th} state, $ma_k^r \in mP \cup \{\Delta\}$,

$mZ^r = (mz_1^r, mz_2^r, \dots, mz_k^r, \dots, mz_m^r)$ – sequence of semaphores of the r^{th} state, $mz_k^r \in mP$ – determines the process (an element of the rule σ_k^1 , ascribed to R_k), which has the access right to the resource R_k ,

$mQ^r = (mq_1^r, mq_2^r, \dots, mq_k^r, \dots, mq_m^r)$ – sequence of semaphore indexes of the r^{th} state, mq_k^r determines the position of the semaphore value mz_k^r in the dispatching rule σ_k^1 , $mz_k^r = s_{k,(mq_k^r)}$, $mq_k^r \in Z$.

Behaviors of the system characterized by various sequences of subsequently reachable states S^r (2) can be illustrated in a graphical form as a state space \mathcal{P} . Fig. 3b shows an example illustrating this possibility for the system from Fig. 2b). If we take the graph-theoretical interpretation of the space \mathcal{P} the digraph corresponding to it is represented by the pair $\mathcal{P} = (\mathbb{S}, \mathbb{E})$, where \mathbb{S} means a set of admissible SCMP states [3], $\mathbb{E} \subseteq \mathbb{S} \times \mathbb{S}$ means a set of arcs representing transitions between SCMP states (transitions take place according to the function $S^f = \delta(S^e)$ described in [3]).

Cyclic behaviors shown in Fig. 3b) are connected with the presence of cyclic subgraphs (e.g. digraph G_1) in the space \mathcal{P} . The initiation of process implementation from an arbitrary state of the digraph G_1 results in attaining (e.g. by means of transition states) the states being part of a cycle (states marked with blue). A set of such states depicted by a cyclic route is called as a **cyclic steady state**.

It must be emphasized that not all of the state space \mathcal{P} result in such a cyclic steady states. Most states lead to deadlock states (marked with the symbol \otimes), which in practice mean an interrupt of the system resulting from the occurrence of a closed chain of requests (e.g. blockade of AGV lines caused by AGV stops occupancy).

In general case the space \mathcal{P} may not include cyclic steady states. Such situations are quite common in the systems with high density of processes (e.g. in transportation systems). In the context of such systems, the question of evaluating the cyclic behavior attainability becomes crucial.

3 Problem Formulation

There is SCMCP with the structure SC (1) including:

- set of resources R ,
- local processes P described by routes p_i , operation sequences O_i and their duration times T_i ,
- multimodal processes mP described by the routes mp_i , operation sequences mO_i and their duration times mT_i ,
- set of dispatching rules Θ .

An answer is sought to the question: Is the space \mathcal{P} of determined cyclic steady states reachable in the SCMCP system with the given structure ?

A positive answer to this question implicates other issues concerning the period of the reachable cyclic steady states, the degree to which the resources are used, methods of transitions between the determined cyclic steady states, the method of avoiding internal disturbances (e.g. breakdowns of modes of transport or traction failures) etc.

4 Method of Generating the State Space

In order to answer the question posed above, we have developed an approach that uses the following properties of the pace of states \mathcal{P} [3]:

1. initial state (in Fig. 3b) marked with the symbol \odot) is a state without input arcs, $SB \subseteq \mathbb{S}$ – means a set of all initial states of the space \mathcal{P} ,
2. deadlock state (\otimes) is a state without output arcs,
3. if in the space \mathcal{P} there is a non-empty set of states $\mathbb{S} \neq \emptyset$, in which no state is an initial state $SB = \emptyset$, then all the states are included into cyclic steady states.

The proposed approach is based on the idea of the iterative elimination of the initial states SB out of the states space \mathcal{P} , which is illustrated by the example below.

In a set of acceptable states \mathbb{S} , of the system with the structure SC , let there be a non-empty subset of initial states $SB_{(1)} \neq \emptyset$ – where the subscript refers to the number of successively implemented steps of the proposed procedure. In the first step, it is assumed that all states $SB_{(1)}$ are inadmissible. It means that constraints excluding states $SB_{(1)}$ are added to constraints determining admissibility. Therefore, a new set $\mathbb{S}_{(1)} \subset \mathbb{S}$ is obtained and it is devoid of initial states $\mathbb{S}_{(1)} \cup SB_{(1)} = \mathbb{S}$. In the graph representation, it means that all digraphs of the space \mathcal{P} have been reduced by the initial states – and that leads to the space $\mathcal{P}_{(1)}$.

It is easily noticeable that the elimination of initial states results in shortening all the transitory routes, and each descendant of the initial state of the space \mathcal{P} becomes an initial state of the space $\mathcal{P}_{(1)}$.

In the subsequent state, a set of initial states $SB_{(2)}$ is determined again, this time, however, in the set of states $\mathbb{S}_{(1)}$ (set of states of the space $\mathcal{P}_{(1)}$). If $SB_{(2)}$ is a non-empty set $SB_{(2)} \neq \emptyset$ then, similarly as in the first step, they are considered inadmissible states. As a result of the reduction of states of the set $SB_{(2)}$, another set

$\mathbb{S}_{(2)} \subset \mathbb{S}_{(1)}$ is obtained (along with the space $\mathcal{P}_{(2)}$) and the whole procedure is repeated all over again for it. The content of the set of initial states is checked by the stop condition. If in the i th state, the set of initial states is empty $SB_{(i)} = \emptyset$, then according to the property (iii), each state of the set $\mathbb{S}_{(i-1)}$ (providing the set is not empty) is a state included into the cyclic steady states. The set $\mathbb{S}_{(i-1)}$ thus obtained is a subset \mathbb{S} consisting only of states included in the steady cyclic steady states.

The stage of determining the set of initial states plays a crucial role in the presented procedure. Therefore, a fundamental question arises concerning the method of determining the sets $SB_{(i)}$. The set $SB_{(i)}$ may be considered as a set of all solutions to the following the constraints satisfaction problem [3, 8]:

$$PS_{SB(i)} = ((S^r, D_S), C_{SB(i)}), \quad (3)$$

where: S^r – decision variable, initial state $S^r \in \mathbb{S}$,

D_S – domain determining admissible values of variables (allocation, semaphores, indexes) characterizing the state S^r ,

$C_{SB(i)} = C_N \cup \{S^r \neq \delta(S^e), \forall S^e \in \mathbb{S}\} \cup C_{SB}(i-1)$ – set of constraints:

C_N – constraints of admissibility of states S^r [3],

$\{S^r \neq \delta(S^e), \forall S^e \in \mathbb{S}\}$ – constraint that guarantees that S^r is an initial state (there is no state S^e leading to state S^r).

$C_{SB}(i-1) = \{S^r \notin SB_{(1)}, S^e \notin SB_{(1)}, S^r \notin SB_{(2)}, S^e \notin SB_{(2)}, \dots, S^r \notin SB_{(i-1)}, S^e \notin SB_{(i-1)}\}$ – a set of constraints excluding initial states determined

in the steps 1 ... $(i-1)$ out of the set of solutions.

A solution to the problem $PS_{SB(i)}$ (3) is the admissible state S^r that no other admissible state S^e : $S^r \neq \delta(S^e), \forall S^e \in SB$ leads to. The state S^r determined in this way meets the assumptions of the initial state (i). Therefore, the set of all thus determined solutions to the problem $PS_{SB(i)}$ is the set $SB_{(i)}$. It means that with use of this problem in the iterative procedure (solution to the problem (3) successively for $i = 1, 2, 3 \dots$) it is possible to determine a set of states constituting all cyclic steady states of the system. The applied algorithm has the following form:

Algorithm 1

```

function CYCLICSTEADYSTATEGENERATION( $D_S, C_N, \delta$ )
     $i \leftarrow 0$        $C_{SB}(0) \leftarrow \emptyset$ 
    do            $i \leftarrow i + 1$ 
         $PS_{SB(i)} \leftarrow ((S^r, D_S), C_N \cup \{S^r \neq \delta(S^e), \forall S^e \in \mathbb{S}\} \cup C_{SB}(i-1))$ 
         $SB_{(i)} \leftarrow \text{SEARCHALL}(PS_{SB(i)})$ 
         $C_{SB}(i) \leftarrow C_{SB}(i-1) \cup \{S^r \notin SB_{(i)}, S^e \notin SB_{(i)}\}$ 
    while  $SB_{(i)} \neq \emptyset$ 
         $PS_C \leftarrow ((S^r, D_S), C_N \cup C_{SB}(i))$ 
         $V \leftarrow \text{SEARCHALL}(PS_C)$ 
    return  $V(DC)$ 
end
```

where:

D_S, C_N, δ – input data, defined as in (3), $PS_{SB(i)}$ – problem of meeting the constraints (3), $SB_{(i)}$ – set of initial states of the i th step,

V – set of states constituting all cyclic steady states of the space \mathcal{P} .

$\text{SearchAll}(PS)$ – function returning the set of all solutions to the problem PS (in case of no solution the function returns an empty set).

The input data for the Algorithm 1 are constraints of admissibility of states C_N , transition function δ and the domain D_S determining values of variables describing states of the system. Each of these values is determined by the accepted form of the structure SC (1) of SCMCP. The set of cyclic steady states V , which does not include initial states, is the result returned by the algorithm.

5 Computational Experiment

Consider a system from Fig. 2a) that represents a transport subsystem of the network from Fig. 1. In the system, transport of materials takes place along routes: north-south and west-east (multimodal processes: mP_1, mP_2).

For thus defined system an answer is sought to the question whether in the system from Fig. 2a) cyclic transport of materials is possible (by modes of local processes P_1, P_2, P_3, P_4). In order to determine all possible cyclic steady states we used Algorithm 1. In the considered case it required the realization of four steps (solving problems $PS_{SB(1)}, PS_{SB(2)}, PS_{SB(3)}, PS_C$). The total time of calculations was no longer than 3 seconds (Oz Mozart, Windows 7, Intel Core Duo2 3.00 GHz, 4 GB RAM). The obtained state space is illustrated in Fig. 4a). It is clear that there are five cyclic steady states $D_{C,1}, D_{C,2}, \dots, D_{C,5}$ reachable in the system.

It means that the analyzed system can work in five modes characterized by various periods of process implementation and, as a result, by different times of passenger transport. In other words, in SCMCP from Fig. 2a) five various forms of behavior are

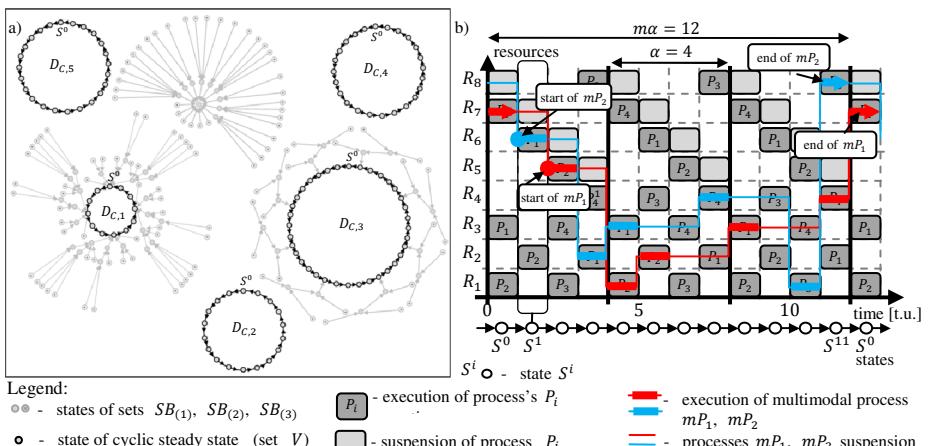


Fig. 4. The state space of SCMCP from Fig. 2a), cyclic schedule for the route $D_{C,1}$

reachable. The periods of the obtained cyclic steady states amount to 12, 18, 36, 24 and 24 t.u. In order to guarantee the shortest transport times, the implementation of processes according to route $D_{C,1}$ should be applied. An example of a schedule corresponding to this route is shown in Fig. 4b).

6 Conclusions

The presented method of prompt prototyping of cyclic steady state space \mathcal{P} (i.e. the states reachable in the given structure of the system), is one of the computer-assisted methods of creating different variants of multimodal networks. The use of the declarative approach makes it possible to evaluate the reachability of cyclic behaviors on a scale that reflects real practice (in the considered case the calculation time was no longer than 3 seconds).

The real-life cases of multimodal networks are strongly affected by the imprecise character of available information (such as time duration, moment of initiation, etc.). The presented expectations specify the development directions for the presented model and method, providing a possibility of taking into consideration the imprecise character of decision variables, e.g. within the Fuzzy Constraints Satisfaction Problem [2, 7] framework.

References

1. Abara, J.: Applying integer linear programming to the fleet assignment problem. *Interfaces* 19, 4–20 (1989)
2. Bach, I., Bocewicz, G., Banaszak, Z., Muszyński, W.: Knowledge based and CP-driven approach applied to multi product small-size production flow. *Control and Cybernetics* 39(1), 69–95 (2010)
3. Bocewicz, G., Banaszak, Z.: Declarative approach to cyclic steady state space refinement: periodic process scheduling. *The International Journal of Advanced Manufacturing Technology* 67(1-4), 137–155 (2013)
4. Bocewicz, G., Wójcik, R., Banaszak, Z.: Multimodal processes scheduling in mesh-like networks composed of periodic systems. In: Grzech, A., Borzemski, L., Świątek, J., Wilimowska, Z. (eds.) *Information Systems Architecture and Technology, Networks Design and Analysis*, Wrocław, pp. 65–76 (2013)
5. Levner, E., Kats, V., Alcaide, D., Pablo, L., Cheng, T.C.E.: Complexity of cyclic scheduling problems: A state-of-the-art survey. *Computers & Industrial Engineering* 59(2), 352–361 (2010)
6. Polak, M., Majdzik, P., Banaszak, Z., Wójcik, R.: The performance evaluation tool for automated prototyping of concurrent cyclic processes. *Fundamenta Informaticae* 60(1-4), 269–289 (2004)
7. Relich, M.: A declarative approach to new product development in the automotive industry. In: *Environmental Issues in Automotive Industry, EcoProduction*, pp. 23–45 (2014)
8. Sitek, P., Wikarek, J.: A hybrid method for modeling and solving constrained search problems. In: *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 385–392 (2013)
9. Song, J.-S., Lee, T.E.: Petri net modeling and scheduling for cyclic job shops with blocking. *Computers & Industrial Engineering* 34(2), 281–295 (1998)
10. Von Kampmeyer, T.: Cyclic scheduling problems, Ph.D. Dissertation, Fachbereich Mathematik/Informatik, Universität Osnabrück (2006)

Using Fuzzy Logic for Improving the Control Performance of Digital Servo Drives with Elastic Coupling

Bogdan Broel-Plater

Faculty of Electrical Engineering, West Pomeranian University of Technology, Szczecin
Bogdan.Broel-Plater@zut.edu.pl

Abstract. In the paper a method for improving the control performance for a typical digital servo drive with a torsionable lead-screw and nonlinear static friction is presented. The method is based on employing an additional feedback the operational intensity of which is determined by means of fuzzy logic. Computer simulations show that the presented solution provides high control accuracy for motions at velocities close to zero and stepwise changes in the position set-point.

Keywords: servo drive, control performance, lead-screw, PI controller, elastic coupling.

1 Introduction

The tool motion control accuracy relative to the workpiece over a wide speed range determines the class of CNC machine tools [3, 9]. Among the most important factors that hinder achieving the high accuracy of the motion mention can be made of variations in mass and moments of inertia of moving objects, backlash and torsionability of lead-screws used in classic machine tool axes [3,9]. Additionally, the control performance is strongly affected by the nonlinear static friction while moving with small varying speeds. That is the reason why work on control algorithms that would improve performance of digital servo drives is still underway [4, 6, 8, 10].

Performance of a servo drive can be assessed by means of different criteria. However for machine tool servo drives the most important goal is to reach the desired position as quick as possible without any overshoot and to maintain that position without any oscillations. Only compliance with these requirements, especially for small speeds and minimal stepwise changes in the position set-point, makes it possible to perform a precise machining and micromachining [2, 3, 9].

In the paper a simple algorithm that provides a significant improvement in the control accuracy of slow motions carried out by a servo drive with a torsionable lead-screw and nonlinear static friction. The presented algorithm is an extension of that described in [2], which minimizes the effect of static friction on the control accuracy of a typical digital servo drive.

The solution presented in the paper has been validated by means of a computer simulated operation of the ACOPOS digital servo drive manufactured by Bernecker & Rainer [1].

2 Basic Structure of the ACOPOS Servo Drive

In a classic servo drive the axially directed motion is governed by a cascade of controllers responsible for position (RX), velocity (RV) and current (RI). In the ACOPOS digital servo drive we have to do with anti-windup PI controllers [1]. In the position controller (RX) of the servo drive use also can be made of an additional feed-forward control. The structure of the servo drive and its component controllers is shown in Figs. 1 – 4.

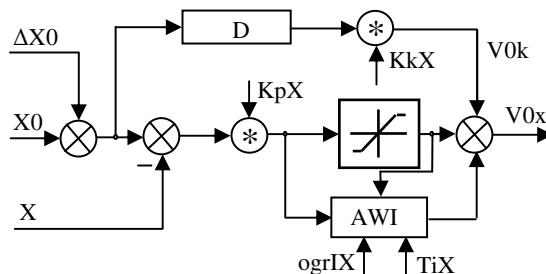


Fig. 1. Structure of the position controller (RX) used in the ACOPOS servo drive: AWI – anti-windup integrator, D - differentiator, X0 and ΔX_0 – basic and auxiliary set-point, X – controlled position, V_{0x} – controller output, V_{0k} – velocity-related feed-forward signal, K_{kX} , K_{pX} , $ogrIX$ and TiX – position controller settings

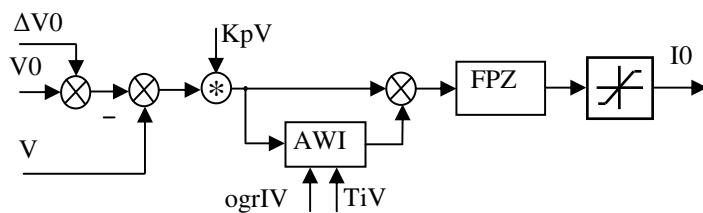


Fig. 2. Structure of the velocity controller (RV) used in the ACOPOS servo drive: AWI – anti-windup integrator, FPZ – band-elimination filter, V_0 and ΔV_0 – basic and auxiliary set-point, V – controlled velocity, I_0 – controller output, K_{pV} , $ogrIV$ and TiV – velocity controller settings

The algorithm of the position controller in the ACOPOS servo drive is repeated every 400 μ s, that of the velocity controller every 200 μ s, and that of the current controller every 100 μ s. With the same frequency the parameter values of any of four controllers also can be altered. This makes it possible to employ the described algorithm to improve the performance of the servo drive.

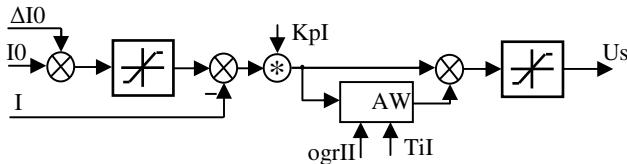


Fig. 3. Structure of the current controller (RI) used in the ACOPOS servo drive: AWI – anti-windup integrator, I0 and Δ I0 – basic and auxiliary set-point, I – controlled current drawn by the motor, Us – controller output, KpI, ogrII and TiI – current controller settings

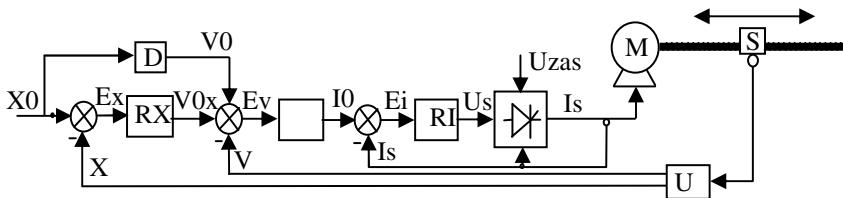


Fig. 4. Output structure of the simulated servo drive: RX, RV, RI – controllers of position, velocity and current respectively, the structure of which is shown in Figs. 1-3, UP – position and velocity measurement system, M – servo drive motor, S – machine tool slide, X0 – slide reference position, V0x, I0 – outputs of position and velocity controllers, Us – current controller output, I – current drawn by the motor, Uzas – supply voltage; the remaining symbols as in Figs. 1-3

3 Motor and Load Model of the Servo Drive

Machine tool servo drives must provide high control accuracy over a wide range of velocities and accelerations. The resultant moment of inertia (J_p) and braking torque (M_h) varying in the process of operation, controller nonlinearities, static friction and dynamic constraint imposed on the current drawn by the servo drive motor make it impossible to optimize effectively the servo drive operation only on the basis of a theoretical analysis. Therefore, computer simulations have been invoked to improve the servo drive performance. Analysis of results obtained enables one to develop a fuzzy algorithm that improves the control performance of the servo drive with a torsionable lead-screw.

In the computer-assisted servo drive simulation a model of a DC motor, shown in Fig. 5, with permanent magnets and an electronic power supply block that constraints the rotor current intensity [5] has been used.

In Figs. 6, 7 and 8 characteristics of the relative driving torque for the modeled motor and those of its own relative braking torque and the relative braking torque referred to the motor shaft produced by the machine tool slide are displayed. These characteristics are related to the maximum resultant braking torque.

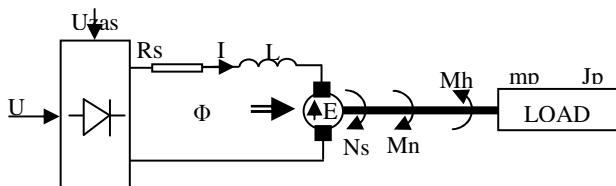


Fig. 5. Equivalent diagram of the motor along with the control block used during tests; U_s – manipulated variable, U_{zas} – supply voltage, m_p i J_p – displaced mass and its moment of inertia referred to the motor axis, N_s – rotational speed of the motor, M_n and M_h – driving and braking torques, Φ and E_s – magnetic flux and induced counter-electromotive force, R and L – resistance and inductance of the rotor circuit, I – current in the rotor circuit

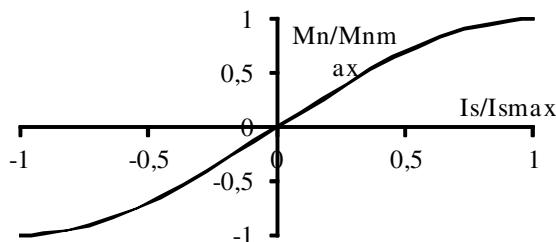


Fig. 6. Characteristic of the relative driving torque developed by the motor as a function of the relative current drawn by the motor

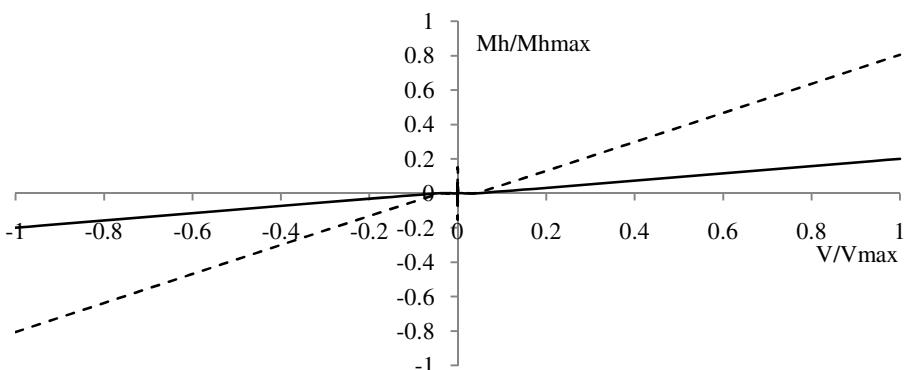


Fig. 7. Full characteristic of the relative braking torque as a function of the relative velocity developed by the motor; solid line – characteristic of the motor self-braking torque, broken line – characteristic of the slide braking torque referred to the motor axis

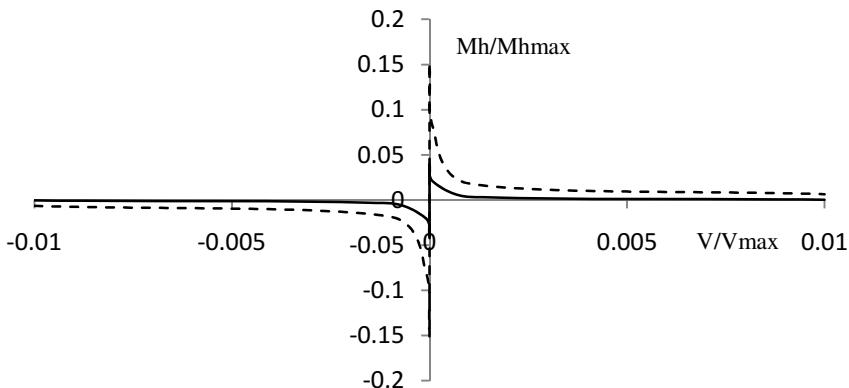


Fig. 8. Relationship between the relative static braking torque of the motor and its relative velocity – a fragment of characteristics of Fig. 7 enlarged for the lowest velocities

Figure 9 shows the torsion of the lead-screw being referred to the motor shaft and dependent on the slide position, at which the static braking torque hindering the machine tool slide motion is overcome and the slide motion starts.

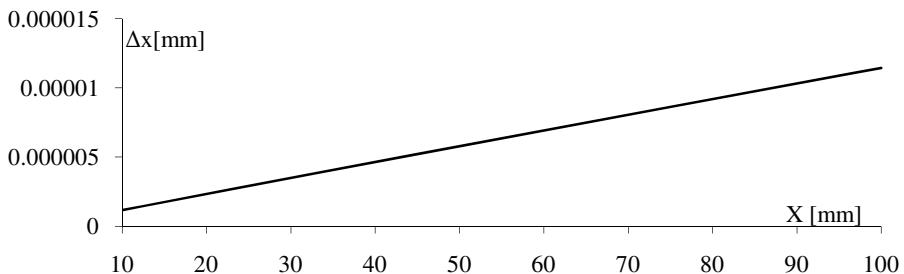


Fig. 9. Characteristic of the lead-screw torsion referred to the motor axis, at which the slide motion starts

4 Algorithm for Improving the Motion Control Accuracy of a Servo Drive with a Torsionable Lead-Screw

As a result of previous work on improving the slow motion control performance for a servo drive with a static braking torque a modified control system structure that makes use of potentialities offered by controllers of the ACOPOS servo drive has been developed [2].

However, an additional corrective action needs to be introduced in view of the fact that the used lead-screw is torsionable. As a result of analysis made on such a servo drive a new structure of the servo drive has been developed (Fig. 10).

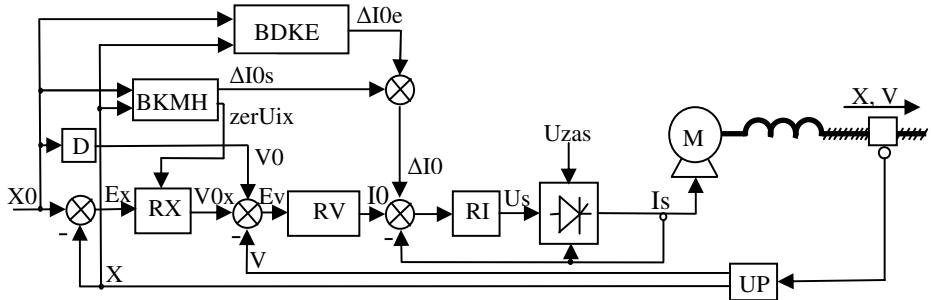


Fig. 10. Structure of the modified servo drive: BDKE – block for dynamic correction of the lead-screw torsionability, BKMH – block for correcting the effect produced by the static braking torque and resetting the integral action of the position controller (RX), zerUIx – signal resetting the integral action of the position controller; the remaining symbols as in Figs. 1–4

It includes two additional feedback paths. The one of them corrects effects that are attributable to the torsionability exhibited by the lead-screw. The nonlinear dynamic block (BDKE) in this feedback path corrects the auxiliary set-point (ΔI_0) of the current controller (RI) of the servo drive on the basis of information about the rate of change of the machine tool slide position error in the vicinity of the desired slide position. The value of the corrective signal ΔI_{0e} is determined every 100 μ s, i.e. as frequent as the algorithm of the current controller (RI) is repeated.

The other feedback path with the corrective block (BKMH) generates a signal ΔI_{0s} in order to minimize the effect of the braking torque at rest on the control performance during slow motions carried out by the servo drive and changes in the direction of the motion. The ΔI_{0s} signal also adjusts the set-point of the current controller (RI). The way the signal is generated and its parameters are determined is described in [2]. Additionally, the BKMH block resets the integral action of the position controller (RX) in the case that the position error ($X_0 - X$) equals zero or changes its sign at the preset velocity being equal to zero. This provides a reduction in overshoots occurring during large changes in the set-point and a reduction in time the set-point is reached. The results of simulation tests indicate that the servo drive with a stiff lead-screw is controlled with very great accuracy over a wide range of its variations in load, and thereby variations in the associated static braking torque [2].

In using the BDKE block of the structure shown in Fig. 11 an empirical relationship between the gain K_e and the machine tool slide position X , which minimizes the overshoot during minimal stepwise changes in the desired slide position occurring in the process of micromachining, has been derived during the tests. The relationship is depicted in Fig. 12.

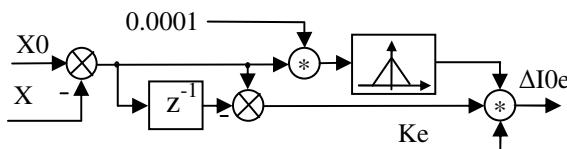


Fig. 11. Structure of the analytical block BDKE; K_1 , K_2 and K_3 - coefficients

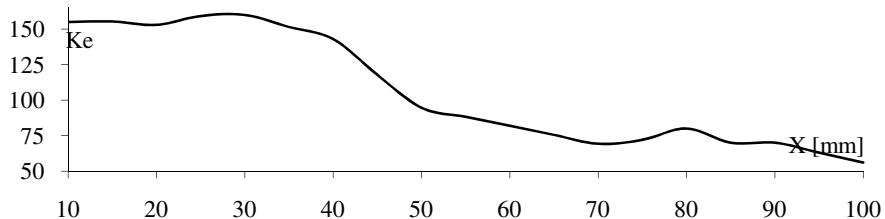


Fig. 12. Relationship between the Ke gain of the BDKE block and the slide position X, which provides a minimal overshoot during minimal stepwise changes in the slide position set-point

Considering the fact that the function $Ke(X)$ is nonlinear, its fuzzy approximation has been determined, and in Fig. 13 the structure of the BDKE block making use of it is shown. The membership functions for its inputs X_1 and X_2 and output Ke are given in Fig. 14.

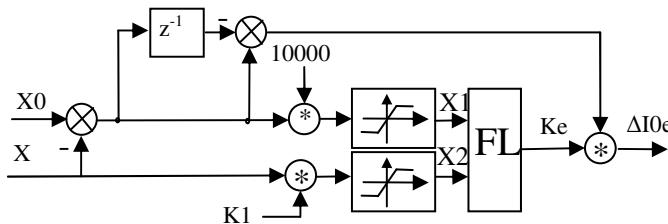


Fig. 13. Structure of the fuzzy block BDKE

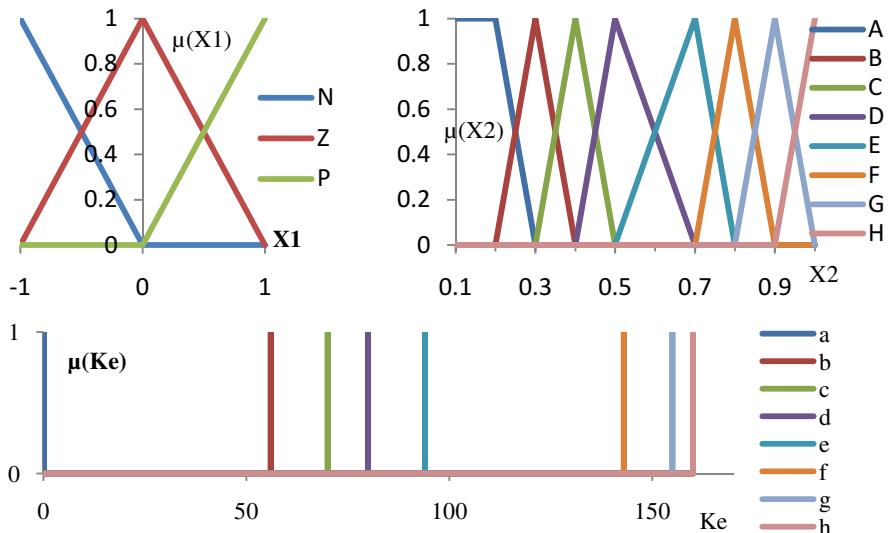


Fig. 14. Membership functions used in the fuzzy block BDKE

The fuzzy inference algorithm of the BDKE block makes use of the control law shown in Table 1 and logic operators defined by Zadeh and the MIN-MAX inference mechanism [7].

Table 1. Inference rules for the BDKE block.

X1 \ X2	A	B	C	D	E	F	G	H
N	a	a	a	a	a	a	a	a
Z	g	h	f	e	d	e	c	b
P	a	a	a	a	a	a	a	a

The result of inference is defuzzified by means of the height method. The K1 coefficient has been chosen so that the entire range of slide positions X be covered by the range of signal X2 variations.

5 Results of Simulation Tests

During the simulation tests the operational effectiveness of the presented servo drive with a fuzzy block to correct dynamically the lead-screw torsionability (BDKE) has been studied.

In Figures 15 and 16 there is shown the behavior of the slide position error (Ex) for several initial positions of the slide during the simulated servo drive motion after a stepwise change in its position set-point by 0.001 mm with and without the BDKE block. In Fig. 17 there is shown the value of the overshoot before and after the BDKE block has been employed for various slide positions X and a stepwise change in its position set-point by 0.001 mm.

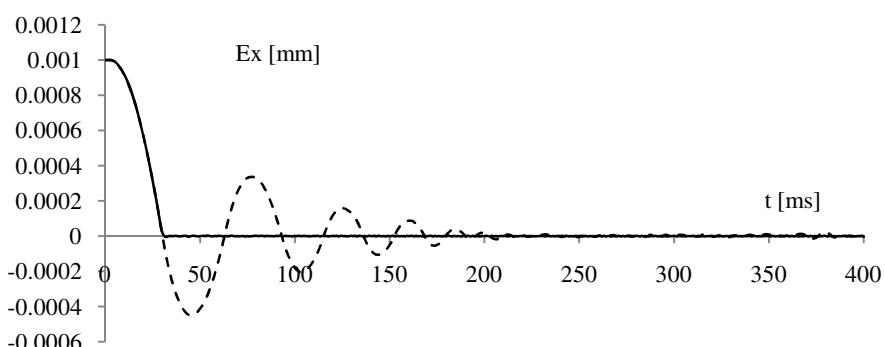


Fig. 15. Tracking errors during the first 400 ms after a stepwise change in the position set-point by $X_0 = 0.001$ mm from the initial value $X=10$ mm of the servo drive before (broken line) and after (solid line) the modification

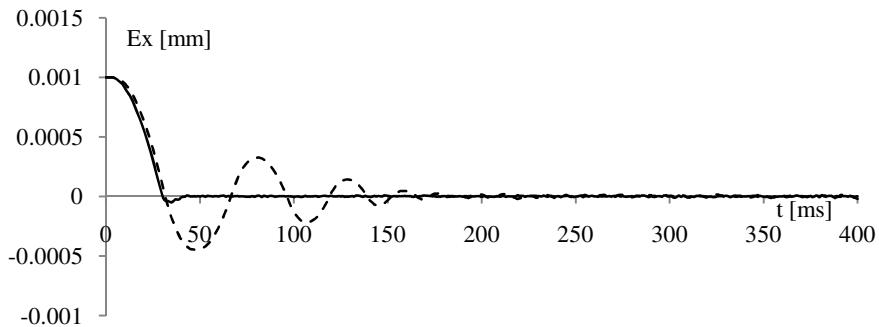


Fig. 16. Tracking errors during the first 400 ms after a stepwise change in the position set-point by $X_0 = 0.001$ mm from the initial value $X=100$ mm of the servo drive before (broken line) and after (solid line) the modification

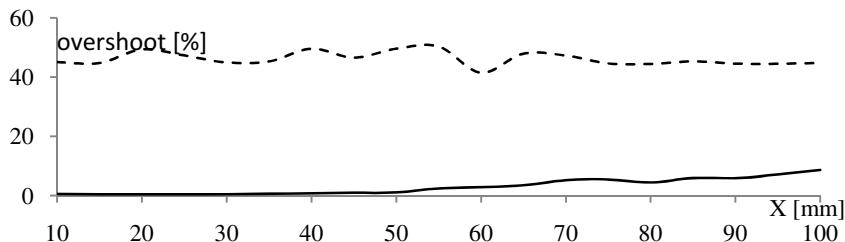


Fig. 17. Overshoots during a stepwise change in the position set-point by 0.001 mm of the servo drive before (broken line) and after the modification (solid line)

In Figs. 18 – 21 there is depicted the behavior of the slide position error during the simulated slide motion with a constant preset velocity of $V_0=0.001$ mm/s for the servo drive with and without the BDKE block and various initial slide positions x . Owing to the use of the BDKE block the highest and the lowest values of the slide position error have decreased over fivefold for a short lead-screw and twofold for a long lead-screw.

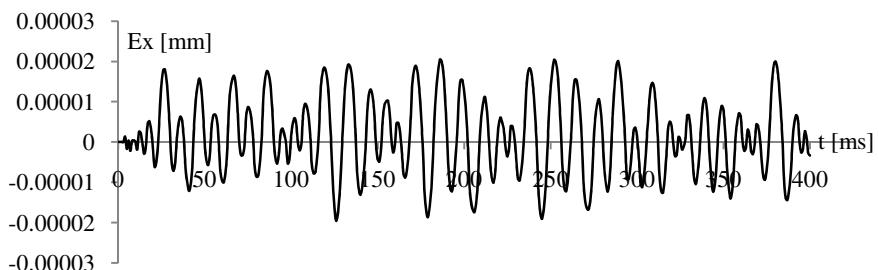


Fig. 18. Tracking error during the first 400 ms after a linear change in the position set-point by $V_0 = 0.001$ mm/s for the initial slide position $X=10$ mm – classic servo drive

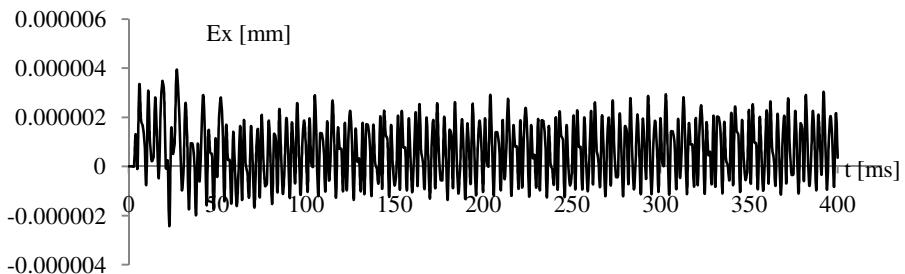


Fig. 19. Tracking error during the first 400 ms after a linear change in the position set-point by $V_0 = 0.001$ mm/s for the initial slide position $X=10$ mm – modified servo drive

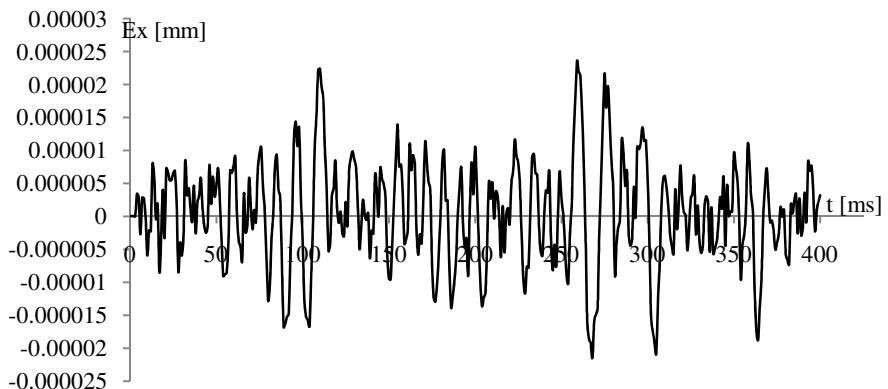


Fig. 20. Tracking error during the first 400 ms after a linear change in the position set-point by $V_0 = 0.001$ mm/s for the initial slide position $X=100$ mm – classic servo drive

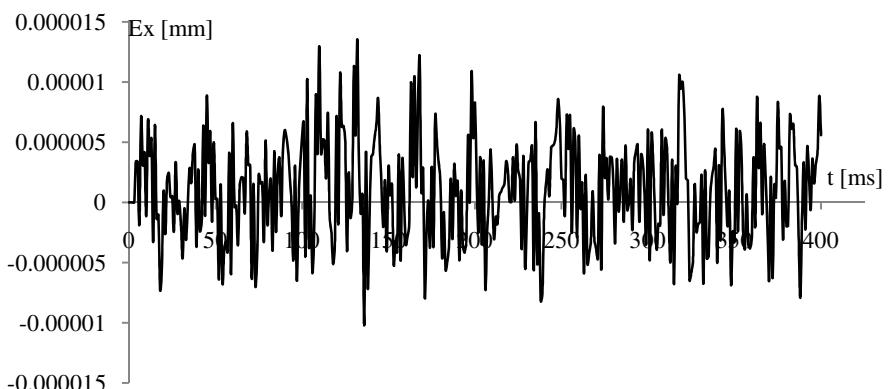


Fig. 21. Tracking error during the first 400 ms after a linear change in the position set-point by $V_0 = 0.001$ mm/s for the initial slide position $X=100$ mm – modified servo drive

All simulations have been carried out with controllers settings (RX, RV and RI) chosen to control the motion of a servo drive motor developing high velocities at no-load.

Comparison of simulation results obtained for the servo drive with a torsionable lead-screw and strong static friction before and after the fuzzy BDKE block has been employed shows that a significant improvement in slow motion control performance has been obtained. The fact that such high control accuracy has been achieved with controller settings chosen for high velocities developed by a servo drive at no-load makes the presented solution even more attractive.

6 Summary

In the paper a simple method for improving the motion control performance of a servo drive with a torsionable lead-screw and a degressive static friction at small reference velocities and small stepwise changes in the reference position is presented. The method consists in introducing an additional nonlinear feedback that adjusts the set-point of the current controller of the servo drive. The simulation tests provide evidence in support of effectiveness of the described algorithm, which also can be applied to digital servo drives of other types.

In the nearest future it is planned to perform practical tests to verify the effectiveness of the algorithm described.

Acknowledgments. This work has been founded by Polish National Science Centre under project N N504 643940

References

1. ACOPOS – Users manual, Bernecker & Rainer
2. Broel-Plater, B.: A method for performance of low-speed control designed for a digital servo drive. *Przegląd Elektrotechniczny* 88(10a), 74–79 (2012) (in polish)
3. Dornfeld, D., Lee, D.-E.: Precision Manufacturing. Springer, Heidelberg (2008)
4. Kamiński, M., Orlowska-Kowalska, T.: Application of Radial Basic Funktion Neural Network in speed control of the driver with elastic connection. *SENE* (2013)
5. Krishnan, R.: Motor Drives – Modeling, analysis and Control. Prentice Hall (2001)
6. Makkapati, V.P., Reinchhartinger, M., Horn, M.: Performed Improvement of a Servo Drives with Mechanical Elasticity via Extended Acceleration Feedback. In: IEEE Intern. Conf. on Control Application, Dubrownik, pp. 1279–1284 (2012)
7. Rutkowski, L.: Computational Intelligence. Methods and Techniques. Springer, Heidelberg (2008)
8. Sato, K., Nakamoto, K., Shimokohbe, A.: Practical control of precision positioning with friction. *Precision Engineering* 28, 426–434 (2004)
9. Suh, S.-H., Kang, S.-K., et al.: Theory and Design of CMC Systems. Springer, Heidelberg (2003)
10. Sato, K., Nakamoto, K., Shimokohbe, A.: Practical control of precision positioning with friction. *Precision Engineering* 28, 426–434 (2004)

Chaos Synchronization of the Modified Van der Pol-Duffing Oscillator of Fractional Order

Mikołaj Busłowicz¹ and Adam Makarewicz²

¹ Białystok University of Technology, Faculty of Electrical Engineering
busmiko@pb.edu.pl

² Doctoral Study, Faculty of Electrical Engineering, Białystok University of Technology
adammakarewicz@interia.pl

Abstract. The paper considers the modified Van der Pol-Duffing oscillator of fractional order. Chaotic behavior of the system is analyzed and the problem of synchronization of two modified Van der Pol-Duffing systems via master/slave configuration with linear coupling is considered. A simple sufficient condition for synchronization is proposed. This condition is based on the chaos stabilization method derived by Jiang et. al. [Chaos Solitons and Fractals, 2003] for the global synchronization of two coupled general chaotic integer order systems with a unidirectional linear error feedback coupling. Numerical simulations show the effectiveness of theoretical considerations.

Keywords: fractional, Van der Pol-Duffing system, chaos, synchronization.

1 Introduction

Dynamical systems described by fractional order differential or difference equations have been investigated in several areas such as viscoelasticity, electrochemistry, diffusion processes, control theory, electrical engineering, etc. The problems of analysis and synthesis of dynamic systems described by fractional order differential (or difference) equations recently have considerable attention, see monographs [8, 13, 15–18], for example.

Many non-linear dynamical systems have behavior known as chaos. Chaos is a very interesting non-linear phenomenon. Synchronization of chaos is a very interesting problem, enjoying a wide interest, for example, in control technology, cryptography, communications [9, 19], etc. The problem of synchronization of chaos recently has been intensively studied in many papers, see for example [3, 4, 7, 10, 11, 24] for systems of integer order and [2, 6, 12, 14, 16, 20, 21] for system of fractional order.

In this paper we consider the modified Van der Pol-Duffing oscillator of fractional order. Chaotic behavior of this system will be analyzed and simple sufficient condition for synchronization of two such systems with non-commensurate fractional orders via master/slave configuration with linear coupling will be given. The proposed condition for synchronization is based on the chaos synchronization method derived in [7] in the case of integer order systems (see also [3]). The method given in

[7] was applied in [12] to chaos synchronization in modified Van der Pol-Duffing system of fractional commensurate order.

2 Preliminaries

The modified Van der Pol-Duffing oscillator is described by following normalized non-linear differential equation

$$\dot{X}(t) = f(X(t)) = \begin{bmatrix} -m[x^3(t) - \mu x(t) - y(t)] \\ x(t) - y(t) - z(t) \\ \beta y(t) \end{bmatrix}, \quad X(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix}. \quad (1)$$

The oscillator (1) recently was considered in [24] (integer order) and [12] (fractional order) for various values of parameters. We consider the following parameters

$$m=100, \mu=0.2 \text{ and } \beta=300. \quad (2)$$

For these parameters the system (1) exhibits a double-scroll chaotic attractor, as shown in Fig. 1 (for initial conditions $x(0)=z(0)=0.05$ and $y(0)=0$). Trajectories of the system were obtained using Simulink package of MATLAB.

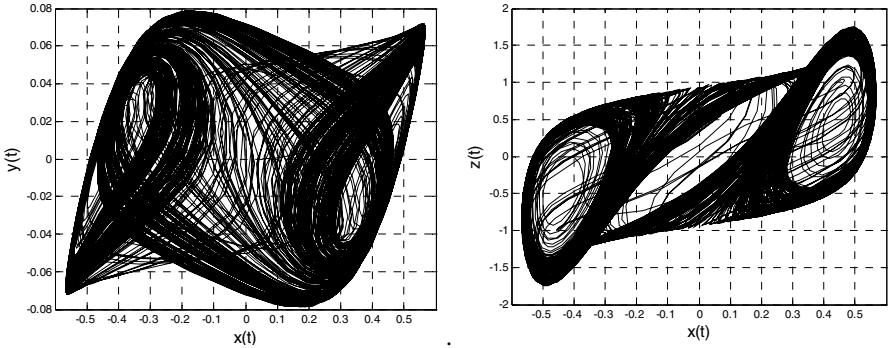


Fig. 1. Chaotic trajectory of the system (1)

In this paper we consider the modified Van der Pol-Duffing oscillator of fractional non-commensurate order, described by the state equation

$$D_t^{\bar{\alpha}} X(t) = f(X(t)) = \begin{bmatrix} -m[x^3(t) - \mu x(t) - y(t)] \\ x(t) - y(t) - z(t) \\ \beta y(t) \end{bmatrix}, \quad D_t^{\bar{\alpha}} X(t) = \begin{bmatrix} D_t^{\alpha_1} x(t) \\ D_t^{\alpha_2} y(t) \\ D_t^{\alpha_3} z(t) \end{bmatrix}, \quad (3)$$

with parameters (2), where derivatives of fractional orders satisfy the inequality $0 < \alpha_i < 1$ for $i = 1, 2, 3$,

$$D_t^{\alpha_i} x(t) = \frac{1}{\Gamma(1-\alpha_i)} \int_0^t \frac{x'(\tau)d\tau}{(t-\tau)^{\alpha_i}}, \quad 0 < \alpha_i < 1, \quad (4)$$

is the Caputo definition for derivative of fractional order α_i , where $x'(t) = dx(t)/dt$ and $\Gamma(\alpha_i) = \int_0^\infty e^{-t} t^{\alpha_i-1} dt$ is the Euler gamma function.

3 Stability Analysis

The fractional order system (3) (similarly as the natural order system (1)) has three equilibrium points. These point are obtained by solution of the non-linear equation $f(X(t)) = 0$, i.e. the set of equations

$$0 = -m[x^3(t) - \mu x(t) - y(t)], \quad 0 = x(t) - y(t) - z(t), \quad 0 = \beta y(t). \quad (5)$$

Solving the equations (5) for parameters (2) one obtains the following equilibrium points

$$E_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad E_1 = \begin{bmatrix} \sqrt{\mu} \\ 0 \\ \sqrt{\mu} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix}, \quad E_2 = \begin{bmatrix} -\sqrt{\mu} \\ 0 \\ -\sqrt{\mu} \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ 0 \\ -\sqrt{2} \end{bmatrix}. \quad (6)$$

Let

$$A_k = \left. \frac{\partial f(X)}{\partial X} \right|_{X=E_k} = \begin{bmatrix} -m(3x^2(t) - \mu) & m & 0 \\ 1 & -1 & -1 \\ 0 & \beta & 0 \end{bmatrix}_{X=E_k}, \quad k = 0, 1, 2, \quad (7)$$

be the Jacobian matrix of the function $f(X)$ evaluated at $X = E_k$ ($k = 0, 1, 2$).

First we assume that fractional orders of the system satisfy the condition $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$. In this case the linearized system (3) about its the equilibrium point E_k ($k = 0, 1, 2$) has the form

$$D_t^\alpha X(t) = A_k X(t), \quad (8)$$

where the matrix A_k is computed form (7) for parameters (2).

It is well known [1] that the fractional order system (8) is stable if and only if its characteristic polynomial of fractional degree has no zeros in the closed right-half of the Riemann complex surface, i.e.

$$w(s) = \det(s^\alpha I - A) \neq 0 \text{ for } \operatorname{Re} s \geq 0, \quad (9)$$

or equivalently, the following condition is satisfied

$$|\arg \lambda_i(A)| > \alpha\pi/2, \quad i = 1, 2, \dots, n, \quad (10)$$

where $\lambda_i(A)$ is the i -th eigenvalues of matrix A .

The condition (10) can be written in the form

$$\alpha < 2\phi/\pi, \quad \phi = \min_i |\arg \lambda_i(A)|. \quad (11)$$

Now, we check the stability of the equilibrium points (6). From (7) we have

$$A_0 = \begin{bmatrix} m\mu & m & 0 \\ 1 & -1 & -1 \\ 0 & \beta & 0 \end{bmatrix} = \begin{bmatrix} 20 & 100 & 0 \\ 1 & -1 & -1 \\ 0 & 300 & 0 \end{bmatrix}, \quad (12)$$

$$A_1 = A_2 = \begin{bmatrix} -2m\mu & m & 0 \\ 1 & -1 & -1 \\ 0 & \beta & 0 \end{bmatrix} = \begin{bmatrix} -40 & 100 & 0 \\ 1 & -1 & -1 \\ 0 & 300 & 0 \end{bmatrix}. \quad (13)$$

Computing eigenvalues of matrices (12), (13) one obtains

- for A_0 : $\lambda_1 = 22.7086$; $\lambda_{2,3} = -1.8543 \pm j16.1486$,
- for $A_2 = A_3$: $\lambda_1 = -42.0745$; $\lambda_{2,3} = 0.5372 \pm j16.8796$.

The matrix A_0 has real positive eigenvalue which lies in instability region. This means that the equilibrium point E_0 is unstable for all $\alpha \in (0, 1)$.

The matrices $A_2 = A_3$ have two eigenvalues with positive real parts. From (11) it follows that the equilibrium points E_1 and E_2 are locally stable for

$$\alpha < \frac{2}{\pi} \arctan \frac{16.8796}{0.5372} = 0.9797. \quad (14)$$

From the above it follows that the system (8) has a chaotic behavior for $\alpha > 0.98$.

Using numerical simulations for $\alpha = 0.985$, $\alpha = 0.98$, $\alpha = 0.9795$ and $\alpha = 0.97$ one obtains the plots shown in Fig. 2. When $\alpha = 0.985$ double-band chaotic behavior is observed. For $\alpha = 0.98$ the system is chaotic with one-band chaos. For $\alpha = 0.9795$ the limit cycle is found and when $\alpha = 0.97$ asymptotically stable attractor exists. Trajectories of the system for $\alpha = 1$ are shown in Fig. 1.

Simulations were performed using the Ninteger Fractional Control Toolbox for MatLab [22]. In this toolbox exists a Simulink block *nid* for fractional derivative and integral. Order and method for rational approximation of fractional derivative/integral can be selected. In simulations we select the Oustaloup's approximation technique (CRONE) of order $n=7$. The block *nid* has the transfer function ks^v , where v is real a number from the interval $(-1, 1)$. In simulations the fractional integrator $1/s^\alpha$ is modeled by series connection of the classical integrator and the block *nid*. Transfer function of this connection is k/s^{v-1} . It is easy to see that $v \in (0, 1)$ for $\alpha \in (0, 1)$.

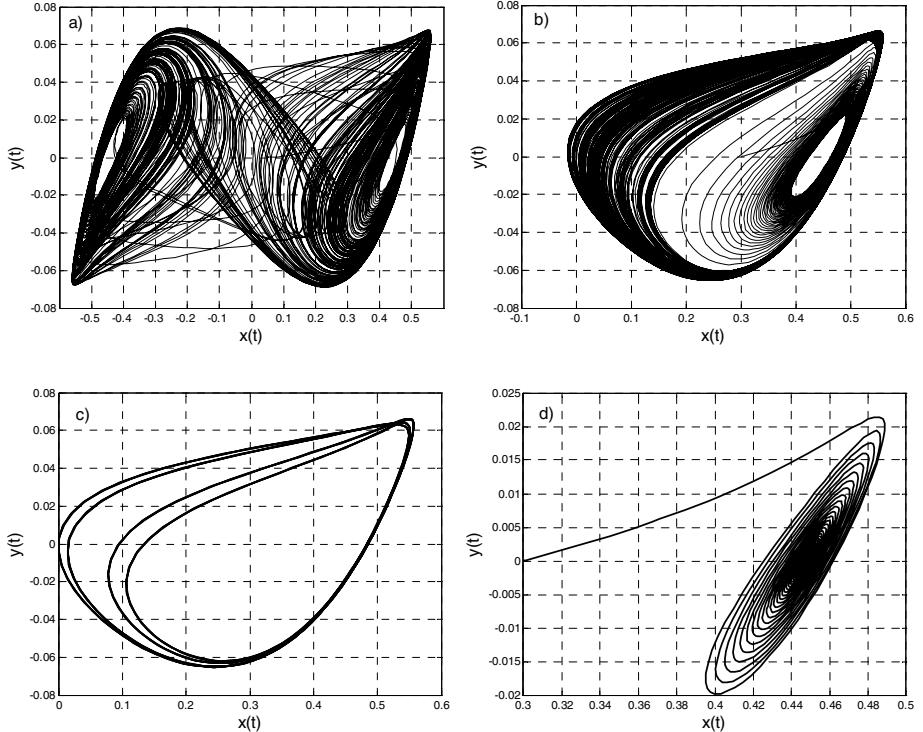


Fig. 2. Trajectories of the system (8): a) $\alpha=0.985$; b) $\alpha=0.98$; c) $\alpha=0.9795$; d) $\alpha=0.97$

Now, using numerical simulations we investigate chaotic behavior of the system (3) with non-commensurate fractional orders and we obtain that the system is chaotic for $\alpha_1 = 0.82$ and $\alpha_2 = \alpha_3 = 0.98$. Chaotic trajectories are shown in Fig. 3.

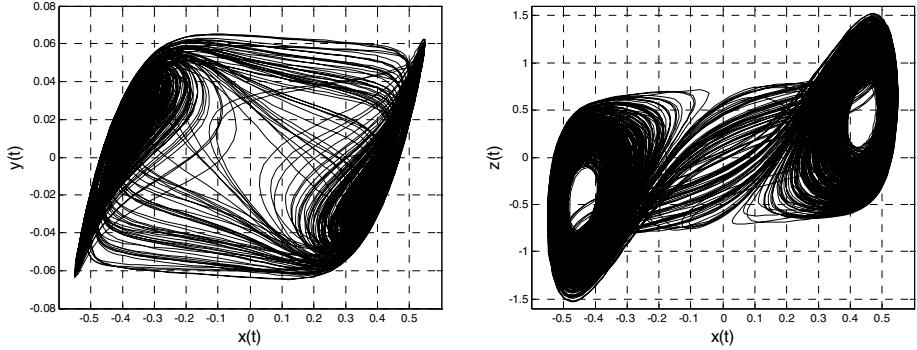


Fig. 3. Chaotic trajectories of the system (3) for $\alpha_1=0.82$ and $\alpha_2=\alpha_3=0.98$

4 Synchronization

Let the fractional system (3) with $\alpha_1=0.82$ and $\alpha_2=\alpha_3=0.98$ be the Master system and the coupled Slave fractional-order system is described by

$$D_t^{\bar{\alpha}} \bar{X}(t) = \begin{bmatrix} -m[\bar{x}^3(t) - \mu\bar{x}(t) - \bar{y}(t)] \\ \bar{x}(t) - \bar{y}(t) - \bar{z}(t) \\ \beta\bar{y}(t) \end{bmatrix} + KE(t), \quad D_t^{\bar{\alpha}} \bar{X}(t) = \begin{bmatrix} D_t^{\alpha_1} \bar{x}(t) \\ D_t^{\alpha_2} \bar{y}(t) \\ D_t^{\alpha_3} \bar{z}(t) \end{bmatrix}, \quad (15)$$

where $K = \text{diag}(k_1, k_2, k_3)$, $k_i \geq 0$, and

$$E(t) = \begin{bmatrix} e_x(t) \\ e_y(t) \\ e_z(t) \end{bmatrix} = X(t) - \bar{X}(t) = \begin{bmatrix} x(t) - \bar{x}(t) \\ y(t) - \bar{y}(t) \\ z(t) - \bar{z}(t) \end{bmatrix} \quad (16)$$

is the synchronization error.

We select a feedback gain K such that $\lim_{t \rightarrow \infty} \|E(t)\| = 0$, where $\|\cdot\|$ denotes the Euclidean norm.

Using (3) and (15), the error dynamics of coupled systems can be written in the form

$$D_t^{\bar{\alpha}} E(t) = \begin{bmatrix} -m[(x_1^3(t) - \bar{x}_1^3(t)) - \mu e_x(t) - e_y(t)] \\ e_x(t) - e_y(t) - e_z(t) \\ \beta e_y(t) \end{bmatrix} - KE(t). \quad (17)$$

Using the relation

$$x^3(t) - \bar{x}^3(t) = [x(t) - \bar{x}(t)][x^2(t) + x(t)\bar{x}(t) + \bar{x}^2(t)] = e_x(t)k_x, \quad (18)$$

where

$$k_x = x^2(t) + x(t)\bar{x}(t) + \bar{x}^2(t), \quad (19)$$

the equation (17) can be written in the form

$$D_t^{\bar{\alpha}} E(t) = (A - K + M_x)E(t), \quad (20)$$

with

$$A = \begin{bmatrix} m\mu & m & 0 \\ 1 & -1 & -1 \\ 0 & \beta & 0 \end{bmatrix}, \quad M_x = \begin{bmatrix} -mk_x & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (21)$$

To select a feedback gain matrix K , we apply the method proposed in [7] (see also [3]) for stabilization of chaos in integer order systems. This method has been applied in [12] to the modified Van der Pol-Duffing system with fractional commensurate order. The method of [7] gives the sufficient condition for stability of coupled natural order chaotic systems. In this method the time-dependent parameter (19) is treated as a interval parameter which values belong to known interval with known bounds, i.e. $k_x \in [a, b]$, where a and b are known.

The system under consideration is a chaotic system with bounded values of state variables. From Fig. 3 it follows that $x(t) \in [-0.06, 0.06]$ and (since the Slave system has the same dynamic as Master) $\bar{x}(t) \in [-0.06, 0.06]$. Hence, from (19) we have that $k_x \in [0, 0.0108]$ for all $t > 0$.

Similarly as in [12], we consider the case of integer order system (20), i.e. the system (20) with $\alpha_1 = \alpha_2 = \alpha_3 = 1$. In this case the equation (21) takes the form

$$\dot{E}(t) = (A - K + M_x)E(t). \quad (22)$$

To select a feedback gain matrix K , we apply the Lyapunov stability theory to the system (22).

Let

$$V(t) = E^T(t)PE(t), \quad (23)$$

where P is a positive definite symmetric constant matrix, be the Lyapunov function for the system (22).

The derivative of (23) along solution of (22) is as follows

$$\dot{V}(t) = \dot{E}^T(t)PE(t) + E^T(t)P\dot{E}(t) = E^T(t)[A_c^T P + PA_c]E(t), \quad (24)$$

where $A_c = A - K + M_x$.

The equation (24) can be written in the form

$$\dot{V}(t) = E^T(t)QE(t), \quad (25)$$

where

$$Q = A_c^T P + PA_c = (A^T - K^T + M_x^T)P + P(A - K + M_x). \quad (26)$$

According to the Lyapunov stability theory, the system (22) is asymptotically stable if and only if $\dot{V}(t) < 0$ for $t \geq 0$, or equivalently, the symmetric matrix Q defined by (26) is negative definite (all eigenvalues have negative real parts).

To check when the matrix (26) has all eigenvalues with negative real parts, we apply the Gershgorin theorem [5, 23].

From (21) it follows that the matrix Q defined by (26) has the following form

$$Q = \begin{bmatrix} m\mu - k_1 - mk_x & 1 & 0 \\ m & -1 - k_2 & \beta \\ 0 & -1 & -k_3 \end{bmatrix} P + P \begin{bmatrix} m\mu - k_1 - mk_x & m & 0 \\ 1 & -1 - k_2 & -1 \\ 0 & \beta & -k_3 \end{bmatrix}. \quad (27)$$

If we choose $P = \text{diag}(p_1, p_2, p_3)$ with $p_i = 1 > 0$, then we obtain

$$Q = \begin{bmatrix} 2(m\mu - k_1 - mk_x) & m+1 & 0 \\ m+1 & -2(1+k_2) & \beta-1 \\ 0 & \beta-1 & -2k_3 \end{bmatrix}. \quad (28)$$

From the Gershgorin theorem we have that all eigenvalues of (28) have negative real parts if the following conditions are satisfied

$$2(m\mu - k_1 - mk_x) + m + 1 < 0,$$

$$-2(1+k_2) + m + 1 + |\beta - 1| < 0,$$

$$-2k_3 + |\beta - 1| < 0.$$

Since $k_x \in [0, 0.0108]$, from the above and (2) we obtain

$$k_1 > 0.5(2m\mu - 2mk_x + m + 1) = 69.42,$$

$$k_2 > 0.5(-2 + 101 + 299) = 199,$$

$$k_3 > 299/2 = 149.5.$$

This means that we can chose $k_1 = 71$, $k_2 = 200$, $k_3 = 150$. For these values of gain matrix $K = \text{diag}(k_1, k_2, k_3)$ the system (22) is asymptotically stable. To check stability of the fractional order system (20), we perform numerical simulations.

Using numerical simulations for $\alpha_1 = 0.82$ and $\alpha_2 = \alpha_3 = 0.98$ we investigate the synchronization problem of two coupled chaotic the modified Van der Pol-Duffing systems (3), (15) for $k_1 = 71$, $k_2 = 200$, $k_3 = 150$ and for lesser feedback gain coefficients. From simulations we obtain that synchronization holds for $k_1 = 71$, $k_2 = 200$, $k_3 = 150$ and also for values of these coefficients essentially smaller, for example, for $k_1 = 12$ and $k_2 = k_3 = 10$. Results of simulations are shown in Fig. 4 with initial conditions for the system (3): $x(0) = z(0) = 0.05$, $y(0) = 0$ and for the system (15): $\bar{x}(0) = 0.5$, $\bar{y}(0) = -0.2$, $\bar{z}(0) = -0.5$. From this figure it follows that synchronization is attained at about 0.1 second for $k_1 = 71$, $k_2 = 200$, $k_3 = 150$ and at about 0.6 second for $k_1 = 12$ and $k_2 = k_3 = 10$.

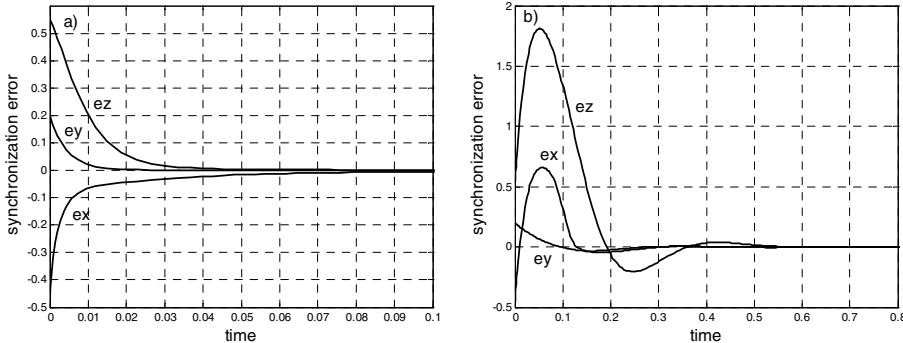


Fig. 4. Synchronization error: a) for $k_1 = 71$, $k_2 = 200$, $k_3 = 150$; b) for $k_1 = 12$, $k_2 = k_3 = 10$

5 Concluding Remarks

Chaotic behavior of the modified Van der Pol-Duffing oscillator of fractional order has been analyzed and the problem of synchronization of two such systems with linear coupling has been considered. A simple necessary condition for chaotic behavior of the system with commensurate orders of derivatives and a simple sufficient condition for synchronization of two coupled systems with non-commensurate orders have been proposed. Numerical simulations showed the effectiveness of theoretical considerations.

Acknowledgement. The work was supported by the National Science Center in Poland under grant N N514 638940.

References

1. Busłowicz, M.: Stability of State-Space Models of Linear Continuous-time Fractional Order Systems. *Acta Mechanica et Automatica* 5, 15–22 (2011)
2. Busłowicz, M., Makarewicz, A.: Synchronization of the Chaotic Ikeda Systems of Fractional Order. In: Mitkowski, W., Kacprzyk, J., Baranowski, J. (eds.) *Theory & Appl. of Non-integer Order Syst. LNEE*, vol. 257, pp. 261–269. Springer, Heidelberg (2013)
3. Deleanu, D.: On a Sufficient Criterion for Global Synchronization in Chaotic Systems. In: Kanarachos, A. (ed.) *Recent Advances in Telecommunications, Signals and Systems*, pp. 95–100. WSEAS Press (2013)
4. Dibakar, G.A., Chowdhury, R., Saha, P.: On the Various Kinds of Synchronization in Delayed Duffing-Van der Pol System. *Commun. Nonlinear Sci. Numer. Simulat.* 13, 790–803 (2008)
5. Gantmacher, F.R.: *The Theory of Matrices*. Nauka, Moscow (1966) (in Russian)
6. He, G.T., Luo, M.: Dynamic Behavior of Fractional Order Duffing Chaotic System and its Synchronization via Singly Active Control. *Appl. Math. Mech.* 33(5), 567–582 (2012)
7. Jiang, G.-P., Tang, W.K.-S., Chen, G.: A Simple Global Synchronization criterion for Coupled Chaotic Systems. *Chaos Solitons and Fractals* 15, 925–935 (2003)
8. Kaczorek, T.: *Selected Problems of Fractional Systems Theory*. LNCIS, vol. 411. Springer, Heidelberg (2011)
9. Kenfack, G., Tiedeu, A.: Secured Transmission of ECG Signals: Numerical and Electronic Simulations. *Journal of Signal and Information Processing* 4, 158–169 (2013)
10. Kimiaeifar, A., Saidi, A.R., Sohouli, A.R., Ganji, D.D.: Analysis of Modified Van der Pol’s Oscillator Using He’s Parameter-Expanding Methods. *Current Applied Physics* 10, 279–283 (2010)
11. Mahmoud, G.M., Aly, S.A., Farghaly, A.A.: On Chaos Synchronization of a Complex Two Coupled Dynamos System. *Chaos, Solitons and Fractals* 33, 178–187 (2007)
12. Matouk, A.E.: Chaos, Feedback Control and Synchronization of a Fractional-Order Modified Autonomous Van der Pol-Duffing Circuit. *Commun. Nonlinear Sci. Numer. Simulat.* 16, 975–986 (2011)
13. Monje, C., Chen, Y., Vinagre, B., Xue, D., Feliu, V.: *Fractional-Order Systems and Controls*. Springer, London (2010)
14. Menacer, T., Hamri, N.: Synchronization of Different Chaotic Fractional-Order Systems via Approached Auxiliary System the Modified Chua Oscillator and the Modified Van der Pol-Duffing Oscillator. *Electronic Journal of Theoretical Physics*, EJTP 8(25), 253–266 (2011)
15. Ostalczyk, P.: *Epitome of the Fractional Calculus, Theory and its Applications in Automatics*. Publishing Department of Technical University of Łódź, Łódź (2008) (in Polish)
16. Petras, I.: *Fractional-Order Nonlinear Systems Modeling, Analysis and Simulation*. Higher Education Press, Springer, Beijing, Heidelberg (2011)
17. Podlubny, I.: *Fractional Differential Equations*. Academic Press, San Diego (1999)
18. Sabatier, J., Agrawal, O.P., Machado, J.A.T. (eds.): *Advances in Fractional Calculus, Theoretical Developments and Applications in Physics and Engineering*. Springer, London (2007)
19. Sheu, L.J., Chen, W.C., Chen, Y.C., Wenig, W.T.: A Two-Channel Secure Communication Using Fractional Chaotic Systems. *World Academy of Science, Engineering and Technology* 65, 1057–1061 (2010)

20. Suchorsky, -M.K., Rand, R.H.: A Pair of Van der Pol Oscillators Coupled by Fractional Derivatives. *Nonlinear Dyn.* 69, 313–324 (2012)
21. Wang, Y., Yin, X., Liu, Y.: Control Chaos in System with Fractional Order. *Journal of Modern Physics* 3, 496–501 (2012)
22. Valério, D.: Ninteger v. 2.3 - Fractional Control Toolbox for MatLab, User and Programmer Manual, Technical University of Lisbona, Lisbona (2005), <http://web.ist.utl.pt/duarte.valerio/ninteger/ninteger.htm>
23. Varga, R.S.: Gershgorin and His Circles. Springer, Berlin (2004)
24. Vincent, U.E., Odunaike, R.K., Laoye, J.A., Gbindinniuola, A.A.: Adaptive Backstepping Control and Synchronization of a Modified and Chaotic Van der Pol-Duffing Oscillator. *J. Control Theory Appl.* 9(2), 273–277 (2011)

Stability Analysis of Descriptor Continuous-Time Two-Term Linear Systems of Fractional Orders

Mikołaj Busłowicz

Białystok University of Technology, Faculty of Electrical Engineering, Białystok, Poland
busmiko@pb.edu.pl

Abstract. The stability problem of fractional continuous-time descriptor linear systems described by the state equation with two differential operators of fractional orders has been considered. Such systems are called the two-term systems. The analytic and frequency domain methods for stability checking of the system with commensurate orders of fractional derivatives have been given. The frequency domain method is based on the Argument Principle and it is simple to apply. The considerations are illustrated by numerical example.

Keywords: linear system, descriptor, continuous-time, fractional, stability.

1 Introduction

In the last decades, the problem of analysis and synthesis of dynamic systems described by fractional order differential or difference equations has been considered in many books (see [9, 15, 18, 20, 21], for example, and references therein).

The problem of stability of linear fractional order systems has been investigated in monographs mentioned above and in the papers [1, 3, 4, 14, 19, 22, 23] for continuous-time systems and in [2, 5, 7, 8, 12] for discrete-time systems.

Recently, the stability problem of the fractional-order systems with double fractional orders has been considered in [6] and [13]. Such systems are also called the two-term systems as a special case of multi-term systems [9].

The state of the art in the descriptor systems theory is given in [10]. The problem of stability of positive descriptor linear systems recently has been addressed in [16].

The aim of the paper is to give the analytic and the frequency domain method for stability analysis of descriptor continuous-time state-space two-term linear systems of fractional commensurate orders.

2 Problem Formulation

Consider a continuous-time two-term descriptor linear system of fractional orders described by the homogeneous state equation

$$E D_t^{k_2\alpha} x(t) + A_1 D_t^{k_1\alpha} x(t) = Ax(t), \quad 0 < \alpha \leq 1, \quad (1)$$

where positive numbers k_1 and k_2 satisfy the inequality $k_2 > k_1$, $x(t) \in R^n$ is the state vector and $E, A, A_l \in R^{n \times n}$ are real matrices.

We will assume that

$$\text{rank } E = n_e < n \quad (2)$$

and

$$\text{rank } H(s) = n \text{ for some } s \in C, \quad (3)$$

where C denotes the field of complex numbers and

$$H(s) = s^{k_2\alpha} E + s^{k_1\alpha} A_l - A \quad (4)$$

is the characteristic matrix of the system.

In (1) the Caputo definition of the derivative of fractional order $0 < \alpha \leq 1$ has been used

$$D_t^\alpha x(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{x'(\tau)d\tau}{(t-\tau)^\alpha}, \quad k=1,2, \quad (5)$$

where $x'(t) = dx(t)/dt$ and $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt$ is the Euler gamma function.

In special case $\alpha=1$ and $k_2=2$, $k_1=1$ the equation (1) describes the descriptor second-order dynamical system considered in [11].

The descriptor linear system (1) will be called stable if for any initial conditions $x(0) = x_0$ the solution $x(t)$ of the state equation (1) satisfies the condition

$$\lim_{t \rightarrow \infty} \|x(t)\| = 0. \quad (6)$$

It is well known that if the fractional order system is stable then norm of the state vector vanishes not exponentially but polynomially [22], i.e. there exist positive constants δ and β such that $\|x(t)\| \leq \delta t^{-\beta} \|x(0)\|$ for $t \geq 0$, where $\|\cdot\|$ denotes the norm.

The stability problem of the standard (i.e. with non-singular matrix E) system (1) has been recently considered in the papers [6, 13]. In these papers have been given the algebraic and frequency domain conditions for stability of the fractional system, described by more general state equation, of the form

$$D_t^{\alpha_2} x(t) + A_l D_t^{\alpha_1} x(t) = Ax(t) + Bu(t), \quad (7)$$

where $0 < \alpha_1 < \alpha_2 \leq 1$.

The aim of the paper to give the analytic and frequency domain conditions for stability of the descriptor fractional order system (1).

3 The Main Results

For the new state-vector

$$\tilde{x}(t) = \begin{bmatrix} x(t) \\ D_t^\alpha x(t) \\ \vdots \\ D_t^{(k_2-1)\alpha} x(t) \end{bmatrix} \in R^{k_2 n} \quad (8)$$

the equation (1) can rewritten in the equivalent form

$$\tilde{E} D_t^\alpha \tilde{x}(t) = \tilde{A} \tilde{x}(t), \quad (9)$$

where

$$\tilde{E} = \begin{bmatrix} I_{(k_2-1)n} & \tilde{0}^T \\ \tilde{0} & E \end{bmatrix} \in R^{k_2 n \times k_2 n}, \quad \tilde{A} = \begin{bmatrix} \tilde{0}^T & I_{(k_2-1)n} \\ A & \tilde{A}_l \end{bmatrix} \in R^{k_2 n \times k_2 n}, \quad (10)$$

$\tilde{0} \in R^{n \times (k_2-1)n}$ is the zero matrix, $(\cdot)^T$ denotes the transposition, $I_{(k_2-1)n}$ is the identity matrix and the block matrix \tilde{A}_l of size $n \times (k_2-1)n$ has k_l -th block equal to $-A_l$ and remaining blocks are equal to zero.

If, $k_2 = 4$ and $k_l = 2$, for example, then

$$\tilde{E} = \begin{bmatrix} I_n & 0_n & 0_n & 0_n \\ 0_n & I_n & 0_n & 0_n \\ 0_n & 0_n & I_n & 0_n \\ 0_n & 0_n & 0_n & E \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} 0_n & I_n & 0_n & 0_n \\ 0_n & 0_n & I_n & 0_n \\ 0_n & 0_n & 0_n & I_n \\ A & 0_n & -A_l & 0_n \end{bmatrix}, \quad (11)$$

where 0_n is $n \times n$ the zero matrix and I_n is $n \times n$ the identity matrix.

From (10) and (2) it follows that

$$\tilde{n}_e = \text{rank } \tilde{E} = (k_2 - 1)n + n_e < k_2 n. \quad (12)$$

Similarly as in [11] we introduce the following definition.

Definition 1. The fractional two-term descriptor system (1) will be called regular if the corresponding one-term descriptor system (9) is regular.

Lemma 1. The fractional two-term descriptor linear system (1) is regular if and only if there exists $s \in C$ such that $w(s) \neq 0$, where

$$w(s) = \det H(s) = \det(s^{k_2 \alpha} E + s^{k_1 \alpha} A_l - A). \quad (13)$$

Proof. It is well known that the fractional system (9) is regular if and only if [17]

$$\det(s^\alpha \tilde{E} - \tilde{A}) \neq 0 \quad (14)$$

for some $s \in C$, where the matrices \tilde{E} and \tilde{A} are defined in (10).

For simplicity of considerations we assume that $k_2 = 4$ and $k_1 = 2$. In this case the matrices \tilde{E} and \tilde{A} have the forms given in (11). Applying the elementary row and column operations to the matrix $s^\alpha \tilde{E} - \tilde{A}$ one obtains

$$\begin{aligned} s^\alpha \tilde{E} - \tilde{A} &= \begin{bmatrix} s^\alpha I_n & -I_n & 0_n & 0_n \\ 0_n & s^\alpha I_n & -I_n & 0_n \\ 0_n & 0_n & s^\alpha I_n & -I_n \\ -A & 0_n & A_l & s^\alpha E \end{bmatrix} \rightarrow \begin{bmatrix} 0_n & -I_n & 0_n & 0_n \\ s^{2\alpha} I_n & 0_n & -I_n & 0_n \\ 0_n & s^{2\alpha} I_n & 0_n & -I_n \\ -A & s^\alpha A_l & A_l + s^{2\alpha} E & s^\alpha E \end{bmatrix} \\ &\rightarrow \begin{bmatrix} 0_n & -I_n & 0_n & 0_n \\ 0_n & 0_n & -I_n & 0_n \\ 0_n & 0_n & 0_n & -I_n \\ -A + s^{2\alpha} A_l + s^{4\alpha} E & s^\alpha A_l + s^{2\alpha} E & A_l + s^{2\alpha} E & s^\alpha E \end{bmatrix} \\ &\rightarrow \begin{bmatrix} 0_n & -I_n & 0_n & 0_n \\ 0_n & 0_n & -I_n & 0_n \\ 0_n & 0_n & 0_n & -I_n \\ -A + s^{2\alpha} A_l + s^{4\alpha} E & 0_n & 0_n & 0_n \end{bmatrix} \end{aligned}$$

Hence

$$\det(s^\alpha \tilde{E} - \tilde{A}) = \det(s^{4\alpha} E + s^{2\alpha} A_l - A). \quad (15)$$

This completes the proof for $k_2 = 4$ and $k_1 = 2$. Proof for other values of k_2 and k_1 is similar.

From Lemma 1 it follows that if (3) holds then the system (1) is regular.

From the theory of stability of linear fractional order systems given by Matignon [19] (see also [3, 22, 23]) we have the following.

Lemma 2. The equivalent fractional order system (9) with non-singular matrix \tilde{E} (for simplicity we assume $\tilde{E} = I_{k_2 n}$), of the form

$$D_t^\alpha \tilde{x}(t) = \tilde{A} \tilde{x}(t), \quad (16)$$

is stable if and only if the fractional degree characteristic polynomial

$$w(s) = \det(s^\alpha I_{k_2 n} - \tilde{A}) \quad (17)$$

has no zeros in the closed right-half of the Riemann complex surface, i.e.

$$w(s) \neq 0 \text{ for } \operatorname{Re} s \geq 0, \quad (18)$$

or equivalently, the following condition is satisfied

$$|\arg \lambda_i(\tilde{A})| > \alpha \frac{\pi}{2}, \quad i = 1, 2, \dots, n, \quad (19)$$

where $\lambda_i(A)$ is the i -th eigenvalues of matrix \tilde{A} .

The stability region of the system (16), described by (19), is shown in Fig. 1.

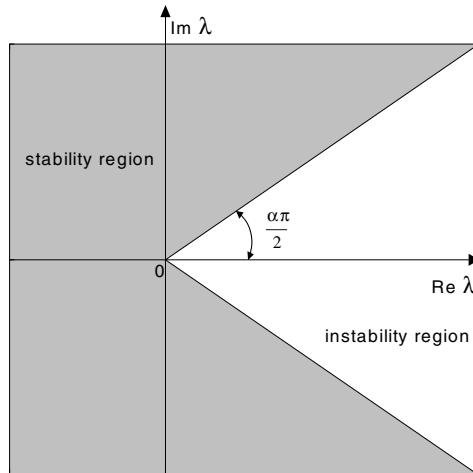


Fig. 1. Stability region of the fractional system (16) in the λ -plane

The condition (19) can be rewritten in the form [3]

$$\gamma > \alpha \frac{\pi}{2}, \text{ where } \gamma = \min_i |\arg \lambda_i(\tilde{A})|. \quad (20)$$

From the theory of stability of linear descriptor systems we have the following lemma [10].

Lemma 3. The equivalent descriptor regular system (9) with $\alpha = 1$, of the form

$$\tilde{E} \dot{\tilde{x}}(t) = \tilde{A} \tilde{x}(t), \quad (21)$$

is stable if and only if

$$\sigma(\tilde{E}, \tilde{A}) \subset \mathbf{C}^- = \{\lambda : \lambda \in \mathbf{C}, \operatorname{Re} \lambda < 0\}, \quad (22)$$

where

$$\sigma(\tilde{E}, \tilde{A}) = \{\lambda : \lambda \in \mathbf{C}, \lambda \text{ is finite, } \det(\lambda \tilde{E} - \tilde{A}) = 0\} \quad (23)$$

is the set of finite poles of the system containing at most \tilde{n}_e (12) complex numbers.

The finite poles of (21) are the finite eigenvalues of the matrix pair (\tilde{E}, \tilde{A}) . These eigenvalues can be computed by using the command $\text{eig}(\tilde{A}, \tilde{E})$ of the MATLAB.

From Lemmas 1 and 2 we have the following theorem.

Theorem 1. The two-term descriptor regular system of fractional orders described by the state equation (1) is stable if and only if all finite roots λ_i of the natural degree characteristic equation

$$\det H(\lambda) = \det(\lambda^{k_2} E + \lambda^{k_1} A_l - A) = 0, \quad (24)$$

or equivalently, all finite eigenvalues λ_i of the pair of matrices (\tilde{E}, \tilde{A}) , of the form given in (10), satisfy the condition

$$|\arg \lambda_i| > \alpha \frac{\pi}{2}, \quad (25)$$

i.e. lie in the stability region shown in Fig. 1.

Proof. Stability of the fractional descriptor two-term regular system (1) if equivalent to stability of the associated fractional descriptor system (9) with commensurate orders. Stability of this system depends from its finite poles which are finite eigenvalues of pair of matrices (\tilde{E}, \tilde{A}) . Finite eigenvalues of (\tilde{E}, \tilde{A}) and finite roots of (24) are the same. Therefore, according to Lemma 3, these roots (eigenvalues) must be located in stability region shown in Fig. 1. This completes the proof.

Note that the equation (24) is obtained by substitution $s^\alpha = \lambda$ in the characteristic polynomial (13).

From the form (10) of matrix \tilde{E} and (12) it follows that the set of finite roots of (24) (finite eigenvalues of pair (\tilde{E}, \tilde{A})) contains at most

$$N = (k_2 - 1)n + n_e < k_2 n \quad (26)$$

complex numbers, where $n_e = \text{rank } E < n$.

From Theorem 1 we have the following lemmas.

Lemma 4. The two-term descriptor regular system (1) is stable if and only if

$$\alpha < \frac{2\gamma}{\pi}, \quad \gamma = \min_i |\arg \lambda_i|, \quad (27)$$

where λ_i are the finite roots of the characteristic equation (24) (finite eigenvalues of pair (\tilde{E}, \tilde{A})).

Lemma 5. If all finite roots λ_i of the characteristic equation (24) (finite eigenvalues of pair (\tilde{E}, \tilde{A})) have negative real parts, then the fractional descriptor two-term regular system (1) is stable for all $\alpha \in (0, 1]$.

The method of Theorem 1 requires computation of eigenvalues of pair (\tilde{E}, \tilde{A}) . Size of these matrices may be very high. If, for example, $\alpha_2 = 0.97$ and $\alpha_1 = 0.7$ in (7), then in (1) we have $\alpha = 0.01$, $k_2 = 97$ and $k_1 = 70$. In this case size of the matrices (10) is $97n \times 97n$. Hence, the investigation of stability of the fractional system (1) by checking the condition of Theorem 1 or the condition (27) can be inconvenient.

Therefore, we apply the frequency domain method to stability checking of the descriptor fractional system (1). This method has been proposed in [1] for asymptotic stability analysis of fractional order continuous-time linear systems described by the transfer function and in [3, 4] for the systems described by state space models. Results of these papers are described in [15, Chapter 9].

Denote by δ the fractional degree of the characteristic polynomial (13). It is easy to see that $\delta \leq N\alpha$, where N is defined by (26).

Let $w_r(s)$ be the reference stable fractional polynomial of degree δ . This polynomial can be chosen in the form

$$w_r(s) = (s + c)^\delta, \quad c > 0. \quad (28)$$

Let us consider the rational function

$$\psi(s) = \frac{w(s)}{w_r(s)} = \frac{\det H(s)}{w_r(s)}, \quad (29)$$

where $\det H(s)$ is defined in (13).

Theorem 2. The fractional order system (1) is stable if and only if

$$\Delta \arg_{\omega \in (-\infty, \infty)} \psi(j\omega) = 0, \quad (30)$$

where $\psi(j\omega) = \psi(s)$ for $s = j\omega$ and $\psi(s)$ is defined by (29).

Proof. From (29) it follows that

$$\Delta \arg_{\omega \in (-\infty, \infty)} \psi(j\omega) = \Delta \arg_{\omega \in (-\infty, \infty)} w(j\omega) - \Delta \arg_{\omega \in (-\infty, \infty)} w_r(j\omega). \quad (31)$$

From the Argument Principle it follows that the fractional degree characteristic polynomial (13) is stable if and only if

$$\Delta \arg_{\omega \in (-\infty, \infty)} w(j\omega) = \Delta \arg_{\omega \in (-\infty, \infty)} w_r(j\omega) = \delta\pi. \quad (32)$$

From (31) it follows that (32) holds if and only if (30) is satisfied.

Satisfaction of (30) means that plot of the function $\psi(j\omega)$ does not encircle or cross the origin of the complex plane as ω runs from $-\infty$ to ∞ .

From (29) and (13) it follows that if the reference polynomial has the form (28) then

$$\psi(0) = \det(-A)/c^\delta \quad (33)$$

and

$$\psi(\infty) = \lim_{\omega \rightarrow \pm\infty} \frac{\det((j\omega)^{k_2\alpha} E + (j\omega)^{k_1\alpha} A_l - A)}{(j\omega + c)^\delta}. \quad (34)$$

From (33) it follows that $\psi(0) \leq 0$ if $\det(-A) \leq 0$. Hence, from Theorem 2 we have that the fractional order system (1) is not stable if $\det(-A) \leq 0$ and $\psi(\infty) \geq 0$.

4 Illustrative Example

Example 1. Consider the system (1) with $n = 3$, $k_2 = 2$, $k_1 = 1$ and the matrices

$$E = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_l = \begin{bmatrix} -0.5 & -0.2 & 0 \\ -0.4 & 0.1 & -0.3 \\ 0 & -0.3 & 0.1 \end{bmatrix}, \quad A = \begin{bmatrix} -0.8 & -0.1 & 0.2 \\ 0.6 & -3.7 & 0.7 \\ 0.1 & 0.8 & -3.3 \end{bmatrix}. \quad (35)$$

Find values of fractional order α for which the system is stable.

From (10) it follows that the matrices of equivalent single-term fractional system (9) are as follows

$$\tilde{E} = \begin{bmatrix} I_3 & 0_3 \\ 0_3 & E \end{bmatrix} \in R^{6 \times 6}, \quad \tilde{A} = \begin{bmatrix} 0_3 & I_3 \\ A & -A_l \end{bmatrix} \in R^{6 \times 6}. \quad (36)$$

Using the command $\text{eig}(\tilde{A}, \tilde{E})$ of the MATLAB we compute the finite eigenvalues of (\tilde{E}, \tilde{A}) : $\lambda_1 = -32.2116$; $\lambda_{2,3} = -0.1845 \pm j2.1154$; $\lambda_{4,5} = 0.2427 \pm j0.7649$.

From (27) we have

$$\gamma = \min_i |\arg \lambda_i| = \arctan(0.7649/0.2427) = 1.2635 \quad (37)$$

and

$$\alpha < 2\gamma/\pi = 0.8044. \quad (38)$$

Hence, according to Lemma 4, the system is stable for $0 < \alpha < 0.8044$.

Now we apply the frequency domain method given in Theorem 2.

From (35), (36) it follows that the characteristic polynomial (13) of the system has the fractional degree $\delta = N\alpha = 5\alpha$, since, according to (26), $N = (k_2 - 1)n + n_e = 5$.

Plot of the function

$$\psi(j\omega) = \frac{\det H(j\omega)}{(j\omega + 4)^{5\alpha}}, \quad \omega \in (-\infty, \infty), \quad (39)$$

is shown in Fig. 2 for $\alpha = 0.75$ and $\alpha = 0.85$.

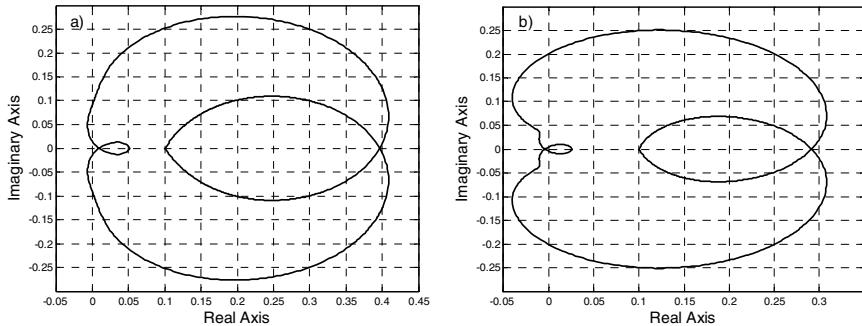


Fig. 2. Plot of the function (39): a) for $\alpha=0.75$; b) for $\alpha=0.85$

From (33), (34) and structure of the matrix $H(s) = s^{2\alpha}E + s^\alpha A_l - A$ it follows that $\psi(0) = 0.0517$ for $\alpha = 0.75$ and $\psi(0) = 0.0258$ for $\alpha = 0.85$. Moreover, $\lim_{s \rightarrow \infty} \det H(s) = 0.1$, which means that $\psi(\infty) = 0.1$ for $\alpha = 0.75$ and $\alpha = 0.85$.

From Fig. 2 it follows that plot of (39) does not encircle the origin of the complex plane for $\alpha = 0.75$ and encircles for $\alpha = 0.85$. This means, according to Theorem 2, that the system (1), (35) is stable for $\alpha = 0.75$ and unstable for $\alpha = 0.85$.

5 Concluding Remarks

The stability problem of continuous-time descriptor two-term linear system (1) with commensurate orders of fractional derivatives has been considered.

It has been shown that stability of the system is equivalent to satisfaction of the condition (25) for all finite roots of the natural degree characteristic equation (24). These roots are the finite eigenvalues of the pair of matrices (\tilde{E}, \tilde{A}) of the form (10) (Theorem 1).

The frequency domain method for stability checking of the system is given in Theorem 2. This method is simpler to apply in the case of large values of k_2 .

Acknowledgement. The work was supported by the National Science Center in Poland under grant N N514 638940.

References

1. Busłowicz, M.: Stability of Linear Continuous-time Fractional Order Systems with Delays of the Retarded type. *Bull. Pol. Acad. Sci., Tech. Sci.* 56, 319–324 (2008)
2. Busłowicz, M.: Robust Stability of Positive Discrete-time Linear Systems of Fractional Order. *Bull. Pol. Acad. Sci., Tech. Sci.* 58, 567–572 (2010)
3. Busłowicz, M.: Stability of State-Space Models of Linear Continuous-time Fractional Order Systems. *Acta Mechanica et Automatica* 5, 15–22 (2011)

4. Busłowicz, M.: Stability Analysis of Continuous-time Linear Systems Consisting of n Subsystems with Different Fractional Orders. *Bull. Pol. Acad. Sci., Tech. Sci.* 60, 279–284 (2012)
5. Busłowicz, M.: Simple Analytic Conditions for Stability of Fractional Discrete-time Linear Systems with Diagonal State Matrix. *Bull. Pol. Ac.: Tech.* 60(4), 809–814 (2012)
6. Busłowicz, M.: Frequency Domain Method for Stability Analysis of Linear Continuous-Time State-Space Systems with Double Fractional Orders. In: Mitkowski, W., Kacprzyk, J., Baranowski, J. (eds.) *Theory & Appl. of Non-integer Order Syst. LNEE*, vol. 257, pp. 31–39. Springer, Heidelberg (2013)
7. Busłowicz, M., Kaczorek, T.: Simple Conditions for Practical Stability of Linear Positive Fractional Discrete-time Linear Systems. *Int. J. Appl. Math. Comput. Sci.* 19(2), 263–269 (2009)
8. Busłowicz, M., Ruszewski, A.: Necessary and Sufficient Conditions for Stability of Fractional Discrete-time Linear State-space Systems. *Bull. Pol. Acad. Sci., Tech. Sci.* 61 (2013) (in print)
9. Diethelm, K.: *The Analysis of Fractional Differential Equations*. LNM. Springer, Heidelberg (2010)
10. Duan, G.-R.: *Analysis and Design of Descriptor Systems*. AMM, vol. 23. Springer, New York (2010)
11. Duan, G.-R., Huang, L.: Robust Pole Assignment in Descriptor Second-order Dynamical Systems. *Acta Automatica Sinica* 33(8), 888–892 (2007)
12. Dzieliński, A., Sierociuk, D.: Stability of Discrete Fractional State-Space Systems. *Journal of Vibration and Control* 14, 1543–1556 (2008)
13. Jiao, Z., Chen, Y.-Q.: Stability Analysis of Fractional-Order Systems with Double Noncommensurate Orders for Matrix Case. *Fractional Calculus Applied Analysis, An Int. J. for Theory and Applications* 14, 436–453 (2011)
14. Kaczorek, T.: Necessary and Sufficient Stability Conditions of Fractional Positive Continuous-time Linear Systems. *Acta Mechanica et Automatica* 5, 52–54 (2011)
15. Kaczorek, T.: *Selected Problems of Fractional Systems Theory*. LNCIS, vol. 411. Springer, Heidelberg (2011)
16. Kaczorek, T.: Stability of Descriptor Positive Linear Systems. *COMPEL* 32(1), 412–423 (2013)
17. Kaczorek, T.: Descriptor Fractional Linear Systems with Regular Pencils. *Asian J. Control* 15(4), 1–14 (2013)
18. Kilbas, A.A., Srivastava, H.M., Trujillo, J.J.: *Theory and Applications of Fractional Differential Equations*. Elsevier, Amsterdam (2006)
19. Matignon, D.: Stability Result on Fractional Differential Equations with Applications to Control Processing. In: IMACS-SMC2, Lille, France, pp. 963–968 (1996)
20. Monje, C., Chen, Y., Vinagre, B., Xue, D., Feliu, V.: *Fractional-Order Systems and Controls*. Springer, London (2010)
21. Podlubny, I.: *Fractional Differential Equations*. Academic Press, San Diego (1999)
22. Sabatier, J., Moze, M., Farges, C.: LMI Stability Conditions for Fractional Order Systems. *Computers and Mathematics with Applications* 59, 1594–1609 (2010)
23. Tavazoei, M.S., Haeri, M.: Note on the Stability of Fractional Order Systems. *Mathematics and Computers in Simulation* 79, 1566–1576 (2009)

Design of Integrated Information Systems for the Security of People and Objects

Małgorzata Cupriak¹, Sławomir Jasiński¹, and Małgorzata Kaliczyńska²

¹ Kontron East Europe sp. z o.o., Warsaw, Poland

{sławomir.jasinski,małgorzata.cupriak}@kontron.pl

² Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland
mkaliczynska@piap.pl

Abstract. The article presents a number of issues concerning the integration of many information systems and automation equipment in order to ensure full control and improve the user experience. Particular attention is paid to the safety systems installed in public utility facilities, including fire protection systems used in urban subways objects (with Warsaw metro as an example). Structure of the integrating system is described and an overview is made of the specialised components used in its design and implementation. Requirements are also presented for obtaining technical approval. Experience gained during the design, implementation and testing of the integrated security system is also summarized.

Keywords: safety system, IT system, detection, object, integration.

1 Introduction

Modern objects of common use, such as factories, shopping malls, airports, train stations, subways are equipped with numerous electronic systems used to ensure automatic control of their proper operation, as well as security of facilities and their users. Often these systems come from different manufacturers, operate independently with various functionalities, installed with the introduction of new technologies. Multiplying systems from different manufacturers, as well as the irresponsible behaviour of users, (often bystanders than service staff) can be a source of unexpected behaviour of individual system components, including unauthorized alarms. In such a situation, the additional master control is required, allowing the supervision carried out by the operator and rapid response, such as switching off the alarms activated accidentally.

2 Overview of Common Solutions

There are several solutions of systems integrating security systems. Some of them are of the general purpose; others are dedicated for operation in particular facilities. Most often used on the domestic market are GEMOS and DMS8000 systems.

GEMOS Facility Management System [1] integrates all safety and automation systems in the building, improving safety standard level of the facility, with the simultaneous reduction of its operating costs. It serves as a master control panel, which collects and logs alarms and other information from safety systems and technical supervision devices, to provide a uniform and consistent information with all the necessary instructions including the actions taken.

The DMS8000 integrating device [2] implements interaction of the building fire protection systems, being used in small objects using a small number of control panels. It can be configured in the point-to-point architecture, providing the redundant connection.

One of the major risks that can disrupt the safe operation of public facilities is fire. It is accompanied by panic, chaos, threatens people with carbon monoxide poisoning, burns, or even death. Therefore, this aspect is particularly seriously analysed, and safety ensuring devices are required.

At most stations of the Warsaw metro, the fire protection system is used supported by the EBL G3 microprocessor control panel designed to work with the analogue addressable smoke detectors, conventional thermal detectors, manual fire alarms, control modules, as well as with the acoustic and optical signalling devices. Each control panel makes it possible to connect up to 4 loops with sensors and signalling devices. The control panels may be connected to form a network of up to 30 control panels, which enlarges the system capacity to more than 15,000 of fire points. The control panels communicate using the doubly terminated bus TLON network (connection redundancy). Each of the control panels has access to all information from other control panels in the same network – the control panels are peer-to-peer. One may monitor system state from any location with Internet access.

3 PEP110 Integrating System – Architecture

Analysis of the functionality and features of the existing solutions contributed to the decision to design the PEP110 integrating system, which will make connection and cooperation possible of various systems, providing a uniform service platform. These may be systems that are used in railway traffic control [5, 6], and above all guarantee safety:

- fire protection system,
 - detection system for objects in the track (people and items),
 - intrusion prevention system [4],
 - traffic situation monitoring system,
 - operation diagnostics system for individual system components and equipment,
 - traveller information system,
 - audible warning system,
 - communication systems,
 - automation systems,
- and/or other required systems.

The integrating system should allow (based on messages, dynamic graphics and activity results of the diagnostic functions) forwarding supervision services necessary information to support decision-making in the state of crisis in the facility. The integrating system should be an open, scalable software-hardware platform, fully independent from the manufacturers of the integrated systems and equipment. It should allow full interoperability of various systems, their easy setup and maintenance. The facility layouts should be presented graphically. The integrating system should not only monitor, visualize, and archive data, it should also permit manual control override of devices and integrated systems, control with the highest priority in relation to automatic control, which is implemented by the fire protection control panel and its I/O modules installed on the loop detection circuits. The system should integrate the various subsystems using open communication standards (CAN, Profibus, Ethernet, LON, etc.), popular network protocols and reliable transmission media.

It should be possible to build a complex management system for the technical equipment of the facility (Fig. 1) containing maximum 120 local and remote monitoring centres.

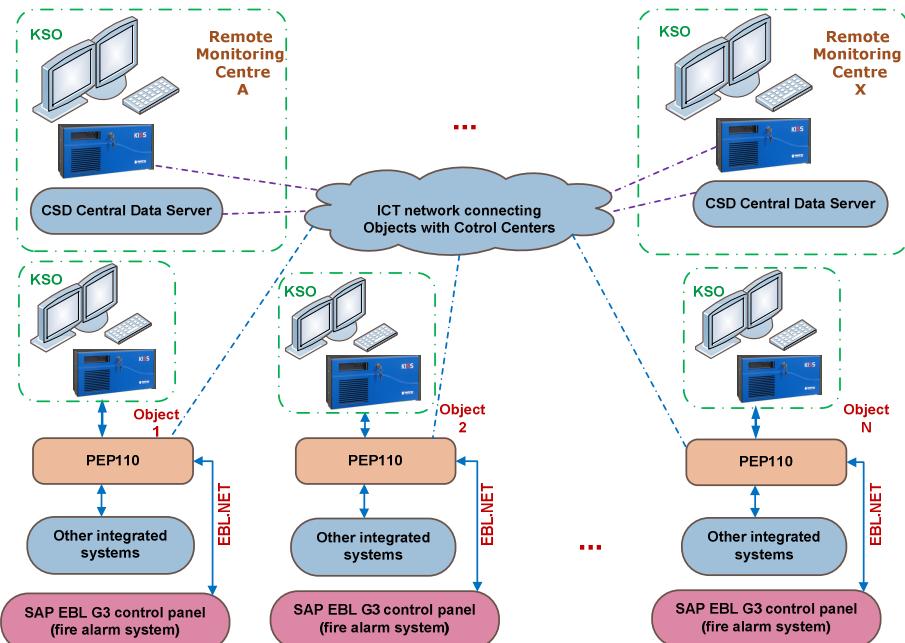


Fig. 1. Functional diagram of the complex management system for the technical equipment of the facility

3.1 Operator's Console

Console (terminal) of the integrating system should be realized in the form of the KSO monitoring and imaging stand. Depending on needs and arrangements with the recipient, the integrating system can be supported by many KSO stands. Every KSO

console is a computer workstation with two monitors. The KSO console software should allow data acquisition, imaging, and supervisory control.

The PEP110 integrating system consists of the following devices:

- PEP110-M – hardware module provided with the IT platform (Fig. 2), housed in the control cabinet. The PEP110-M consists of modular drivers for the following purpose:
 - KS – control computer for technical devices not connected directly with fire protection devices;
 - KWD_x – data exchange computer ensuring two-way communication with the SAP EBL G3 system control panel (based on Ethernet technology); installed software (using components of the EBL.NET library) enables the overarching scenarios and manual control. To ensure redundancy of a connection with the EBL G3 control panel, the PEP110-M uses at least two KWD_x modules (x stands for the next KWD computer).
- KSO – console (terminal) for the PEP100 integrating system is the monitoring and imaging stand. Depending on needs, PEP100 can be supported by many KSO stands. Every KSO console is a computer workstation with two LCD monitors. Its software allows data acquisition, imaging, and supervisory control.
- KD20 – local diagnostic KD20 computer, designed to record data (events and states of the system) and for diagnostics, ensures transmission access to the facility signals.

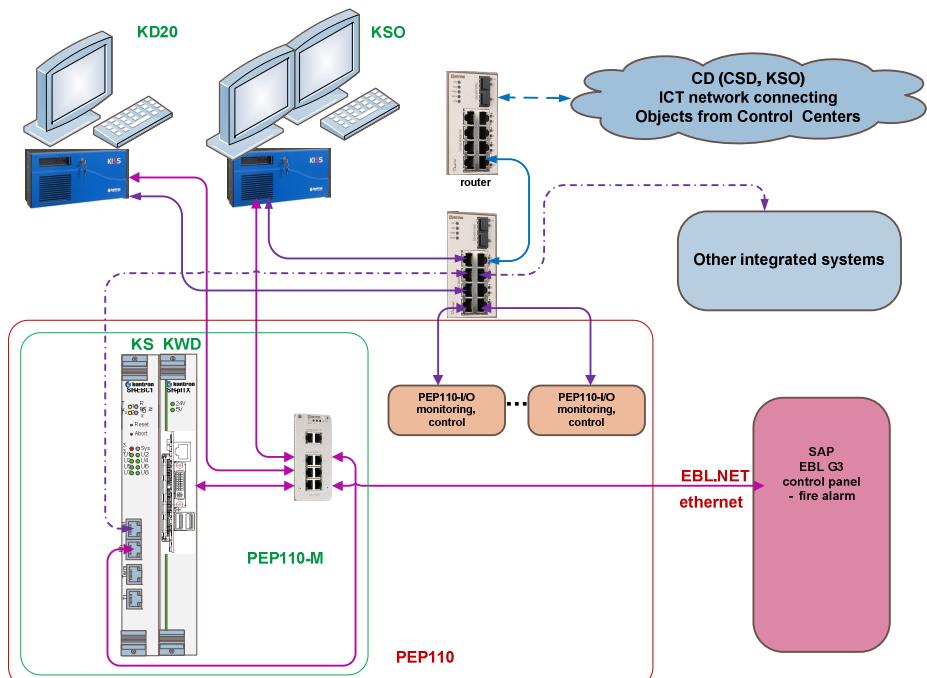


Fig. 2. Functional diagram of the integrating system

Software requirements of the KSO control desk:

- a) operating system – Windows XP Embedded or newer,
- b) the system should allow for the collection of information on alerts, notifications and actual parameters transferred from the connected control units and controls for fire protection equipment and display them in a transparent manner on one or more KSOs,
- c) system use (operation) should be implemented using standard peripherals (mouse, touchpad, keyboard),
- d) the system should be adapted to display information on multiple monitors, one of them, selected from the connected monitors, should display the fire information only,
- e) all events taking place throughout the system should be automatically recorded in real time; a function is required to print this information on request, also including the use of appropriate filters,
- f) information about failures signalled in the connected control panels of the fire protection system should also be indicated visually and acoustically by the KSO,
- g) information about any disruption of data transmission should be signalled visually and acoustically by the KSO,
- h) disruptions of EBL G3 operation should not affect the operation of the integrating system, disruptions of KSO operation should not affect the work of EBL G3,
- i) imaging system should enable easy discovery of the detector element based on designations consistent with those used when handling EBL G3 and display of the current status of the element, with automatic indication of its location on the layout,
- j) in the absence of communication between the facility device and fire alarm system., information about such a situation should be indicated in the computer system; the notification time depends on the delays characteristic of the signalling system communication protocol and must not exceed 300 seconds after the connection is interrupted,
- k) all messages received by the KSO should be confirmed after their safe recording by the recorder; the messages may not be processed before they are secured,
- l) new alarm message should be displayed until it is confirmed by the stand staff member or another authorized person,
- m) damage messages should be processed in the same way as alarm messages.
- n) messages should be retrieved from the message queue in the order they are received,
- o) the system should allow the on-demand presentation of messages sent from each EBL G3 control panel connected to the system,
- p) the system should provide in real time all unapproved fire alarm messages,
- q) the system should allow for the acceptance of messages; the message signalisation ends with the acceptance of the presented warning indication,

- r) damage messages should include specific data, at least: the type of damage (eg power failure), the event (loss of primary power supply), the date and time of the event,
- s) event registration should be implemented in at least two registers (redundant write), located in the KD20 diagnostic computers on the stands and in an additional CSD central data server in the Control Centre,
- t) registry damage (bad sectors, lost messages) should be detected by the system,
- u) all messages received by the KSO should be recorded with the date and time of confirmation (with an accuracy of 1 s),
- v) all information about damage and local information generated by the KSO should be recorded with the date and time of generation,
- w) capacity of the memory used should guarantee a long-term recording (1–2 years),
- x) register shall enable the preservation of data (messages and local information) in the event of a total loss of power,
- y) in the case of a corrupted registry warning alarm should be generated within 10 seconds of receiving a message by the KSO that could not be registered because of a corrupted registry.

3.2 Imaging the Status of all Integrated Systems

The system presents graphical information about the status of all monitored/integrated systems and their constituent components (individual inputs, outputs, sensors, zones, collective sensors informing about the states, etc.). Imaging system takes into account the states reported by the device:

- normal,
- damage,
- alarm/excitation.

The presentation of states of each component is implemented in a manner agreed upon with the customer.

3.3 Functional Requirements for the KD20 Diagnostic Computer

KD20 diagnostic computers should allow for:

- registration of data (events and states of the system) in three independent registers; lack of registration should be signalled to the system operator,
- diagnostics of the hardware, system, software, communication with other systems,
- transmission access to facility signals,
- supervision of the system, as well as processing and analysis of all the information from the facility – from field devices and SAP Control Panel.

4 Functional Requirements of PEP110 System

PEP110 system is an open software-hardware platform carrying out functions of registration of events and states of the individual system elements and its environment, data acquisition, imaging and reporting. It also allows the implementation of the master script, affecting the SAP central panel, according to the prepared scenarios for emergency purposes.

At the inputs of the SAP EBL G3 control panel signals are given from the optical detectors and manual call points located in the route/track fan units chambers and at stations, and from heat detectors installed only at stations. Depending on which of the sensors gets excited, the appropriate script is launched consisting in controlling the pressure-suction fans and dampers.

The prepared sample fire fighting scenarios are carried out both by the central fire protection control panel, and at the supervisory level – run by the service centre staff or the authorized persons in the Control Centre. Each element of fire protection system can also be actuated manually.

Integration of systems and devices implemented in PEP110, aims to, among others, to unify the presentation method of information from different systems made by different manufacturers. Designed and manufactured PEP110 system guarantees integration of many systems:

- fire alarm and indicating control panels (SAP control panels),
- audible warning systems (DSO),
- desmoking- and fire venting control systems,
- natural ventilation systems (dampers and smoke extraction vents),
- emergency lighting systems,
- fire division elements (doors, curtains, gates),
- devices and systems of the fixed extinguishing devices,
- other systems, installations and equipment used or operated during the state of the fire alarm (fire lifts, escalators, walkways covered with access control, etc.),
- equipment and video monitoring systems and visual alarm verification,
- alarm transmission systems,
- control and measurement control panels for gas detection systems.

The PEP110 system guarantees the information transmission safety. IP protocol is used for data transmission on the fibre optic loop. Communication in the loop fibre is secured with the encryption protocol – in accordance with the user requirements.

5 Technical Approval – Required Tests

The required tests were carried out at the approving entity premises to obtain the technical approval for the developed system. The initial type testing is the test confirming the required technical and operational properties, performed before introducing the devices into the market.

The initial type testing includes:

- a) cold resistance,
- b) resistance to stable moist heat,

- c) impact strength,
- d) resistance to power supply voltage fluctuation,
- e) resistance to electrostatic discharges,
- f) resistance to transmitted disturbances induced by radiofrequency fields,
- g) resistance to series of electrical fast transients,
- h) resistance to voltage surges,
- i) resistance to radiofrequency electromagnetic fields radiation,
- j) resistance to dynamic decays and fluctuations of power supply voltage.

Routine tests were carried out of finished products such as checking their appearance and labelling.

The tests, which were the basis of the approval procedure to determine the technical and operational characteristics of the products, are also the initial type testing in the assessment of conformity.

Design and installation work on the PEP110 integrating system is conducted and documented in accordance with AQAP 2110 – NATO requirements concerning quality assurance in design, development and manufacture.

6 Summary

The designed and tested system is the IT information system that integrates security systems. Its primary purpose is to inform the system operator of critical situations, controlling them and launching the previously developed scenarios.

Implementation of systems integration helps to reduce the number of false alarms. This system can be developed and modified, depending on the nature of the facility, its purpose and installed automation systems.

Currently, the system has undergone numerous trials for a variety of typical and critical situations. It has also obtained the approval of the Building Research Institute AT159106/2013.

References

1. GEMOS System,
<http://www.automatyka-budynkowa.com/produkty/kat/440/>
2. DMS8000 Danger Management System,
<http://www.buildingtechnologies.siemens.com>
3. Panasonic EBL G3. Analogue Addressable Fire Alarm System,
<http://www.raj-international.net/index.php?nid=2>
4. Jasiński, S.: Detection of objects on railway stations. *Pomiary Automatyka Robotyka* 15(10), 64–67 (2011) (in Polish)
5. Grochowski, K., Jasiński, S., Maciejewski, M., Sitek, I.: Development of a reliable railway traffic control system. *Pomiary Automatyka Robotyka* 12(11), 19–23 (2008) (in Polish)
6. Dyduch, J., Kornaszewski, M.: Railway traffic control systems. WPR, Radom (2007) (in Polish)

The Systematized Data Structures Oriented Towards Diagnosis and Prediction

Mariusz Piotr Hetmańczyk, Jerzy Świder, and Grzegorz Wszolek

The Silesian University of Technology, Faculty of Mechanical Engineering,
Institute of Engineering Processes Automation and Integrated Manufacturing Systems,
44-100 Gliwice, Konarskiego 18A Street, Poland
{mariusz.hetmanczyk,jerzy.swider,grzegorz.wszolek}@polsl.pl

Abstract. The structured notation using an identification of relations between key process data and functional characteristics of distributed systems has been presented in the paper. Such approach allows obtaining optimal form mapping the structures stored in a PLC's memory and also forming reciprocal relationships between hardware devices. Described notation is used at the stage connected with elaboration of diagnostic and prognostic functions. The authors have proposed an universal notation in a domain of sets, which can be used in identification of operational data indispensable in phase connected with a definition of program structures. A particular attention was given to stages connected with optimization necessity of number of main parameters (especially an isolation of a minimal data set required at control and diagnosis stages in case of distributed drives). Revised data sets combine both quantitative and qualitative features, allowing simultaneously for a reconstruction of the structural relationships between real components of distributed systems (a minimization of technical documentation).

Keywords: data optimization, expert systems, monitoring, diagnosis, distributed drives.

1 Introduction

The majority of automated production systems used in industrial applications are supported by Programmable Logic Controllers (PLCs). Simultaneously there has been a noticeable phenomenon of a growing share of industrial computers [7].

For small applications units belonging to Micro or Nano groups are used. What arises from the beneficial ratio of price to performance and the characteristics of serviced processes (limitation of measurement points to several). Situation is changing diametrically when the analysis concerns a system with superior control unit (a MASTER device) and SLAVE units connected via network interface.

In this case it is reasonable to use modular controllers with the possibility of extension of the control unit functionality adjusted to current system needs.

The advantages of modular units include among others:

- wide scope of hardware configuration varieties (including ease of modification, expansion or replacement),
- minimization or elimination of downtimes by means of a redundant system,
- a high computing power of modern processors that allows an implementation of additional features (implementation of data conditioning),
- the ease of implementation of industrial networks and mutual communication between different PLC units (due to the standardization of network interfaces).

The optimization problem within assumed criteria related to a minimization of algorithm volume and data acquisition rate is a significant task in all industrial applications [1,2].

Distributed systems come across considerable difficulties in the diagnosis phase. The difficulties concern the need of determination of the interrelationships between the functional elements (especially distributed drives, arranged on large areas or in hard to reach locations).

The proposed structure includes process data used for control, monitoring, diagnosis and prognosis of complex mechatronic devices equipped in industrial networks interfaces.

2 The Classification of Control and Diagnostic Data

The range of analysed relations (i.e. the set of signals gathered from machines environment in the form of production lines or machining centres), input functions and their effects has been marked as S_{S_ROZ} .

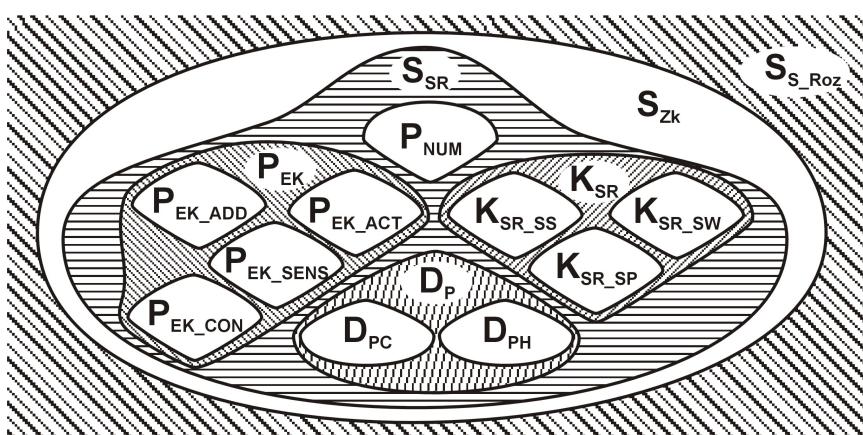
The set of descriptive information has been denoted as S_{SR} (among others including quantitative and qualitative data) and has been divided into several disjoint groups:

- quantitative parameters (P_{num}) – the unique order number,
- operational parameters and design features of: actuators (P_{EK_ACT}), control equipment (P_{EK_CON}), sensors (P_{EK_SENS}), auxiliary devices (P_{EK_ADD}),
- the dependency structure (K_{SR}) describing features of hardware (K_{SR_S}) and software (K_{SR_P}) in configurations related to:
 - the individual control: superior devices (K_{SR_SS}), actuators (K_{SR_SW}),
 - the group control (necessary consideration of the configuration of used industrial networks K_{SR_SP}),
- process data (D_P): current data D_{PC} (an application level – SCADA, control and diagnostic), historical data D_{PH} (higher levels used in a production control and management).

Table 1. Denotations of isolated sets with description of contents

Denotation		Contents
P_{Num}		A quantitative representation of the distributed system (a systematization and a definition of individual identification numbers assigned to identified components)
P_{EK}	$P_{\text{EK_ACT}}$	Drive parameters (electric motor + reduction gear + items mediating in motion), inputs and outputs of diagnosed modules
	$P_{\text{EK_CON}}$	PLC controller parameters, description of operation principle of the work and the hardware configuration
	$P_{\text{EK_SENS}}$	Sensor parameters
	$P_{\text{EK_ADD}}$	Additional elements, e.g. power cables (cross-sectional area, length), communication cables (length, type, terminations etc.)
K_{SR}	$K_{\text{SR_SS}}$	Individual settings of the control units, responsible for the formation of the control and response signals (control of operational states of drives)
	$K_{\text{SR_SW}}$	Hardware settings affecting a nature of the work of drives (continuous control of rotational speed during operation or a START/STOP mode with a constant rotational speed in a steady state)
	$K_{\text{SR_SP}}$	The nature of relations between units of the considered system (MASTER/SLAVE devices), configurations and physical structures of industrial networks
D_p	D_{PC}	Number and type of used variables (BOOL, INT, WORD, DWORD, etc.) dependent on types of controlled or monitored devices
	D_{PH}	Historical data, the immediate values of D_{PB} cells subjected to backup (contents depending on the type of monitored units)

Graphical illustration of isolated relationships is shown in Figure 2.

**Fig. 1.** Schematic representation of correlations between isolated sets

Each PLC controller operates on the basis of conversion of memory contents. In the case of complex control algorithms difficulties regarding the addresses identification (both physical and virtual) are frequent. The use of the proposed recording allows quick reproduction of searched relationships.

In any case, excessive memory load of control units slows down their operation what in critical applications may disturbed the correct operation of the designed algorithm [4].

3 The Data Notation Using the Proposed Approach Dedicated to Distributed Control Units

3.1 Data Vectors of Actuators and Sensors

Presented data structures $P_{EK_ACT_{(1,2,\dots,i,\dots,k)}}$ have been developed for frequency converters. In this case parameters are defined as follows:

$$P_{EK_ACT_{(1,2,\dots,i,\dots,k)}} = \langle P_{MOT_i}, P_{MP_i} \rangle \quad (1)$$

$$P_{MOT_i} = \cup \langle P_{ENG_i}, P_{GEM_i} \rangle \quad (2)$$

$$P_{MP_i} = \cup \langle P_{IO_i}, P_{INV_i} \rangle \quad (3)$$

where: P_{MOT_i} – set of gear motor attributes, P_{MP_i} – set of field module attributes, P_{ENG_i} – set of electric drive attributes, P_{GEM_i} – set of gear attributes, P_{IO_i} – set of IOs module attributes, P_{INV_z} – set of frequency inverter attributes.

Contents of enumerated sets of actuators were isolated in the following form:

- gear motor attributes:

$$\begin{aligned} P_{MOT_i} = & \langle M_{MOT_TP_i}, M_{MOT_SN_i}, M_{MOT_MT_i}, M_{MOT_HR_i}, M_{MOT_OM_i}, \\ & M_{MOT_cos\varphi_i}, M_{MOT_n_i}, M_{MOT_PSV_i}, M_{MOT_CL_i}, M_{MOT_IC_i}, \\ & M_{MOT_PT_i}, M_{MOT_BV_i}, M_{MOT_BT_i}, M_{MOT_GR_i}, M_{MOT_M_i} \rangle \end{aligned} \quad (4)$$

where: $M_{MOT_TP_i}$ – motor type (a unique supplier code), $M_{MOT_SN_i}$ – serial number, $M_{MOT_MT_i}$ – mounting type, $M_{MOT_HR_i}$ – power [kW], $M_{MOT_OM_i}$ – mode of operation, $M_{MOT_cos\varphi_i}$ – power factor, $M_{MOT_n_i}$ – nominal rotational speed [rpm], $M_{MOT_PSV_i}$ – required supply voltage [V], $M_{MOT_CL_i}$ – nominal rated current [A], $M_{MOT_IC_i}$ – insulation class, $M_{MOT_PT_i}$ – degree of protection, $M_{MOT_BV_i}$ – supply voltage of a brake circuit [V], $M_{MOT_BT_i}$ – braking torque [Nm], $M_{MOT_GR_i}$ – gear ratio, $M_{MOT_M_i}$ – total weight [kg].

- field module attributes:

$$P_{IO_i} = \langle P_{IO_TP_i}, P_{IO_SN_i}, P_{IO_SIN_i}, P_{IO_IN_num_i}, P_{IO_OUT_num_i} \rangle \quad (5)$$

where: $P_{IO_TP_i}$ – type, $P_{IO_SN_i}$ – serial number, $P_{IO_SIN_i}$ – types of supported industrial networks, $P_{IO_IN_num_i}$ – number of inputs, $P_{IO_OUT_num_i}$ – number of outputs.

- frequency inverter attributes:

$$P_{INV_i} = \langle P_{INV_TP_i}, P_{INV_SN_i}, P_{INV_PSV_i}, P_{INV_CW_i}, P_{INV_CL_i}, P_{INV_OF_i}, P_{INV_MP_i} \rangle \quad (6)$$

where: $P_{INV_TP_i}$ – type, $P_{INV_SN_i}$ – serial number, $P_{INV_PSV_i}$ – required supply voltage [V], $P_{INV_CW_i}$ – connection of motor windings (G-star, T-triangle), $P_{INV_CL_i}$ – current load [A], $P_{INV_OF_i}$ – output frequency [Hz], $P_{INV_MP_i}$ – allowable power of the connected motor [kW].

Another element binding the structure of distributed systems is the attributes subset of sensors:

$$P_{EK_SENS_{(1,2,...,j,...,d)}} = \langle P_{TP_j}, \langle P_{EL_SENS_j} \rangle, \langle P_{MECH_SENS_j} \rangle \rangle \quad (7)$$

$$P_{EL_SENS_j} = \langle N_{SV_j}, I_{NC_j}, U_{VD_j}, I_{LC_j}, I_{CC_j} \rangle \quad (8)$$

$$P_{MECH_SENS} = \langle \langle W_{D_j} \rangle, M_{M_j}, K_{CT_j} \rangle \quad (9)$$

where: P_{TP_j} – type (a vendor name), $P_{EL_SENS_j}$ – basic electrical parameters, $P_{MECH_SENS_j}$ – basic mechanical parameters, N_{SV_j} – supply voltage [V], I_{NC_j} – nominal rated current [mA], U_{VD_j} – voltage drop [V], I_{LC_j} – leakage current [mA], I_{CC_j} – current consumption [A], K_{CT_j} – connection type, M_{M_j} – construction material, W_{D_j} – dimensions.

3.2 Vectors of Auxiliary Data

Configuration of relations between network components K_{SR} has been described in the form of following sets:

$$K_{SR} = \langle \langle K_{SR_SS} \rangle, \langle K_{SR_SW} \rangle, \langle K_{SR_SP} \rangle \rangle \quad (10)$$

$$K_{SR_SS} = \langle L_{INV_IW_i}, L_{IINV_OW_i} \rangle \quad (11)$$

$$K_{SR_SW} = \langle T_{INV_OM_i}, T_{I/O_i}, \langle O_{n_SLV_i} \langle O_{n_Sens_1}, O_{n_Sens_2}, \dots, O_{n_Sens_d} \rangle \rangle \rangle \quad (12)$$

$$K_{SR_SP} = \left\langle O_{n_MSTR_1}, \left\langle O_{n_SLV_1}, O_{n_SLV_1}, \dots, O_{n_SLV_k} \right\rangle \right\rangle \quad (13)$$

where: $L_{INV_IW_i}$ – number of input words (feedback information), $L_{INV_OW_i}$ – number of output words (control information), $T_{INV_OM_i}$ – work mode (0 – the Start/Stop mode, 1 – stepless speed control), T_{I/O_i} – I/Os service (notation: T/number of sensors – the use of I/Os, N – I/Os inactive), $O_{n_Sens_j}$ – ordinal number of a sensor supported by the SLAVE station with the $O_{n_SLV_i}$ number, $O_{n_MSTR_1}$ – ordinal number of MASTER station, $O_{n_SLV_i}$ – ordinal number of the SLAVE station supervised by the MASTER unit with $O_{n_MSTR_1}$ number.

Subsets of data cells and diagnostic program data were divided into two separate groups, in the form:

$$D_P = \langle\langle D_{PC} \rangle, \langle D_{PH} \rangle \rangle \quad (14)$$

$$D_{PB} = \langle\langle D_{PC_t}, D_{PC_16} \rangle, \langle D_{PC_t}, D_{PC_32} \rangle, \langle D_{PC_t}, D_{PC_1} \rangle \rangle \quad (15)$$

$$D_{PH} = \langle\langle D_{PH_t}, D_{PH_16} \rangle, \langle D_{PH_t}, D_{PH_32} \rangle, \langle D_{PH_t}, D_{PH_1} \rangle \rangle \quad (16)$$

where: $D_{PC_t_z/s}$ ($D_{PH_t_z/s}$) – allocation of words and bits (respectively control, diagnostics, notation S/D, S/0, 0/D) of current and historical data, $D_{PC_16_z/s}$ ($D_{PH_16_z/s}$) – number of 16-bit words (divided into types: WORD, INT, REAL, STRING), $D_{PC_32_z/s}$ ($D_{PH_32_z/s}$) – number of 32-bit words (divided into types: DWORD, DINT, REAL, TIME), $D_{PC_1_z/s}$ ($D_{PH_1_z/s}$) – number of bit cells (the BOOL type).

3.3 Data Vector of the MASTER Station

Described vectors are applied to reduce the notation of complex distributed systems and minimize the technical documentation volumes.

Structural correlations were brought to the form of ordered vectors containing data classified according to the scheme presented in Figure 1.

As an example the authors presented the system based on the ProfiBus DP network with eight SLAVE devices supervised by the MASTER unit (Fig. 2).

Each unit is identified by a unique ordinal number O_n (dependent on the network address) allowing an unambiguous identification of all devices of the considered control system [3].

In order to identify the relations between constituent elements the authors have used the n-dimensional dependences in the following form:

- The relationship of quantitative parameters assigned for the unit I_d ,
- Structures of subsystem and in the consequence created system,
- The structure of information exchange (assignment of SLAVE units up to the MASTER station),
- Assigning the relations of sensors to the relevant network module.

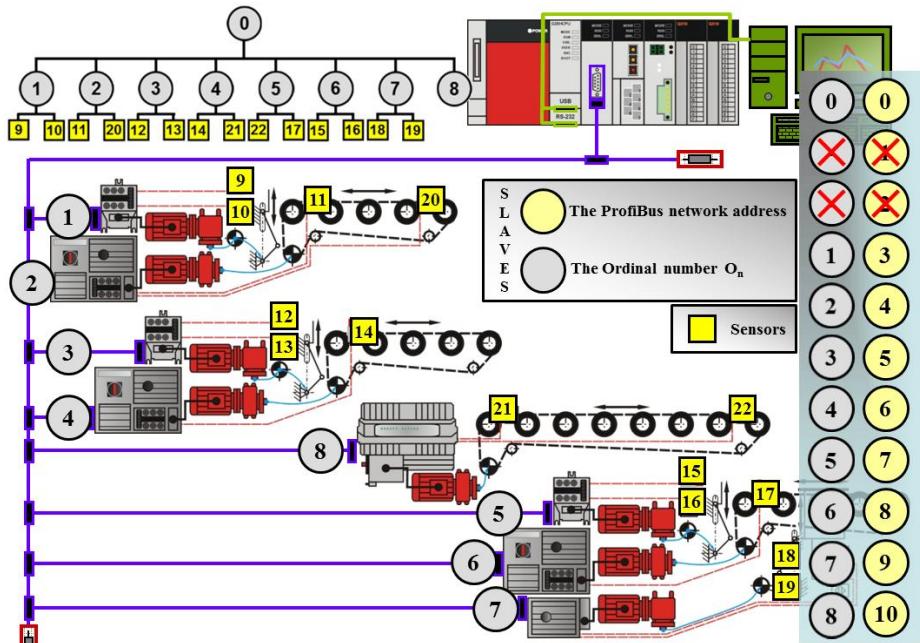


Fig. 2. The simplified diagram of a distributed system with usage of benchmarks determining reciprocal interdependencies

Subsets of attributes related to the MASTER unit P_{EK_CON} (in the form of a PLC controller) have been defined as:

$$P_{E K _C O N _(1,2,...,l,...c)} = \langle P_{T P _C O N_1}, \langle P_{C O N F _C O N_1} \rangle, \langle P_{E L _C O N_1} \rangle, \langle P_{M E C H _C O N_1} \rangle \rangle \quad (17)$$

$$P_{C O N F _C O N_1} = \langle P_{C O N F _M o d_1}, L_{I O _D_1}, L_{I O _A_1}, L_{N M_1}, T_{N M_1}, K_{R M_1}, L_{R M_1} \rangle \quad (18)$$

$$P_{E L _C O N_1} = \langle P_{V S_1}, P_{V S T_1}, P_{E L _I O_1} \rangle \quad (19)$$

$$P_{M E C H _C O N_1} = \langle \langle W_{D _C O N_1} \rangle, T_{F T _C O N_1} \rangle \quad (20)$$

where: P_{T P_C O N_1} – PLC controller type, P_{E L_C O N_1} – main electrical parameters, P_{M E C H_C O N_1} – mechanical parameters, P_{C O N F_C O N_1} – configurations of basic and additional modules (P_{C O N F_M o d}) or remote units, L_{I O_D} – number of digital inputs/outputs, L_{I O_A} – number of analog inputs/outputs, L_{N M_1} – number of network modules, T_{N M_1} – types of network modules, L_{R M_1} – number of remote modules, K_{R M_1} – configurations of remote modules, P_{V S_1} – supply voltage [V], P_{V S T_1} – type of power supply, P_{E L_I O_1} – electrical parameters of inputs/outputs, W_{D_C O N_1} – dimensions, T_{F T_C O N_1} – a fixing type.

By means of the definition a generalized product of sets an ordered set of elements containing attributes of a distributed system was created. The described vector has been formulated as follows:

$$P_{PAR_CON_I} = J_{IDEN_CON_I} \times P_{EK_CON_I} \times D_{P_I} \times K_{SR_SP_I} \quad (21)$$

This procedure enables obtaining a complete vector that defines the structure of control unit (the MASTER device):

$$\begin{aligned} P_{PAR_CON_I} = & \left\langle 1, \left\langle P_{TP_CON_I}, P_{KONF_Mod}, L_{IO_D_I}, L_{IO_A_I}, L_{NM_I}, T_{NM_I}, K_{RM_I}, L_{RM_I}, \right. \right. \\ & P_{VS_I}, P_{VST_I}, P_{EL_IO_I}, \dim_x, \dim_y, \dim_z, T_{FT_CON_I}, D_{PC_t}, D_{PC_16}, \\ & D_{PC_t}, D_{PC_32}, D_{PC_t}, D_{PC_I}, D_{PH_t}, D_{PH_16}, D_{PH_t}, D_{PH_32}, D_{PH_t}, \\ & \left. \left. D_{PH_I}, O_{n_MSTR_I}, O_{n_SLV_I}, O_{n_SLV_I}, \dots, O_{n_SLV_k} \right\rangle \right\rangle \end{aligned} \quad (22)$$

After a substitution of physical parameters to the given structure the authors obtained vector presented in Figure 3.

In the presented case an encapsulation of data was applied, which involved the extraction of data into subsets [8, 9] classified according to their suitability for the purpose of control, diagnosis, management, etc.

The set of functional interdependencies (in case of the considered example) includes twenty three vectors describing the relations and parameters of the analysed components (the MASTER unit, eight SLAVE units and fourteen sensors).

The presented form of notation is clear and is characterized by great adaptation possibilities for the analysis of systems with significant numbers of data with the ability of control and historical data isolation.

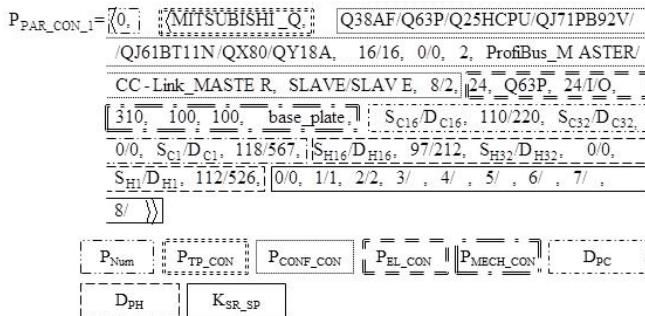


Fig. 3. View of the features vector of a PLC MASTER unit

Presented structure is based on maintenance manuals of network modules and also on the IEC 61158 standard ("Digital Data Communications for Measurements and Control - Fieldbus for Use in Industrial Control Systems") and the IEC 61784 standard ("Profiles Sets for Continuous and Discrete Manufacturing Relative to Fieldbus Use in Industrial Control Systems").

The exemplary set for actuators (i.e. frequency inverters, Softstart modules, gear motors) is presented in the Figure 4.

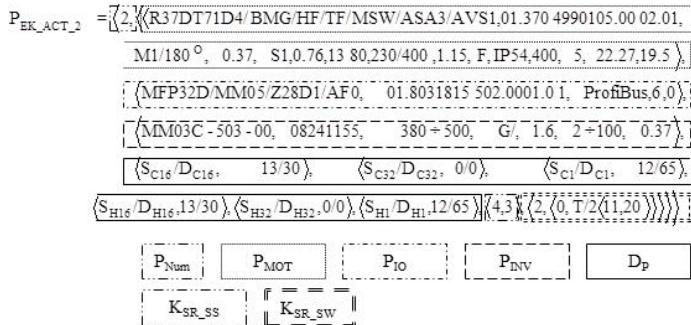


Fig. 4. View of the gear motor vector

In case of industrial sensors the vector can be written in accordance to the scheme shown in Figure 5.

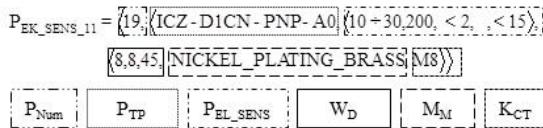


Fig. 5. View of the vector with described features of the sensor unit

With respect to all of the sets of the distributed system the relations between the identified subsystems were described.

The legibility of the record notation, allowing rapid identification of the quantitative parameters of the system and the functional parameters of individual units, constitutes an additional advantage.

4 Conclusions

Systems which use industrial networks allow the collection of three types of data in the form of diagnostic information, extended diagnostic information and alarm information.

The proposed structure is a sufficient source of information for control and diagnostic purposes, especially with the use of industrial networks. A data classification allows standardization and elaboration of the approach dedicated to rational resource management [5, 6], especially the memory of control units (with the main assumption of required high reliability).

The main criterion used in the data type identification phase (especially in processing algorithm) is maintaining the control algorithm duty by adjusting the number of required variables for each considered case.

In spite of the definition of the framework there still exists the problem of classifying the process data to the particular set. The described operation is not subjected to algorithmization, and the final decision must be made by the designer of the control system.

References

1. Bloch, H.P., Geitner, F.K.: Practical Machinery Management for Process Plants – Machinery Component Maintenance and Repair. Elsevier Linacre House, Oxford (2005)
2. Dhillon, B.S.: Engineering Maintenance - A Modern Approach. CRC Press, New York (2002)
3. Hetmanczyk, M.P., Michalski, P., Swider, J.: Utilization of advanced self-diagnostic functions implemented in frequency inverters for the purpose of the computer-aided identification of operating conditions. *Journal of Vibroengineering* 14(1), 117–122 (2012)
4. Korbicz, J., Kościelny, J., Kowalcuk, Z., Cholewa, W.: Processes diagnostic: models, methods of the artificial intelligence, applications. WNT, Warsaw (2002)
5. Krenzczyk, D., Kalinowski, K., Grabowik, C.: Integration Production Planning and Scheduling Systems for Determination of Transitional Phases in Repetitive Production. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012, Part II. LNCS, vol. 7209, pp. 274–283. Springer, Heidelberg (2012)
6. Grabowik, C., Krenzczyk, D., Kalinowski, K.: The Hybrid Method of Knowledge Representation in a CAPP Knowledge Based System. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012, Part II. LNCS, vol. 7209, pp. 284–295. Springer, Heidelberg (2012)
7. Legierski, T.: PLC Controllers Programming. Jacek Skalmierski Publishing, Gliwice (1998)
8. Ryabinin, I.: Reliability of engineering systems. MIR Publishers, Moscow (1976)
9. Hetmańczyk, M.P.: The Multilevel Prognosis System Based on Matrices and Digraphs Methods. In: Gosiewski, Z., Kulesza, Z. (eds.) Solid State Phenomena, vol. 199, pp. 79–84. Trans Tech Publications, Switzerland (2013)

The Analysis of the Registration Accuracy of Distributed Drives Parameters

Mariusz Piotr Hetmańczyk, Jerzy Świder, and Grzegorz Wszołek

The Silesian University of Technology, Faculty of Mechanical Engineering,
Institute of Engineering Processes Automation and Integrated Manufacturing Systems,
44-100 Gliwice, Konarskiego 18A Street, Poland
{mariusz.hetmanczyk,jerzy.swider,grzegorz.wszolek}@polsl.pl

Abstract. The article includes a description of the validation used for an assessment of accuracy of measured symptoms in case of distributed drives connected via the ProfiBus DP network. Particular attention has been given to acquiring the diagnostic data especially in case of currents intensities in the start-up and braking phases. The authors also raise a problem of the usefulness and precision of measured data obtained with usage of recording methods by means of dedicated system in comparison with the solutions delivered by the manufacturer of examined drives. Presented results of practical measurements contain values based on the registration of output currents of frequency converters and the measured values of current intensities in the motor windings. The described parameters are significant for the evaluation of electrical damages, caused by various factors (including excessive loads, improper selection, etc.).

Keywords: distributed systems, state diagnosis, expert systems, industrial networks.

1 Introduction

Modern driving systems especially distributed drives with frequency converters are characterized by a high degree of structural complexity. The mentioned feature generates difficulties concerning access to the diagnostic socket, which is equivalent to the need of developing alternative methods of data collection and centralized access to resources. Another issue concerns the analysis of process data and the definition of set necessary for states estimation. The identification and an analysis of diagnostic variables depend on employee qualifications, elaborated diagnostic inference structure and diagnostic premises.

In most cases technical personnel uses dedicated computer systems or solutions provided by the manufacturers of the used devices. Expert systems facilitate the diagnosis process but in case of obtaining poor data quality results can lead to incorrect diagnoses.

The described facts, significant in case of diagnosis and implementation of corrective actions (enclosed in the adopted maintenance model), require the analysis of their advantages and disadvantages (Table 1).

Table 1. Differences between most popular approaches in relation to the diagnosis of distributed drives

Considered feature	Common solution	Proposed approach [6]
Used equipment	Highly specialized measuring equipment	Data collector (a PLC controller and additional databases)
Measurements and control	Periodically (machines checked at the distinct request)	Continuous control (continuous observation with browsing of history and forecasting of significant states)
Possibility of adaptation to changing requirements	Low (require hardware changes or analysers modernization)	Very high (usage of specialized modules, changes in the source code, replacement of sensors)
The degree of reliability	High (based on hardware)	Very high (possible use of redundant hardware and software)
Number of monitored variables	Usually one (sometimes more than one)	Number limited only by addresses of I/O expansion modules or hardware limitations of connected devices
Sampling frequency	High	Dependent on used processor (a high performance requires optimization of the program in terms of a volume and a cycle time)
Characteristics of the controlled variables	Dependent on used analysers	Dependent on additional devices (velocity, acceleration, displacement, temperature, motor current intensities, query of frequency converters)
Application flexibility	High (analysers can be used for testing multiple devices)	Low (system dedicated to the specific application, designed up to concrete operational requirements)

On the basis of the enumerated advantages the Computer Control and Diagnostic System of Distributed Drives (CCaDSoDD) [6–8] was developed. The described system enables diagnostics by means of the central MASTER (PLC controller), distributed modules and data exchange realized via the ProfiBus DP network (Fig. 1).

The ProfiBus network (Process Field Bus) is an open digital communication system with a vast array of industrial applications, based on the OSI layered model (Open Systems Interconnection Reference Model). The ProfiBus DP (Decentralized Peripherals) is a solution used to communicate modules of distributed systems based on the MASTER/SLAVE model [2].

The DP-V2 version (dedicated for industrial applications with motion control) allows synchronic data transfer and communication between network drives (with particular emphasis on the driving axes) at high speeds. The reliability and a high data

rate of the ProfiBus DP network enable a control and collection of diagnostics data [1]. This approach is a necessary for combining the data exchange with other units especially with dedicated advisory systems that are developed for the implementation of advanced features (focused on a specific application).

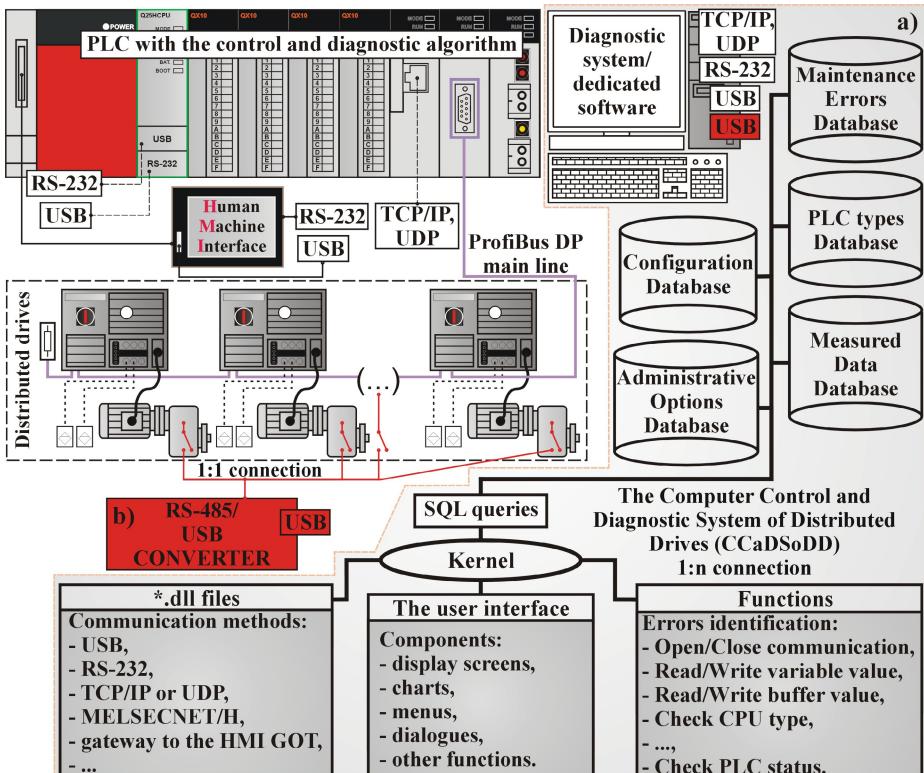


Fig. 1. Schematic representation of hardware and functional structures of: a) the CCaDSoDD system, b) measurements realized by unit delivered by the producer

The main problem concerns a data accuracy provided by the expert system through a network.

2 Identification of Failures in Case of Measurements of Current Intensities

The analysis shows that most of the failures in relation to electrical machines appears as a results of improper selection of work parameters to real conditions.

The most unreliable parts include rolling bearings. However there are no any alternative solutions to the decrease of the friction coefficient in the rotational motion.

The concepts presented in the literature are usually connected with diagnosis of individual electric drives and are based on two different variations of diagnostic experiments:

- testing of electric drives in laboratory conditions with simultaneous comparison of responses with a numerical model,
- the diagnostic analysis (*in situ*) based on the use of specialized equipment (separated measurement and control paths), measurements in order to check the degree of machine components wear.

It is required to pay attention to the identification of diagnostic premises allowing the prevention of failures that occur in a consequence of exceeding the rated current, due to various external factors. Depending on the operating conditions the phenomenon of excessive amounts of thermal energy release may lead to distortion or interruption of the operation (damage or total destruction).

3 The Practical Verification of Data Accuracy Evaluation

3.1 Limitations and Assumptions of the Developed Method

On the basis of the restrictive assumptions (adopted by the authors) electrical faults are evaluated by processing parameters of the distributed drives gathered by the PLC controller (the MASTER station).

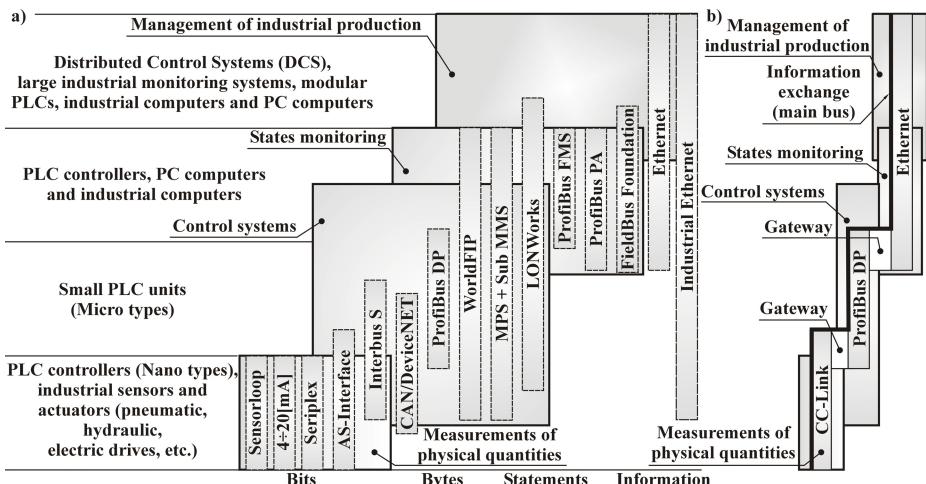


Fig. 2. Application range of industrial communication protocols: a) the most common protocols, b) interpretation of the main route in relation to the case described by the authors

Such approach causes limitations regarding the number of useful diagnostic relations, forcing a necessity of implementation of simplifying assumptions [3].

Based on literature analysis and taking into account restrictive assumptions [1,3] there was formulated a basis for the definition of functions allowing the parameters diagnosis with the usage of the central control, in the following form:

- mechanical components of electric drives can be monitored by observing the parameters of vibrations generated by the drive during operation,
- an identification and an assessment of the failure impact having the electrical nature, are related to the values of real and nominal output current intensities, drop-down torque value, load and braking torques,
- in the discussed case, data readout are performed by means of the dedicated algorithm (consolidation of control and diagnostic functions) saved in a PLC memory [6].

3.2 Evaluation of Data Accuracy in Case of the Fieldbus Data Exchange

In order to verify the developed algorithm dedicated for diagnosis and control, compliance tests were performed (replacement of in-situ measurements). Described test pertain to accuracy of the measured values gathered within the following assumptions:

- connection to diagnostic sockets of the producer device (single point measurement),
- data exchange via the ProfiBus DP network and analysis with the usage of tools implemented in the CCaDSOoD system,
- control is realized with the usage of the CCaDSOoD system,
- recorded values are collected using two independent data collectors: the CCaDSOoD system – diagnostic node, the MOVITOOLS-MotionStudio (or the MT-Manager) – the dedicated software for SEW units,
- parallel operations of two independent data collectors eliminate errors concerning the divergence of environmental conditions (a synchronous recording).

The registration were subjected an output current intensity of the inverter and the current intensity measured from electric drives by the frequency converter. All measured variables are expressed as a percentage value of the nominal current I_n . The aim of the study was to determine the extension of measurements compliance of both diagnostic tools and also a definition of the boundaries of effective sampling.

The study was conducted during transient states (start-up phase and braking). Data stored in a PLC controller were updated with a period equal to the time of the PLC cycle (or the network cycle). The consecutive step was recording the information (a computer with installed expert system - the processing and analysis).

Measurements were performed in sampling periods between 5 ms to 500 ms (the paper presents only selected results). In addition, each measuring cycle was carried out with the same parameters of the drive allowing comparison of results.

The selected sampling rate was determined based on current changes on the inverter output (in case of presented measurement the rotational speed has been limited to 138 rpm, giving the inverter output frequency change of 5 Hz).

In accordance to the Nyquist-Shannon sampling theorem:

$$f_s = \frac{1}{T} \quad (1)$$

where: f_s – signal frequency [Hz], T – signal period [1/s].

Minimal frequency rate f_p allowing reconstruction of the real waveform:

$$f_p \geq 2f_s \quad (2)$$

In the considered case the f_p frequency was defined as 10 Hz. Measurements made with 100 ms sampling period, allow for accurate representation of the course of the current intensity. On both characteristics, at the start-up (Fig. 3) and braking (Fig. 4), are visible waveform peaks resulted from the need to overcome motion resistance (putting the masses in motion or deceleration).

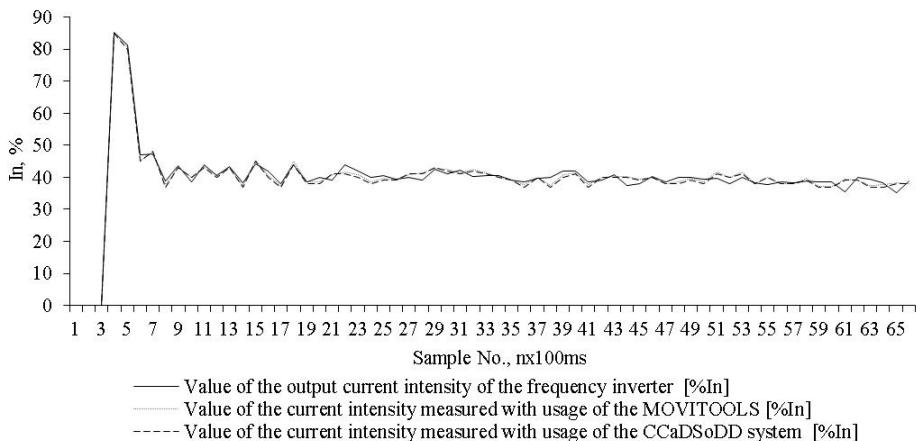


Fig. 3. A comparison of the current intensities recorded by the CCaDSoDD system and the MOVITOOLS-Scope (start-up phase, the sampling period equals to 100 ms, speed equal 10% and the value of ramp time equals to 1000 ms)

Recording with sampling time equals 500 ms (the most common value of data acquisition in SCADA systems) leads to the loss of a significant amount of information. In the start-up phase the peak current has not been recorded (Fig. 5), while in the stop phase the value was still observable, and the peak in the phase of multiple repetitions of the measurement is nondeterministic.

On the basis of the measurements the scope of the minimum and maximum sampling periods of the CCaDSoDD system (10 to 100 ms) for the measurement of transient states an analyzed distributed drive system has been defined. It should be also kept in mind that operating systems cause the limitations of the recording times related to the database of the expert system. A selection of the minimum registration

measurement times, with minimum value of 10 ms, results from limitations of the Windows system.

The necessity of sampling with higher frequencies required development of data buffering method in the cache memory of the PLC logic controller, and then transferring the contents of the buffers to an expert system [4,5].

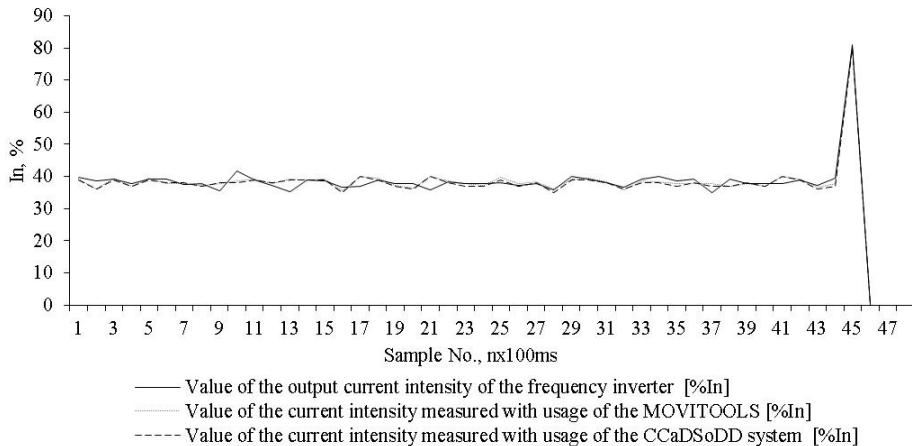


Fig. 4. A comparison of the current intensities recorded by the CCaDSOoD system and the MOVITOOLS-Scope (braking phase, the sampling period equals to 100 ms, speed equal 10% and the value of ramp time equals to 1000 ms)

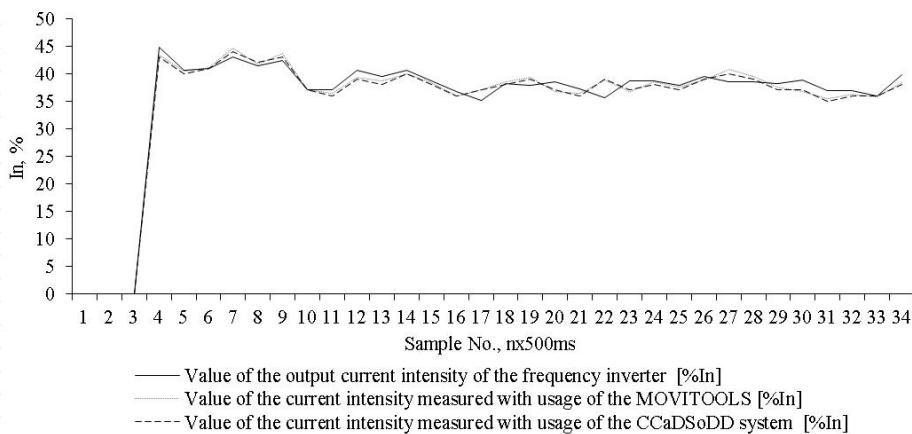


Fig. 5. A comparison of the current intensities recorded by the CCaDSOoD system and the MOVITOOLS-Scope (start-up phase, the sampling period equal 500 ms, speed equal 10% and the value of ramp time equal 1000 ms)

4 Summary

Evaluation method presented by the authors in case of the recording accuracy related to currents intensities of frequency converters is an important element confirming the usefulness of the application of dedicated expert systems in case of industrial applications.

As a measure of the assessment the values registered by software and hardware provided by manufacturer of considered drives were adopted. Tests realized on the industrial objects shows that the registration with the CCaDSoDD system [6], in case of the described values, do not differ much from the reference values. An important advantage of this system is the monitoring possibility with time periods limited only by the capacity of data storage medium on which the measured values of selected parameters are recorded.

The use of real-time computer allows data measurement and transmission in a packet system. Such approach significantly improves the measurements resolution. However, in the considered case the given minimum sampling frequency is the value that allows obtaining satisfactory results, and maintaining simultaneously the stability of the expert system.

References

1. Bloch, H.P., Geitner, F.K.: Practical Machinery Management for Process Plants – Machinery Component Maintenance and Repair. Elsevier Linacre House, Oxford (2005)
2. Hetmańczyk, M.P., Michalski, P., Świder, J.: Utilization of advanced self-diagnostic functions implemented in frequency inverters for the purpose of the computer-aided identification of operating conditions. *Journal of Vibroengineering* 14(1), 117–122 (2012)
3. Ryabinin, I.: Reliability of engineering systems. MIR Publishers, Moscow (1976)
4. Świder, J., Hetmańczyk, M.: Hardware and software integration of mechatronic systems for an example measurement path for temperature sensors. In: Gosiewski, Z., Kulesza, Z. (eds.) Solid State Phenomena, vol. 147-149, pp. 676–681. Trans. Tech. Publications, Switzerland (2009)
5. Świder, J., Hetmańczyk, M.: The visualization of discrete sequential systems. *Journal of Achievements in Materials and Manufacturing Engineering* 34(2), 196–203 (2009)
6. Świder, J., Hetmańczyk, M.: The computer integrated system of control and diagnosis of distributed drives. Silesian Technical University Publishing, Gliwice (2011)
7. Hetmańczyk, M.P.: The Multilevel Prognosis System Based on Matrices and Digraphs Methods. In: Gosiewski, Z., Kulesza, Z. (eds.) Solid State Phenomena, vol. 199, pp. 79–84. Trans. Tech. Publications, Switzerland (2013)
8. Hetmańczyk, M.P.: The Reliability Model of a AC-Asynchronous Drive Based on the Multilevel Prognosis System Based on Matrices and Digraphs Methods. In: Gosiewski, Z., Kulesza, Z. (eds.) Solid State Phenomena, vol. 199, pp. 85–90. Trans. Tech. Publications, Switzerland (2013)

CPDev Engineering Environment for Modeling, Implementation, Testing, and Visualization of Control Software

Marcin Jamro, Dariusz Rzońca, Jan Sadolewski, Andrzej Stec,
Zbigniew Świderek, Bartosz Trybus, and Leszek Trybus

Rzeszow University of Technology,
al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland,
`{mjamro,drzonca,js,astec,swiderzb,btrybus,ltrybus}@kia.prz.edu.pl`

Abstract. The paper presents the CPDev engineering environment, which is a comprehensive set of integrated tools for programming PLCs, PACs, and DCSs, according to the IEC 61131-3 standard. The environment supports various parts of the project development, including modeling (using SysML diagrams), implementation (with IEC 61131-3 textual, graphical, and mixed languages), testing (by textual, table, and performance tests), simulation (in off-line mode), debugging (with tracing variable values and breakpoints), hardware resources configuration, execution (either on the virtual machine or FPGA-based hardware machine), as well as visualization. The CPDev project is designed in a way supporting portability. Therefore, prepared control programs can be run on any supported target platform, such as AVR, ARM, and x86. The environment has a few applications, both industrial and laboratory.

Keywords: control software, engineering environment, IEC 61131-3, modeling, implementation, simulation, debugging, testing, visualization.

1 Introduction

The control software is an important and complex kind of the real-time software that is frequently used to control various devices and processes. It has many industrial applications, such as in PLCs (Programmable Logic Controllers), PACs (Programmable Automation Controllers), and DCSs (Distributed Control Systems). Nowadays, such a kind of software becomes more and more complex, as well as sometimes replaces some parts that have been earlier developed as hardware solutions [1]. For these reasons, it is important to propose a suitable development approach that is convenient for engineers and can make the whole process more robust, easier, and faster. One of the possible solutions may combine the MDD (Model-Driven Development) approach [2] and precise testing [3] with implementation and other necessary operations, such as configuration.

In this paper, the authors present the comprehensive solution that supports modeling, implementation, testing, visualization, and deployment of the control software, in the form of the integrated development environment named CPDev

(Control Program Developer)¹. It is created in the Department of Computer and Control Engineering at Rzeszow University of Technology (Poland).

The paper is organized as follows. In the second section, overview of the CPDev engineering environment is presented. The next section describes features related to modeling the control software using SysML graphical diagrams. The fourth part explains implementation functionalities, including support for textual and graphical languages. It presents also the simulator and debugging features. The fifth section is related to testing capabilities of the environment, including support for unit testing using textual and table tests, as well as checking performance requirements. The next part describes the multiplatform visualization mechanism integrated with CPDev. A few industrial and laboratory applications are briefly presented in the seventh section.

2 Overview of the CPDev Environment

The CPDev engineering environment allows to create control applications according to the third part of the worldwide IEC 61131 standard [4]. The norm defines five programming languages, namely ST (Structured Text), IL (Instruction List), FBD (Function Block Diagram), LD (Ladder Diagram), and SFC (Sequential Function Chart). The first two languages are textual, the next two are graphical, while the last one is mixed. The standard also supports division of the control software into smaller parts named POU (Program Organization Units), i.e. functions, function blocks, and programs. Each POU is created in a particular IEC 61131-3 language, however, the control software can be composed from POUs developed in different languages.

There are already some processors which can directly execute code in IEC 61131-3 languages [5], but it is not a common approach. Significantly more often, the controllers are based on multi-purpose processors. One of possible solutions is translation of POUs created in any language into the machine code for the target processor. In a common scenario, the temporary code is generated during the compilation stage. Then, it is used while linking to prepare the machine code, depending on the target platform configuration [6]. Unfortunately, this approach has a few disadvantages, such as no portability of the code, a necessity of recompilation to make even minor configuration modifications, as well as a necessity of changing the compiler to support another microprocessor. The next solution is based on the idea of translating all POUs to some high-level language, such as C or C++. Then, it is possible to compile the control software to the machine code for a particular target processor [7]. Another approach is based on the virtual machine, which is run on the target controller and interprets the intermediate code. Such a code is generated during compilation of control programs. In [8], the intermediate code is based on the IL language.

The solution with the virtual machine has been chosen also in case of the CPDev engineering environment, as proposed in [9]. In this case, the additional dedicated intermediate code, named VMASM (Virtual Machine Assembler), is

¹ <http://cpdev.kia.prz.edu.pl/>

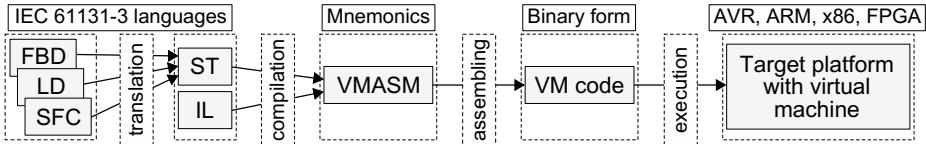


Fig. 1. Process of translation, compilation, and execution of the code in CPDev

used. The authors prepared a set of instructions for the virtual machine [10] that perform operations common for control programs, including simple (e.g. adding two numbers) and complex (e.g. calculating a day of week for a given date).

The ST language is chosen as the base language for the environment. Thus, POU_s prepared in graphical and mixed languages are automatically translated into the textual ST code. Then, such code is compiled to VMASM and assembled to the VM code in the binary form (Fig. 1). It is worth mentioning that this code is universal and can be executed on various target platforms, without recompilation. The supported platforms include AVR, ARM, x86, as well as the FPGA hardware implementation.

Each target platform has some specific features, such as a way of binding variables with inputs and outputs, as well as a definition of communication tasks. For this reason, it is necessary to configure hardware resources. In the CPDev solution, the executable code must be equipped with the platform-dependent hardware resources assignment map. The virtual machine, which is included in the controller's embedded software (firmware), interprets the code and performs control operations. The CPDev virtual machine is prepared in ANSI C language, thus adding support for additional target platforms is straightforward. The machine interfaces the hardware via a set of platform-dependent functions [10]. They handle hardware resources, such as system and real-time clocks, random number generator, and flash memory. They also perform time counting and maintain the controller cycle. To support another controller platform, the engineer should implement interface functions for a particular hardware. An additional PC tool is needed to upload CPDev programs to the controller, configure binary or analogue inputs and outputs, as well as map them to program variables.

The current version of the CPDev engineering environment supports not only the implementation and execution stages, but also modeling (see Sec. 3) and testing (see Sec. 5). All of these parts are cooperating together to compose a comprehensive solution for developing the control software, together with ensuring its high quality and easier maintenance.

3 Modeling with SysML Diagrams

Due to still increasing complexity of control software, it is important to propose various approaches to the design stage of the project development. The current results in this area are reviewed in [11]. Some concepts are related to moving the development process into a higher level of abstraction. It can be achieved

by the MDD (Model-Driven Development) approach. There is some research regarding applying such a concept to the industrial domain, such as conducted by Thramboulidis and Frey [12], as well as by Chiron and Kouiss [13]. The modeling stage of project development often uses a graphical modeling language, such as UML (Unified Modeling Language) [14] or SysML (Systems Modeling Language) [15]. The latter is based on UML, but provides engineers with possibility of modeling a wider range of system components.

The modeling features and tools in the CPDev engineering environment are described in [2]. They are based on the SysML modeling language and use four its diagrams to model the following parts of the project:

- functional and non-functional requirements with the Requirement Diagram
- resources and tasks with the Package Diagram
- a structure of POU with the Block Definition Diagram
- state machine-based POU with the State Machine Diagram

The authors indicated rules for proper use of the proposed methodology in [2]. For instance, a single Block Definition Diagram should be used for modeling a particular POU, which is represented by the **block** stereotype. All inputs and outputs are shown as flow ports. The programming language is specified as a value of the **lang** element, while the POU type as the POU stereotype. Other data (such as internal variables and tests dedicated to POU) are stored as values in the compartments. Complying to proposed rules is important, because diagrams are used to automatically generate an implementation part of the project. The structure of POUs is created based on data from Block Definition Diagrams, while their implementation – using associated State Machine Diagrams.

The environment has been equipped with the dedicated modeling tool, named CPModel. Its main window (Fig. 2) is divided into two parts – a list of diagrams on the left, and the main workspace on the right. Each supported SysML diagram can be created with the usage of build-in graphics editors.

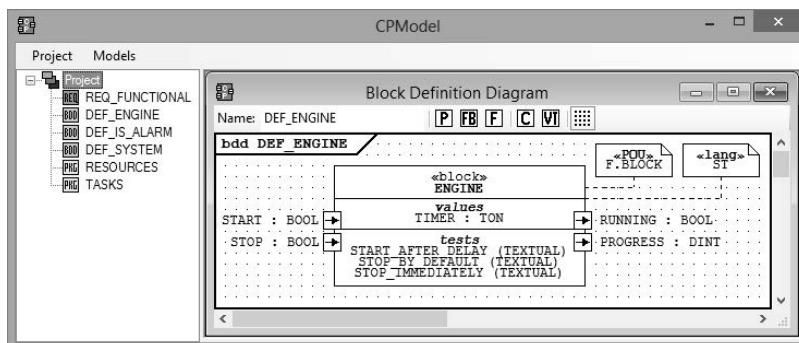


Fig. 2. Modeling tool in the CPDev environment

4 Implementation, Simulation, and Debugging

The CPDev engineering environment supports developing POU s in all languages from the IEC 61131-3 standard. The software contains textual and graphical editors [16], which are created as modules of the CPDev IDE (Fig. 3, at the top). All of them are equipped with many advanced features, including code completion in case of the ST editor, as well as automatic connection finding mechanism [17] in case of FBD, LD, and SFC editors. The IDE supports creation of function blocks, functions, and programs for their further reuse. The engineer can create own library or include default ones – with the standard IEC 61131-3 blocks, as well as with simple and complex blocks (such as PID regulators).

After preparing the implementation, it is useful to check the control program using the off-line simulator. Therefore, the engineer does not need to configure external hardware devices and can easily simulate the behavior of other system parts. To support such a functionality, the CPDev environment has been equipped with the CPSim application (Fig. 3, at the bottom). It supports starting, stopping, pausing, and resuming execution of the control program, as well as showing values of variables. What is more, it allows to present information about a particular task and generate reports from simulation.

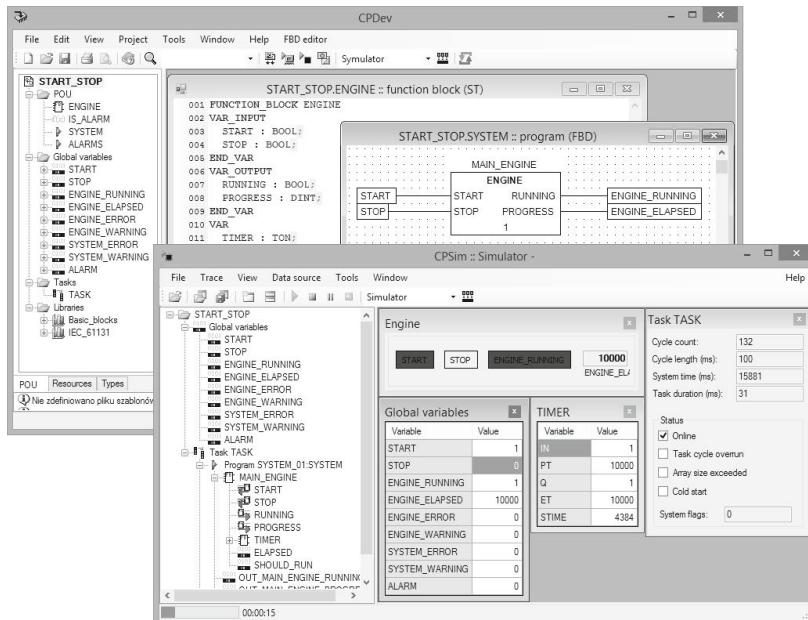


Fig. 3. Integrated development environment and simulator in the CPDev software

Apart from the simulator, the CPDev package has been equipped with the debugging capabilities in case of textual and graphical editors of IEC 61131-3 languages. The engineer can trace values of variables and place breakpoints.

5 Testing Framework

Testing should perform an important role in the development process, regardless of the software type. However, in case of control applications, testing is less systematic and less organized. As stated by Krapfenbauer et al. in [18], the control software is tested mainly manually. It is really surprising, because such a kind of software often performs crucial role while controlling processes and external devices. As observed by Hametner et al. [19], systematic, automation-supported, and agile testing approaches should be used while developing the control software. There are some approaches to solve the problem of insufficient testing, such as proposed by Hametner et al. in [20], where the Test-Driven Development process, dedicated to industrial automation, is proposed.

To create the complete solution for control projects development, the CPDev engineering environment has been equipped with the dedicated testing framework [3]. It allows to develop and execute unit and performance tests. The testing features are accessible by the CPTest application (Fig. 4). A list of available tests for the project is presented on the left, while the main workspace (with windows representing particular tests or test runs) is shown on the right.

5.1 Unit Testing

The first group of testing techniques is related to checking small parts of the overall project to ensure that they work according to requirements. In case of

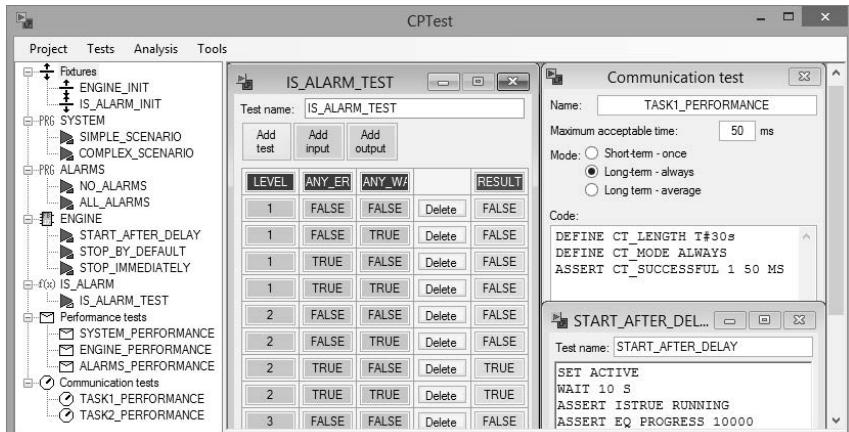


Fig. 4. Testing application in the CPDev environment

the IEC 61131-3 standard, such testing can be provided to a single POU. It is especially beneficial in case of function blocks and functions, because these elements can be used in many parts of the projects or even exported into external library and reused in other solutions. By providing precise testing of these parts, a quality of associated projects can be also improved.

The CPDev engineering environment supports two ways of unit testing – either using tables or a dedicated test definition language. Both approaches are briefly presented below, while their descriptions and examples are shown in [3].

The first method of unit testing is dedicated to simple scenarios of checking whether a POU produces correct values of outputs for particular values of inputs. Such a kind of testing does not require to write any code by engineers, because they just need to fill the table. While execution, the testing framework forces values of inputs and checks whether the actual output values are equal to the expected ones. The exemplary table test is shown in Fig. 4 (**IS_ALARM_TEST**).

In case of more advanced scenarios, tests can be created using the CPTTest+ dedicated test definition language. This approach has some features common with keyword-driven methods [19]. CPTTest+ supports a few instructions that performs operations useful while executing tests, such as setting or resetting a value of logical variable (**SET**, **RESET**), setting a value of variable of any kind (**ASSIGN**), saving a particular information to the test run log (**LOG**), holding the test execution for a given period of time (**WAIT**), as well as checking whether a particular condition is met (**ASSERT**). The last instruction is the most complex and supports various operators, such as checking equality or inequality. While execution, the system interprets instructions sequentially. If all assertions are met, the test is passed. The exemplary test created in the CPTTest+ language is named **START_AFTER_DELAY** and is shown in the bottom-right corner of Fig. 4.

5.2 Performance Testing

Apart from checking whether a POU produces correct values of outputs in the following cycles, it is important to verify some performance requirements. In case of the CPDev engineering environment, two kinds of such tests are available – dedicated to POUs and to communication tasks.

The first kind of performance tests is proposed to ensure that execution of a particular POU does not exceed the expected maximum time. The test can ensure that this condition is always met or only in case of the average execution time. The solution supports performing tests on various target platforms, as well as promotes taking the performance measurement results into account also during the initial stages of development.

Another way of testing performance requirements is dedicated to master-slave communication in DCSs, for instance between a controller and I/O modules. Untested performance requirements for communication tasks can cause problems by performing operations on not up-to-date values. The concept of Communication Performance Tests is proposed in [21]. It allows to check whether communication transactions are finished (with or without checking correctness) within a given period of time.

6 Multiplatform Visualization Mechanism

Apart from creating implementation of control application, it is often necessary to allow its operators to observe a state of the control process, as well as to set operation data. To accomplish these tasks, control systems may be equipped with HMI (Human-Machine Interface) graphical panels that present important information, as well as allow to entry data or adjust settings. A role of such devices becomes more and more important, for instance due to increased possibilities. Nowadays, HMI panels support multicolor touchable screens with high resolution that makes it possible to present even complex graphics. What is more, graphical panels can be integrated with controllers.

To support creating visualization of control programs, the CPDev engineering environment has been equipped with the CPVis multiplatform visualization mechanism [22]. It is a solution that allows to prepare visualization displays using the dedicated graphics editor and then to present them on real devices equipped with displays. The CPVis software supports multiple platforms, what means that prepared visualization displays can be presented on all supported devices, including ARM-based panels and PCs.

The CPVis solution is oriented towards components that compose visualization displays. Each component can be adjusted by setting values of its properties, either to fixed values or ones obtained from global variables of a control program. The latter allows to present not only static screens, but also animated objects. What is more, the project supports programming graphics using the IEC 61131-3 languages, both textual and graphical. Thus, the engineer can use exactly the same languages and tools to program the control software and the associated visualization panel. The last assumption, taken while designing and developing CPVis, is related to supporting devices with limited resources. For this reason, the visualization displays can be presented on the target device in two modes – optimized by used memory or by a number of CPU operations.

7 Applications

During CPDev development, the authors cooperated with a few companies and received useful feedback that indicated a direction of further work. Due to the successful cooperation, the CPDev engineering environment has a few applications, both industrial and laboratory.

CPDev is used in the Mega-Guard ship control and monitoring system developed by Praxis Automation Technology B.V. company (the Netherlands)². The system consists of several controllers and HMI panels exchanging data via Ethernet. Another industrial application involves deployment for LUMEL S.A. company (Poland)³ for programming their controllers, including SMC that acts as a central point in small DCSs. In these systems, data between SMC and external

² <http://www.praxis-automation.nl/>

³ <http://www.lumel.com.pl/en/>

I/O modules are exchanged via a communication protocol, such as Modbus. Recently, a cooperation with Nauka i Technika Sp. z o.o. [Ltd.] company (Poland)⁴ has been started. The firm uses CPDev for programming StTr-760-PLC controller that is a part of systems for pumping stations and transportation.

Apart from industrial applications, CPDev has some laboratory and research scenarios. One of such applications is the fast FPGA controller that acts as the hardware version of the CPDev virtual machine. Another solution is named softPLC and makes it possible to use PC as a controller. Such a project supports external I/O modules from National Instruments⁵ and Inteco⁶ companies.

8 Conclusion

The CPDev project was initiated at the end of 2006. At the beginning, it contained only a simple ST editor and compiler. In a few following years, the project has been significantly expanded. Currently, it is a comprehensive engineering environment supporting various stages of the project development.

The engineers can model the project using SysML diagrams and automatically generate an implementation template. The system supports all IEC 61131-3 languages, namely ST, IL, FBD, LD, and SFC. The integrated development environment contains graphics editors with advanced mechanisms, for instance to automatically find connections between elements. After adjusting the implementation, the engineer may simulate the control program, debug it, as well as create various kinds of tests. The CPTest framework supports unit testing (with textual and table tests) and performance testing (for checking POU and communication requirements). The package is also equipped with the CPVis multiplatform visualization mechanism to prepare visualization displays and present them on target devices. The environment supports configuration of hardware resources and assists the engineer with commissioning on the target platform.

Despite the fact that the CPDev environment seems to be a complete solution for programming PLCs, PACs, and DCSs, the authors plan to continue work on this software. One of the closest plans is related to preparing an updated version of the ST compiler with better support for arrays and structured data types.

References

1. Kormann, B., Vogel-Heuser, B.: Automated test case generation approach for PLC control software exception handling using fault injection. In: 37th Annual Conference on IEEE Industrial Electronics Society, IECON 2011, pp. 365–372 (2011)
2. Jamro, M., Trybus, B.: An approach to SysML modeling of IEC 61131-3 control software. In: 2013 18th International Conference on Methods and Models in Automation and Robotics (MMAR), pp. 217–222 (2013)
3. Jamro, M., Trybus, B.: Testing Procedure for IEC 61131-3 Control Software. In: 12th IFAC/IEEE International Conference on Programmable Devices and Embedded Systems (PDeS), pp. 192–197 (2013)

⁴ <http://www.nit.pl/en/>

⁵ <http://www.ni.com/>

⁶ <http://www.inteco.com.pl/>

4. IEC: IEC 61131-3 - Programmable controllers - Part 3: Programming languages (2003)
5. Okabe, M.: Development of processor directly executing IEC 61131-3 language. In: SICE Annual Conference, pp. 2215–2218 (2008)
6. Ferreira, E., Paulo, R., Cruz, D.D., Henriques, P.: Integration of the ST language in a model-based engineering environment for control systems: An approach for compiler implementation. Computer Science and Information Systems 5(2), 87–101 (2008)
7. Tisserant, E., Bessard, L., de Sousa, M.: An Open Source IEC 61131-3 Integrated Development Environment. In: 2007 5th IEEE International Conference on Industrial Informatics, vol. 1, pp. 183–187 (2007)
8. Chunjie, Z., Hui, C.: Development of a PLC Virtual Machine Orienting IEC 61131-3 Standard. In: International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2009, vol. 3, pp. 374–379 (2009)
9. Rzonca, D., Sadolewski, J., Stec, A., Swider, Z., Trybus, B., Trybus, L.: Open environment for programming small controllers according to IEC 61131-3 standard 10(3) (2009)
10. Trybus, B.: Development and Implementation of IEC 61131-3 Virtual Machine. Theoretical and Applied Informatics 23(1), 21–35 (2011)
11. Colla, M., Leidi, T., Semo, M.: Design and implementation of industrial automation control systems: A survey. In: 7th IEEE International Conference on Industrial Informatics, INDIN 2009, pp. 570–575 (2009)
12. Thramboulidis, K., Frey, G.: An MDD process for IEC 61131-based industrial automation systems. In: 2011 IEEE 16th Conference on Emerging Technologies Factory Automation (ETFA), pp. 1–8 (2011)
13. Chiron, F., Kouiss, K.: Design of IEC 61131-3 function blocks using SysML. In: Mediterranean Conference on Control Automation, MED 2007, pp. 1–5 (2007)
14. OMG: Unified Modeling Language (OMG UML), Infrastructure, V2.4.1 (2011)
15. OMG: Systems Modeling Language (OMG SysML), V1.3 (2012)
16. Jamro, M.: Graphics editors in CPDev environment. Journal of Theoretical and Applied Computer Science 6(1), 13–24 (2012)
17. Jamro, M., Rzonca, D.: Automatic connections in IEC 61131-3 Function Block Diagrams. In: 2013 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 463–469 (2013)
18. Krapfenbauer, H., Ertl, D., Zoitl, A., Kupzog, F.: Improving Component Testing of Industrial Automation Software. In: Fourth International Multi-Conference on Computing in the Global Information Technology, ICCGI 2009, pp. 259–262 (2009)
19. Hametner, R., Winkler, D., Zoitl, A.: Agile testing concepts based on keyword-driven testing for industrial automation systems. In: IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, pp. 3727–3732 (2012)
20. Hametner, R., Winkler, D., Ostreicher, T., Biffl, S., Zoitl, A.: The adaptation of test-driven software processes to industrial automation engineering. In: 2010 8th IEEE International Conference on Industrial Informatics (INDIN), pp. 921–927 (2010)
21. Jamro, M., Rzonca, D., Trybus, B.: Communication Performance Tests in Distributed Control Systems. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) CN 2013. CCIS, vol. 370, pp. 200–209. Springer, Heidelberg (2013)
22. Jamro, M., Trybus, B.: IEC 61131-3 Programmable Human Machine Interfaces for Control Devices. In: 2013 6th International Conference on Human System Interactions (HSI), pp. 48–55 (2013)

Development and Execution of POU-Oriented Performance Tests for IEC 61131-3 Control Software

Marcin Jamro

Rzeszow University of Technology,
al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
mjamro@kia.prz.edu.pl

Abstract. Due to performing complex and crucial tasks in industry, the control software should be created in a way ensuring possibly the highest quality. It can be increased by modeling, standardized implementation, and precise testing. The latter should verify not only functional requirements, but also nonfunctional, such as performance. In the paper, a concept of agile POU-oriented performance testing is proposed, which is dedicated to the IEC 61131-3 standard. The approach allows to measure and analyze execution times of particular Program Organization Units. The concept supports multiple target platforms and allows to take performance into account during early development stages, as well as to perform regression testing. The proposed process consists of a few stages, namely modeling (using SysML diagrams), implementation (in a dedicated test definition language), and execution (by a dedicated testing framework). The approach has been introduced into the CPDev engineering environment for programming industrial controllers.

Keywords: control software, IEC 61131-3, performance, testing.

1 Introduction

Nowadays, control systems often perform crucial and complex tasks in industry. Their improper work may cause serious negative consequences, such as impact on operator's health or damages of machines. What is more, complexity and size of such systems are still increasing, what is caused by many factors, including moving some hardware implementations into software [1]. For these reasons, it is important to provide engineers with a suitable methodology of modeling, implementation, and testing of control software. Regarding testing, some new methods are still necessary. They should be systematic, automation-supported, and agile, as stated by Hametner et al. [2]. What is more, testing should verify not only functional, but also nonfunctional requirements, such as performance, which is a main topic of this paper.

Chapman and Kirby [3] indicate that poor performance can make the system useless, similarly as in case of functional errors. Apart from it, performance problems may be even more expensive to solve. As described by Barna et al. [4],

performance testing is fundamental to estimate performance of various software components and the whole system. Perathoner et al. [5] indicate that analysis of system level performance is crucial while designing hard real-time embedded systems. For the reasons shown above, performance should be tested commonly while developing various kinds of systems. As explained by Seelig et al. [6], performance issues can be addressed in two ways, namely late or early in the overall development process.

In this paper, the author proposes an approach to performance testing, which is dedicated to solutions created according to the IEC 61131-3 standard [7]. Such a norm defines five languages for programming controllers. The languages can be divided into three groups, namely textual (Structured Text – ST, Instruction List – IL), graphical (Function Block Diagram – FBD, Ladder Diagram – LD), and mixed (Sequential Function Chart – SFC). Apart from languages, the norm specifies a software structure. Each control system is composed from Program Organization Units (POUs), namely programs, function blocks, and functions. Resources (such as controllers) have tasks assigned. Each task contains a list of programs that are executed with a given cycle time.

The proposed concept is orientated towards POU, what allows to measure and analyze execution times of a particular POU. A process of testing consists of a few stages, namely modeling (using SysML diagrams), implementation (in a dedicated test definition language), and execution (by a dedicated testing framework). The approach supports agile and iterative development process, as well as makes it easier to find performance problems introduced while creating the following versions of the project, as in the idea of regression testing.

To demonstrate the concept, the example is described in the following parts of this chapter. It is a system for monitoring performance of the engine, which calculates standard deviation and means (arithmetic, geometric, harmonic) of the last 1000 values of RPM, ignoring 5 lowest and 5 highest. Obtained results can be later shown on an HMI panel for visualization, used for further calculations, as well as used to start alarm in case of incorrect results. The control system project is composed from three POUs, namely SD for calculating standard deviation, MEANS for calculating means, and MONITORING for checking performance of the engine and preparing data for presentation on the HMI panel. The proposed concept is a black-box testing technique, thus implementation code of the system is neither presented nor explained in this chapter.

The features related to performance tests development and execution have been introduced into the CPDev engineering environment¹, which is developed in the Department of Computer and Control Engineering at Rzeszow University of Technology (Poland). The environment makes it possible to develop control software for PLCs (Programmable Logic Controllers), PACs (Programmable Automation Controllers), and DCSs (Distributed Control Systems), in all languages defined in the IEC 61131-3 standard, both textual and graphical [8]. Currently, CPDev is a complex environment that supports not only implementation of the control software, but also its modeling (with SysML diagrams) [9], testing (by

¹ <http://cpdev.kia.prz.edu.pl/>

table, unit, communication and POU-oriented performance tests) [10,11], visualization [12], as well as configuration, commissioning, and simulation. CPDev has a few applications, both industrial and laboratory. It is used by Praxis Automation Technology B.V. company (the Netherlands)² in systems for controlling and monitoring ships. Other industrial applications involve programming controllers used by LUMEL S.A. (Poland)³ and Nauka i Technika Sp. z o.o. (Poland)⁴. As for the laboratory usage, the CPDev package allows to create software for softPLC. The fast FPGA-based controller has been developed as well.

The paper is organized as follows. In the next section, related work is presented. In the third section, a concept of performance tests is described. The following stages of the associated process, i.e. modeling, implementation, and execution, are explained in three next parts.

2 Related Work

A subject of this paper is related to a few research areas, including modeling and various ways of testing. However, only topics related to performance testing are presented in this section. More information about modeling and testing can be found in [9] and [10], respectively. The proposed concept uses some assumptions and solutions similar to the approach of communication performance tests in distributed control systems, which are proposed by the author in [11].

Johnson et al. [13] introduce performance testing into a TDD approach. They call this technique as test-first performance and propose two sets of test cases, namely running for a short and long period of time. The first kind is executed together with unit tests, while the second is dedicated to run in a performance testing lab. Such a feature is similar to a mechanism of short-running and long-running performance tests, existing in the concept proposed in this paper.

Regarding other research, Barna et al. [4] present an adaptive framework for performance testing of transactional systems. Jin et al. [14] show a time performance analysis method for testing embedded systems, while Perathoner et al. [5] specify benchmarks for evaluating abstractions for performance analysis of distributed embedded systems.

Software performance engineering is another research problem. Its aim is to construct software systems to meet performance requirements [15]. This approach provides principles, patterns and antipatterns, as well as procedures and guidelines. It is a complex process involving many stages, such as modeling [16], gathering data, verification, and validation. This problem is analyzed by many researchers. As an example, Seelig et al. [6] show a methodology for performance engineering that is based on simulation and allows to estimate system performance early in the development process.

² <http://www.praxis-automation.nl/>

³ <http://www.lumel.com.pl/en/>

⁴ <http://www.nit.pl/en/>

3 Concept of POU-Oriented Performance Tests

An aim of the POU-oriented Performance Tests (PPTs) is to measure performance of a particular POU, what allows to find parts of the system that could have a negative impact on the overall performance (also named bottlenecks). Such tests can be used in an agile and iterative development process to monitor performance in the following iterations. That allows to perform regression testing to ensure that no performance problems are introduced after changing a related implementation part. What is more, the concept makes it possible to take performance into account just from starting development and to adjust the solution based on performance results. The approach introduces also a possibility of integration with the existing Model-Driven Development (MDD) process, as well as supports multiple target testing platforms, such as a simulator and an ARM-based controller.

The concept involves a methodology of PPT development and execution, as shown in Fig. 1. At the beginning, a developer models POU (together with their structure), tasks, and assignment of constraints to particular POU (see Sec. 4). A definition of POU, tasks, and constraints is platform-independent, while assignments are dependent on the target testing platform and are based on an expert's knowledge. After preparing the model, it is used to automatically generate implementation in a dedicated test definition language (see Sec. 5). Then, the generated code is executed by a testing framework in one out of two modes – short-running or long-running (see Sec. 6). The first is dedicated to simple performance tests that return results in a short period of time, while the others require more time to complete. Thus, the first type of performance tests should be run each time when unit tests for a particular POU are executed, while the other only while performing a complete test run.

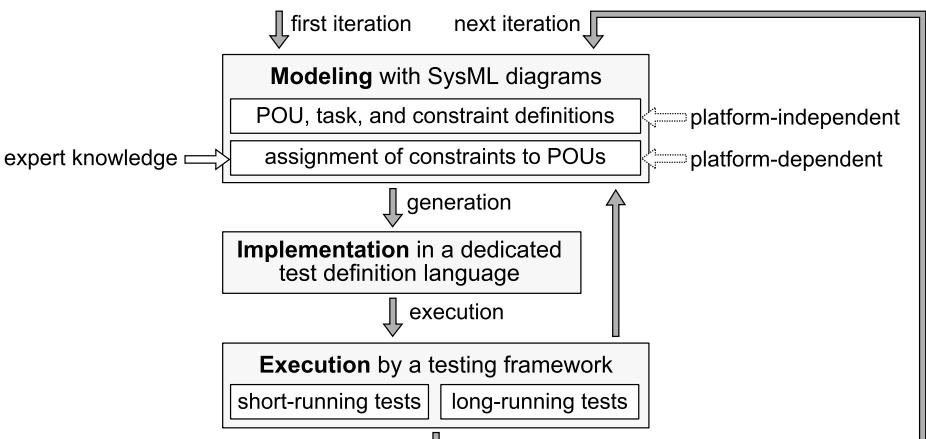


Fig. 1. Overall concept of POU-oriented Performance Tests development and execution

4 Modeling

The first stage of the process of PPTs development and execution is modeling. During this part, the developer prepares a set of diagrams presenting a structure of POUs and performance requirements. In this approach, Systems Modeling Language (SysML) [17] is chosen. It is based on Unified Modeling Language (UML) [18] and consists of nine diagram types to model requirements, structure, and behavior. Its proper use allows to move the development process into a higher level of abstraction, model a wide range of software and hardware components, as well as to generate implementation of the system or some its parts.

A structure of POUs is modeled using Block Definition Diagrams, as shown in Fig. 2 (at the top and bottom) and described in [9]. It allows to specify a name of the POU, its type, language, inputs, outputs (as flow ports), internal variables, as well as a set of tests (various kinds). In case of the running example, the SD function block is modeled, which will be implemented in the ST language and has one input (VALUE of REAL type), one output (RESULT of REAL), an array as internal variable (VALUES), as well as four unit tests (U indicator at the end) and two performance tests (P indicator) – one short-running (PS in the name) and one long-running (PL). Apart from the SD function block, the MONITORING program is modeled (Fig. 2, at the bottom). It will be implemented in the FBD language and has an instance of the SD function block as an internal variable. The program will be tested using two performance tests.

Apart from designing POUs, the developer should model tasks, assign programs to them, and specify proper cycle times. Regarding the running example, on the Package Diagram (Fig. 2, on the right) only one task is specified (TASK_MAIN). It will execute the MONITORING program every 100 ms.

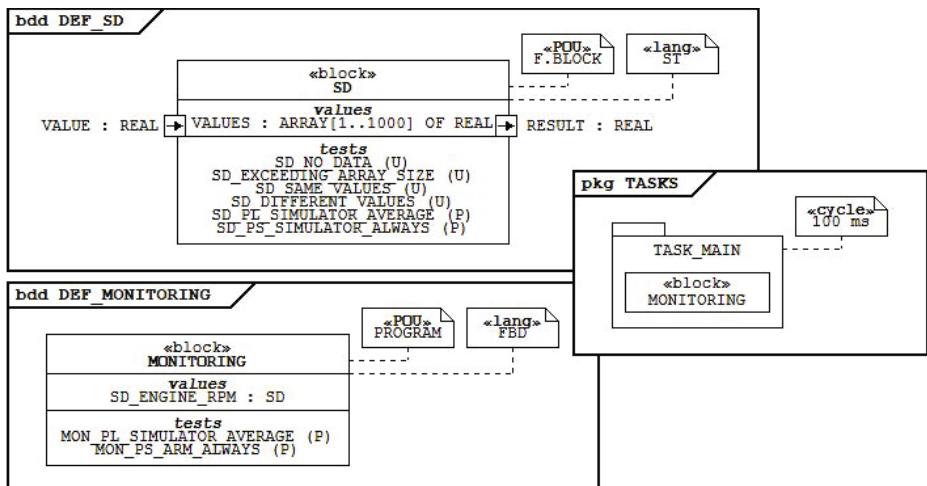


Fig. 2. Definition of POUs and a task using SysML diagrams

In the next stage of modeling, the developer assigns constraints to particular POUs. Such a task is accomplished using the predefined PT_PASSED constraint, which definition is shown in Fig. 3 (at the top). It specifies two conditions, namely ASSERT_AVERAGE and ASSERT_ALWAYS. The first verifies whether the actual average execution time (#V_AVG#) of a particular POU is smaller or equal to the expected time. The other checks whether the actual maximum execution time (#V_MAX#) is smaller or equal to the expected time. It is worth mentioning that each condition is taken into account only in case of a proper testing mode (explained below). Constraints can be adjusted using four parameters, namely POU (a POU that should be tested), EXPECTED_TIME (an expected average or maximum execution time), MODE (a mode of testing – AVERAGE or ALWAYS), as well as LENGTH (a period of time while the test should be running).

Assignment of constraints to particular POU is made on the Parameter Diagram by configuring constraint instances, which represent test cases. This process should be conducted by an expert, who has knowledge of performance requirements and possibilities in a given scenario. Fig. 3 (at the bottom) presents the diagram for the exemplary system, which defines two performance tests, namely SD_PL_SIMULATOR_AVERAGE and SD_PS_SIMULATOR_ALWAYS. Both are designed for the simulator as a target platform and check performance of the SD function block. The first test will be executed for 100 seconds, in the AVERAGE mode, with

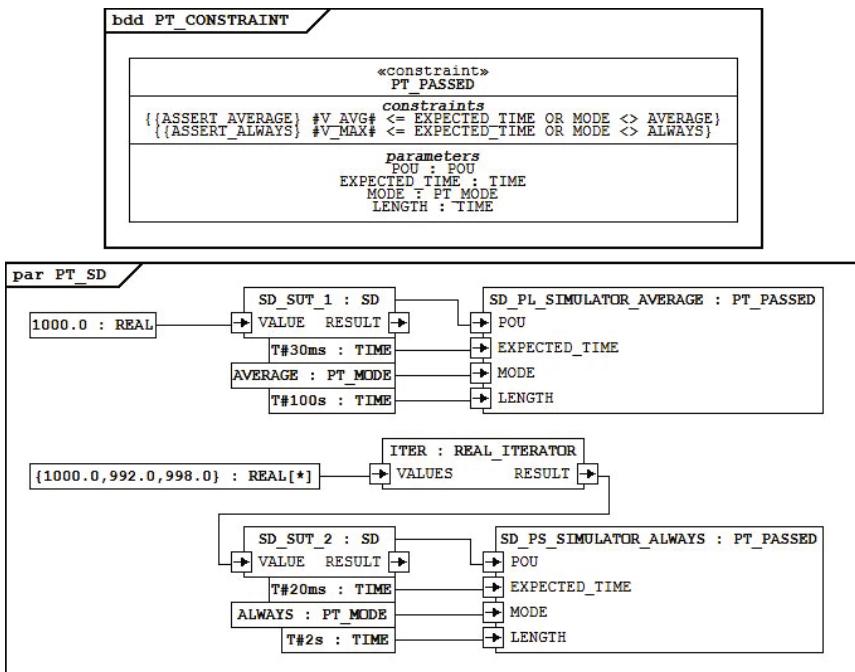


Fig. 3. Model of the PT_PASSED constraint and its assignment to the SD function block

an expected average time smaller or equal to 30 milliseconds. It uses 1000.0 as a value of the **VALUE** input of the **SD** function block. A slightly different scenario is presented in the second test, which is designed for the **ALWAYS** mode, an expected time smaller or equal to 20 milliseconds, and testing time period set to 2 seconds. However, the most visible difference is related to values of the input of the **SD** instance. In this example, an additional predefined block is used (**REAL_ITERATOR**) that returns the following items in a cyclical way. In this scenario, in consecutive cycles, a value of the **VALUE** input will be equal to 1000.0, 992.0, 998.0, 1000.0, 992.0, and so on.

5 Implementation

The implementation code for PPTs is automatically generated from SysML diagrams. It is prepared in a textual form in the CPTTest+ dedicated test definition language [10], which has been proposed by the author as a way of defining various kinds of test, including unit and performance ones. This approach has some similarities with a method of creating test cases using keywords from IEC 61131-3 and IEC 61499 standards, as shown by Hametner et al. [2].

The CPTTest+ language syntax consists of the basic part and a set of extensions related to particular kinds of testing. The basic part specifies a few instructions that represent operations performed often during manual testing, such as setting a value of logical variable to **TRUE** or **FALSE** (**SET** and **RESET**, respectively), as well as setting a value of any variable (**ASSIGN**). Other instructions allow to hold execution of the test for a given period of time (**WAIT**) and write some comments into the test run log (**LOG**). The last instruction is named **ASSERT** and checks whether a condition is met. If it is not, the test fails. Otherwise, the test continues till the next assertion or till its end. The **ASSERT** instruction can check various conditions, such as equality or a few types of inequality.

The basic instructions are not sufficient for describing PPTs configuration. For this reason, the language has been extended with some new functionalities, including the **DEFINE** instruction. It allows to specify a configuration setting, such as a mode, a name of the POU, a test length, an expected time, as well as values

a)

```
DEFINE PT_POU SD
DEFINE PT_INPUT[VALUE] 1000.0
DEFINE_PT_LENGTH T#100s
DEFINE PT_MODE AVERAGE
DEFINE PT_EXPECTED_TIME T#30ms
ASSERT PT_PASSED
```

b)

```
DEFINE PT_POU SD
DEFINE PT_INPUT[VALUE] 1000.0,992.0,998.0
DEFINE_PT_LENGTH T#2s
DEFINE PT_MODE ALWAYS
DEFINE PT_EXPECTED_TIME T#20ms
ASSERT PT_PASSED
```

Fig. 4. Exemplary code generated from the model for (a) **SD_PL_SIMULATOR_AVERAGE** and (b) **SD_PS_SIMULATOR_ALWAYS** performance test cases

of particular inputs that can be set either as fixed values or a list of values set in a cyclical way. Apart from adding a new instruction, the **ASSERT** one has been extended by a new operator (**PT_PASSED**). It is used to start, monitor, and stop execution of the PPT (configured earlier by **DEFINE**), as well as to check whether it meets performance requirements. A code generated for two performance tests from the running example is presented in Fig. 4.

6 Execution

When PPTs are designed and their code is automatically generated, the tests can be executed using the dedicated testing framework. A way of executing tests is presented in Alg. 1. It consists of three stages, namely arrangement of the system environment, execution, and presentation of results.

In the first stage, the testing framework prepares the environment where tests will be executed. This step depends on a type of the POU. In case of function blocks and functions, the system prepares temporary programs ("stubs") to test performance of a particular function block or function. Here, suitable values of inputs or global variables are set, based on data from the Parametric Diagram. Then, a testing task with the program to test (or its "stub" version) is prepared.

Next, the system proceeds to the execution stage. Here, the framework executes one cycle, calculates the actual test result, and presents real-time feedback to the developer. Such operations are repeated multiple times. In case of short-running tests, only a few iterations are conducted. However, to prepare a comprehensive solution for testing a particular module, many long-running tests

Algorithm 1. Execution process of the POU-oriented Performance Test

Require: *type* is a type of the POU (*program*, *block*, or *function*)
 {Arrangement of the system environment}

1. **if** *type* = *program* **then**
2. force values of global variables
3. **else if** *type* = *block* **then**
4. create a "stub" program with an instance of the function block
5. set values of inputs
6. **else if** *type* = *function* **then**
7. create a "stub" program with the function call
8. set values of inputs
9. **end if**
10. create and configure a testing task with the program (or the "stub" program)
 {Execution}
11. **repeat**
12. execute one cycle
13. present real-time feedback to the developer
14. **until** testing time is elapsed **or** test is stopped
 {Presentation of results}
15. present final results

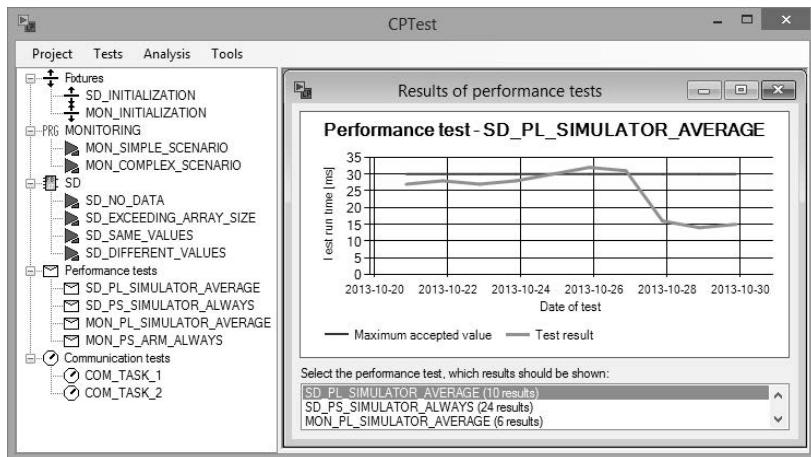


Fig. 5. CPTest testing platform with the performance analysis features

should be created as well. In such cases, thousands or even millions of iterations are executed to ensure that the system works correctly also in a longer period of time and do not cause some performance-related problems.

The last stage of execution involves presentation of the final results either as passed ("green") or failed ("red"). Additional information about the test run is available in the log. To make analysis of performance easier, results are also presented as charts, as shown in Fig. 5.

7 Conclusion

Nowadays, control systems perform many complex and important tasks in industry. To limit a chance of their incorrect work, various testing methods can be applied. However, apart from checking that the software works according to functional requirements, it is crucial to ensure that performance requirements are satisfied as well. This topic is very important in case of control software, because overlapping the cycle time could have serious impact on the whole system.

To make performance testing easier and more convenient, the author proposes the POU-oriented approach to testing performance. It has the agile characteristic, as well as allows to perform regression testing and to introduce measuring performance just from the beginning of developing the product. The proposed concept supports multiple target platforms and can be introduced into the overall Model-Driven Development approach for modeling various system components.

The proposed concept consists of a few stages, namely modeling (using SysML diagrams), implementation (automatic, in the dedicated test definition language),

as well as execution (by the testing framework). Together, they form a complete solution for testing performance of particular POU's in short-running and long-running modes. The concept has been introduced into the CPDev engineering environment for programming various kinds of industrial controllers.

References

1. Kormann, B., Vogel-Heuser, B.: Automated test case generation approach for PLC control software exception handling using fault injection. In: IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society, pp. 365–372 (2011)
2. Hametner, R., Winkler, D., Zoitl, A.: Agile testing concepts based on keyword-driven testing for industrial automation systems. In: IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, pp. 3727–3732 (2012)
3. Chapman, N., Kirby, L.: Performance engineering. *BT Technology Journal* 15(3), 19–25 (1997)
4. Barna, C., Litoiu, M., Ghanbari, H.: Model-based performance testing: NIER track. In: 2011 33rd International Conference on Software Engineering (ICSE), pp. 872–875 (2011)
5. Perathoner, S., Wandeler, E., Thiele, L., Hamann, A., Schliecker, S., Henia, R., Racu, R., Ernst, R., González Harbour, M.: Influence of different abstractions on the performance analysis of distributed hard real-time systems. *Design Automation for Embedded Systems* 13(1-2), 27–49 (2009)
6. Seelig, M., Schaffner, J., Decker, G.: Performance Engineering for Enterprise Applications. In: Castillo, O., Xu, L., Ao, S.I. (eds.) *Trends in Intelligent Systems and Computer Engineering. LNEE*, vol. 6, pp. 557–574. Springer US (2008)
7. IEC: IEC 61131-3 - Programmable controllers - Part 3: Programming languages (2013)
8. Jamro, M.: Graphics editors in CPDev environment. *Journal of Theoretical and Applied Computer Science* 6(1), 13–24 (2012)
9. Jamro, M., Trybus, B.: An approach to SysML modeling of IEC 61131-3 control software. In: 2013 18th International Conference on Methods and Models in Automation and Robotics (MMAR), pp. 217–222 (2013)
10. Jamro, M., Trybus, B.: Testing Procedure for IEC 61131-3 Control Software. In: 12th IFAC/IEEE International Conference on Programmable Devices and Embedded Systems (PDeS), pp. 192–197 (2013)
11. Jamro, M., Rzońca, D., Trybus, B.: Communication Performance Tests in Distributed Control Systems. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) *CN 2013. CCIS*, vol. 370, pp. 200–209. Springer, Heidelberg (2013)
12. Jamro, M., Trybus, B.: IEC 61131-3 Programmable Human Machine Interfaces for Control Devices. In: 2013 6th International Conference on Human System Interactions (HSI), pp. 48–55 (2013)
13. Johnson, M., Maximilien, E., Ho, C.W., Williams, L.: Incorporating Performance Testing in Test-Driven Development. *IEEE Software* 24(3), 67–73 (2007)

14. Jin, H., Chen, L.Y., Zeng, L.M., Li, B.L.: Performance Testing Based on Time Complexity Analysis for Embedded Software. In: International Conference on Embedded Software and Systems, ICESS 2008, pp. 243–247 (2008)
15. Smith, C.U.: Introduction to Software Performance Engineering: Origins and Outstanding Problems. In: Bernardo, M., Hillston, J. (eds.) SFM 2007. LNCS, vol. 4486, pp. 395–428. Springer, Heidelberg (2007)
16. Singleton, P.: Performance Modelling - What, Why, When and How. BT Technology Journal 20(3), 133–143 (2002)
17. OMG: Systems Modeling Language (OMG SysML), V1.3 (2012)
18. OMG: Unified Modeling Language (OMG UML), Infrastructure, V2.4.1 (2011)

A New Formulation and Solution of the Minimum Energy Control Problem of Positive 2D Continuous-Discrete Linear Systems

Tadeusz Kaczorek

Bialystok University of Technology, Faculty of Electrical Engineering
Wiejska 45D, 15-351 Bialystok, Poland
kaczorek@isep.pw.edu.pl

Abstract. A new formulation of the minimum energy control problem for the positive 2D continuous-discrete linear systems is proposed. Necessary and sufficient conditions for the reachability of the systems are established. Conditions for the existence of the solution to the minimum energy control problem and procedures for computation of an input minimizing the given performance index are given. Effectiveness of the procedure is demonstrated on numerical example.

Keywords: 2D continuous-discrete, linear, positive system, reachability, minimum energy control.

1 Introduction

A dynamical system is called positive if its trajectory starting from any nonnegative initial state remains forever in the positive orthant for all nonnegative inputs. An overview of state of the art in positive theory is given in the monographs [5, 10]. Variety of models having positive behavior can be found in engineering, economics, social sciences, biology and medicine, etc.

The positive 2D continuous-discrete linear systems have been introduced in [14], positive hybrid linear systems in [11] and the positive fractional 2D hybrid systems in [13]. Different methods of solvability of 2D hybrid linear systems have been discussed in [25] and the solution to singular 2D hybrids linear systems has been derived in [27]. The realization problem for positive 2D hybrid systems has been addressed in [15]. Some problems of dynamics and control of 2D hybrid systems have been considered in [4, 6]. The problems of stability and robust stability of 2D continuous-discrete linear systems have been investigated in [1–3, 9, 26, 28, 29] and of positive fractional 2D continuous-discrete linear systems in [12]. Recently the stability and robust stability of general model and of Roesser type model of scalar continuous-discrete linear systems have been analyzed by Busłowicz in [2, 3]. Stability of continuous-discrete 2D linear systems has been considered in [17]. The minimum energy control problem for standard linear systems has been formulated and solved by Klamka [21–24] and for 2D linear systems with variable coefficients in

[20]. The controllability and minimum energy control problem of linear systems with distributed delays has been investigated by Klamka in [24]. The minimum energy control of fractional positive continuous-time linear systems has been addressed in [7] and for descriptor positive discrete-time linear systems in [8].

In this paper a new formulation and solution to the minimum energy control problem for positive 2D continuous-discrete linear systems will be presented.

The paper is organized as follows. In section 2 necessary and sufficient conditions for the positivity of 2D continuous-discrete linear systems are established. The reachability and the problem formulation are given in section 3. Problem solution and a procedure for solving the minimum energy control problem are given in section 4. Concluding remarks are given in section 5.

The following notation will be used: \mathfrak{R} – the set of real numbers, $\mathfrak{R}^{n \times m}$ – the set of $n \times m$ real matrices, $\mathfrak{R}_+^{n \times m}$ – the set of $n \times m$ matrices with nonnegative entries and $\mathfrak{R}_+^n = \mathfrak{R}_+^{n \times 1}$, M_n – the set of $n \times n$ Metzler matrices (real matrices with nonnegative off-diagonal entries), I_n – the $n \times n$ identity matrix.

2 Positivity of 2D Continuous-Discrete Systems

Consider the 2D continuous-discrete linear system

$$\dot{x}(t, i) = Ax(t, i) + Bu(t, i) \quad (1)$$

where $\dot{x}(t, i) = \frac{\partial x(t, i)}{\partial t}$, $x(t, i) \in \mathfrak{R}^n$, $u(t, i) \in \mathfrak{R}^m$ are the state and input vectors and

$A \in \mathfrak{R}^{n \times n}$, $B \in \mathfrak{R}^{n \times m}$ ($n \geq m$) and $t \in \mathfrak{R}_+$ is continuous variable (usually time) and $i \in \mathbb{Z}_+$ is discrete variable.

Definition 1. The system (1) is called (internally) positive if $x(t, i) \in \mathfrak{R}_+^n$, $t \in \mathfrak{R}_+$, $i \in \mathbb{Z}_+$ for any boundary conditions $x_{0i} \in \mathfrak{R}_+^n$, $x_{t0} \in \mathfrak{R}_+^n$, $\dot{x}_{t0} \in \mathfrak{R}_+^n$ and all inputs $u(t, i) \in \mathfrak{R}_+^m$, $t \in \mathfrak{R}_+$, $i \in \mathbb{Z}_+$.

Theorem 1. The system (1) is positive if and only if

$$A \in M_n \text{ and } B \in \mathfrak{R}_+^{n \times m}. \quad (2)$$

Proof. Necessity. Let $u(t, i) = 0$, $t \geq 0$ and $x(0, i) = e_i$ (i -th ($i = 1, \dots, n$) column of the identity matrix I_n). The trajectory does not live the orthant \mathfrak{R}_+^n only if the derivative $\dot{x}(0, i) = Ae_i \geq 0$, what implies $a_{ij} \geq 0$, $i \neq j$. Therefore, the matrix A has to be the Metzler matrix. For the same reasons for $x(0, i) = 0$ we have $\dot{x}(0, i) = Bu(0, i) \geq 0$, what implies $B \in \mathfrak{R}_+^{n \times m}$, since $u(0, i) \in \mathfrak{R}_+^m$ may be arbitrary for $i \in \mathbb{Z}_+$.

Sufficiency. The solution of the equation (1) is given by

$$x(t,i) = e^{At}x(0,i) + \int_0^t e^{A(t-\tau)}Bu(\tau,i)d\tau. \quad (3)$$

It is well-known [10, 16] that $e^{At} \in \Re_+^{n \times n}$, $t \in \Re_+$ if and only if $A \in M_n$. From (3) it follows that if the conditions (2) are met and $x(0,i) \in \Re_+^n$, $u(t,i) \in \Re_+^m$, $t \in \Re_+$, $i \in Z_+$ then $x(t,i) \in \Re_+^n$, $t \in \Re_+$, $i \in Z_+$. Hence by Definition 1 the system (1) is positive. ■

3 Reachability and Problem Formulation

Consider the 2D continuous-discrete linear system (1).

Definition 2. The positive system (1) is called reachable in segment of line $\{[t_f, 0], [t_f, q]\}$ if for any given final state vector $x_f \in \Re^n$ there exists an input $u(t,i) \in \Re_+^m$, $0 \leq t \leq t_f$, $0 \leq i \leq q$ that steers the state vector $x(t,i)$ of the system from $x(0,i) = 0$, $i = 0, 1, \dots, q$ to $x_f = x(t_f, 0) + x(t_f, 1) + \dots + x(t_f, q)$.

Theorem 2. The positive system (1) is reachable on the segment of line $\{[t_f, 0], [t_f, q]\}$ if and only if the matrix $A \in M_n$ is diagonal and the matrix $B \in \Re_+^{n \times n}$ is monomial.

Proof. Using (3) for $t = t_f$, $i = 0, 1, \dots, q$ and $x(0,i) = 0$, $i = 0, 1, \dots, q$ we obtain

$$x_f = x(t_f, 0) + x(t_f, 1) + \dots + x(t_f, q) = \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} \bar{u}(\tau) d\tau \quad (4a)$$

where

$$\begin{aligned} \bar{B} &= [B \quad B \quad \dots \quad B] \in \Re_+^{n \times \bar{m}}, \\ \bar{u}(\tau) &= \begin{bmatrix} u(\tau, 0) \\ u(\tau, 1) \\ \vdots \\ u(\tau, q) \end{bmatrix} \in \Re_+^{\bar{m}}, \quad \bar{m} = n(q+1). \end{aligned} \quad (4b)$$

It is well-known [10, 16] that if $A \in M_n$ is diagonal then $e^{At} \in \Re_+^{n \times n}$ is also diagonal and if $B \in \Re_+^{n \times n}$ is monomial then $\bar{B}\bar{B}^T \in \Re_+^{n \times n}$ is also monomial. In this case the matrix

$$R(t_f, q) = \int_0^{t_f} e^{A\tau} \bar{B} \bar{B}^T e^{A^T \tau} d\tau \in \Re^{n \times n}_+ \quad (5)$$

is also monomial and $R^{-1}(t_f, q) \in \Re^{n \times n}_+$. The input

$$\hat{u}(t) = \bar{B}^T e^{A^T(t_f - t)} R^{-1}(t_f, q) x_f \quad (6)$$

steers the state of the system (1) from $x(0, i) = 0$, $i = 0, 1, \dots, q$ to the segment of line $\{[t_f, 0], [t_f, q]\}$.

Using (4) and (3) we obtain

$$\begin{aligned} x(t_f, q) &= \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} \hat{u}(\tau) d\tau = \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} \bar{B}^T e^{A^T(t_f - \tau)} d\tau R^{-1}(t_f, q) x_f \\ &= \int_0^{t_f} e^{A\tau} \bar{B} \bar{B}^T e^{A^T\tau} d\tau R^{-1}(t_f, q) x_f = x_f. \end{aligned} \quad (7)$$

Necessity. From the Cayley-Hamilton theorem we have

$$e^{At} = \sum_{k=0}^{n-1} c_k(t) A^k \quad (8)$$

where $c_k(t)$, $k = 0, 1, \dots, n - 1$ are some nonzero functions of time depending on the matrix A . Substitution of (8) into

$$\int_0^{t_f} e^{A(t_f - \tau)} \bar{B} \bar{u}(\tau) d\tau \quad (9)$$

yields

$$x_f = [\bar{B} \quad A\bar{B} \quad \dots \quad A^{n-1}\bar{B}] \begin{bmatrix} v_0(t_f) \\ v_1(t_f) \\ \vdots \\ v_{n-1}(t_f) \end{bmatrix} \quad (10)$$

where

$$v_k(t_f) = \int_0^{t_f} c_k(\tau) \bar{u}(t_f - \tau) d\tau, \quad k = 0, 1, \dots, n - 1. \quad (11)$$

For given $x_f \in \Re^n_+$ it is possible to compute nonnegative $v_k(t_f)$, $k = 0, 1, \dots, n - 1$ if and only if the matrix

$$[B \ AB \ \dots \ A^{n-1}B] \quad (12)$$

has n linearly independent monomial columns and this takes place only if the matrix $[A \ B]$ contains n linearly independent monomial columns [10, 16].

Note that for nonnegative $v_k(t_f)$, $k = 0, 1, \dots, n - 1$ it is possible to find a nonnegative input $\bar{u}(t) \in \mathfrak{R}_+^{\bar{m}}$ only if the matrix $B \in \mathfrak{R}_+^{n \times n}$ is monomial and the matrix $A \in M_n$ is diagonal. ■

If the positive system (1) is reachable on the segment of line $\{[t_f, 0], [t_f, q]\}$, then usually there exists many different inputs $\bar{u}(\tau) \in \mathfrak{R}_+^{\bar{m}}$ that steers the state of the system from $x(0, i) = 0$, $i = 0, 1, \dots, q$ to $x_f = x(t_f, 0) + x(t_f, 1) + \dots + x(t_f, q) \in \mathfrak{R}_+^n$.

Among these inputs we are looking for an input $\hat{u}(t) \in \mathfrak{R}_+^{\bar{m}}$ for $t \in [0, t_f]$ that minimizes the performance index

$$I(u) = \int_0^{t_f} u^T(\tau) Q u(\tau) d\tau \quad (13)$$

where $Q \in \mathfrak{R}_+^{\bar{m} \times \bar{m}}$ is a symmetric positive defined matrix and $Q^{-1} \in \mathfrak{R}_+^{\bar{m} \times \bar{m}}$.

The minimum energy control problem for can be stated as follows: Given the matrices $A \in M_n$, $B \in \mathfrak{R}_+^{n \times \bar{m}}$, $Q \in \mathfrak{R}_+^{\bar{m} \times \bar{m}}$ and $x_f \in \mathfrak{R}_+^n$, find an input $\bar{u}(t) \in \mathfrak{R}_+^{\bar{m}}$ for $t \in [0, t_f]$ that steers the state vector of the system from $x(0, i) = 0$, $i = 0, 1, \dots, q$ to $x_f = x(t_f, 0) + x(t_f, 1) + \dots + x(t_f, q) \in \mathfrak{R}_+^n$ and minimizes the performance index (13).

4 Problem Solution

To solve the problem we define the matrix

$$W = W(t_f, Q) = \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} Q^{-1} \bar{B}^T e^{A^T(t_f - \tau)} d\tau \quad (14)$$

where \bar{B} is defined by (4b).

By Theorem 2 the matrix (14) is monomial and $W^{-1} \in \mathfrak{R}_+^{n \times n}$ if and only if the positive system (1) is reachable at the segment of line $\{[t_f, 0], [t_f, q]\}$. In this case we may define the input

$$\hat{u}(t) = Q^{-1} \bar{B}^T e^{A^T(t_f - t)} W^{-1} x_f \text{ for } t \in [0, t_f]. \quad (15)$$

Note that $\hat{u}(t) \in \mathfrak{R}_+^{\bar{m}}$ for $t \in [0, t_f]$ if

$$Q^{-1} \in \Re_+^{\bar{m} \times \bar{m}} \text{ and } W^{-1}x_f \in \Re_+^{n \times n}. \quad (16)$$

Theorem 3. Let $\bar{u}(t) \in \Re_+^{\bar{m}}$ for $t \in [0, t_f]$ be an input that steers the state of the positive system (1) from $x(0, i) = 0, i = 0, 1, \dots, q$ to $x_f = x(t_f, 0) + x(t_f, 1) + \dots + x(t_f, q) \in \Re_+^n$. Then the input (15) also steers the state of the system from $x(0, i) = 0, i = 0, 1, \dots, q$ to $x_f = x(t_f, 0) + x(t_f, 1) + \dots + x(t_f, q) \in \Re_+^n$ and minimizes the performance index (13), i.e. $I(\hat{u}) \leq I(\bar{u})$. The minimal value of the performance index (13) is equal to

$$I(\hat{u}) = x_f^T W^{-1} x_f. \quad (17)$$

Proof. If the conditions (16) are met then $\hat{u}(t) \in \Re_+^{\bar{m}}$ for $t \in [0, t_f]$. We shall show that the input steers the state of the system from $x(0, i) = 0, i = 0, 1, \dots, q$ to $x_f \in \Re_+^n$. Substitution of (15) into (4a) for $t = t_f$ yields

$$x(t_f) = \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} \hat{u}(\tau) d\tau = \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} Q^{-1} \bar{B}^T e^{A^T(t_f - \tau)} d\tau W_f^{-1} x_f = x_f \quad (18)$$

since (14) holds. By assumption the inputs $\bar{u}(t)$ and $\hat{u}(t), t \in [0, t_f]$ steer the state of the system from $x(0, i) = 0, i = 0, 1, \dots, q$ to $x_f \in \Re_+^n$. Hence

$$x_f = \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} \bar{u}(\tau) d\tau = \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} \hat{u}(\tau) d\tau \quad (19a)$$

or

$$\int_0^{t_f} e^{A(t_f - \tau)} \bar{B} [\bar{u}(\tau) - \hat{u}(\tau)] d\tau = 0. \quad (19b)$$

By transposition of (19b) and postmultiplication by $W^{-1}x_f$ we obtain

$$\int_0^{t_f} [\bar{u}(\tau) - \hat{u}(\tau)]^T \bar{B}^T e^{A^T(t_f - \tau)} d\tau W^{-1} x_f = 0. \quad (20)$$

Substitution of (15) into (20) yields

$$\int_0^{t_f} [\bar{u}(\tau) - \hat{u}(\tau)]^T Q \hat{u}(\tau) = 0 \quad (21)$$

since

$$Q \hat{u}(\tau) = \bar{B}^T e^{A^T(t_f - \tau)} W^{-1} x_f = x_f. \quad (22)$$

Using (21) it is easy to verify that

$$\int_0^{t_f} \bar{u}(\tau)^T Q \bar{u}(\tau) d\tau = \int_0^{t_f} \hat{u}(\tau)^T Q \hat{u}(\tau) d\tau + \int_0^{t_f} [\bar{u}(\tau) - \hat{u}(\tau)]^T Q [\bar{u}(\tau) - \hat{u}(\tau)] d\tau. \quad (23)$$

From (23) it follows that $I(\hat{u}) < I(\bar{u})$ since the second term in the right-hand side of the inequality is nonnegative. To find the minimal value of the performance index (13) we substitute (15) into (13) and we obtain

$$I(\hat{u}) = \int_0^{t_f} \hat{u}^T(\tau) Q \hat{u}(\tau) d\tau = x_f^T W^{-1} \int_0^{t_f} e^{A(t_f - \tau)} \bar{B} Q^{-1} \bar{B}^T e^{A^T(t_f - \tau)} d\tau W^{-1} x_f = x_f^T W^{-1} x_f \quad (24)$$

since (14) holds. ■

Theorem 4. If the diagonal matrix Q is a scalar matrix

$$Q = \text{diag}[q_1, \dots, q_n] \in \Re_+^{n \times n} \quad (25)$$

then the input (15) is independent of Q and has the form

$$\hat{u}(t) = \bar{B}^T e^{A^T(t_f - t)} \left[\int_0^{t_f} e^{A\tau} \bar{B} \bar{B}^T e^{A^T\tau} d\tau \right]^{-1} x_f \in \Re_+^{n}, \quad t \in [0, t_f] \quad (26)$$

for any $x_f \in \Re_+^n$.

Proof. If (25) holds then from (14) we have

$$W = \frac{1}{q_1} \int_0^{t_f} e^{A\tau} \bar{B} \bar{B}^T e^{A^T\tau} d\tau \quad (27)$$

and from (15)

$$\hat{u}(\tau) = Q^{-1} \bar{B}^T e^{A^T(t_f - \tau)} W^{-1} x_f = \bar{B}^T e^{A^T(t_f - \tau)} \left[\int_0^{t_f} e^{A\tau} \bar{B} \bar{B}^T e^{A^T\tau} d\tau \right]^{-1} x_f \in \Re_+^{n} \quad (28)$$

for any $x_f \in \Re_+^n$. ■

From the above considerations we have the following procedure for solving the minimum energy control problem.

Procedure 1.

- Step 1. Using for example the Sylvester formula compute the matrix e^{At} .
- Step 2. Knowing the matrices A, B, Q, t_f and using (14) compute the matrix W .
- Step 3. Using (15) compute $\hat{u}(t)$.
- Step 4. Using (18) compute minimal value of the performance index $I(\hat{u})$.

Example 1. Consider the positive system (1) with matrices

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (29)$$

and the performance index (13) with the matrix

$$Q = \text{diag}[2, 2, 2, 2, 2, 2, 2, 2] \quad (30)$$

and $t_f = 1$, $q = 2$. Compute the optimal input $\hat{u}(t) \in \mathfrak{R}_+^9$ for $t \in [0,1]$ that steers the state of the system from $x(0,i) = 0$, $i = 0, 1, 2$ to $x_f = [1 \ 1 \ 1]^T \in \mathfrak{R}_+^3$ (T denote the transpose).

By Theorem 2 the positive system (1) is reachable at the segment of line $\{[1,0], [1,2]\}$ since $A \in M_3$ is diagonal and B is monomial. It is easy to check that the conditions (16) are met and the minimum energy control problem has a solution. Using the Procedure 1 we obtain the following:

Step 1. In this case we have

$$e^{At} = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{-t} \end{bmatrix}. \quad (31)$$

Step 2. Using (14), (31) we obtain

$$\begin{aligned} W &= \int_0^{t_f} e^{A(t_f-\tau)} \bar{B} Q^{-1} \bar{B}^T e^{A^T(t_f-\tau)} d\tau = \int_0^1 \begin{bmatrix} 1.5e^{-2\tau} & 0 & 0 \\ 0 & 1.5e^{4\tau} & 0 \\ 0 & 0 & 1.5e^{-2\tau} \end{bmatrix} d\tau \\ &= \begin{bmatrix} \frac{3}{4}(1-e^{-2}) & 0 & 0 \\ 0 & \frac{3}{8}(e^4-1) & 0 \\ 0 & 0 & \frac{3}{4}(1-e^{-2}) \end{bmatrix}. \end{aligned} \quad (32)$$

Step 3. Using (15) and (31) we obtain

$$\begin{aligned} \hat{u}(t) &= Q^{-1} \bar{B}^T e^{A^T(t_f - t)} W^{-1} x_f = [\hat{u}_1(t) \quad \hat{u}_2(t) \quad \hat{u}_3(t) \quad \hat{u}_4(t) \quad \hat{u}_5(t) \quad \hat{u}_6(t) \quad \hat{u}_7(t) \quad \hat{u}_8(t) \quad \hat{u}_9(t)]^T \\ &= \text{diag}[0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5] \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^{-(1-t)} & 0 & 0 \\ 0 & e^{2(1-t)} & 0 \\ 0 & 0 & e^{-(1-t)} \end{bmatrix} \\ &\times \begin{bmatrix} \frac{3}{4}(1-e^{-2}) & 0 & 0 \\ 0 & \frac{3}{8}(e^4-1) & 0 \\ 0 & 0 & \frac{3}{4}(1-e^{-2}) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned} \quad (33a)$$

where

$$\begin{aligned} \hat{u}_1(t) &= \frac{4e^{2(1-t)}}{3(e^4-1)}, \quad \hat{u}_2(t) = \frac{2e^{-(1-t)}}{3(1-e^{-2})}, \quad \hat{u}_3(t) = \frac{2e^{-(1-t)}}{3(1-e^{-2})}, \\ \hat{u}_4(t) &= \frac{4e^{2(1-t)}}{3(1-e^{-2})}, \quad \hat{u}_5(t) = \frac{2e^{-(1-t)}}{3(1-e^{-2})}, \quad \hat{u}_6(t) = \frac{2e^{-(1-t)}}{3(1-e^{-2})}, \\ \hat{u}_7(t) &= \frac{4e^{2(1-t)}}{3(1-e^{-2})}, \quad \hat{u}_8(t) = \frac{2e^{-(1-t)}}{3(1-e^{-2})}, \quad \hat{u}_9(t) = \frac{2e^{-(1-t)}}{3(1-e^{-2})}. \end{aligned} \quad (33b)$$

Step 4. Using (17) and (30) we obtain

$$\begin{aligned} I(\hat{u}) &= x_f^T W^{-1} x_f = [1 \quad 1 \quad 1] \begin{bmatrix} \frac{3}{4}(1-e^{-2}) & 0 & 0 \\ 0 & \frac{3}{8}(e^4-1) & 0 \\ 0 & 0 & \frac{3}{4}(1-e^{-2}) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \frac{8}{3(e^4-1)} + \frac{8}{3(1-e^{-2})}. \end{aligned} \quad (34)$$

Using the method presented in [18, 19] the considerations can be extended to 2D continuous-discrete linear systems with bounded inputs.

5 Concluding Remarks

A new formulation and solution to the minimum energy control problem have been proposed. New necessary and sufficient conditions for the positivity of a class of 2D continuous-discrete linear systems have been established (Theorem 1). A new notion of the reachability to the segment of line has been introduced and necessary and sufficient conditions for the reachability have been presented (Theorem 2). Conditions for the existence of the solution to the minimum energy control problem (Theorem 3) and procedures for computation of an input minimizing the given performance index have been proposed. Effectiveness of the procedure have been demonstrated on numerical example.

An open problem is an extension of the minimum energy control problem to standard and positive 2D continuous-discrete linear systems described by the equation [10, 16]

$$\dot{x}(t, i+1) = A_0 x(t, i) + A_1 \dot{x}(t, i) + A_2 x(t, i+1) + B_0 u(t, i) + B_1 \dot{u}(t, i) + B_2 u(t, i+1)$$

where $\dot{x}(t, i) = \frac{\partial x(t, i)}{\partial t}$, $x(t, i) \in \mathbb{R}^n$, $u(t, i) \in \mathbb{R}^m$ are the state and input vectors and $A_k \in \mathbb{R}^{n \times n}$, $B_k \in \mathbb{R}^{n \times m}$, $k = 0, 1, 2$.

Acknowledgment. This work was supported under work S/WE/1/11.

References

1. Bistritz, Y.: A stability test for continuous-discrete bivariate polynomials. In: Proc. Int. Symp. on Circuits and Systems, vol. 3, pp. 682–685 (2003)
2. Busłowicz, M.: Stability and robust stability conditions for a general model of scalar continuous-discrete linear systems. Measurement Automation and Monitoring 56(2), 133–135 (2010)
3. Busłowicz, M.: Robust stability of the new general 2D model of a class of continuous-discrete linear systems. Bull. Pol. Acad. Sci. Techn. 57(4) (2010)
4. Dymkov, M., Gaishun, I., Rogers, E., Gałkowski, K., Owens, D.H.: Control theory for a class of 2D continuous-discrete linear systems. Int. J. Control 77(9), 847–860 (2004)
5. Farina, L., Rinaldi, S.: Positive Linear Systems; Theory and Applications. J. Wiley, New York (2000)
6. Gałkowski, K., Rogers, E., Paszke, W., Owens, D.H.: Linear repetitive process control theory applied to a physical example. Int. J. Appl. Math. Comput. Sci. 13(1), 87–99 (2003)
7. Kaczorek, T.: Minimum energy control of fractional positive continuous-time linear systems. In: Proc. of Conf. MMAR Międzyzdroje, Poland, August 26–29 (2013)
8. Kaczorek, T.: Minimum energy control of descriptor positive discrete-time linear systems. Compel 3(4) (2013) (in press)
9. Kaczorek, T.: New stability tests of positive standard and fractional linear systems. Circuit and Systems 2(4), 261–268 (2011)
10. Kaczorek, T.: Positive 1D and 2D systems. Springer, London (2001)

11. Kaczorek, T.: Positive 2D hybrid linear systems. *Bull. Pol. Acad. Sci. Tech.* 55(4), 351–358 (2007)
12. Kaczorek, T.: Positive fractional 2D continuous-discrete linear systems. *Bull. Pol. Acad. Tech.* 59(4), 575–579 (2011)
13. Kaczorek, T.: Positive fractional 2D hybrid linear systems. *Bull. Pol. Acad. Tech.* 56(3), 273–277 (2008)
14. Kaczorek, T.: Reachability and minimum energy control of positive 2D continuous-discrete systems. *Bull. Pol. Acad. Sci. Tech.* 46(1), 85–93 (1998)
15. Kaczorek, T.: Realization problem for positive 2D hybrid systems. *Compel* 27(3), 613–623 (2008)
16. Kaczorek, T.: Selected Problems of Fractional Systems Theory. LNCIS, vol. 411. Springer, Heidelberg (2011)
17. Kaczorek, T.: Stability of continuous-discrete linear systems described by general model. *Bull. Pol. Acad. Sci. Tech.* 59(2), 189–193 (2011)
18. Kaczorek, T.: Minimum energy control of positive continuous-time linear systems with bounded inputs. *Int. J. Appl. Math. Comput. Sci.* (2013) (in press)
19. Kaczorek, T.: Necessary and sufficient conditions for the minimum energy control of positive discrete-time linear systems with bounded inputs. *Bull. Pol. Acad. Sci. Tech.* (2013) (in press)
20. Kaczorek, T., Klamka, J.: Minimum energy control of 2D linear systems with variable coefficients. *Int. J. of Control* 44(3), 645–650 (1986)
21. Klamka, J.: Controllability and minimum energy control problem of fractional discrete-time systems. In: Baleanu, D., Guvenc, Z.B., Tenreiro Machado, J.A. (eds.) *New Trends in Nanotechnology and Fractional Calculus*, pp. 503–509. Springer, New York (2010)
22. Klamka, J.: *Controllability of Dynamical Systems*. Kluwer Academic Press, Dordrecht (1991)
23. Klamka, J.: Minimum energy control of 2D systems in Hilbert spaces. *System Sciences* 9(1-2), 33–42 (1983)
24. Klamka, J.: Relative controllability and minimum energy control of linear systems with distributed delays in control. *IEEE Trans. Autom. Contr.* 21(4), 594–595 (1976)
25. Kaczorek, T., Marchenko, V., Sajewski, Ł.: Solvability of 2D hybrid linear systems - comparison of three different methods. *Acta Mechanica et Automatica* 2(2), 59–66 (2008)
26. Narendra, K.S., Shorten, R.: Hurwitz stability of Metzler matrices. *IEEE Trans. Autom. Contr.* 55(6), 1484–1487 (2010)
27. Sajewski, Ł.: Solution of 2D singular hybrid linear systems. *Kybernetes* 38(7/8), 1079–1092 (2009)
28. Xiao, Y.: Stability test for 2-D continuous-discrete systems. In: Proc. 40th IEEE Conf. on Decision and Control, vol. 4, pp. 3649–3654 (2001)
29. Xiao, Y.: Stability, controllability and observability of 2-D continuous-discrete systems. In: Proc. Int. Symp. on Circuits and Systems, vol. 4, pp. 468–471 (2003)

The Impact of an ERP System on the Technical Preparation of Production

Sławomir Kłos

University of Zielona Góra, Licealna 9, 65-417 Zielona Góra, Poland
S.Kłos@iizp.uz.zgora.pl

Abstract. The area of the technical preparation of production includes tasks related to the development of design, technology, tools and any equipment necessary to create new products. This functional area has key strategic importance for the development of almost every aspect of a production company. This is an integral area associated with the implementation of an ERP system, because data for materials, structures, devices, technologies, tools, etc., are created in the technical preparation of production. Based on an analysis of the functionality of ERP systems and associated case studies, a methodology can be proposed for assessing the impact of the functionality of an ERP system on the efficiency of the processes performed in the technical preparation of production area. The proposed methodology allows, already at the stage of the implementation of an ERP system, for the selection of features and an evaluation of the potential benefits in the area of the technical preparation of production. An analysis of the impact of the effectiveness of an ERP system on technical preparation of production activities is generally concerned with the processes involved in creating bills of materials, manufacturing technologies, cost calculation, variant operations and variant part lists, etc.

Keywords: ERP, technical preparation of production, process analyze.

1 Introduction

Implementing an ERP system is an expensive and time-consuming process for every manufacturing company [13]. The benefits of implementing an ERP system are chiefly concerned with reducing inventory levels, improving customer service and increasing the efficiency of business processes by relieving an organisation of unnecessary bureaucracy which can be automated by an ERP system. However, the implementation of an ERP system, in a different way, affects the execution of processes in different areas of the company. One of the key areas for the implementation of an ERP system, which generates most of the data concerning a new product is the area of the technical preparation of production (TPP). Figure 1 shows a simple model of the links between the operational areas of a manufacturing enterprise with regards to aspects of the generation and use of data. From the model, it is evident that all functional areas of the company benefit from the data generated in the area of the technical preparation of production. Functional areas related to

logistics, production and sales all use the data created by engineers (material indexes, bills of materials, operation times, technologies, etc.) to create sales offers, purchasing orders or production schedules. Therefore, the implementation of an ERP system makes the area of the TPP an essential source of large amounts of data, which can be used in different functional areas of a manufacturing enterprise. On the basis of this, the following question can be formulated: What is the impact of an ERP system on the efficiency of the processes carried out in the area of the technical preparation of production? The system requires engineers to register large amounts of data that are used by other functional areas. How can the ERP system automate processes realized in the area of the technical preparation of production, what routine activities can be eliminated and which critical decisions can be supported?

The implementation and operation of an ERP system in the area of the TPP requires additional efforts to be made by engineers. The majority of the results and registered data are used in other functional areas of an enterprise. Table 4 shows examples of the tasks carried out in the TPP and any functional areas using the results of these actions.

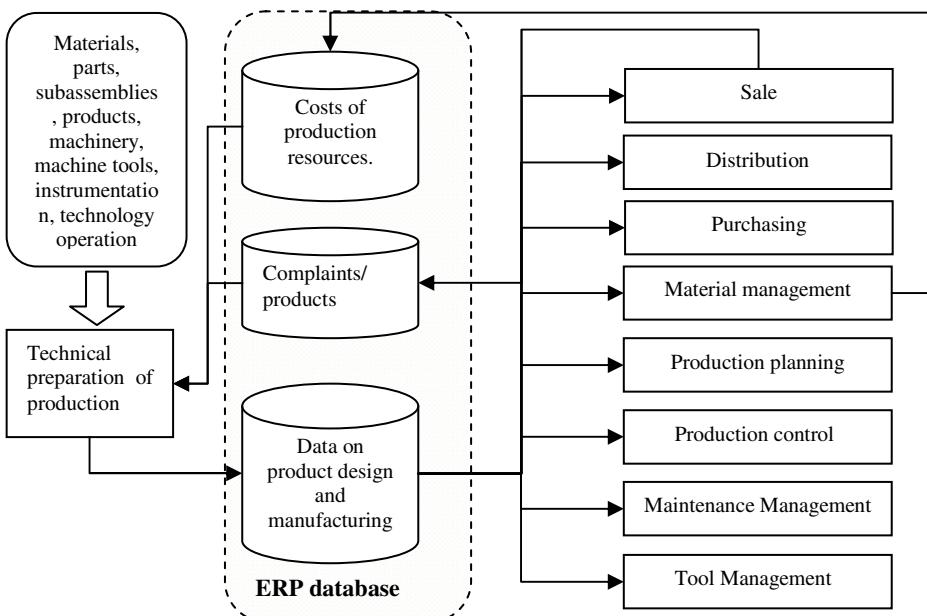


Fig. 1. A Model of data flow between the TPP and other functional areas of a manufacturing company

For the presented assumptions, the following research problem can be formulated: On the basis that there is a manufacturing company which includes the area of the TPP in its structure, we may ask: What are the quantitative and qualitative indicators that allow us to assess the impact of the implementation and operation of an ERP system on the processes carried out in the technical preparation of production?

The research problem has been formulated in very general terms. Because of the multiplicity and diversity of the kinds, forms and types of production that should be taken into account when considering the implementation and exploration of an ERP system in different companies, and the fact that it will have different effects on the efficiency of the processes carried out in the area of the technical preparation of production.

Table 1. The main ERP tasks in the area of the TPP and the end-use of the results

Tasks in TPP	Other areas that profit from the results of the work
Creation of databases of materials, parts, intermediates, subassemblies and finished products.	Sales, distribution, material management, purchasing, production planning/
Verification of operations times and bills of materials.	Production planning and control,
Building of product structures.	Material management, purchasing, production planning,
Definition of technology and production routes.	Production planning and control, maintenance.
Development of variant designs and technologies.	Sales, distribution, material management, purchasing, production planning.
Calculation of structural variants.	Sales, purchasing.
Creating a database of manufacturing resources.	Production planning and control, maintenance and tool management.
Creating tools and equipment databases.	Production planning and control, maintenance and tool management.

In companies carrying out mass make-to-stock (MTS) production for example in the food industry, the ERP system will have a completely different impact on pre-production than in a company which is carrying out engineer-to-order production, for example production lines for the heat treatment of metals. In the second case, the role of engineers regarding the use of an ERP in the area of the TPP would be much greater.

2 ERP Systems and the Technical Preparation of Production

Davenport et al. defined an ERP system as a packaged software application that connects and manages information flows within and across a complex organization, allowing managers to make decisions based on information that truly reflects the current state of their business [3]. The support of product development by ERP systems is the subject of much research and numerous scientific studies. Ou-Yang and Chang proposed concepts to bridge the gaps between PDM and ERP systems and to support the collaboration of activities between the two systems. To develop the agent-based collaboration system, a three-stage framework was proposed [11]. In the implementation stage, a PDM/ERP collaboration system was developed to support the designer in making decisions about the replacement parts requirement analysis. Wei et al. analyzed the characteristics of ERP and PDM integration and presented the technology of business-related remote function calls in order to better support ERP and PDM system integration for independent enterprise network manufacturing systems. The design & engineering stage is especially important for engineer-to-order and design-to-order companies[14]. Rudberg and Wikner provided a rare contribution

when they proposed a framework to forecast the lead time required for design and engineering activities using a database of historical activities and by considering the current workload. While valuable, discussion of the framework is limited; there is insufficient detail for others to apply the method in practice [12]. Bellatreche et al. [1] presented the ontology-based data integration modeling approach. It ensures the automation of the integration process when all sources reference a shared ontology, and possibly extends it by adding their concept specializations. Lee et. al. proposed digital manufacturing as a key tool of data integration between PDM and ERP systems [9]. They proposed digital manufacturing as the generation and verification method for the manufacturing of bills of materials (MBOM), data of primary importance that need to be managed in the manufacturing industry, and also presented the methodology of PDM and ERP system integration through digital manufacturing. Hicks and McGovern conducted a study on the potential functionality of PLM for engineer-to-order companies. They selected specific modules of PLM (e.g. design change control and capability maturity models) and pointed out that it can help ETO companies to manage the product life cycle process [4]. Olhager and Selldin carried out questionnaire surveys addressing 190 Swedish companies and concentrating on selection, pre-implementation audits, implementation, benefits and future prospects for developing ERP systems [10]. Similarly, Ehie and Madsen, based on a survey among 36 companies, analysed critical factors pertaining to the implementation of ERP systems [5]. Some of the factors included project management, feasibility studies, the development of human resources, the re-engineering of business processes, board support, project budget management, IT infrastructure, and support from consultants. Any analysis of ERP implementation effects should include the influence of corporate culture and the attitude of the management [6].

An analysis of the literature shows that ERP implementation in a manufacturing company is a very complex project that requires the full involvement of the employees. [2], [15]. However, the workload associated with the implementation of an ERP system is different for different departments of a manufacturing company and typically depends on the kind of production in question. The implementation and operation of an ERP system in the area of the TPP can play an extremely significant role on the variety of manufactured products, the number of variants of design and technology, and their levels of complexity. For the engineer-to-order business model, prototype production and the creation of BOMs should be monitored by a permanent cost analysis. Planned costs of prototypes (budget of a project) should be continuously compared to the costs of construction of the product. Ideally, engineers (designers and technologists) should be responsible for the comparison of the costs. The cost monitoring process is relatively simple to automate provided that the schema of the manufacturing cost calculation is implemented in the ERP system. In the case of make-to-order (MTO) and assembly-to-order production, product customization is of key importance (the introduction of material replacements into the structure of the product). For make-to-stock production (MTS), it is more important to have the possibility to use different technological variants (different technological routes of product manufacture). Table 2 specifies the importance of different data sets created in the area of the TPP for different kinds of production (MTS, MTO, ATO, ETO). The evaluation is made on the basis of an analysis of many case studies of the implementation of ERP systems in different branches of manufacturing companies [8]. The following evaluation scale is used: 1 – medium

importance, 2 – high importance, 3 – critical importance. The analysis of the evaluation presented in the table shows that a significant amount of critical data is generated in the area of the TPP for the MTS models of production. This is due to the need for the efficient management of tools and special equipment in large-scale production. Data which relate to material management is of primary importance. Details of the technology used are also more important for mass-scale, repetitive production models than for small batch or individual manufacturing models.

Table 2. The importance of data sets created in the technical preparation of production

Data sets	MTS	MTO	ATO	ETO	Data importance
raw materials, parts	3	3	3	3	12
intermediates, subassemblies	3	3	3	3	12
ready products	3	3	3	3	12
product structure	3	3	3	3	12
design drawings	1	1	1	2	5
services	2	2	2	2	8
structural variants, variant part list	1	2	2	1	6
calculation variants	1	1	2	2	6
technological operations	3	2	1	1	7
technological variants,	2	1	1	1	5
machines, production cells	3	2	1	1	7
tools	3	2	1	1	7
instruments, special devices	3	2	1	1	7
Importance according to kinds of production	31	27	24	24	

In the case of ETO models of production, production scheduling is often replaced by a specific work plan on the basis of project management methodology. Variants of product structures have greater importance for MTO and ATO type production than for ETO manufacturing, because of the possibility of the rapid configuration of products (rapid product customization).

Table 3. Labor-intensity of data registration in an ERP system regarding the technical preparation of production

Data sets	MT S	MTO	AT O	ETO	Effort according to data sets
raw materials, parts	1	1	1	2	5
intermediates, subassemblies	1	2	2	3	8
ready products	2	2	2	1	7
product structure	1	2	2	3	8
design drawings	1	2	2	2	7
services	1	1	1	1	4
structural variants, variant part list	1	2	2	1	6
calculation variants	1	2	2	3	8
technological operations	2	2	2	2	8
technological variants, technological routes	2	2	1	1	6
machines, production cells	1	1	1	1	4
tools	1	1	1	1	4
instruments, special devices	1	1	1	1	4
Effort according to kinds of production	16	21	20	22	

Table 3 presents a similar analysis, however, it is one which is related to the labor-intensity of data registration in an ERP system. The following evaluation scale used: 1 – medium labor-intensity, 2 – high labor-intensity, 3 – very high labor-intensity. An analysis of the results from table 2 shows that the most labor-intensive process is data registration carried out in the enterprises which use an ETO method of production.

The level of labor-intensity is determined by the number of new intermediates and the complexity of the product structure manufactured as single prototype production. Based on the tables, it is possible to identify the critical data sets that are important during the implementation and operation of an ERP system in the area of the technical preparation of production. The data encompassed in table 4 are calculated as a simple result of the multiplication of data from tables 1 and 2.

Table 4. The identification of critical data sets related to the implementation and operation of an ERP system in the technical preparation of production

Data sets	MTS	MTO	ATO	ETO
raw materials, parts	3	3	3	6
intermediates, subassemblies	3	6	6	9
ready products	6	6	6	3
product structure	3	6	6	9
design drawings	1	2	2	4
services	2	2	2	2
structural variants, variant part list	1	4	4	1
calculation variants	1	2	4	6
technological operations	6	4	2	2
technological variants, technological routes	4	2	1	1
machines, production cells	3	2	1	1
tools	3	2	1	1
instruments, special devices	3	2	1	1

An analysis of the data presented in table 4 shows that for ETO production models, a significant influence on ERP-system effectiveness is created by the complexity of the structures, the number of components and calculation variants. For example in the case of manufacturing companies that construct CNC machines or production lines on the basis of customer specifications, the number of new intermediates created by designers and the complexity of product structures help to decide about the effectiveness of the whole area of the technical preparation of production. In addition, engineers are additionally engaged by the sales area through the preparation of the technical specifications required in sales offers. A good solution for the problems of the excessive labor-intensity of engineers is the automation of the process of data registration. Automatic generation of product structures on the basis of CAD drawings can provide very good results. The integration of ERP and CAD systems can enable the automatic creation of bills of materials on the basis of the geometrical dimensions and physical features of the objects generated in the CAD systems.

3 Processes in the Area of the Technical Preparation of Production

The main processes carried out in the area of the TPP deal with the products of design and technology. Figure 2 shows the basic structure of the processes carried out in the

area of engineering activities supported by an ERP system. The processes carried out in the area of the TPP include five general subprocesses: product design, technological development, cost calculation, documentation and production support. In MTS and MTO production, the decision about the production of a new product is made, if the product design and technology is ready and the calculation of production costs is acceptable.

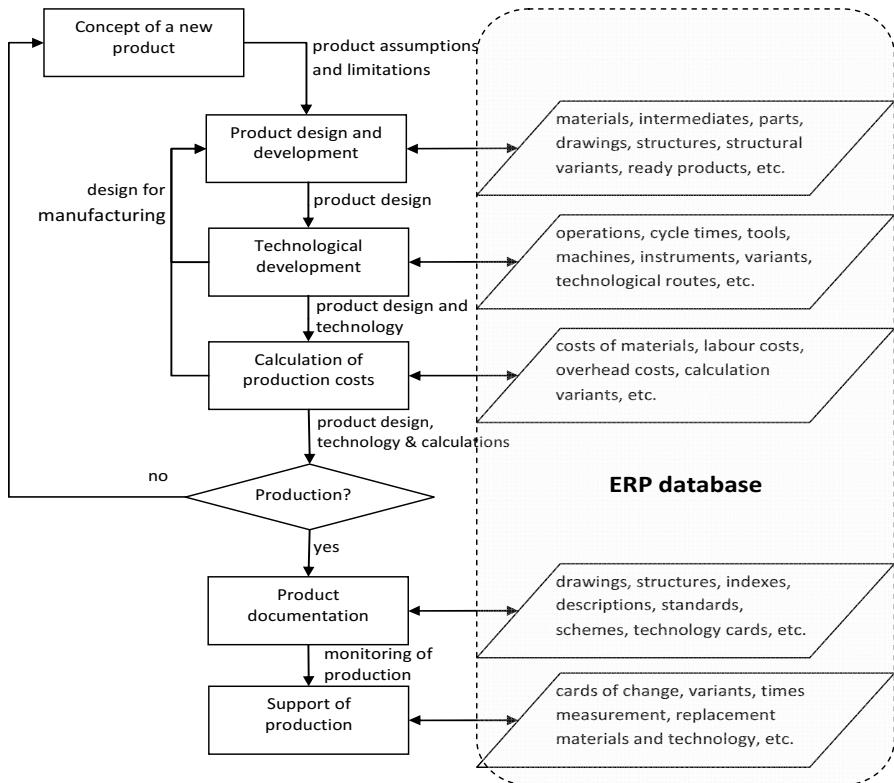


Fig. 2. Processes of new product development in the TPP area

The process of product and technological design in the case of ETO production or ATO montage is carried out according to customer specifications and the specification of product assumptions and limitations. The effectiveness of the presented process is influenced by the following factors: average time of construction registration in ERP – T_i^c , number of new material indexes – m_i , average time of material index definition – t_m , number of new parts, intermediates, subassemblies – p_i , average time of part index definition – t_p , number of construction variants – v_i^c , average time of a construction variant definition – t_{vc} , average time of technology registration in ERP – T_{ti} , number of new technological operations – o_i , average time of technological operation definition – t_o , number of new manufacturing resources – r_i , average time of manufacturing resource definition – t_r , number of new instruments

and tools – s_i , average time of instruments definition – t_s , number of alternative technological routes – v_i^t , average time of a technological variant definition – t_{vt} . The index denotes the number of a new product. The total time T_{TPP} of process realization in the area of the TPP (construction and technology data registration) can be calculated on the basis of formula 3.

$$T_D = \sum_{i=1}^N (m_i \cdot t_m + p_i \cdot t_p + v_i^c \cdot t_{vc}) \quad (1)$$

$$T_T = \sum_{i=1}^N (o_i \cdot t_o + r_i \cdot t_r + s_i \cdot t_s + v_i^t \cdot t_{vt}) \quad (2)$$

$$T_{TPP} = T_D + T_T \quad (3)$$

where N – the number of new products created in the area of the TPP in a period of time. The value of new product sales in a period of time can be calculated as follows:

$$V = \sum_{i=1}^N V_i \quad (4)$$

The productivity of the ERP system in the area of the TPP could be calculated as follows:

$$P_{TPP} = \frac{V}{T_{TPP}} \quad (5)$$

The reduction of time could be a result of the implementation of material replacements and a configurable finished product structure, but also it could be the result of a new functionality of the ERP system in the sales area. Theoretically, the best way to estimate the impact of an ERP system on the various functional areas of a manufacturing enterprise is by periodically switching off a functionality and time measurement. Such experiments can be very interesting because it makes it possible to find useless or inefficient functions of an ERP system.

To improve the process of data registration in the area of the technical preparation of production, the integration of an ERP database and CAD system is required [7]. The designers can use an ERP database with material and intermediate indexes as a drawing library. Such integration allows a saving of 30% of the time that designers need to create material indexes, bills of materials and product structures. The material quantities can be directly calculated on the basis of the geometrical dimensions of the CAD objects automatically, this also allows the avoidance of unique errors by data registration.

4 A Methodology of the Analysis of ERP Impact on the Technical Production Preparation

The data generated in the area of the TPP which is registered in an ERP system can create a know-how database for manufacturing enterprises. To analyze the impact of an ERP system, it is important to evaluate the role of the engineers employed in the area of the TPP and who are responsible for data registration. On the other hand, it is important to consider the advantages, such as: decision support, reporting or the automation of routine activities. The activities of engineers in the ERP system could

be directly measured on the basis of workflows (if it is built into the ERP system). The proposed analysis is based both on quantitative and qualitative factors. The qualitative factors could be determined only on the basis of a research survey or direct interviews with engineers employed in the area of the technical preparation of production. Figure 3 presents the methodology of an analysis of ERP impact on the technical preparation of production.

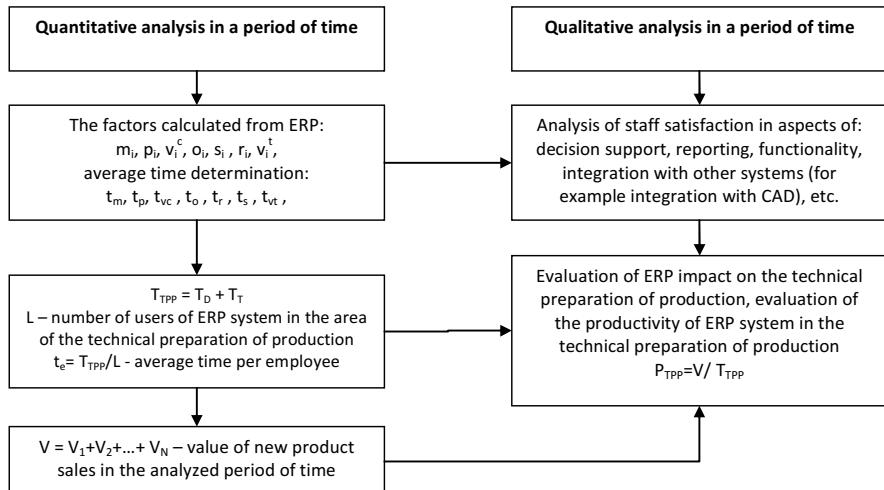


Fig. 3. Methodology of the analysis of ERP impact on the technical preparation of production

The productivity of an ERP system in the TPP should improve together with the development of an ERP database, the standardization of materials and intermediates, tools and instruments, etc. For ETO manufacturing companies, it is very important to rapidly calculate and make offers. The support of the sales area by making offers and cost monitoring during the construction process are very important and time-consuming tasks carried out by engineers in such types of enterprises. Very good results are provided by the function of ERP systems that enables the copying of the structure of an old product onto a new one. The procedure is not easy because it requires the automatic generation of new indexes for subassemblies or intermediates. However, it enables the rapid preparation of an offer for a new product and calculates material and labor costs immediately. Another very important function of an ERP system, with regards to the effective management of the technical preparation of production, is the support it may grant to the motivation system. In an ERP system, it is possible to register the activity times of engineers, results of timely work (constructed products) and failures (cards of changes). On the basis of the registered data, a motivation system can be proposed that enables the evaluation of the direct activities of the engineers of the TPP area.

5 Conclusions

In the article, the impact of an ERP system on the area of the TPP is analyzed. The main problem is the variety of manufacturing systems and functions of ERP systems. A general model of data flow between the TPP and other functional areas of a manufacturing company is proposed. For different kinds of manufacturing systems (MTS, MTO, ATO and ETO); a procedure of the evaluation of the importance and labor-intensity of the data entry is proposed. On the basis of the analysis, a critical set of data for different kinds of production systems is made. For typical tasks carried out in the area, a methodology is proposed that enables the evaluation of productivity of using an ERP system. The methodology requires an analysis of an amount of ERP data which needs to be generated periodically, and averages the times of the data registration and the values of new product sales in a defined period of time. Further research work will be focused on improving the efficiency of processes in the area of the technical preparation of production.

References

1. Bellatreche, L., Dung, N.X., Pierra, G., Hondjack, D.: Contribution of ontology-based data modeling to automatic integration of electronic catalogues within engineering databases. *Computers in Industry* 57(8-9), 711–724 (2006)
2. Chou, S.-W., Chang, Y.-C.: The implementation factors that influence the ERP (enterprise resource planning) benefits. *Decision Support Systems* 46, 149–157 (2008)
3. Davenport, T.H., Harris, J.G., Cantrell, S.: Enterprise systems and ongoing process change. *Business Process Management Journal* 10, 16–26 (2004)
4. Hicks, C., McGovern, T.: Product life cycle management in engineer-to-order industries. *International Journal of Technology Management* 48(2), 153–167 (2009)
5. Ehie, I., Madsen, M.: Identifying critical issues in enterprise resource planning (ERP) implementation. *Computers in Industry* 56, 545–557 (2005)
6. Ke, Q., Wei, K.K.: Organizational culture and leadership in ERP implementation. *Decision Support Systems* 45, 208–218 (2008)
7. Kłos, S.: Methodology of CAD and ERP systems integration for project-driven manufacturing. In: *Virtual Design and Automation: 2nd International Conference: New Trends in Collaborative Product Design*, Poznań, Poland (2005)
8. Kłos, S.: Evaluation methodology of ERP system implementation in manufacturing enterprises. Oficyna Wydawnicza Uniwersytetu Zielonogórskiego, Zielona Góra (2010)
9. Lee, C., Leem, C.S., Hwang, I.: PDM and ERP integration methodology using digital manufacturing to support global manufacturing. *International Journal of Advanced Manufacturing Technology* 53, 399–409 (2011)
10. Olhager, J., Selldin, E.: Enterprise resource planning survey of Swedish manufacturing firms. *European Journal of Operational Research* 146, 365–373 (2003)
11. Ou-Yang, C., Chang, M.J.: Developing an agent-based PDM/ERP collaboration system. *International Journal of Advanced Manufacturing Technology* 30, 369–384 (2006)

12. Rudberg, M., Wikner, J.: Mass customization in terms of the customer order decoupling point. *Production Planning & Control* 15(4), 445–458 (2004)
13. Sarkis, J., Gunasekaran, A.: Enterprise resource planning – modeling and analysis. *European Journal of Operational Research* 146(2), 229–232 (2003)
14. Wei, Z., Tan, J., Feng, Y.: Integration technology of ERP and PDM based on business remote function call. *International Journal of Advanced Manufacturing Technology* 40, 1044–1052 (2009)
15. Wu, L.-C., Ong, C.-S., Hsu, Y.W.: Active ERP implementation management: A Real Options perspective. *The Journal of Systems and Software* 81, 1039–1050 (2008)

Minimum Energy Control of Fractional Discrete-Time Linear Systems with Delays in State and Control

Rafał Kociszewski

Bialystok University of Technology, Faculty of Electrical Engineering, Bialystok, Poland
rafko@pb.edu.pl

Abstract. In the paper the problem of minimum energy control of fractional discrete-time linear system with multiple delays in state and control are addressed. General form of solution of the state equation of the system is given and conditions for reachability and minimum energy control are established. The considerations are illustrated by numerical example.

Keywords: fractional, linear systems, discrete-time, time-delay, minimum energy control.

1 Introduction

Dynamical systems described by fractional order differential or difference equations have been investigated in several areas such as viscoelasticity, electrochemistry, diffusion processes, automatic control, etc. (see [4, 8, 19, 20, 22], for example). The problem of controllability and reachability of dynamical systems without delays or with delays for standard or fractional order systems have been considered in [2, 3, 5, 15, 18, 21, 23]. The problem of minimum energy control for standard systems has been firstly introduced and solved in [12]. This problem has been investigated in [6, 13, 14] for standard systems and in [3, 9, 10, 16, 17] for fractional order systems. The problem of minimum energy control with bounded inputs has been recently examined in [3, 11].

The main purpose of the paper is to give the general form of solution of the state equation of fractional discrete-time linear system with multiple delays in state and control and solution of the minimum energy control problem for this systems.

2 Problem Formulation

Let us consider the discrete-time linear system with delays described by the state equation

$$\Delta^\alpha x_{i+1} = A_0 x_i + \sum_{k=1}^h A_k x_{i-k} + B_0 u_i + \sum_{j=1}^q B_j u_{i-j}, \quad (1)$$

with the initial conditions

$$x_{-k} \in \Re^n, \quad k = 0, 1, \dots, h; \quad u_{-j} \in \Re^m, \quad j = 1, 2, \dots, q \quad (2)$$

where h and q are positive integers (number of delays), $x_i \in \Re^n$, $u_i \in \Re^m$ are the state and input vectors, respectively, $A_k \in \Re^{n \times n}$ ($k = 0, 1, \dots, h$), $B_j \in \Re^{n \times m}$ ($j = 0, 1, \dots, q$),

$$\Delta^\alpha x_i = \sum_{k=0}^i (-1)^k \binom{\alpha}{k} x_{i-k} \quad (3)$$

is the fractional difference of order $\alpha \in \Re$ of the discrete-time function x_i and

$$\binom{\alpha}{k} = \frac{\alpha!}{k!(\alpha - k)!} \quad (4)$$

Substituting of (3) for $i+1$ into (1) we obtain

$$x_{i+1} = F_0 x_i + \sum_{k=1}^h A_k x_{i-k} + \sum_{k=1}^i c_k(\alpha) x_{i-k} + \sum_{j=0}^q B_j u_{i-j}, \quad (5)$$

where

$$F_0 = A_0 + I_n \alpha, \quad (I_n \text{ is the } n \times n \text{ identity matrix}) \quad (6)$$

and

$$c_k(\alpha) = (-1)^k \binom{\alpha}{k+1}, \quad k = 1, 2, \dots \quad (7)$$

The coefficients (7) we can compute by the use of the formulas [3]:

$$c_{k+1}(\alpha) = c_k(\alpha) \frac{k+1-\alpha}{k+2}, \quad k = 1, 2, \dots \quad c_1(\alpha) = 0.5\alpha(1-\alpha). \quad (8)$$

Now we shall formulate the fundamental definitions for the reachability of the system (1) which is necessary to the further considerations about the minimum energy control problem of the fractional system (1).

Definition 1. A state $x_f \in \Re^n$ is called reachable in N steps if there exists a sequence of inputs $u_i \in \Re^m$, $i = 0, 1, \dots, N-1$, that transfers the fractional system with delays (1) from zero initial conditions (2) to the state x_f .

If every state $x_f \in \Re^n$ is reachable in N steps, according to the above definition, then we can say that the fractional system (1) is reachable in N steps.

Definition 2. If for every state $x_f \in \Re^n$ there exists a natural number N such that the state x_f is reachable in N steps then the system is called reachable.

The general problem of minimum energy control of the fractional system (1) we can formulate in the same manner as for the fractional discrete-time systems without delays. This problem can be stated as follows:

Find a control sequence $u_i \in \Re^m$, $i = 0, 1, \dots, N - 1$, which transfers the fractional system (1) from zero initial conditions (2) to the desired final state $x_f \in \Re^n$ and minimizes the performance index

$$I(u) = \sum_{i=0}^{N-1} u_i^T Q u_i \quad (9)$$

where $Q \in \Re^{m \times n}$ is a symmetric positive definite weighting matrix.

The control sequence $u_i \in \Re^m$, $i = 0, 1, \dots, N - 1$ that minimizes the performance index (9) is called minimal one.

The aim of this paper is to give the general form of solution of the state equation (1) of the fractional discrete-time linear system with delays, the condition of reachability, and in consequence the solution of the minimum energy problem of the fractional system with delays in state and control (1).

3 Problem solution

Taking the \mathcal{Z} -transform (similarly as in [1]) to both sides of the equation (5) with (2) we obtain

$$\begin{aligned} zX(z) - zx_0 &= F_0 X(z) + \sum_{k=1}^h A_k z^{-k} [X(z) + \sum_{r=-k}^{-1} x_r z^{-r}] + \sum_{k=1}^i c_k z^{-k} X(z) + \\ &+ B_0 U(z) + \sum_{k=1}^q B_k z^{-k} [U(z) + \sum_{i=-k}^{-1} u_i z^{-j}] \end{aligned} \quad (10)$$

where

$$X(z) = \mathcal{Z}\{x_i\}, \quad U(z) = \mathcal{Z}\{u_i\}. \quad (11)$$

The equation (10) can be written in the form

$$\Delta(z)X(z) = zx_0 + \sum_{k=1}^h A_k z^{-k} \sum_{r=-k}^{-1} x_r z^{-r} + B_0 U(z) + \sum_{k=1}^q B_k z^{-k} [U(z) + \sum_{i=-k}^{-1} u_i z^{-i}] \quad (12)$$

where $\Delta(z)$ is the characteristic matrix and has the form [3]

$$\Delta(z) = zI_n - F_0 - \sum_{k=1}^h A_k z^{-k} - \sum_{k=1}^i I_n c_k(\alpha) z^{-k} \quad (13)$$

Solving the equation (13) for $X(z)$ we obtain

$$\begin{aligned} X(z) = & [\Delta^{-1}(z)z]x_0 + [\Delta^{-1}(z)z] \sum_{k=1}^h A_k \sum_{r=0}^{k-1} x_{r-k} z^{-r-1} + \\ & + [\Delta^{-1}(z)z]B_0 U(z) z^{-1} + [\Delta^{-1}(z)z] \sum_{k=1}^q B_k \sum_{r=0}^{k-1} z^{-r-1} u_{r-k} + [\Delta^{-1}(z)z] \sum_{k=1}^q B_k z^{-r-1} U(z) \end{aligned} \quad (14)$$

Taking the inverse \mathcal{Z} -transform to (14) gives the solution of the equation (5) (and the state equation (1)) in the form

$$x_i = \Phi_i x_0 + \sum_{k=1}^h \sum_{r=0}^{k-1} \Phi_{i-r-1} A_k x_{r-k} + \sum_{k=1}^q \sum_{r=0}^{k-1} \Phi_{i-r-1} B_k u_{r-k} + \sum_{j=0}^{i-1} \sum_{k=0}^q \Phi_{i-1-k-j} B_k u_j \quad (15)$$

where

$$\Phi_i = \mathcal{Z}^{-1}\{z\Delta^{-1}(z)\} \quad (16)$$

is the state-transition matrix for the equation (5). From (16) and (13) it follows that the state-transition matrix satisfies the equation

$$\Phi_{i+1} = F_0 \Phi_i + \sum_{k=1}^h A_k \Phi_{i-k} + \sum_{k=1}^i c_k(\alpha) \Phi_{i-k} \quad (17)$$

with the initial conditions

$$\Phi_0 = I_n, \quad \Phi_i = 0 \text{ for } i < 0. \quad (18)$$

From (16) it follows that the solution of the equation (5) for $i = N$ with the zero conditions (2) has the form

$$x_N = R_N u^N \quad (19)$$

where

$$R_N = [\Psi_{N-1} \quad \Psi_{N-2} \quad \dots \quad \Psi_1 \quad \Psi_0] \quad (20)$$

is called the reachability matrix, and

$$\Psi_i = \sum_{k=0}^q \Phi_{i-k} B_k, \quad i = 0, 1, \dots, N-1 \quad (21)$$

$$u^N = \begin{bmatrix} u_{N-1} \\ u_{N-2} \\ \vdots \\ u_0 \end{bmatrix} \in \Re^{Nm}. \quad (22)$$

The matrix Φ_i in (21) has the form (17).

The following condition for the fractional system (1) can be proved in the same manner as for the positive system with $\alpha = 1$ (see [18], for example).

Theorem 1. The fractional system with delays (1) is reachable in N steps if and only if there exists integer number N such that rank of the reachability matrix (20) is equal to n . If this holds, then control u^N which transfers the fractional system (1) from zero initial conditions (2) to the desired final state $x_f \in \Re^n$, can be computed from the formula

$$u^N = R_N^T [R_N R_N^T]^{-1} x_f + (I_{Nm} - R_N^T [R_N R_N^T]^{-1} R_N) K = \bar{x}_f + HK, \quad (23)$$

where $\bar{x}_f = R_N^T [R_N R_N^T]^{-1} x_f$, $H = I_{Nm} - R_N^T [R_N R_N^T]^{-1} R_N$ and $K \in \Re^{Nm \times n}$ is an arbitrary matrix, but such that $\det[R_N K] \neq 0$. ■

The formula (23) is based on the right-inverse of the reachability matrix (20). Using another forms of the right-inverse of the rectangular matrix (see [7], for example) we can write the following formulas

$$u^N = K_1 [R_N K_1]^{-1} x_f, \quad (24)$$

or

$$u_0^N = G \begin{bmatrix} -\frac{1}{\bar{a}_0} (\bar{A}_1^{m-1} \bar{a}_{m-1} \bar{A}_1^{m-2} + \dots + a_1 I_n) (I_n - \bar{A}_2 K_2) \\ K_2 \end{bmatrix} x_f \quad (25)$$

where $K_1 \in \Re^{Nm}$ ($K_1 \in \Re^{Nm \times n}$) is an arbitrary vector (matrix), $K_2 \in \Re^{(Nm-n) \times n}$ is also an arbitrary matrix, G is a permutation matrix of columns of matrix (21) such that

$$R_N G = [\bar{A}_1 \quad \bar{A}_2], \quad \bar{A}_1 \in \Re^{n \times n}, \quad \bar{A}_2 \in \Re^{n \times (Nm-n)} \quad (26)$$

and $\bar{a}_0, \dots, \bar{a}_{n-1}$ are coefficients of polynomial of matrix \bar{A}_1 in the form

$$\det[I_n z - \bar{A}_1] = z^n + \bar{a}_{m-1} s^{m-1} + \dots + \bar{a}_1 z + \bar{a}_0 \quad (27)$$

Optimal control which minimizes performance index (9) depends on the weighting matrix Q ($Q = Q^T \succ 0$ – symmetric and positive defined). If we assume that

$$Q = \text{diag}(q_1, \dots, q_m), \quad q_1 = \dots = q_m = v^2 \quad (28)$$

then we can write the following condition.

Theorem 2. Let the fractional system (1) be reachable in N steps. The control sequence \hat{u}_0^N that minimizes the performance index

$$I(u) = v^2 \sum_{i=0}^{N-1} u_i^T u_i, \quad v > 0 \quad (29)$$

which steers the state of the system (1) from zero initial conditions to any desired final state x_f has the form

$$\hat{u}_0^N = R_N^T [R_N R_N^T]^{-1} x_f. \quad (30)$$

Proof. The performance index (30) for (23) we can write in the form

$$\begin{aligned} I(u) &= v^2 (u_0^N)^T u_0^N = v^2 \left\{ u^N = R_N^T [R_N R_N^T]^{-1} x_f + (I_{Nm} - R_N^T [R_N R_N^T]^{-1} R_N) K \right\}^T \times \\ &\quad \times \left\{ R_N^T [R_N R_N^T]^{-1} x_f + (I_{Nm} - R_N^T [R_N R_N^T]^{-1} R_N) K \right\} = \\ &= v^2 \left(\bar{x}_f^T \bar{x}_f + K^T H^T H K + K^T H^T \bar{x}_f + \bar{x}_f^T H K \right) \end{aligned} \quad (31)$$

Let us notice that

$$K^T H^T \bar{x}_f = K^T (I_{Nm} - R_N^T [R_N R_N^T]^{-1} R_N)^T R_N^T [R_N R_N^T]^{-1} x_f = 0, \quad \bar{x}_f^T H K = 0 \quad (32)$$

Taking into account the above formulas we obtain the performance index in the form

$$I(u_0^N) = v^T (\bar{x}_f^T \bar{x}_f + K^T H^T H K) \quad (33)$$

The performance index (33) achieves minimum for $K = 0$, therefore for the sequence of controls (30). ■

Let us notice that if $K_1 = R_N^T$ then from (24) we obtain (30).

The minimum energy problem can be solved by the use of the following procedure.

Procedure

Step1. Knowing $A_k \in \Re^{n \times n}$, ($k = 0, 1, \dots, h$), $B_j \in \Re^{n \times m}$ ($j = 0, 1, \dots, q$), α , N and using (20) find the matrix R_N (20).

Step 2. Knowing the matrix R_N (20), and using (30) compute sequence of inputs $\hat{u}_0, \dots, \hat{u}_{N-1}$.

Step 3. Knowing v and using (29) compute the value of the index $I(\hat{u})$.

4 Example

Consider fractional system (1) with $h = q = 2$ delays, $\alpha = 0.8$ and the matrices

$$A_0 = \begin{bmatrix} 0.1 & 0 & 0 \\ -0.1 & 0 & 0.2 \\ 0.1 & 0 & -0.2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 0 & 0.1 \\ -0.3 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 0 \\ -0.2 & 0 & 0 \\ -0.1 & 0 & 0.2 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (34)$$

Find an optimal control (sequence of inputs) that steers the state of the system from zero initial condition (2) to the state $x_f = [9 \ 8 \ 9]^T$ in $N = 8$ steps and minimizes the performance index (29) for $v = 0.4$.

Using Procedure we obtain the following.

Step 1. The condition of reachability in $N = 8$ is satisfied because the reachability matrix R_N (20)

$$R_8 = [\Psi_7 \ \Psi_6 \ \Psi_5 \ \Psi_4 \ \Psi_3 \ \Psi_2 \ \Psi_1 \ \Psi_0] =$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.02 & 0 & 0 & 0 & 0.2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.01 & -0.01 & 0 & 0 & 0.1 & -0.08 & 0 & -0.25 & 1.0 & -0.14 & 0 & 0.2 & 0 & 0 \\ 0 & 0.01 & 0 & -0.01 & 0 & 0.01 & 0 & 0.03 & 0 & -0.05 & 0 & 0.24 & 0 & -0.2 & 0 & 1 \end{bmatrix} \quad (35)$$

has rank equal to 3. The matrices $\Psi_0 \dots \Psi_7$ we compute from (21), where $\Phi_1 \dots \Phi_7$ are computed from recursive formula (17).

Step 2. The optimal sequence $\hat{u}_0, \dots, \hat{u}_7$ computed from (30) with (35) has the form

$$\begin{aligned}\hat{u}_0 &= \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}, & \hat{u}_1 &= \begin{bmatrix} 0.06 \\ -0.14 \end{bmatrix}, & \hat{u}_2 &= \begin{bmatrix} 0 \\ 0.06 \end{bmatrix}, & \hat{u}_3 &= \begin{bmatrix} 0.62 \\ 0.03 \end{bmatrix} \\ \hat{u}_4 &= \begin{bmatrix} 0 \\ -2.04 \end{bmatrix}, & \hat{u}_5 &= \begin{bmatrix} 6.22 \\ 3.29 \end{bmatrix}, & \hat{u}_6 &= \begin{bmatrix} 0 \\ 8.34 \end{bmatrix}, & \hat{u}_7 &= \begin{bmatrix} 0 \\ 9.78 \end{bmatrix}\end{aligned}\quad (36)$$

Step 3. The minimal value of the performance index is equal to

$$I(\hat{u}) = (0.4)^2 [\hat{u}_0^T \hat{u}_0 + \hat{u}_1^T \hat{u}_1 + \hat{u}_2^T \hat{u}_2 + \hat{u}_3^T \hat{u}_3 + \hat{u}_4^T \hat{u}_4 + \hat{u}_5^T \hat{u}_5 + \hat{u}_6^T \hat{u}_6 + \hat{u}_7^T \hat{u}_7] = 35.08 \quad (37)$$

Let us check obtained results. Computing of the solution (17) for $i = 0, 1, \dots, 7$ with (34) and zero initial conditions (2) we obtain

$$\begin{aligned}x_1 &= \begin{bmatrix} 0 \\ 0 \\ 0.1 \end{bmatrix}, & x_2 &= \begin{bmatrix} 0.1 \\ 0.02 \\ -0.16 \end{bmatrix}, & x_3 &= \begin{bmatrix} -0.1 \\ -0.02 \\ 0.12 \end{bmatrix}, & x_4 &= \begin{bmatrix} 0.04 \\ 0.05 \\ 0 \end{bmatrix}, \\ x_5 &= \begin{bmatrix} 0.03 \\ 0.05 \\ -2.05 \end{bmatrix}, & x_6 &= \begin{bmatrix} -2.01 \\ 0.15 \\ 3.64 \end{bmatrix}, & x_7 &= \begin{bmatrix} 2.85 \\ 0.95 \\ 7.12 \end{bmatrix}, & x_8 &= \begin{bmatrix} 9 \\ 8 \\ 9 \end{bmatrix}\end{aligned}\quad (38)$$

The optimal control sequence was computed correctly, because $x_f = x_8$. The trajectory of the considered fractional system is shown on Fig 1, where x^1, x^2, x^3 are components of the vectors (38) ($x_k = [x^1 \ x^2 \ x^3]^T$, $k = 1, 2, \dots, 8$).

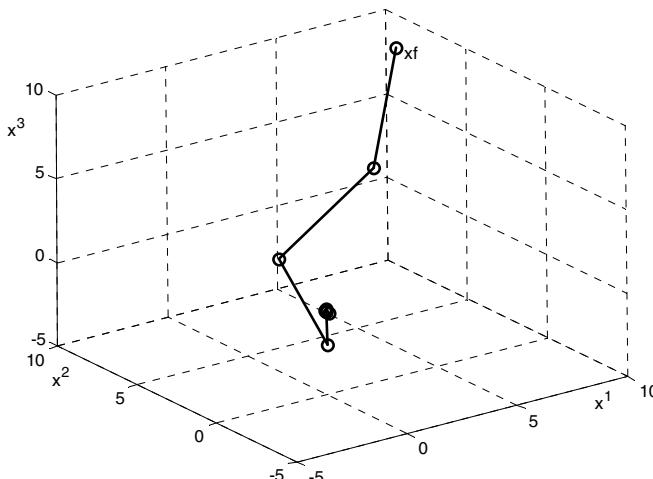


Fig. 1. Trajectory of the considered fractional system with order $\alpha = 0.8$, and the matrices (34) and delays ($h = q = 2$) ('o' are the states (38))

5 Concluding Remarks

The problem reachability and minimum energy control of fractional discrete-time system with delays (1) have been addressed. The general form of solution of state equation (1) is given. Necessary and sufficient conditions for reachability and minimum energy control have been established and illustrated by numerical example.

The considerations can be extended to fractional positive discrete-time linear systems with delays and for the minimum energy control for that class of dynamical systems with bounded controls.

Acknowledgement. This work was supported by the National Science Center in Poland under grant NN 514 6389 40.

References

1. Busłowicz, M.: On some properties of the solution of state equation of discrete-time systems with delays. *Zesz. Nauk. Polit. Bial., Elektrotechnika* (1), 17–29 (1983) (in Polish)
2. Busłowicz, M., Kaczorek, T.: Reachability and minimum energy control of positive linear discrete-time systems with multiple delays in state and control. *Measurement Automation and Monitoring* 53(10), 40–45 (2007)
3. Busłowicz M.: Controllability, reachability and minimum energy control of fractional discrete-time linear systems with multiple delays in state. *Bull. Pol. Acad.: Tech.* (in print)
4. Dzieliński, A., Sierociuk, D., Sarwas, G.: Some applications of fractional order calculus. *Bull. Pol. Acad. Scie.: Tech.* 58(4), 583–592 (2010)
5. Guermach, S., Djennoune, S., Bettayeb, M.: Controllability and observability of linear discrete-time fractional-order systems. *Int. J. Appl. Math. Comput. Sci.* 18(2), 213–222 (2008)
6. Kaczorek, T., Klamka, J.: Minimum energy control of 2D linear systems with variable coefficients. *Int. J. of Control* 44(3), 645–650 (1986)
7. Kaczorek, T.: Vectors and matrices in automatics and electrotechnics. WNT, Warsaw (1998)
8. Kaczorek, T.: Selected Problems of Fractional Systems Theory. LNCIS, vol. 411. Springer, Heidelberg (2011)
9. Kaczorek, T.: Minimum energy control of fractional positive continuous-time linear systems. In: Proc. 18th Int. Conference Methods and Models in Automation and Robotics, Międzyzdroje, Poland, pp. 622–626 (2013) (CD-ROM)
10. Kaczorek, T.: Minimum energy control of fractional positive continuous-time linear systems with bounded inputs. *Int. J. Appl. Math. Comput. Sci.* (in print)
11. Kaczorek, T.: An extension of Klamka's method of minimum energy control to fractional positive discrete-time linear systems with bounded inputs. *Bull. Pol. Acad.: Tech.* 62(1) (2014) (in print)
12. Klamka, J.: Relative controllability and minimum energy control of linear systems with distributed delays in control. *IEEE Trans. on Automatic Control* AC-21(4), 594–595 (1976)
13. Klamka, J.: Minimum energy control of discrete systems with delays in control. *Int. J. Control* 26(5), 737–744 (1977)

14. Klamka, J.: Minimum energy control of 2D systems in Hilbert spaces. *Systems Science* 9(1-2), 33–42 (1983)
15. Klamka, J.: Controllability of Dynamical Systems. Kluwer Acad. Publ., Dordrecht (1991)
16. Klamka, J.: Controllability and minimum energy control problem of infinite dimensional fractional discrete-time systems with delays. In: Proc. First Asian Conference on Intelligent Information and Database Systems, Dong Hoi City, Vietnam, pp. 398–403 (2009)
17. Klamka, J.: Controllability and minimum energy control problem of fractional discrete-time systems. In: Baleanu, D. (ed.) New Trends in Nanotechnology and Fractional Calculus Applications, pp. 503–509. Springer, New York (2010)
18. Kociszewski, R.: Controllability and observability of linear time-invariant positive discrete-time systems with delays. PhD Dissertation, Faculty of Electrical Engineering, Bialystok University of Technology, Bialystok (2008) (in Polish)
19. Luo, Y., Chen, Y.-Q.: Fractional Order Motion Controls. John Wiley & Sons Ltd, Chichester (2013)
20. Monje, C.A., Chen, Y.Q., Vinagre, B.M., Xue, D., Feliu-Batlle, V.: Fractional-order Systems and Controls Fundamentals and Applications. Springer, London (2010)
21. Mozyrska, D., Pawłuszewicz, E.: Local controllability of nonlinear discrete-time fractional order systems. *Bull. Pol. Ac.: Tech.* 61(1), 251–256 (2013)
22. Podlubny, I.: Fractional Differential Equations. Academic Press, San Diego (1999)
23. Trzasko, W.: Reachability and controllability of positive fractional discrete-time systems with delay. *Journal of Automation, Mobile Robotics & Intelligent Systems* 2(3), 43–47 (2008)

Efficient Mechanism of Output Constraint Handling for Analytical Predictive Controllers Based on Hammerstein Models

Piotr M. Marusak

Institute of Control and Computation Engineering, Warsaw University of Technology,
ul. Nowowiejska 15/19, 00–665 Warszawa, Poland
P.Marusak@ia.pw.edu.pl

Abstract. The proposed mechanism of output constraint handling uses a method of output prediction generation, which originates from Model Predictive Control (MPC) algorithms and is based on Hammerstein models. Therefore the mechanism can give very good results in control systems of nonlinear plants. It is relatively easy to use and, at the same time, very efficient, because in the output constraint handling the influence of the control action many sampling instants ahead can be taken into consideration. Moreover, the proposed method is flexible – it is possible to choose how many future predicted output values are constrained.

Keywords: predictive control, nonlinear control, constrained control, Hammerstein models.

1 Introduction

The Model Predictive Control (MPC) algorithms, thanks to very good control performance, are widely used in practice. They can be successfully applied in control systems of plants with difficult dynamics (large delay or inverse response) and with constraints; see e.g. [1,3,10,12]. The source of advantages of the MPC algorithms is prediction of behavior of the control system during generation of the control signals. This prediction is done using a model of the control plant. Standard MPC algorithms are based on linear models. However, application of such algorithms to nonlinear control plants may result in unsatisfactory behavior of the control system. Therefore MPC algorithms based on nonlinear models were proposed.

Many processes can be successfully modeled using Hammerstein models; see e.g. [2]. In Hammerstein models a nonlinear static block precedes a linear dynamic block. The efficient method of prediction generation using a Hammerstein model and a linear approximation of this model was proposed in [5] and used to formulate a numerical MPC algorithm (with quadratic optimization problem solved at each iteration). Unfortunately, for fast control plants a numerical MPC algorithm may generate control signal too slowly. Therefore analytical MPC algorithms were also designed. Such a numerically efficient analytical MPC algorithm based on Hammerstein models was proposed in [7] and is used in the

example discussed in Sect. 4. In this algorithm the main part of calculations needed to derive the control value can be done offline. Therefore, even solving of the quadratic programming problem is avoided and the algorithm can be applied in cases when the control signal must be generated frequently.

In some applications of analytical MPC controllers, output constraints can be important. It is because the output constraints often decide the safety and the economic efficiency of the process. The constraints can be handled using the control structure with additional predictive set-point optimizer (see e.g. [8]) which unfortunately leads to complication of the control system. However, the output constraint handling mechanism should be as simple as possible in order not to complicate the algorithm too much. The method proposed in the paper fulfills this condition. It uses the idea developed for standard DMC controllers in [4], but is designed for the predictive controllers based on Hammerstein models. Despite the constraint mechanism exploits the nonlinear model (as the prediction is obtained using the Hammerstein model), it is relatively simple. It is also efficient, as it will be demonstrated in the example control system.

The very basic idea of the proposed method of output constraint handling is to modify the control action generated by a controller in such a way that the predicted output does not violate the constraints. The significant advantage of the proposed method is possibility of taking into consideration the values of predicted output many sampling instants ahead, during constraint handling. Such an approach results in very good performance offered by the method.

Although the proposed mechanism of output constraint handling uses predictions generated with the Hammerstein model in a way described in [5], it can be applied with different analytical predictive controllers, also with the one described in [6] and even with controllers not necessarily based on Hammerstein models (but the Hammerstein model used in the constraint handling should reflect behavior of the control plant).

In the next section, advanced prediction method used in the analytical predictive controllers, based on Hammerstein models is briefly reviewed. In Sect. 3 the output constraint handling mechanism is described in details. Section 4 contains presentation of simulation experiments done in the example control system of a nonlinear control plant with delay, illustrating very good performance of the proposed approach. The paper is summarized in the last section.

2 Advanced Prediction Using Hammerstein Models

Assume that the Hammerstein model is used to describe the process. In such a model a nonlinear static block is followed by a linear dynamic block; see Fig. 1. In general, in the static block a nonlinear function is used:

$$z_k = f(u_k) , \quad (1)$$

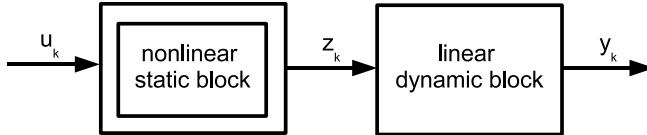


Fig. 1. Structure of the Hammerstein model; u_k – input, y_k – output, z_k – output of the nonlinear static block

where z_k is the output of the static block. In the example described in Sect. 4, it is assumed that a fuzzy Takagi-Sugeno static part is used, i.e.:

$$f(u_k) = \sum_{j=1}^l w_j(u_k) \cdot z_k^j = \sum_{j=1}^l w_j(u_k) \cdot (b_j \cdot u_k + c_j) , \quad (2)$$

where $w_j(u_k)$ are weights obtained using fuzzy reasoning (for details see e.g. [9,11]), z_k^j are outputs of local models in the fuzzy static model, l is the number of fuzzy rules in the model, b_j and c_j are parameters of the local models in the fuzzy static part of the model.

Assume that the dynamic part of the model has the form of the step response:

$$\hat{y}_k = \sum_{n=1}^{p_d-1} a_n \cdot \Delta z_{k-n} + a_{p_d} \cdot z_{k-p_d} , \quad (3)$$

where \hat{y}_k is the output of the fuzzy Hammerstein model, a_i are coefficients of the step response of the linear dynamic block, p_d is the horizon of the process dynamics (equal to the number of sampling instants after which the step response can be considered as settled).

The prediction is obtained in a way described in [5]. In this prediction method the Hammerstein model (3) is used to obtain the free response (the response to the control action applied in the previous sampling instants, i.e. it contains future values of the output variable calculated assuming that the control signal does not change in the future):

$$\tilde{y}_{k+i|k} = \sum_{n=i+1}^{p_d-1} a_n \cdot \Delta z_{k-n+i} + a_{p_d} \cdot z_{k-p_d+i} + d_k , \quad (4)$$

where $\tilde{y}_{k+i|k}$ is the element of the free response for the $(k+i)^{\text{th}}$ sampling instant, predicted at the k^{th} sampling instant, $d_k = y_k - \hat{y}_k$ is the DMC-type disturbance model (it is assumed the same for all future sampling instants). In order to simplify further calculations, the influence of the future control changes is derived using a linear approximation of the fuzzy Hammerstein model (3) at each algorithm iteration:

$$\hat{y}_k^L = d z_k \cdot \left(\sum_{n=1}^{p_d-1} a_n \cdot \Delta u_{k-n} + a_{p_d} \cdot u_{k-p_d} \right) , \quad (5)$$

where dz_k is a slope of the static characteristic near the z_k . It can be calculated analytically or numerically. In the latter case, which can be applied for instance if non-differentiable membership functions are used in the fuzzy static model, the following formula can be used:

$$dz_k = \frac{\sum_{j=1}^l (w_j(u_k + du) \cdot (b_j \cdot (u_k + du) + c_j) - w_j(u_k) \cdot (b_j \cdot u_k + c_j))}{du}, \quad (6)$$

where du is a small number.

Finally, the following prediction is obtained [5]:

$$\mathbf{y} = \tilde{\mathbf{y}} + \mathbf{A}_k \cdot \Delta \mathbf{u}, \quad (7)$$

where $\mathbf{y} = [y_{k+1|k}, \dots, y_{k+p|k}]$, $y_{k+i|k}$ is a value of the output for the $(k+i)^{\text{th}}$ sampling instant, predicted at the k^{th} sampling instant, $\tilde{\mathbf{y}} = [\tilde{y}_{k+1|k}, \dots, \tilde{y}_{k+p|k}]$, p is the prediction horizon, $\Delta \mathbf{u} = [\Delta u_{k|k}, \dots, \Delta u_{k+s-1|k}]$, $\Delta u_{k+i|k}$ are future changes in manipulated variable, s is the control horizon,

$$\mathbf{A}_k = dz_k \cdot \mathbf{A} = dz_k \cdot \begin{bmatrix} a_1 & 0 & \dots & 0 & 0 \\ a_2 & a_1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_p & a_{p-1} & \dots & a_{p-s+2} & a_{p-s+1} \end{bmatrix}, \quad (8)$$

where \mathbf{A} is the constant matrix.

3 Mechanism of Output Constraint Handling

The analytical controllers generate the control value for the current sampling instant and then, in the next sampling instant, they must calculate the new value. Thus, for further considerations, it is assumed that the control signal remains unchanged. By such an assumption, the elements of the prediction vector \mathbf{y} given by (7) can be expressed as:

$$y_{k+i|k} = dz_k \cdot a_i \cdot \Delta u_k + \tilde{y}_{k+i|k}, \quad (9)$$

where $\Delta u_k = \Delta u_{k|k}$. The form of (9) dependent on the value of the control signal is as follows:

$$y_{k+i|k} = dz_k \cdot a_i \cdot u_k + \tilde{y}_{k+i|k} - dz_k \cdot a_i \cdot u_{k-1}. \quad (10)$$

Usually, the following output constraints should be fulfilled:

$$y_{\min} \leq y_{k+i|k} \leq y_{\max}, \quad (11)$$

where y_{\min} and y_{\max} are lower and upper output limits, respectively. Application of the prediction (10) in the output constraints allows taking them many sampling instants ahead:

$$y_{\min} \leq dz_k \cdot a_i \cdot u_k + \tilde{y}_{k+i|k} - dz_k \cdot a_i \cdot u_{k-1} \leq y_{\max}. \quad (12)$$

Split the output constraints (12) into two sets of constraints: the first set grouping lower constraints imposed on the currently derived control value:

$$dz_k \cdot a_i \cdot u_k \geq y_{\min} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1} , \quad (13)$$

and the second one – grouping upper constraints:

$$dz_k \cdot a_i \cdot u_k \leq y_{\max} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1} . \quad (14)$$

Now, the following rules of control value modification can be formulated:

- for lower constraints:
 - if $dz_k \cdot a_i \cdot u_k \geq y_{\min} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1}$ then

$$u_k = \frac{y_{\min} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1}}{dz_k \cdot a_i} ; \quad (15)$$
- for upper constraints:
 - if $dz_k \cdot a_i \cdot u_k \leq y_{\max} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1}$ then

$$u_k = \frac{y_{\max} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1}}{dz_k \cdot a_i} . \quad (16)$$

Note that not all constraints (13) and (14) must be imposed. It is often sufficient to use constraints for only a few sampling instants ahead. Thus, the proposed mechanism can be easily scaled to the current problem.

In the prediction used in the output constraint handling mechanism the DMC-type model of disturbance, containing also influence of modeling errors, was used. However, if it is possible to assess the values of modeling errors, such a knowledge can be easily used in the proposed mechanism.

Assume that the output prediction with uncertainty is described as follows:

$$\begin{aligned} \check{y}_{k+i|k} &= y_{k+i|k} + r_{k+i|k} \\ &= dz_k \cdot a_i \cdot u_k + \tilde{y}_{k+i|k} - dz_k \cdot a_i \cdot u_{k-1} + r_{k+i|k} , \end{aligned} \quad (17)$$

where $r_{k+i|k}$ represent influence of the modeling error on the prediction (usually unknown). Assume that the minimum and maximum values of the $r_{k+i|k}$ can be assessed:

$$r_{k+i|k}^{\min} \leq r_{k+i|k} \leq r_{k+i|k}^{\max} , \quad (18)$$

where $r_{k+i|k}^{\min} \leq 0$ and $r_{k+i|k}^{\max} \geq 0$, then the rules given above should be modified as follows:

- for lower constraints:
 - if $dz_k \cdot a_i \cdot u_k \geq y_{\min} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1} - r_{k+i|k}^{\min}$ then

$$u_k = \frac{y_{\min} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1} - r_{k+i|k}^{\min}}{dz_k \cdot a_i} ; \quad (19)$$

– for upper constraints:

— if $dz_k \cdot a_i \cdot u_k \leq y_{\max} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1} - r_{k+i|k}^{\max}$ then

$$u_k = \frac{y_{\max} - \tilde{y}_{k+i|k} + dz_k \cdot a_i \cdot u_{k-1} - r_{k+i|k}^{\max}}{dz_k \cdot a_i} . \quad (20)$$

Remark. Note that if there is a delay in the control plant, then changing the current value u_k of the control signal influences output of the control plant with the delay. In other words, it will influence future output values beginning from the $(k + d + 1)^{th}$ one. Such a situation takes place in the example discussed in the next section.

4 Simulation Experiments

4.1 Control Plant

The experiments were done in the control system of a highly nonlinear control plant with large time delay – the ethylene distillation column used for tests also in [5,7]. The Hammerstein model of the control plant is shown in Fig. 2. The static part of the Hammerstein model has the form of the fuzzy Takagi-Sugeno model with three local models of the form:

$$z_k^j = b_j \cdot u_k + c_j , \quad (21)$$

where $b_1 = -2222.4$, $b_2 = -1083.2$, $b_3 = -534.4$, $c_1 = 9486$, $c_2 = 4709.3$, $c_3 = 2408.7$. The assumed membership functions are shown in Fig. 3. The

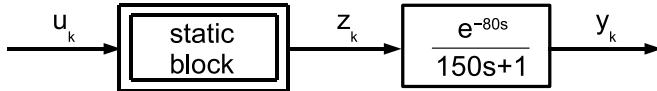


Fig. 2. Hammerstein model of the distillation column

output variable y_k is the impurity of the product counted in ppm. The following output constraint, being a result of demanded quality of the product, is assumed:

$$y_k \leq 400 \text{ ppm} . \quad (22)$$

Exceeding the impurity limit causes wasting of the product. The control variable u_k is the reflux to product ratio; the higher it is the purer product is obtained. During the experiments it was assumed that the reflux is constrained:

$$4.05 \leq u_k \leq 4.4 . \quad (23)$$

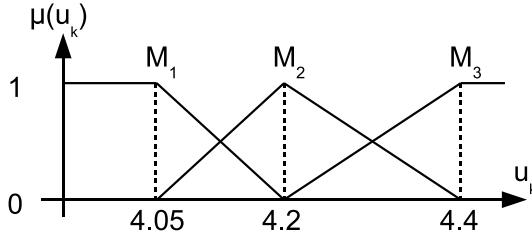


Fig. 3. Membership functions of the static part of the Hammerstein model

4.2 Analytical MPC Controller Based on Hammerstein Models (HMPC)

In the MPC algorithms, control signals are generated using prediction of future behavior of the control plant many sampling instants ahead. The values of control variables are calculated in such a way that the prediction fulfills assumed criteria. Usually, minimization of the following performance function is demanded [1,3,10,12]:

$$\min_{\Delta \mathbf{u}} \left\{ J_{\text{MPC}} = \sum_{i=1}^p (\bar{y}_k - y_{k+i|k})^2 + \sum_{i=0}^{s-1} \lambda_k \cdot (\Delta u_{k+i|k})^2 \right\}, \quad (24)$$

where \bar{y}_k is a set-point value, $\lambda_k \geq 0$ is a weighting coefficient. Thus, minimization of future control errors is demanded (the first part of the performance function) but at the same time, it is demanded that the control signal does not change too rapidly (the second part of the performance function). Assume that the predicted values of the output variable $y_{k+i|k}$ are derived using the method reviewed in Sect. 2.

The performance function (24) can be expressed as:

$$J_{\text{MPC}} = (\bar{\mathbf{y}} - \mathbf{y})^T \cdot (\bar{\mathbf{y}} - \mathbf{y}) + \Delta \mathbf{u}^T \cdot \mathbf{A}_k \cdot \Delta \mathbf{u}, \quad (25)$$

where $\bar{\mathbf{y}} = [\bar{y}_k, \dots, \bar{y}_k]$ is the vector of length p , $\mathbf{A}_k = \lambda_k \cdot \mathbf{I}$ is the $s \times s$ matrix. After application of prediction (7) to the performance function (25) one obtains:

$$J_{\text{HMPC}} = (\bar{\mathbf{y}} - \tilde{\mathbf{y}} - \mathbf{A}_k \cdot \Delta \mathbf{u})^T \cdot (\bar{\mathbf{y}} - \tilde{\mathbf{y}} - \mathbf{A}_k \cdot \Delta \mathbf{u}) + \Delta \mathbf{u}^T \cdot \mathbf{A}_k \cdot \Delta \mathbf{u}. \quad (26)$$

The performance function (26) depends quadratically on decision variables $\Delta \mathbf{u}$. Thus, if constraints are not taken into consideration, the vector minimizing the performance function (26) at each iteration is described by the following formula:

$$\Delta \mathbf{u} = \left(\mathbf{A}_k^T \cdot \mathbf{A}_k + \lambda_k \cdot \mathbf{I} \right)^{-1} \cdot \mathbf{A}_k^T \cdot (\bar{\mathbf{y}} - \tilde{\mathbf{y}}). \quad (27)$$

After assuming that $\lambda_k = dz_k^2 \cdot \lambda$ and using (8) one obtains (see [7] for details):

$$\Delta \mathbf{u} = \frac{1}{dz_k} \cdot \mathbf{K} \cdot (\bar{\mathbf{y}} - \tilde{\mathbf{y}}). \quad (28)$$

where the matrix $\mathbf{K} = (\mathbf{A}^T \cdot \mathbf{A} + \lambda \cdot \mathbf{I})^{-1} \cdot \mathbf{A}^T$ does not change and thus the main part of calculations can be performed offline.

In the analytical MPC algorithms the control constraints are taken into consideration by using a mechanism of control projection on constraint set; see e.g. [12] for details. For good performance of the controller exploiting the control projection mechanism it is essential to use the modified control values (those used to control the plant) in the subsequent iterations, during control signal calculation by the controller. Otherwise, control performance may be degraded. This remark is also valid in the case when the output constraint handling mechanism presented in Sect. 3 is used.

4.3 Results

The proposed method of output constraint handling was tested in the control system with analytical MPC controller based on the Hammerstein model described in Sect. 4.2. In comparison to the example described in [7], the controller was detuned in order to obtain overshooting needed to test the output constraint handling mechanism. The values of parameters of the controller were assumed as follows: prediction horizon $p = 7$, control horizon $s = 3$ and $\lambda = 6 \cdot 10^5$. In order to test influence of modeling uncertainty a Hammerstein model with polynomial static model was used to simulate the control plant. Despite that, the basic version of the proposed mechanism (without additional modeling uncertainty taken into consideration as in (17)) proved to be sufficient.

The example responses obtained in the control system of distillation column are shown in Fig. 4. The set-point value $\bar{y} = 395$ ppm was assumed. In the case when the mechanism of output constraint handling is not used in the controller the constraint is violated (dashed lines in Fig. 4). The situation changes after application of the proposed mechanism. If only one future predicted output value is constrained, situation improves much, though the constraint is slightly violated (gray lines in Fig. 4). In the case when two (solid lines in Fig. 4) or three (dotted lines in Fig. 4) future predicted output values are constrained, the constraint is fulfilled. The mechanism enforces modification of the control signal in advance what improves operation of the controller. In all three cases when the proposed mechanism was applied, the obtained responses are less oscillatory than the original ones.

5 Summary

The effective and easy to use mechanism of output constraint handling for analytical MPC controllers based on Hammerstein models is proposed in the paper. In the constraint handling mechanism of prediction of process behavior many sampling instants ahead can be taken into consideration. By virtue of this approach, the control signal can be modified in advance, before potential constraint violation. Thus, it is easier to ensure fulfillment of constraints and obtain good performance of the control system.

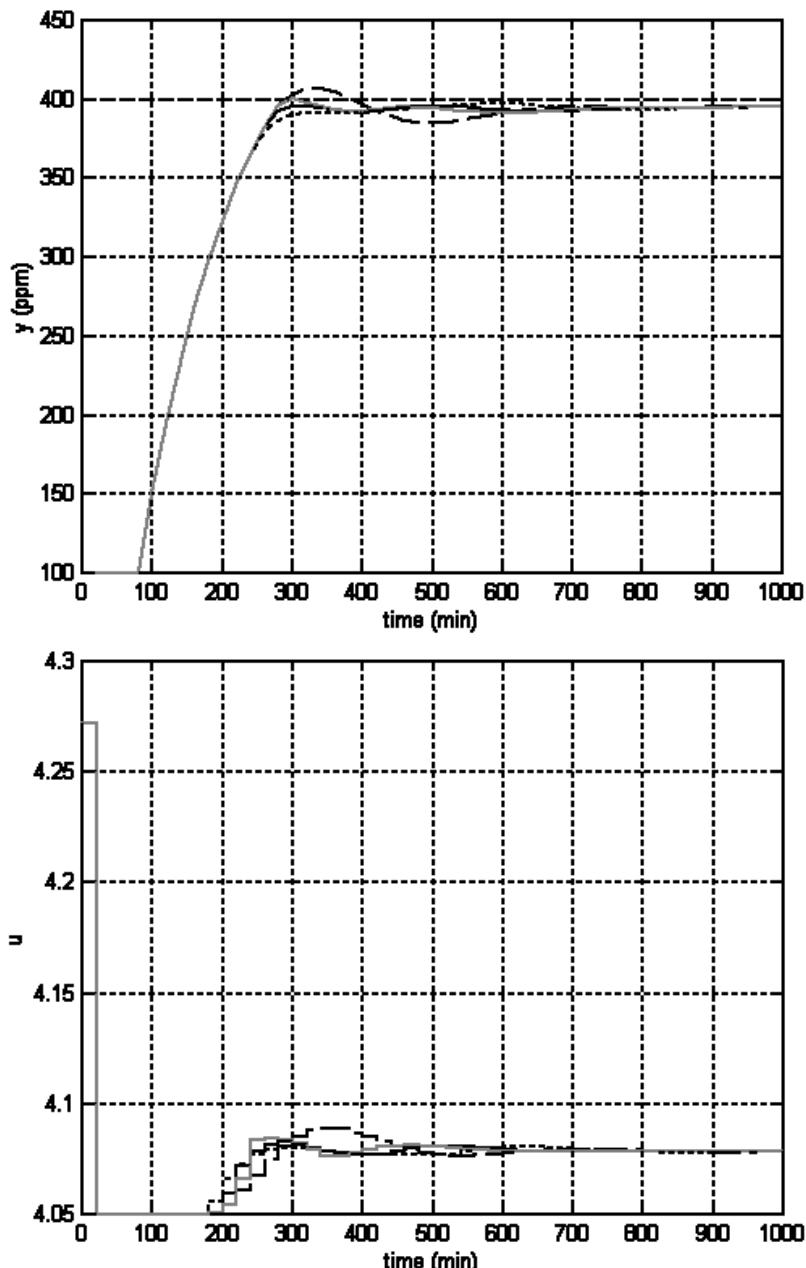


Fig. 4. Responses of the control system with analytical HMPC controller to the change of the set-point value to $\bar{y} = 395$ ppm; output constraints: not taken into consideration (dashed lines), taken into consideration in only one next sampling instant (gray lines), taken into consideration in next two sampling instants (solid lines), taken into consideration in next three sampling instants (dotted lines)

References

1. Camacho, E.F., Bordons, C.: *Model Predictive Control*. Springer (1999)
2. Janczak, A.: *Identification of nonlinear systems using neural networks and polynomial models: a block-oriented approach*. Springer, Heidelberg (2005)
3. Maciejowski, J.M.: *Predictive control with constraints*. Prentice Hall, Harlow (2002)
4. Marusak, P.: Analytical predictive controllers with efficient handling of output constraints. In: Malinowski, K., Rutkowski, L. (eds.) *Recent Advances in Control and Automation*, pp. 131–140. Academic Publishing House EXIT, Warszawa (2008)
5. Marusak, P.: On prediction generation in efficient MPC algorithms based on fuzzy Hammerstein models. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) *ICAISC 2010, Part I. LNCS (LNAI)*, vol. 6113, pp. 136–143. Springer, Heidelberg (2010)
6. Marusak, P.M.: Efficient predictive control algorithm based on fuzzy Hammerstein models: a case study. In: Gao, X.-Z., Gaspar-Cunha, A., Köppen, M., Schaefer, G., Wang, J. (eds.) *Soft Computing in Industrial Applications*. AISC, vol. 75, pp. 11–20. Springer, Heidelberg (2010)
7. Marusak, P.M.: Numerically efficient analytical MPC algorithm based on fuzzy Hammerstein models. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) *ICANNGA 2011, Part II. LNCS*, vol. 6594, pp. 177–185. Springer, Heidelberg (2011)
8. Marusak, P., Tatjewski, P.: Actuator fault tolerance in control systems with predictive constrained set-point optimizers. *International Journal of Applied Mathematics and Computer Science* 18, 539–551 (2008)
9. Piegat, A.: *Fuzzy Modeling and Control*. Physica-Verlag, Berlin (2001)
10. Rossiter, J.A.: *Model-Based Predictive Control*. CRC Press, Boca Raton (2003)
11. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. Systems, Man and Cybernetics* 15, 116–132 (1985)
12. Tatjewski, P.: *Advanced Control of Industrial Processes; Structures and Algorithms*. Springer, London (2007)

Piezoceramic Transformer Based Ionization-Deionization System

Piotr Mateusiak¹ and Roman Szewczyk²

¹ Industrial Research Institute for Automation and Measurements PIAP,
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
pmateusiak@piap.pl

² Warsaw University of Technology, Faculty of Mechatronics,
św. A. Boboli 8, 02-525 Warsaw, Poland

Abstract. This paper describes a principle of operation and construction of the miniature ionization-deionization system based on multilayer piezoelectric transformer. The device was designed for ultra-precise scales in order to decrease the influence of electrostatic charge. There was a significant increase in device efficiency and decrease in its volume and weight.

Keywords: transformer, ionization, deionization.

1 Introduction

Laboratory scales operate in a strictly controlled environment, which covers a vast number of different environmental parameters and among them the amount of electrostatic charge.

Unbalanced charge results in increase of the electrostatic force, which causes slow drift of shown mass value, significant dispersion in series of measurements and problems with returning to zero state after removing object from scale.

According to Coulomb's law the magnitude of the electrostatic force of interaction between two point charges is directly proportional to the scalar multiplication of the magnitudes of charges and inversely proportional to the square of the distance between them, accordingly to following equation:

$$F = k \cdot \frac{q_1 \cdot q_2}{r^2} \quad (1)$$

where:

F – magnitude of electrostatic force

k – Coulomb's constant

q_1, q_2 – signed magnitudes of charge

r – distance between the charges

The charge vary mostly depending on the kind of weighed material, air humidity, distance to other electronic devices and quality of anti-electrostatic protection chosen

by the user. The most popular way to eliminate or compensate unbalanced charge is to deionize the measured object and surroundings using high voltages.

2 Principle of Operation

2.1 High Voltage Ionization-Deionization

Ionization is conducted by generating positive aeroions (cations) and/or negative aeroions (anions). It is important that the ions produced are charged opposed to the charge we want to eliminate. Production of such ions results in the process of recombination between the particles in the air and on the measured object surface.

In the equilibrium conditions the pace of ionization is equal to the pace of recombination, which causes the formation of a constant level of ionization, equal to complete elimination of free charge.

2.2 Output Control

There are two basic methods of high voltage output control on the working electrode in order to ionize the area:

- direct current (DC),
- alternating current (AC).

DC method is executed by simply applying a high DC voltage to the output electrode and maintaining this state until the state of equilibrium is obtained.

AC method includes changing the polarity of high voltage applied to the output electrode. This is usually the matter of low and medium frequencies due to the time of recombination. In higher frequencies recombination occurs mostly between generated ions instead of harmful charge. What is more we can use sinusoidal as well as square wave.

In general the AC method is more effective due to the fact, that it allows almost complete neutralization of the target area instead of just polarizing it with negative or positive charge like the DC method. Due to generation alternating between positive and negative potential at an appropriate pace the produced ions quickly overwhelm unbalanced charge in target area, which leads to the stable equilibrium state.

3 Hardware

3.1 Requirements

There were three basic requirements for the hardware for deionization in weighing scales – possibility of both AC and DC type ionization, miniature size and high output power.

In order to fulfill all those requirements there was a need to apply new kind of high voltage generation method. It was necessary to find or build a device to vastly amplify

input voltage while keeping miniature size. The multilayer piezoelectric transformer seems to be most suitable for this application.

3.2 Simplified Structure of the Device

In general the deionization system consists of supply source, two high voltage branches (positive and negative), automatically controlled switch and output electrode, as it is presented in figure 1.

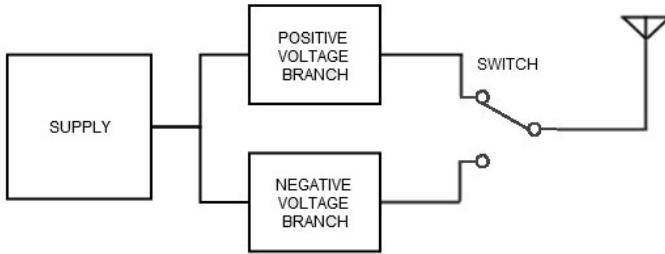


Fig. 1. Block diagram of the ionization-deionization system

This kind of structure allows performing ionization using AC or DC method according to the user needs.

The method of ionization is completely determined by the switch working mode. The switch is bistable and may be externally controlled by timer circuit or microcontroller and additionally by buttons available for users (in order to choose the output control method).

3.3 High Voltage Generation

Schematic block diagram of a single high voltage generation branch is presented in figure 2. The whole system contains two branches in total, the positive and the negative one.

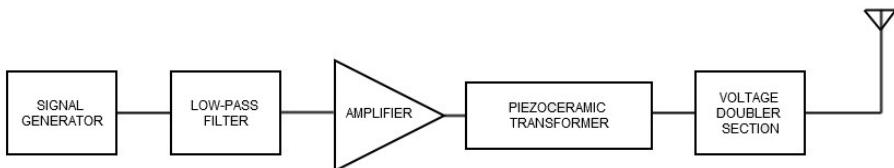


Fig. 2. Block diagram of a single branch of the high voltage generator

The core of the high voltage generation section is the piezoelectric transformer. Its equivalent circuit is presented in figure 3. This element is very sensitive to working conditions and requires precisely set input signal frequency, while the signal should

be sinusoidal to cause the least energy loss. To ensure that transformer's operating point is suitable, there is a necessity of impedance matching between transformer and generated signal supply.

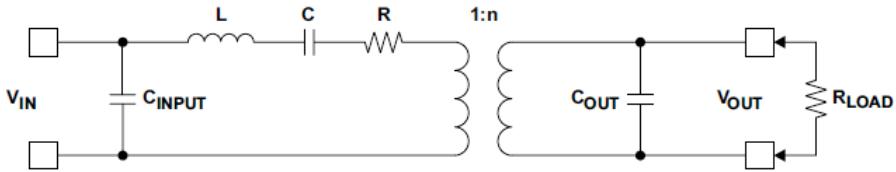


Fig. 3. Equivalent piezoelectric transformer circuit model [4]

In order to minimize transformer temperature rise there is an analog low-distortion sinusoidal signal source with a wide frequency sweep range used. The need of variable frequency signal supply is justified by the gain dependence on both frequency and load magnitude; the characteristic is presented in figure 4.

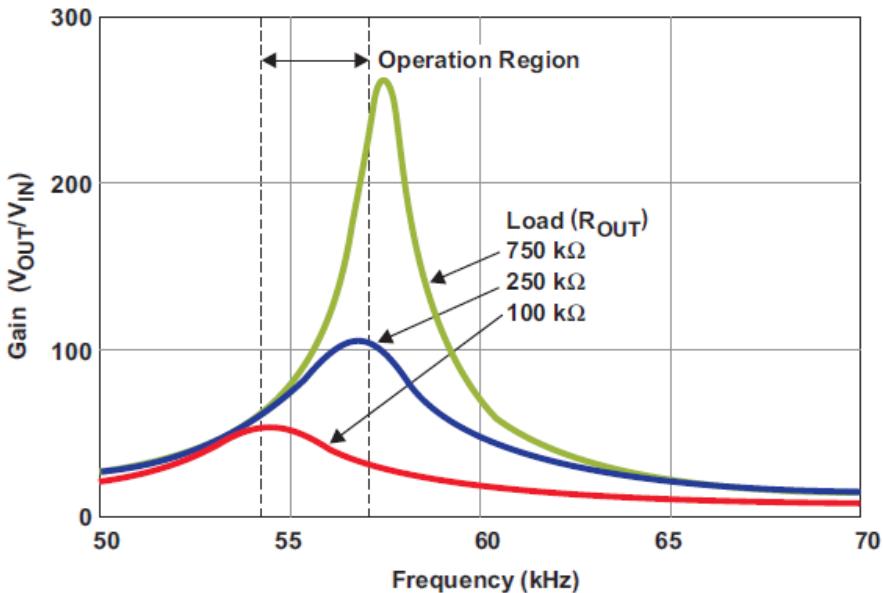


Fig. 4. Typical piezoelectric gain characteristics vs. frequency and load [4]

The use of variable frequency generator and the additional low-pass filter with cutoff frequency slightly above piezoceramic transformer resonant frequency ensures high quality of signal fed to amplifier and further to the transformer. The important issue is also signal generation section and, as mentioned before, transformer impedance matching. The input impedance value of the piezoelectric transformer was measured and taken into account, while designing the power amplifier.

Since the transformer output voltage magnitude is not big enough and its resonant frequency is too high to use for ionization (50-70 kHz depending on the output electrode load), there is a multistage voltage doubler circuit in the Greinacher topology, shown in figure 5, used so that the output voltage is multiplied and at the same time rectified. Rectification of the output signal makes both AC and DC deionization methods available.

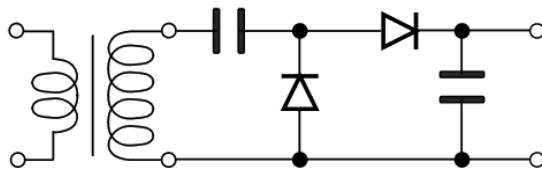


Fig. 5. Greinacher voltage doubler circuit

DC voltage on the output enables easy switching between negative and positive voltage with any frequency the user wants. As an automatically controlled switch there are two flip-flops and timer circuit used. There was also an alternative circuit with microcontroller driven relay instead of flip-flops tested. Both circuits are equally well suited for this application.

3.4 Output Electrode

The output electrode may come in different shapes, sizes and materials. The ionization-deionization system was tested with point, edge and string electrodes. There are two most important features that make an effective electrode. Firstly electrode should be characterized by appropriate load to be suitable for high voltage source and secondly, the shape of the electrode is critical too, because it determines homogeneity and gradient of generated electrostatic field.

The string electrode was chosen as it can be easily installed inside high precision scale's weighing space. Due to its flatness, placing such electrode on one of the walls of the scale wouldn't cause any noticeable change in the device structure and there won't be a decrease in the weighing box volume.

4 Developed Solution

The prototype has been built based on described design. The construction was executed on two layer PCB plus voltage doubler ladder. The PCB is 8 cm × 8 cm and weighs about 100 g.

The sinusoidal signal with variable frequency is generated by analog integrated circuit XR-2206, which enables precise frequency, amplitude, sine-shaping and

symmetry adjustment. After generation the signal is conditioned and amplified by power amplifier LM3886 and fed directly to piezoceramic transformer. Whole circuit is supplied by dual output DC power supply set in series mode.

The generation PCB is connected directly to a voltage doubler ladder and an output electrode. This forms the complete ionization-deionization system.

5 Conclusion

Multilayer piezoceramic transformer based ionization-deionization system presented in this paper enables both ionizing the measured object and surrounding air with positive or negative polarity according to the user needs, as well as deionizing the object to minimize any deviation from the state of charge equilibrium. The method of ionization may also be adjusted to the user needs.

It should be indicated that the system proved to eliminate almost completely the unbalanced charge and therefore removed any noticeable influences of electrostatic charge on the result of precise laboratory scales measurements mentioned earlier.

Presented device was inspired by the systems available for high precision balances. As the result of the research there are significant improvements easily noticeable in comparison to the traditional inductive transformer, while maintaining the same or better ionization performance. While the power of the device was increased 100 percent versus the previously mentioned old system, there was also a decrease in both weight and volume. The comparison is presented in table 1. Those improvements enable placing the ionization-deionization system inside the scale as an optional module instead of offering it as a completely separate device. As it can be also seen in table 1, every single requirement mentioned before was fulfilled.

Table 1. Comparison between piezoceramic transformer and traditional inductive transformer deionization systems.

Physical quantity	Change
Power	Increase 100%
Weight	Decrease 80%
Volume	Decrease 70%

Acknowledgment. This work was partially supported by The National Centre of Research and Development (Poland).

Authors would like to express their gratitude to RADWAG (www.radwag.pl) for kind support during experiments with electrostatics influence on the weighing process as well as sharing the knowledge of ionizing techniques.

References

1. Rosen, C.A.: Ceramic Transformers and Filters. In: Proceedings of the Electronic Components Symposium (1956)
2. Lin, C.-Y.: Design and Analysis of piezoelectric Transformer Converters. Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University (1997)
3. Lin, R.-L.: Piezoelectric Transformer Characterization and Application of Electronic Ballast. Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University (2001)
4. Day, M., Lee, B.S.: Understanding piezoelectric transformers in CCFL backlight applications. Texas Instruments Incorporated (2005)
5. Borucki, S.: Diagnostic measurements of power transformers using the vibroacoustic method. Pomiary Automatyka Robotyka 14(12), 110–114 (2010) (in Polish)

The Rapid Prototyping of Active Magnetic Bearings

Paulina Mazurek and Maciej Henzel

The Faculty of Mechatronics and Aerospace, Military University of Technology,
Warsaw, Poland

{paulina.mazurek,maciej.henzel}@wat.edu.pl

Abstract. The paper deals with rapid prototyping of active magnetic bearings. Modern trends in aviation and possibilities of magnetic suspension using are characterized. Geometric analysis of active axial magnetic bearing are also described. Designing input parameters are defined and calculation results are presented. MES model and MES analysis of designed magnetic bearing are also shown. Active magnetic bearing control system with PD controller are presented as well as its time characteristics.

Keywords: magnetic suspension, axial magnetic bearing, MEA concept, aircraft engine.

1 Introduction

Modern aircrafts onboard systems are design according to More Electric Aircraft (MEA) concept. The concept makes assumption that pneumatic, hydraulic and mechanic systems will be replaced by their electric equivalent. Environmental control systems, electrical power systems, aircraft engines, anti-icing systems, hydraulic systems, flight control and actuation systems were designed in compliance with MEA technology [2].

MEA concept development also give opportunities of active magnetic bearing (AMB)¹ utilization in aircraft on-board systems. That technology eliminates friction forces between co-operating kinematic pairs and make possible to monitoring and detecting airborne systems. Magnetic suspension can perform wide range of function in technical systems. They could work as shaft support of aircraft engine, electric drives, starter/generator, gyroscopes, etc. These new systems can work in rugged conditions, e.g. high and low temperature, low pressure, vacuum [3].

Active magnetic bearings can be used in suspension system of aircraft engine turbine shaft. That shaft can do linear movement in three directions. So bearing system must consist of one radial and two homopolar axial magnetic bearing at least.

Designing of axial active magnetic bearing for aircraft engine shaft is subject of this paper. Designing algorithm was presented in Fig. 1.

¹ AMB are devices, which use attraction and repulsion forces to provide stable rotor levitation at work point.

First step of presented designing process of axial magnetic bearing is definition of designing input parameters (electromagnetic force, suspended rotor diameter and weight, etc.). The next step is preliminary calculations of bearing physical parameters in MATLAB software. Thereafter MATLAB code should be implemented in LabVIEW application. The application generates dedicated files with physical parameters of bearing for Comsol Multiphysics software and Inventor software as well. Presented algorithm also includes bearing dynamic analysis.

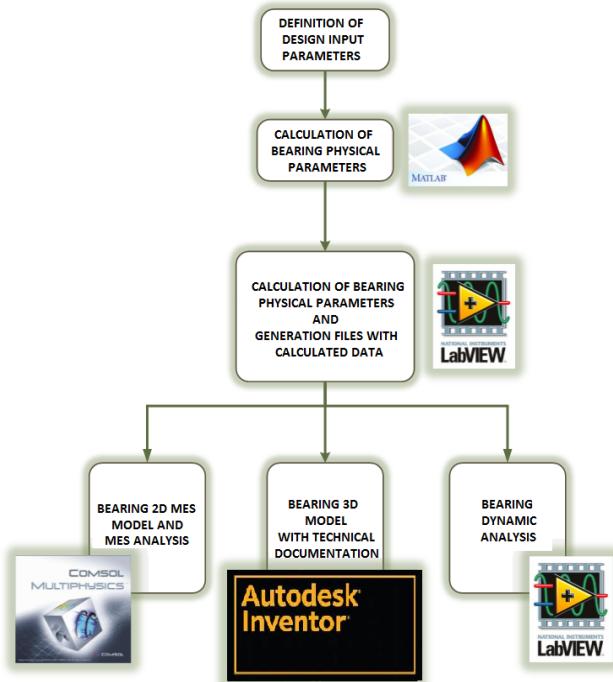


Fig. 1. Design algorithm

2 Bearing Geometric Analysis

Difference electromechanical actuator of axial magnetic bearing consists of two electromagnets. Rings made of soft-magnetic material is located between electromagnets. It is made of two disk separated by air gap. That configuration minimize coupling between electromagnets. In Fig. 2 axial bearing scheme was presented.

Estimating of actuator maximal electromagnetic force is first step of magnetic bearing designing process. That force acts parallel to shaft axis and counteract its axial movement. Electromagnetic force value determined bearing physical characteristics. Designing input parameters are shown in Table 1 [2, 3].

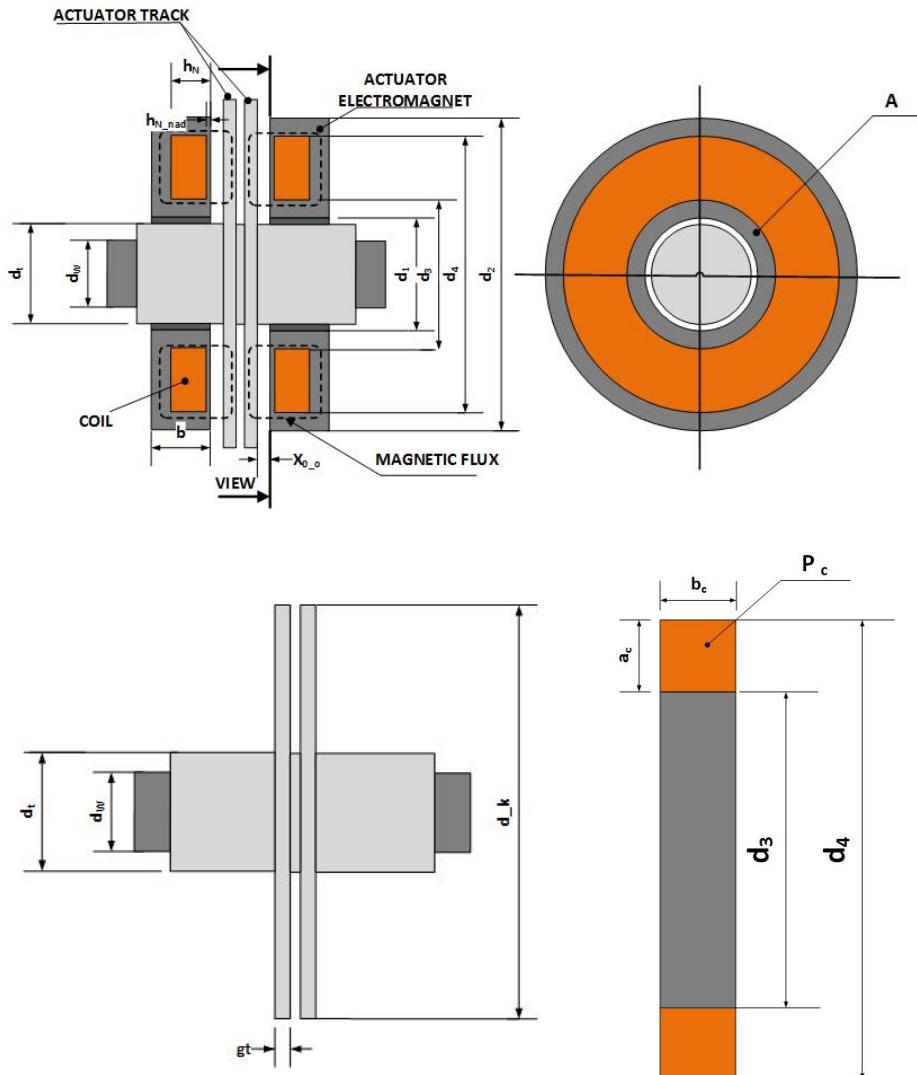
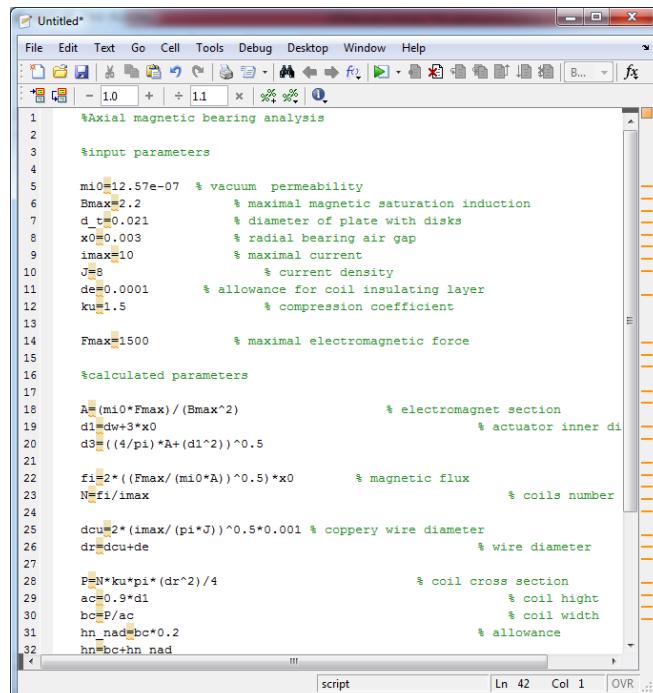


Fig. 2. Axial bearing scheme

Bearing physical characteristics were calculated by application wrote in MATLAB software (Fig. 3). Then MATLAB code was implemented in LabVIEW (Fig. 4). LabVIEW application consists of user interface (Front Panel) and program code (Block Diagram). Front panel window (Fig. 4a) contains interactive input (controls) and output terminals (indicators). Controls are used to enter input designing parameters and indicators are used to show estimated bearing parameters. Block diagram window (Fig. 4b) contains code graphical representation.

Table 1. Designing input parameters

Parameter	Icon	Value	Unit
Vacuum permeability	μ_0	$12.57 \cdot 10^{-7}$	(Vs)/(Am)
Maximal electromagnetic force	F_{max}	1500	N
Maximal magnetic saturation induction	B_{max}	1.6	T
Rotor diameter	d_w	15	mm
Diameter of sleeve with disks	d_t	21	mm
Radial bearing air gap	x_0	0.3	mm
Maximal current	i_{max}	5	A
		10	
Current density	J	8	A/mm ²
Allowance for coil insulating layer	d_e	0.1	mm
Compression winding coefficient	k_u	1.5	
		1.7	
Axial bearing air gap	x_{0_o}	0.5	mm
Shaft weight	m	1.27	kg



```

Untitled*
File Edit Text Cell Tools Desktop Window Help
Untitled.m - 1.0 + 11 × ⓘ fx
1 %Axial magnetic bearing analysis
2
3 %input parameters
4
5 mu=12.57e-07 % vacuum permeability
6 Bmax=2.2 % maximal magnetic saturation induction
7 dt=0.021 % diameter of plate with disks
8 x0=0.003 % radial bearing air gap
9 imax=10 % maximal current
10 J=8 % current density
11 de=0.0001 % allowance for coil insulating layer
12 ku=1.5 % compression coefficient
13
14 Fmax=1500 % maximal electromagnetic force
15
16 %calculated parameters
17
18 A=(mu0*Fmax) / (Bmax^2) % electromagnet section
19 d1=dw+3*x0 % actuator inner dia
20 d3=((4*pi)*A+(d1^2))^0.5
21
22 fi=2*((Fmax/(mu0*A))^0.5)*x0 % magnetic flux
23 N=fi/imax % coils number
24
25 dcu=2*(imax/(pi*J))^0.5*0.001 % coppery wire diameter
26 dr=dcu+de % wire diameter
27
28 E=N*ku*pi*(dr^2)/4 % coil cross section
29 ac=0.9*dr % coil height
30 bc=P/ac % coil width
31 hn_nad=bc*0.2 % allowance
32 hn=bc+hn_nad

```

Fig. 3. Calculation of bearing physical parameters in MATLAB software

Into two different electronic spreadsheet was saved evaluated data. Spreadsheets were formatted for Comsol Multiphysics (Fig. 5a) and Inventor (Fig. 5b).

In Fig. 6 axial magnetic bearing 3D model was presented (Inventor). Bearing physical parameters were loaded from dedicated spreadsheet.

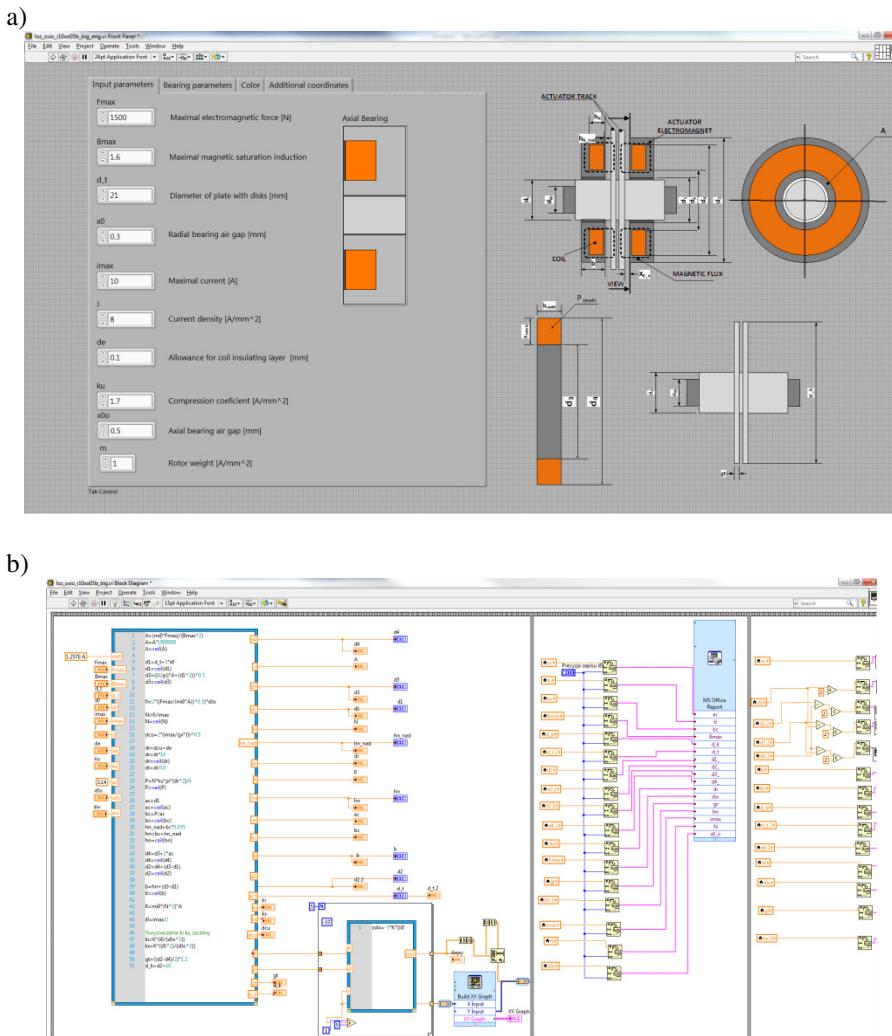


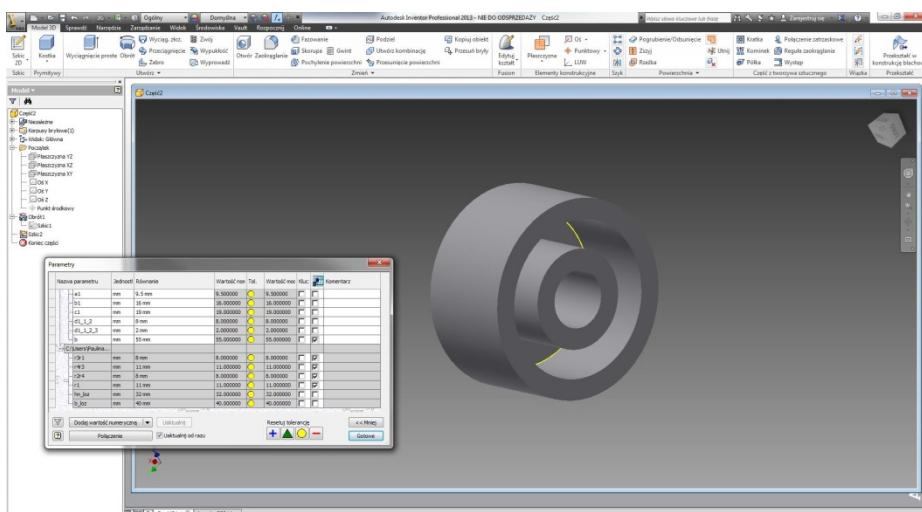
Fig. 4. Application in LabVIEW, a) front panel, b) block diagram

a)

	A	B	C	D
1	dw	15 [mm]	Shaft diameter	
2	d1	22 [mm]	Bearing inner diameter	
3	d3	38 [mm]	Coil inner diameter	
4	d4	82 [mm]	Coil external diameter	
5	d2	98 [mm]	Bearing external diameter	
6	hn	18 [mm]	Electromagnet hight	
7	b	34 [mm]	Bearing width	
8	ac	22 [mm]	Coil hight	
9	bc	16 [mm]	Coil width	
10	N	128	Coils number	
11	Bmax	1.6 [T]	Maximal saturation induction	
12	imax	10 [A]	Maximal current	
13	d_t	21 [mm]	Plate diameter	
14	d_k	108 [mm]	Disk diameter	
15	dr	1.4 [mm]	Wire diamater	
16	gt	9.6 [mm]	Disk width	
17	x0_o	0.5 [mm]	Axial bearing air gap	

b)

	A	B	C
1	r3r1	8.000000mm	
2	r4r3	22.000000mm	
3	r2r4	8.000000mm	
4	r1	11.000000mm	
5	hn_loz	18.000000mm	
6	b_loz	34.000000mm	
7	d_t	21.000000mm	
8	dw	15.000000mm	
9	d_k	108.000000mm	
10	dr	1.400000mm	
11	ac1	22.000000mm	
12	bc	16.000000mm	
13	gt	9.600000mm	
14	d44	82.000000mm	

Fig. 5. Spreadsheet, a) data for Comsol Multiphysics, b) data for Inventor**Fig. 6.** Bearing 3D model in Inventor

3 MES Analysis

MES model of axial magnetic bearing was made in Comsol Multiphysics. Because the geometry has symmetries the studies could be made for 2D axisymmetric space dimension. Bearing physical parameters (Fig. 7) was loaded from spreadsheet (generated in LabVIEW application). As shown in Fig. 8a bearing geometry consist of cross-section of electromagnets and rings as well as coils. In Fig. 8b was presented bearing triangle mesh was presented.

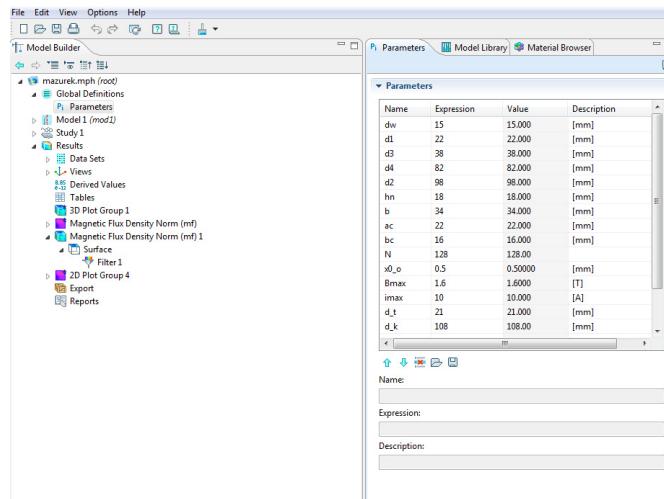


Fig. 7. Parameters definition in Comsol Multiphysics

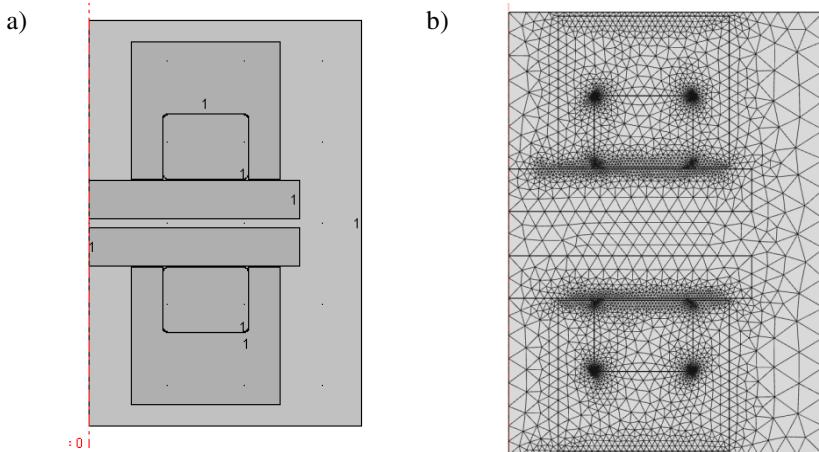


Fig. 8. 2D axisymmetric model of bearing, a) geometry, b) triangle mesh in Comsol Multiphysics software

In Fig. 9 and Fig. 10 (3D) magnetic flux density norm was presented. Maximal and minimal value of magnetic flux density norm was also marked (Fig. 10). The problem was solved with Magnetic Fields interface (mf). Each coils have 128 turns and carries current of 1 A.

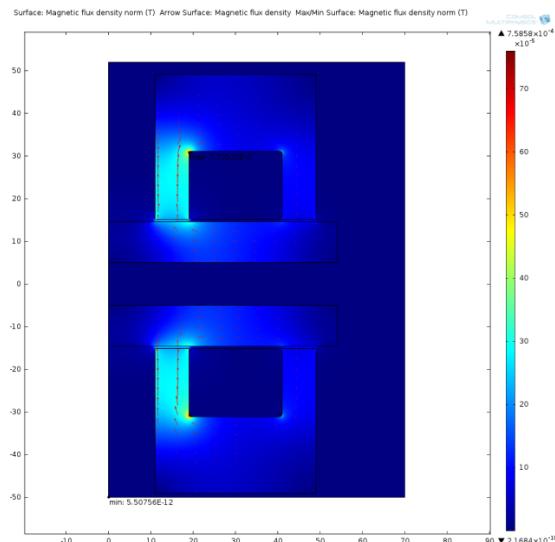


Fig. 9. Magnetic flux density norm for current of 1 A

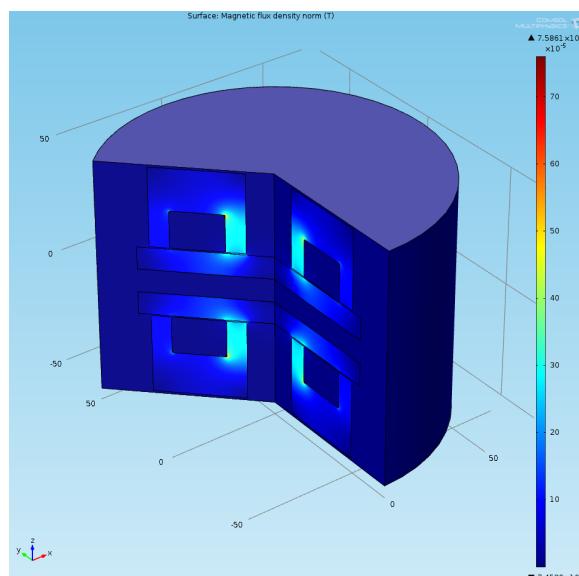


Fig. 10. Magnetic flux density norm 3D for current of 1 A

4 Bearing Dynamic Analysis

Transfer function of homopolar active magnetic bearing is presented in equation (1).

$$Z(s) = \frac{2k_i}{ms^2 - 2k_x} i(s) + \frac{1}{ms^2 - 2k_x} F_z(s) \quad (1)$$

where:

k_i – current stiffness;

k_x – displacement stiffness;

m – shaft weight;

$i(s)$ – control current;

$F_z(s)$ – disturbing force;

$Z(s)$ – shaft position.

Active magnetic bearing is an unstable structure. Therefore, it requires suitable feedback control system with proper controller. In Fig. 11 active magnetic suspension model with PD controller was presented. Transfer function of PD controller was shown in equation (2) [1].

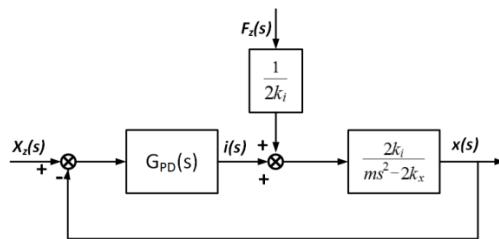


Fig. 11. Active magnetic suspension model with PD controller

$$G_{PD}(s) = K_p + K_d s \quad (2)$$

where:

K_p – proportional gain;

K_d – derivative gain.

PD controller parameters (K_p , K_d) were calculated by LabVIEW application (Fig. 12). It was based on equations (1)–(2) and pole-placement method. Input parameters were assumed setting time and damping coefficient (also current stiffness k_i , displacement stiffness k_x and shaft weight m). Application has determined proportional gain K_p , derivative gain K_d and generated step response characteristic of closed loop system (Fig. 13).

```

m          22 %EQUATION POLES
DBL[m]    23
i1        24
ki        25 w_n=(3.2)/(dzeta*tr)
i2        26 a=-w_n*dzeta
lo        27 b=(1-(dzeta^2))^0.5
tr        28 i=(-1)^0.5
dzeta     29 p1=a-(i*(w_n*b));
           30 p2=a+(i*(w_n*b))
M          31 %PD PARAMETERS
DBL[M]    32 Kd0=((-p1+p2)*m)/(ki)
           33 Kp0=(((p1*p2*m)+lo)/(ki))
           34
           35 %Kd1=29.2456
           36 %Kp1=33272
           37 %TRANSFER FUNCTION OF CLOSED LOOP SYSTEM
           38
           39 L=[ki*Kd0 Kp0*ki]
           40 M=[m Kd0*ki (Kp0*ki-lo)]
           41 sys=tf(L,M)
           42 %STEP RESPONSE
           43
           44 figure(1)
           45 [y,t]=step(sys);
           46 %A=ones(1,56);
           47 y1=0.0001*y;
           48 %sim(sys,y1,t)
           49 plot(t,y1,'linewidth',3)
           50 Y = axes;
           51 set(Y, 'YLim', [0,0.00015], 'XLim', [0, 0.0125])
           52 grid on;
           53 hold on
           54 y2=0.0001
           55 plot(t,y2,'k','linewidth',3)
           56 legend('RESPONSE','EXTRACTION')
           57 xlabel('TIME [s]')
           58 ylabel('AMPLITUDE [m]')
           59 figure(2)
           60 pzmap(sys)
           61 grid on
           62
           63
           64
           65

```

Fig. 12. PD controller parameters calculating in LabVIEW

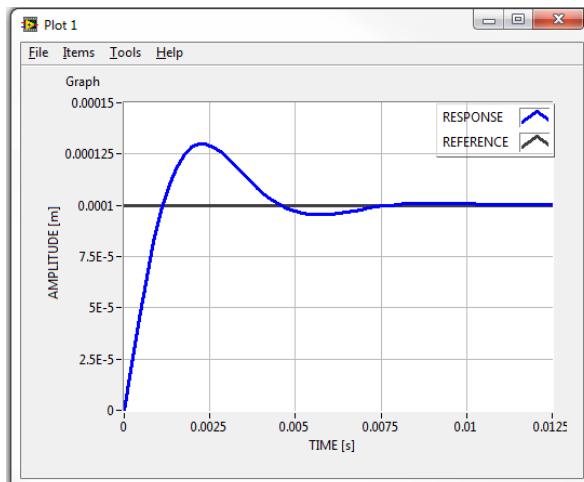


Fig. 13. Step response characteristic of closed loop system with PD controller

5 Summary

In the paper, axial magnetic bearing design for aircraft engine was shown. There was presented the rapid prototyping of axial magnetic bearing with used of LabVIEW application. Input designing parameters were presented in Table 1. Their values should be estimated according to bearing destination. Application results were saved into files. First was loaded to Inventor and second was loaded to Comsol Multiphysics. In that way, 3D geometry model and MES model of bearing were built. Additionally, dynamic analysis of magnetic bearing was effectuated. There was presented controller designed and step response characteristic of closed loop with PD controller.

Presented applications can be used to design and analyze different axial magnetic bearing. It can be also used to build applications for radial magnetic bearing analysis.

Acknowledgements. The work was financed in part from the government support of scientific research for years 2012–2015, under grant PBS1/B6/7/2012.

References

1. Falkowski, K., Gosiewski, Z.: Multifunctional magnetic bearings. Institute of Aviation Scientific Library, Warsaw (2003)
2. Falkowski, K., Gosiewski, Z.: The differential passive magnetic bearing for high-speed flexible rotor. Solid State Phenomena 144, 273–278 (2009)
3. Falkowski, K., Matuszewski, L.: Mathematical model of radial passive magnetic bearing. Polish Maritime Research 17(3), 37–44 (2010)
4. Mazurek, P.: Review of bearingless constructions for aviation applications. In: Advances in Chemical and Mechanical Engineering. Faculty of Mechanical Engineering, Gdansk University of Technology, vol. II/II, pp. 978–983 (2012) ISBN 978-83-88579-62-2

Application of the MFC Method in Electrohydraulic Servo Drive with a Valve Controlled by Synchronous Motor

Andrzej Milecki, Dominik Rybarczyk, and Piotr Owczarek

Institute of Mechanical Technology, Poznan University of Technology,
ul. Piotrowo 3, 60-965 Poznań, Poland

{Andrzej.milecki, Dominik.rybarczyk, Piotr.owczarek}@put.poznan.pl

Abstract. The article describes the use of the Model Following Control method to control of electrohydraulic servo drive. In the drive a new, designed by us, proportional valve with a synchronous motor controlled by dedicated power electronics is used. The model of the electrohydraulic servo drive controlled by MFC method prepared in MATLAB-Simulink is described. The laboratory test stand is described and step responses of the drive are shown.

Keywords: electrohydraulic servo drive, proportional valve, model following control, synchronous motor.

1 Introduction

Electrohydraulic servo drives are highly non-linear devices. Nevertheless, as their controller classic one-loop PID controllers are most frequently used. They work well in situations, in which no large changes of the drive parameters occur. Therefore, it is difficult to design a suitable PID control system of electrohydraulic drives when their parameters changes in time. For this reason, it is necessary to look for new control methods and solutions, which are not susceptible to big changes of drive parameters or are able to adapt to such changes. One of the technique recently used in control of non-linear system is the Model Following Control (MFC). This method allows to ensure high quality control and the resistance to external interference in quite easy way [5, 6], therefore it is worth to test the use of this method in electrohydraulics.

In the article the Authors describe the application of the MFC method for control of the electrohydraulic servo drive with proportional valve controlled by modern low power synchronous motor.

2 Model Following Control Method

Over the last 20 years many attempts have been taken to use the Model Following Control in different devices and processes [1, 3–5, 9, 19]. In the paper [13] a research results of the use of MFC technology for testing of the pump control with synchronous motor. The MFC method is a two-loop structure, which is noted for its

simplicity and relatively high robustness to disturbances and stable perturbations. In design of MFC system the object is mostly represented by transfer function. In some cases also the state variable approach has also been employed. However the state variable system representation seems to be improper in order to clarify basic universal properties of the MFC structure.

There is difficult to find in the literature papers, which deal with the application of MFC method in control of electrohydraulic servo drives. In the paper [2] a special model-following control scheme is developed in order to force a linear system behavior. The linearization is achieved by feed forward compensation and proportional feedback. The study [11] was focused on the compensation of the impact of the load changes in the electrohydraulic drive. There are no publications which described the use of MFC method in electrohydraulic servo drive with proportional valve controlled by synchronous motor.

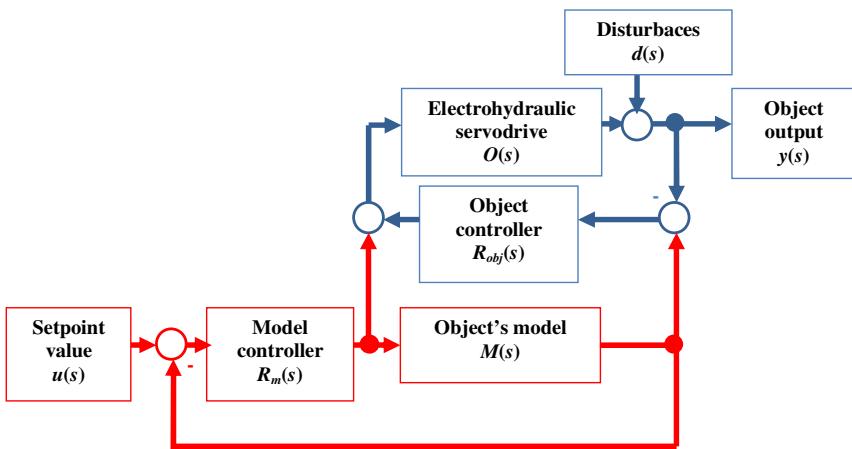


Fig. 1. Model Following Control structure

The principle of MFC method is based on the idea of Model Based Control (MBC). It means, that a part of the control system structure is a mathematical model of the object. Model Following Control structure bases on two regulators: the regulator of the model (main regulator) and the regulator of the object (correcting regulator). So, the whole structure consists of two loops with feedback: main object loop and model loop. The mathematical model of the object is identified by a mathematical analysis or by experimental test. The first loop consists of the model of a controlled object and the PID controller. The difference between output of the model and the output of the real object and output signal from the PID regulator used in model loop is used for the control signal correction. In both loops the same PID regulator is applied. If the model perfectly fits the real object, the system works like a classical one-loop PID controller (signals from correcting loop is equal to zero). Due to the principle of operation of the described system, on the input of correcting regulator (in contrast to the control model), a step function signal will be never occur.

This allows to tuning the correcting regulator with much larger reserves of stability (for example: with a higher gain coefficient). This allows to make more effective correction of disturbances. In hydraulic drive systems, the disturbances can be: load, mass, stiffness, temperature and supply pressure changes [2]. As usual, also friction may very during movement.

3 Electrohydraulic Servo Drive with Synchronous Motor

3.1 Proportional Valve with Synchronous Motor

Due to the development of modern control methods and production technology both, the properties and the prices of modern synchronous motors significantly dropped. Therefore it has become possible to use them in applications where even 15 years ago this would be economically unprofitable. Modern synchronous motors ensure high positioning accuracy with high dynamics. In the proposed here valve (see Fig. 2), the spool is actuated by a low-power Permanent Magnet Synchronous Motor (PMSM). The motor (1) was connected to the spool (3) by flexible coupling bellows (2). Applying the electrical power to the motor causes rotation and simultaneously axial translation of the spool. Control edge openings x are proportional to the angular motor position and to the pitch of used thread (5). Direction of rotation determines direction of spool translation and opening or closing of valve gaps.

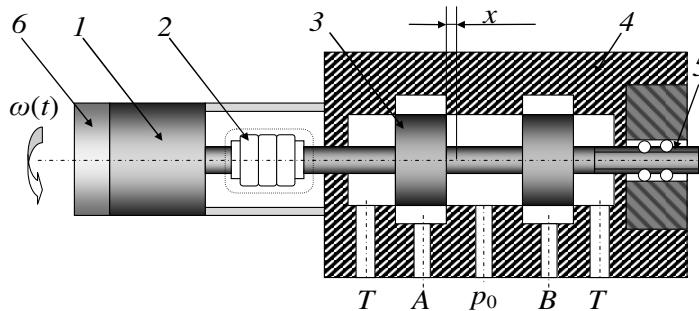


Fig. 2. Proportional valve with synchronous motor

In the valve, the low power PMSM motor type B&R 8LVA23 was applied. Its basic parameters are: rated speed 3000 rev./min, rated current 2.9 A and stall torque 0.680 Nm. The engine was equipped with an absolute encoder type EnDat, providing a continuous information about the current position, even after a power failure, which assures a very high positioning accuracy and resolution of 262144 pulses on one revolution.

3.2 Modelling of the Servo Drive Controlled by MFC Method

In order to test, assure proper design and to implement the MFC method it was necessary to build at first a model of the electrohydraulic servo drive based on the described above valve. Model of servo drive was constructed based on the mathematical description of the hydraulic drive given in:

$$Q_a(t) = Q_{sa}(t) + Q_{ha}(t) + Q_v(t) \quad (1)$$

$$Q_b(t) = Q_{sb}(t) + Q_{hb}(t) - Q_v(t) - Q_{vb}(t) \quad (2)$$

$$Q_b(t) = K_{qp} x(t) - K_l p_b(t) \quad (3)$$

$$Q_{ha}(t) = A \frac{dy(t)}{dt} \quad (4)$$

$$Q_{hb}(t) = aA \frac{dy(t)}{dt} \quad (5)$$

$$Q_{sa}(t) = \frac{V_a}{E_o} \frac{dp_a(t)}{dt} \quad (6)$$

$$Q_{sb}(t) = -\frac{V_b}{E_o} \frac{dp_b(t)}{dt} \quad (7)$$

$$Q_v(t) = K_v [p_a(t) - p_b(t)] \quad (8)$$

$$Q_{vb}(t) = K_{vb} p_b(t) \quad (9)$$

$$m \frac{d^2 y(t)}{dt^2} + D \frac{dy(t)}{dt} = A [p_a(t) - ap_b(t)] \quad (10)$$

where: Q_a, Q_b – flow, Q_{ha}, Q_{hb} – absorption of the actuator chambers, Q_{sa}, Q_{sb} – flow of the covering losses due to compressibility, Q_{vb} – leakage flow on the piston rod, p_a, p_b – the pressure in the chambers of the actuator, A_a, A_b – active surfaces of the piston, V_a, V_b – the volume of liquid in the chambers of the actuator.

Unknown parameters such as capacity of the wires and the flow rate has been identified with used of the Kalman filter. The model was implemented in the model loop of the MFC controller. Control system implemented on PLC controller enable to switch between PID and MFC controller.

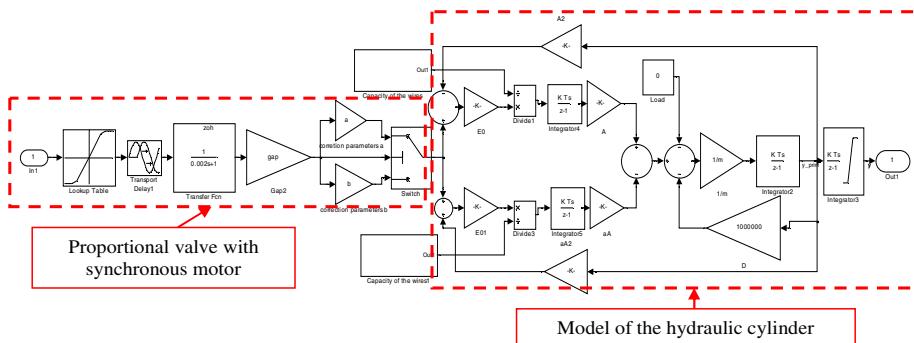


Fig. 3. Model of the electrohydraulic servo drive used in MFC controller (the model was built based on the mathematical description in section 3.2 of the article)

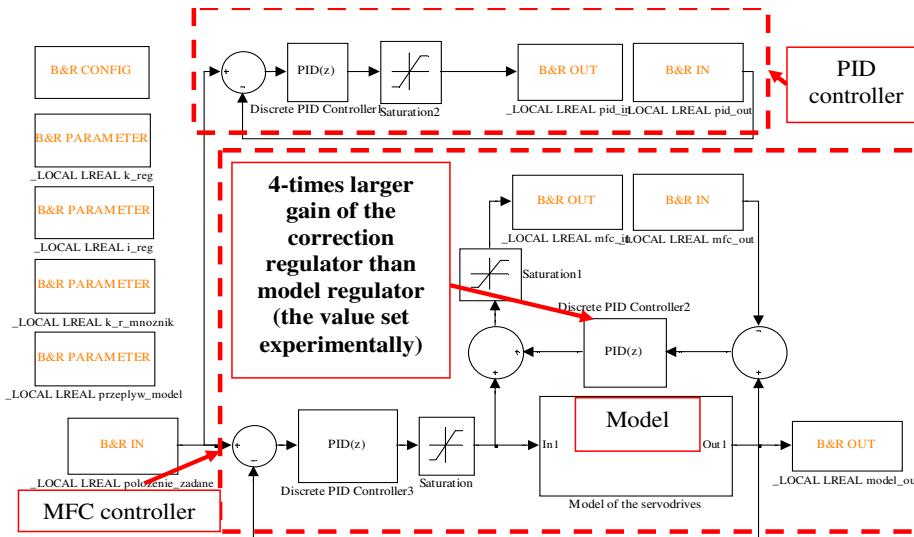


Fig. 4. Control system implemented on PLC

4 Experimental Test Stand and Investigations

In Fig. 5 the scheme of the experimental stand for testing of the electrohydraulic servo drive based on proportional valve driven by synchronous motor, is shown. The stand enables to check up the effectiveness of the control system based on the MFC method. The test stand consists of a single-acting cylinder with a proportional valve controlled by synchronous motor and control system based on PLC technology. The piston of the cylinder was coupled to the position sensor. A control system was based on master

controller which was PLC with touch panel and servo-inverter which acts as a slave controller. As a communication interface between the PLC and the servo-inverter, a Powerlink network was used. The PLC was running under the control of real-time operating system. The complete application enabled real-time control and the system was working with sampling frequency of 1250 Hz.

In the described system, Authors used the rapid prototyping technique for control system design and implementation. This means that the model of the control system has been built previously in software environment and recompiled for the C code and then implemented it as one of the task class in the CPU unit. Due to the discrete work of the industrial controller, it was required to discretize the continuous model of control system. The discrete time base was $T = 0.0008$ s.

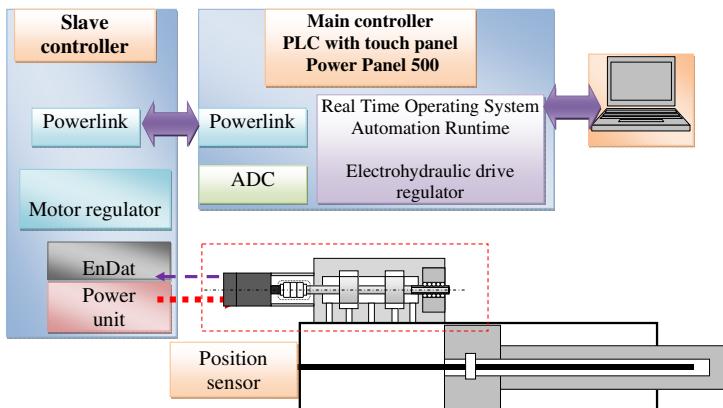


Fig. 5. Control system scheme

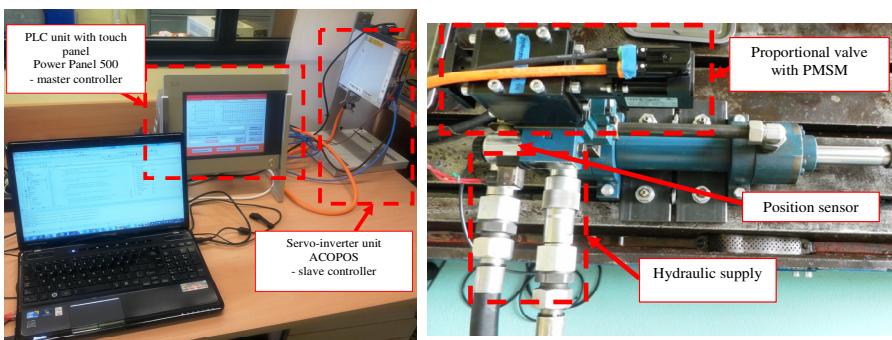


Fig. 6. Electrohydraulic servo drive and control stand view

5 Experimental Investigation

During the investigations on the system input a step signal was given. The experiment results are shown in the charts below (Fig. 8). The study was performed for two types

of regulator: PID controller and MFC controller. The both regulator's gain coefficients, derivative and integrative time constants were established experimentally and were as matched to achieve the step response as fast as possible, but without oscillation. In the next step the hydraulic supply pressure was changed from 8 MPa to 2 MPa. The PID and MFC coefficients remain unchanged. The investigation results shown that recorded step response of the servo drive controlled by MFC was significantly better in comparison to the step response obtained when the drive was controlled by PID controller.

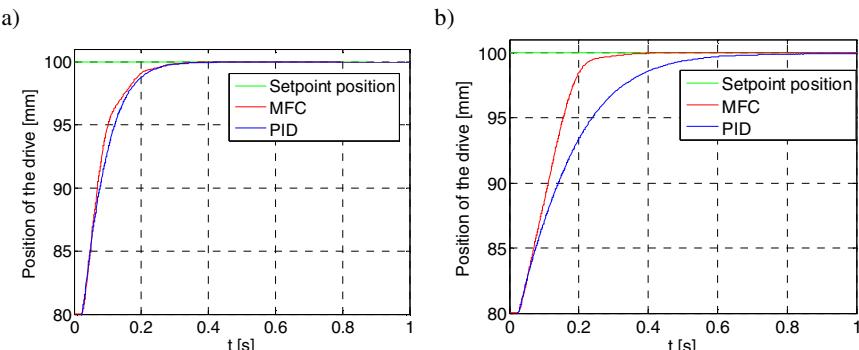


Fig. 7. Position of drive with use of PID controller and MFC controller for pressure supply:
a) 8 MPa, b) 2 MPa

6 Conclusion

MFC method is commonly used in process control that require automatic adjustment. The presented here research results were executed to verify the use of MFC method capabilities in control of electrohydraulic servo drive.

The main advantage of the described here MFC method is a significant simplification of the controller structure, in comparison with other types of advance method of control, like for example adaptive control. Taking into account the collected results, MFC method seems to be more robust and more resistant on external disturbances in comparison with classical PID control method.

The research of the usage of the MFC method in control of electrohydraulic servo drives will be continued in near future. The currently ongoing studies are aimed at checking the effectiveness of the MFC method with load impact changes.

References

1. Chern, T.L., Chang, G.K.: Automatic voltage regulator design by modified discrete integral variable structure model following control. *Automatica* 34(12), 1575–1582 (1998)
2. Lausch, H., Lierschaft, K., Schwarz, H.: Structure-variable model-following control for hydraulic drives (Conference Paper). In: IFAC Symposia Series – Proceedings of a Triennial World Congress, vol. 4, pp. 545–550 (1991)

3. Lee, S.H., Song, J.B.: Control of linear motor-based arm motion generator using active impedance implementation. In: Proceedings of 14th Triennial World Congress, Beijing, China, C-2a-16-3, pp. 551–556 (1999)
4. Li, G., Tsang, K.M., Ho, S.L.: A novel model following scheme with simple structure for electrical position servo systems. International Journal of Systems Science 29, 959–969 (1998)
5. Park, K., Kim, S.H., Kwak, Y.K.: Adaptive control for time optimal tracking systems. Int. J. Control 63(5), 951–964 (1996)
6. Rybarczyk, D.: Application of MFC method in electrohydraulic servo driver. Pneumatika a Hydraulika, Słowacja
7. Skoczowski, S., Domek, S., Pietrusiewicz, K.: Model following PID control system. Kybernetes 32(5/6) (2003)
8. Sugai, H., Nonami, K.: Reference model following sliding mode control for hydraulic mine detection hexapod robot. Nihon Kikai Gakkai Ronbunshu, C Hen/Transactions of the Japan Society of Mechanical Engineers, Part C 72(9), 2829–2837 (2006)
9. Tsang, K.M., Li, G.: Nonlinear nominal-model following control to overcome deadzone nonlinearities. IEEE Transactions on Industrial Electronics 48(1), 177–184 (2001)
10. Wiegandt, M.: Development of a servomotor driven proportional valve. In: 7th International Fluid Conference Aachen
11. Wu, H., Kawabata, H.: Robust model following control with zero-tracking error for a class of uncertain linear systems. In: Proceedings of the 1999 IEEE Hong-Kong Symposium on Robotics and Control, pp. 85–90 (1999)
12. Yun, J.S., Cho, H.S.: Adaptive model following control of electrohydraulic velocity control system subjected to unknown disturbances. IEE Proceedings D: Control Theory and Applications 135(2), 149–156 (1988)
13. Zhang, H., Quan, L.: Research on disturbance observer for permanent magnet synchronous motor driving hydraulic pump. Zhongguo Jixie Gongcheng/China Mechanical Engineering 21(24), 2914–2917 (2010)

Electromechanical Actuators – Selected Safety-Related Problems

Tadeusz Missala

Industrial Research Institute for Automation and Measurements PIAP
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
tmissala@piap.pl

Abstract. This paper presents the results of considerations and calculations on the role of electromechanical actuators in a path of realization of automated functions and on the requirements resulting from the intended application. The safety-related functions and process-related functions are discussed. A special situation of actuators working in 1oo1 architecture, as well as consequences of that architecture on satisfaction of requirements, are presented. Some procedures of conduct resulting from the good engineering practice are also described.

Keywords: electromechanical actuators, safety requirements, life cycle requirements, inherent safety design.

1 Introduction

Electromechanical end actuators are very important devices for manufacture- and process-automation. They proved irreplaceable for movement-related actions, e.g. open and close the control valves used in chemical and petrochemical process industry as well as in power plants, moving robot arms, automatically open or close doors, barriers, etc. To them belong also relays, contactors and other electromechanical elements used in electrical circuits, e.g. in machinery, elevators and various motion systems.

The widespread use of them is the reason to consider some of safety-related problems concerning electromechanical end actuators.

2 Realization Path of Automated Functions

The simplified, general description of realization path of any automated function can be reduced to three subsystems, connected in line: sensor subsystem (all sensors, transmitters, etc.), logic subsystem (from simple feedback to complicated computer logic data logger) and final element subsystem, mostly electromechanical actuators.



Fig. 1. Simplified block diagram of function realization path (S – sensor subsystem, L – logic subsystem, FE – final element subsystem)

3 Setting of the Problem

Two general kinds of functions can be recognized: process-related functions and safety-related functions (safety functions). The task of process-related functions is the control of process parameters and provision of guarantee that the end product will comply with the technological and client requirements.

The task of safety functions is to ensure that the related to the process risk for people, environment and property will be acceptable or tolerable.

Each function shall comply with suitable requirements. There are two groups of requirements:

- requirement for function functionality;
- requirements for function dependability properties over a plant or machine lifetime.

Requirements for safety functions dependability properties are the result of risk assessment. The relationships between results of risk assessment and requirement for safety functions properties are established in EN 61508-1 [13a], EN 61784-3 [14] and EN-ISO 13849-1 [11]. If the electromechanical actuator is the end element of a chain of a safety function, it shall comply with the appropriate safety requirements.

Requirements for process-related functions properties result from the technological conditions of the process and cannot be precise on the general way.

On the other hand, for the both groups of functions, requirements for function dependability can be formulated as a result of consideration of the lifetime of the plant [10].

From the Fig. 1 and remarks made in p. 1 results the conclusion: while sensor system and logic system can be designed as redundant systems, the electromechanical actuator cannot have any redundancy – it is not possible to manipulate e.g. a control valve with aid of two actuators in, e.g. 1oo2 system.

The result of the considerations as above is the following thesis:

Electromechanical actuators, as they cannot contain any redundancy (they work in architecture 1oo1), require a special design and realization procedures with the aim of ensuring they suitable dependability.

4 Risk Reduction and Safety Requirements for Safety Functions

Process of risk reduction is realized according to the ALARP principle (As Low As Rationally Practicable). The idea of this principle is demonstrated in the Fig. 2.

To rich the region of broadly acceptable risk, appropriate safety functions complying with suitable safety requirements shall be used. The requirements, expressed as probabilistic quantities and grouped into four safety integrity levels (SIL) from 1 to 4, are established in EN 61508-1 [13a] and are demonstrated in Table 1.

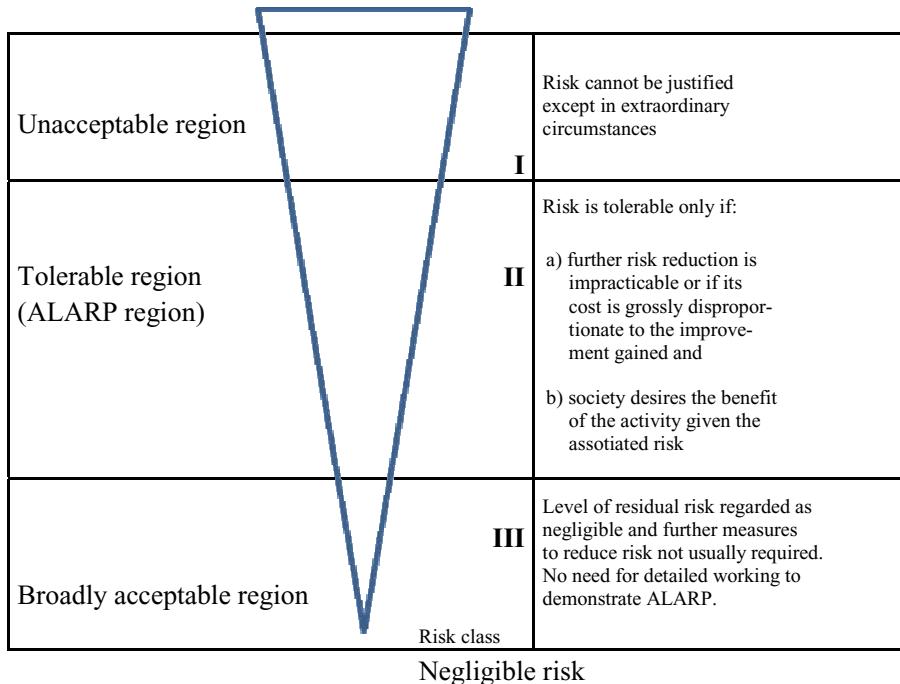


Fig. 2. Illustration of the ALARP principle

Table 1. Safety integrity levels – target failure measures for a safety function operating in high demand mode of operation or continuous mode of operation [13a]

Safety integrity level (SIL)	Average frequency of dangerous failure of safety function [h^{-1}] (PFH)
4	$\geq 10^{-9}$ to $< 10^{-8}$
3	$\geq 10^{-8}$ to $< 10^{-7}$
2	$\geq 10^{-7}$ to $< 10^{-6}$
1	$\geq 10^{-6}$ to $< 10^{-5}$

If in realization of safety functions is used any transmission line (Fieldbus, Ethernet, wireless, etc.), the supplementary requirements, concerning above mentioned line, are important, as demonstrated in Table 2 [14].

Standard IEC 61508 has the status of basic safety standard, in this way in other standards concerning safety-related systems its requirement shall be introduced. These requirements set producers themselves, e.g. producers of MEMS are worrying – see [19].

To determine which safety integrity level (SIL) is required for the particular safety function, one needs to use the procedures of risk assessment – descriptions of such procedures are available, e.g. in the standards [11, 12, 13e]. Methods of design and analysis as well as methods of verification of the end project of safety function are presented in [13f].

Table 2. Relationship of residual error rate of transmission protocols to SIL level [14]

Applicable for safety functions up to SIL	Average frequency of dangerous failure of the functional safety communication system [h^{-1}]	Maximum permissible residual error rate for the functional safety communication system
4	from $\geq 10^{-11}$ to $< 10^{-10}$	from $\geq 10^{-11}$ to $< 10^{-10}$
3	from $\geq 10^{-10}$ to $< 10^{-9}$	from $\geq 10^{-10}$ to $< 10^{-9}$
2	from $\geq 10^{-9}$ to $< 10^{-8}$	from $\geq 10^{-9}$ to $< 10^{-8}$
1	from $\geq 10^{-8}$ to $< 10^{-7}$	from $\geq 10^{-8}$ to $< 10^{-7}$

The parameter which shall comply with the requirements established in Table 1 is the average frequency of dangerous failure per hour of a safety function – this parameter is defined in [13f] and will be used in the further part of this study.

Four meanings of this parameter are distinguished:

- PFH_{SYS} – concerning the full system,
- PFH_S – concerning the sensor subsystem,
- PFH_L – concerning the logic subsystem,
- PFH_{FE} – concerning the final element subsystem.

From limitations of each subsystem architecture, as well as from the architecture of the whole path, the following rules are valid [11, 13b]:

- SIL of subsystem or any its path can be at most equal the lowest SIL of elements used,
- if the path is composed from maximum 3 elements having the same SIL, the path is representing this SIL,
- if the path is composed from 4 to 6 elements having the same SIL, SIL of the path is one lower.

5 Lifecycle Requirement for Process-Related Functions

The second group of automated functions are process-related functions, this means all safety non-related functions. These functions shall be also realized with satisfactory dependability in satisfactory time intervals. They can be derived from lifetimes of industry plants, as presented in Table 3.

The PFH_{FE} values derived from required lifetime are located:

- in case of energy and nuclear power plants, as well as machine building in space of SIL 1,
- in case of chemical batch plants below SIL 1,
- in case of chemical continuous plant on fringe about SIL 1.

Table 3. Lifetimes of industry plants (source [15])

Item	Plant	Lifetime desired for industrial plant		PFH _{FE} h ⁻¹	Remarks
		Years	Hours		
1	Energy	> 25	> 217 500	$< 4.6 \cdot 10^{-6}$	No product change
2	Nuclear power	> 40	> 348 000	$< 2.86 \cdot 10^{-6}$	No product change
3	Chemical, batch	5 – 10	43 500 – 87 000	$2.3 \cdot 10^{-5}$ to $1.15 \cdot 10^{-5}$	Product change possible
4	Chemical continuous	10 – 25	87 000 – 217 500	$1.15 \cdot 10^{-5}$ to $4.6 \cdot 10^{-6}$	Product change possible
5	Machine building	25	217 500	$4.6 \cdot 10^{-6}$	Product change possible

6 Conclusions for Electromechanical Actuators

The main space of use of electromechanical actuators are:

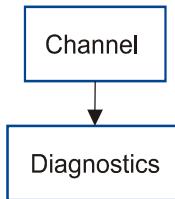
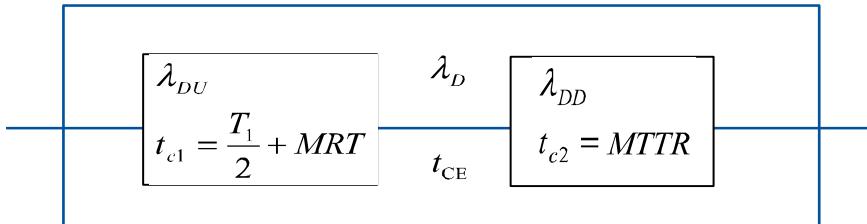
- energy power plants,
- nuclear power plants,
- chemical continuous plant,
- machine tool building.

From analysis of data in Table 3, it follows that actuators shall be built in such a way so as to comply with last safety integrity level SIL 1.

This entitles for consideration of all electromechanical actuators as elements of safety functions.

7 Calculation of PFH_{FE}

As mentioned above, electromechanical actuators work in the architecture 1oo1. The relevant block diagrams are presented in Fig. 3 and Fig. 4 [13f]:

**Fig. 3.** 1oo1 physical block diagram**Fig. 4.** 1oo1 reliability block diagram

On the basis of these diagrams the following equations are valid:

$$\begin{aligned} \lambda_D &= \lambda_{DU} + \lambda_{DD} \\ t_{CE} &= \frac{\lambda_{DU}}{\lambda_D} \left(\frac{T_1}{2} \right) + MRT + \frac{\lambda_{DD}}{\lambda_D} MTTR \\ \lambda_{DU} &= \lambda_D (1 - DC); \lambda_{DD} = \lambda_D DC \end{aligned} \quad (1)$$

where: T_1 – proof test interval [hour]; $MTTR$ – Mean time to restoration [hour] – typically 8 h; MRT – mean repair time [hour] – typically 8 h; DC – diagnostic coverage [fraction or percentage] – typically (0, 60, 90, 99) %; λ – total failure rate (per hour) of a channel in a subsystem; λ_{DD} – the sum of detected dangerous failure rate (per hour) of a channel in a subsystem; λ_{DU} – the sum of undetected dangerous failure rate (per hour) of a channel in a subsystem – typically $(0,05; 0,25; 0,5; 2,5; 5; 25) \times 10^{-6}$; (for electromechanical actuator – final element notation λ_{DUFE} will be used); t_{CE} – channel equivalent mean down time (hour) for 1oo1, 1oo2, 2oo2 and 2oo3 architectures.

If it is assumed that the safety system puts the EUC into a safe state on detection of any failure, for an electromechanical actuator having 1oo1 architecture the following is obtained:

$$PFH_{FE} = \lambda_{DUFE} \quad (2)$$

Electromechanical actuator is composed, as minimum, of two main parts: electric motor and mechanical gearbox connected in series. The consequence is the relation:

$$\lambda_{DUFE} = \lambda_{DUM} + \lambda_{DUG} \quad (3)$$

Parameter λ_{DU} can be calculated in two ways [5, 6, 12, 13a] under the assumption that electromechanical actuator is an element which is required to operate for a defined mission time during which no repair can take place.

First method [13a, 7.6.2.9 note 7] presents, the required safety integrity level for a safety function can be derived as follows. Determine the probability of failure of the safety function during the mission time and divide this by the mission time to obtain a probability of failure per hour, then use table 1 to derive the safety integrity level. On this basis one can obtain:

$$\lambda_{DU} = \frac{p_{tm}}{T_m} \quad (4)$$

where: T_m – is the mission time, p_{tm} – is the probability of failure during the mission time.

Second method [16, 17], developed for evaluation of electromechanical relays and referred in [12] for safety-related electrical control systems (SRECS) introduce a parameter B_{10} defined as a expected time to failure of 10 % of elements tested. Under assumption that the parameter λ_{DU} is constant and sufficiently low (e.g. $1 \gg \lambda_{DU} \times T_l$), the following basic equations can be used:

$$\begin{aligned} \lambda &= \frac{1}{MTTF} \\ \lambda &= 0,1 \times \frac{C}{B_{10}} \end{aligned} \quad (5)$$

Verification of parameters values obtained from equations (4) and (5) is carried out using probabilistic methods [4, 17]. The heart of them is the **long time work test** [4]. The test results can be evaluated by means of various statistical models.

A kind of model applied is connected from the store of data concerning the actuator considered. If we have no information on probabilistic distribution of failures, then it is necessary to apply the non-parametric evaluation. In this case [4] the limits, the upper – $R_u(t)$ and the lower – $R_l(t)$, of the probability $R(t)$ the mission time will be working can be evaluated. Here:

$$R(t) = 1 - p_{tm} \quad (6)$$

The values of these limits depend on the number n of specimens used in the long time work test and of the number $m(t)$ of specimens failed during the test. The corresponding relations are presented in Table 4.

Table 4. The values of probability $R(t)$

Number of specimens n	$m(t) = 1$		$m(t) = 2$	
	$R_l(t)$	$R_u(t)$	$R_l(t)$	$R_u(t)$
20	0,8189	0,9947	0,7547	0,9731
25	0,8528	0,9958	0,8004	0,9785
30	0,8765	0,9965	0,8322	0,9821
40	0,9060	0,9973	0,8723	0,9866
50	0,9241	0,9979	0,8971	0,9893

To assess λ_{DU} the lower limits shall be applied. Taking into account that typical brushless electric motors have the mission time of 20 000 h, the values λ_{DU} obtained on the basis of (4) are presented in Table 5.

Table 5. λ_{DUm} values of electric motors at $T_m = 20000$ h as a function of $R_l(t)$

$m(t) = 1$	$R_l(t)$	λ_{DU}	$m(t) = 2$	$R_l(t)$	λ_{DU}
	0.8189	9.06×10^{-6}		0.7547	12.26×10^{-6}
	0.8528	7.36×10^{-6}		0.8004	9.98×10^{-6}
	0.8765	6.18×10^{-6}		0.8322	8.39×10^{-6}
	0.9060	4.70×10^{-6}		0.8723	6.38×10^{-6}
	0.9241	3.80×10^{-6}		0.8971	5.15×10^{-6}

These estimations are quite satisfactory from the point of view of the requirements for process-related functions (see Table 3).

If we have collected information on electromechanical actuators failure times during long-time work tests, there is a possibility to estimate an adequate distribution model. The 20-ty years of Author's experience by certification process of electric motors enabled to apply the parametric method – normal distribution cut on $x = 0$ (time cannot be negative). In case of 25 specimens and 2 failed during the test, the following assessment was obtained:

$$R(t) = 0.9003$$

i.e. the higher value than presented in Table 4. If introduced to (3), the value:

$$\lambda_{DUm} = 5.0 \times 10^{-6}$$

lower and more adequate is obtained.

It is necessary to underline the long-time work test is time-consuming and expensive. 20 000 h this is almost 3 years, therefore it is recommended to develop the model of failure arising and on basis of it, to elaborate the procedure of accelerated long-time work test.

The second method (4) [12, 16, 17] is referring the Weibull analysis [18] to estimate parameter B_{10} . The essence of it is the long-time work test carried out to the moment, the 7-th specimen will fail. The time of that test could be probably several time longer than mission time of electromechanical actuator and for such a devices is completely impracticable. This test is appropriate for relays, contactors and other similar elements, e.g. end-breakers, when the time of test can be shortened by accelerating the process of switching.

The problem to solve is, how reach the satisfactory values of PFH_D ?

8 Way to Satisfactory Design, Development, Manufacturing and Use

The way to satisfactory design, development and use can be determined after Petroski as in “Checking In & Checking Out Design” [8]. This means:

- ensuring safety inherency [1, 2, 10],
- properly checking and testing [2, 5, 7],
- ensuring proof tests [13d] during revision cycles,
- complying with the work and service instructions and procedures.

Safety inherency as well as properly checking and testing can be ensured by:

- use of well-proven parts and elements, including well-proven motors and gearboxes from trusted suppliers,
- lead design and development according to procedures based on good engineering practice and safety life cycle [13a, 2] and,
- execute full verification and validation [13a] by analysis and experiment, especially in conditions of environmental influences (e.g. cold, heat, shocks, vibrations, changes of temperature and humidity),
- execute all proof tests in periods resulting from revision cycles presented in Table 6.

Table 6. Revision cycles in industry (source [15])

Item	Plant	Modernization cycle (years)	Revision cycle (months)
1	Energy	15 – 20	12
2	Nuclear power	15 – 20	12
3	Chemical bath	5	< 12
4	Chemical continuous	15	60
5	Machine building	10	12

The procedures resulting from good engineering practice and safety life cycle corresponding to safety inherent design are presented in Fig. 5 for conception stage and in Fig. 6 for design stage.

It is to underline that the insufficient adherence to procedures at design and development leads to very serious consequences – see [19].

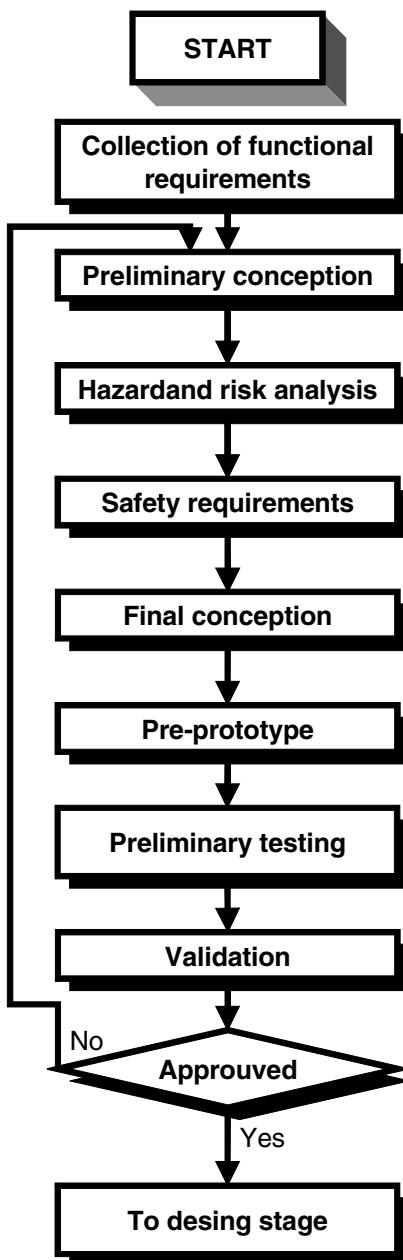


Fig. 5. Proceedings during conception stage

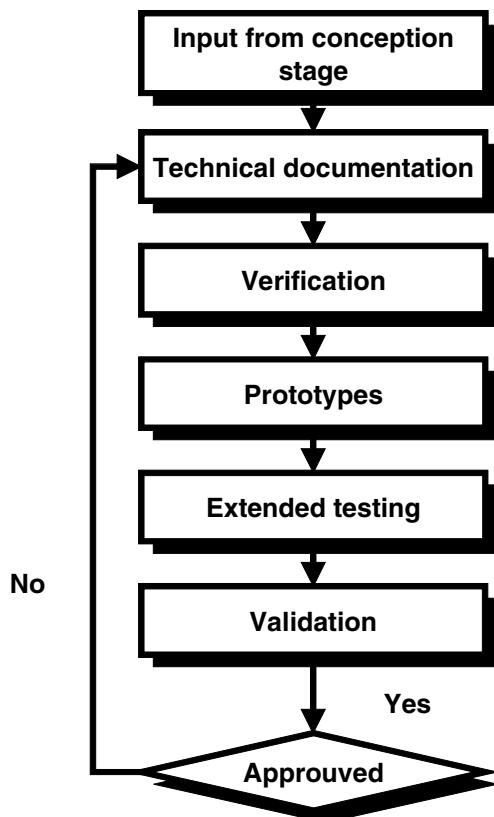


Fig. 6. Proceedings during design stage

9 Conclusions

Presented methodology enables to achieve fulfillment of the requirements resulting from the domain of application of the electromechanical actuators: safety functions or process-related functions, and is the solution of the thesis presented in the paper.

References

1. Missala, T.: Bezpieczeństwo funkcjonalne w procesie projektowania urządzeń mechatroniki (Functional safety in mechatronics devices design) Politechnika Warszawska, Prace naukowe. Konferencje, z. 14. In: III Krajowa Konferencja Naukowo-Techniczna ‘Mechatronika’ 1997, Warszawa, pp. 541–546 (September 1997) (in Polish)
2. Missala, T.: Ocena a priori i a posteriori bezpieczeństwa robota (A priori and a posteriori assessment of robot safety). Prace Naukowe Instytutu Cybernetyki Stosowanej Politechniki Wrocławskiej, Nr 99. In: Prace VI Krajowej Konferencji Robotyki, Monograph, Wrocław, vol. 2, pp. 133–148 (September 1998) (in Polish)

3. Missala, T.: Mechatronics Elements in Safety-Related Circuits. *Mechatronics* 2004 Warsaw University of Technology, Faculty of Mechatronics and Brna University of Technology, Faculty of Mechanical Engineering. *Elektronika* (8-9), 183–185 (2004)
4. Missala, T.: Próba pracy długotrwałej jako narzędzie diagnostyki technicznej (Long-time work test as the tool of technical diagnostics). *Pomiary Automatyka Kontrola* (9) 321–323 (2005) (in Polish)
5. Missala, T.: Metodologia oceny nienaruszalności bezpieczeństwa elementów o ustalonej trwałości (Methodology of safety integrity assessment for the elements with defined mission time). In: *Materials of Conference Automation 2006*, Warszawa, pp. 111–118 (2006) (in Polish)
6. Missala, T.: Porównanie metod oceny nieuskadzalności elementów podanych w PN-EN 61508-1 i PN-EN 61810-2 (Comparison of reliability assessment methods presented in EN 61508-1 and EN 61810-2) *Bezpieczeństwo przemysłowe*. Monograph, t. 3. Klub Paragraf 34. Warszawa (2009)
7. Petroski, H.: Petroski on Engineering: Checkin. In: & Checking Out Design, http://www.designnews.com/author.asp?section_id=1365&doc_id=248247&print=yes
8. Titus, J.: Careful designers get the most from brushless DC motors, <http://www.ecnmag.com/articles/2012/08/careful-designers-get-most-brushless-dc-motors>
9. Dunn, M.: Bad design and its consequences: US legal ruling on Toyota's "killer firmware". *EE Times Europe* (October 30, 2013), http://www.automotive-eetimes.com/_includes/print.php?1g=en&cnp_id=17&safe_mode
10. EN-ISO 12000: 2010, Safety of machinery – General principles for design – Risk assessment and risk reduction
11. EN ISO 13849-1:2008, Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design (ISO 13849-1:2006)
12. EN 62061: 2005 + am 1:2012-11, Safety of machinery – Functional safety of safety-related electrical, electronic and programmable electronic control systems
13. PN-EN 61508 (IEC 61508): Functional safety of electrical/electronic/ programmable electronic safety-related systems: PN-EN 61508-1:2010P Part 1: General requirements,
14. PN-EN 61508-2:2010E Part 2: Requirements for electrical/electronic/programmable electronic safety-related systems
15. PN-EN 61508-3:2010E Part 3: Software requirements
16. PN-EN 61508-4:2010E Part 4: Definitions and abbreviations
17. 13e PN-EN 61508-5:2010E Part 5: Examples of methods for the determination of safety integrity levels
18. 13f PN-EN 61508-6:2010E Part 6: Guidelines for the application of IEC 61508-2 and IEC 61508-3
19. PN-EN 61508-7:2010E Part 7: Overview of techniques and measures
20. PN-EN 61784-3:2009, Industrial communication networks – Profiles – Part 3: Functional safety fieldbuses – General rules and profile definitions
21. DIS IEC 62890 (65/526/NP), Life-cycle management for systems and products used in industrial-process measurement, control and automation (in development)
22. PN-EN 61810-2:2011-02, Electromechanical elementary relays – Part 2: Reliability
23. PN-EN 61810-2-1:2011-02, Electromechanical elementary relays – Part 2-1: Reliability – Procedure for the verification of B10 values
24. PN-EN 61649:2008E, Weibull analysis
25. Clarke, P.: Safety Standards Could Disrupt MEMS in Automotive. *EE Times*, <http://www.designnews.com/>
26. Missala, T.: Dobra praktyka inżynierska ułatwia deklarowanie zgodności europejskiej. *Pomiary Automatyka Robotyka* (2), 294–303 (2008) (in Polish)

Immune Algorithm for Fuzzy Models Generation

Bogumiła Mrozek

Instytut Informatyki, Politechnika Krakowska
bmrozek@pk.edu.pl

Abstract. This paper presents fuzzy models which rules were extracted from numerical data using clonal selection, subtractive, fuzzy C-means, Gustafson–Kessel clustering algorithms, implemented in the MATLAB code. These algorithms were used for the identification of parameters in the fuzzy model Sugeno-type. There are two testing examples: *Trip* data and *DWP* data set from the multi-detector sensor. Fuzzy model of the fire risk index was built based on the laboratory data measurements. The results are shown in tables and graphs.

Keywords: data clustering, fuzzy model, clonal selection algorithm, immune algorithm, fire hazard.

1 Introduction

The measurement data are affected by errors arising, for example from the inaccuracy of measuring devices or accepted measurement techniques. These errors result in failure to meet theoretical equations describing the operation of the modeled object or phenomena.

Fuzzy models are able to reproduce the relationship between physical quantities that are difficult to describe by mathematical relationships. They are applied if it is difficult to create a mathematical model, but the situation can be described by fuzzy rules in a qualitative way. The fuzzy models can be identified sequentially. The sequential mode may determine the structure and antecedents first by chosen clustering algorithm, and then optimize the consequent and/or antecedent parameters by evolutionary optimization methods [12].

The majority of developed Artificial Immune Systems (AIS) algorithms are based on immunological theories of clonal selection, immune networks and negative selection. The clonal selection algorithm was used in identification of parameters of Sugeno fuzzy model [11] extracted from numerical data. Clonal selection algorithm can be applied both for optimization problems as well as issues of clustering and pattern recognition [10]. This means that the clonal selection algorithm may be used to automatically generate Sugeno-type fuzzy models, both at the initial stage and at the final tuning.

The clonal selection algorithms were used for the fuzzy clustering problems. As an application the clonal selection algorithm for optimization of parameters in fuzzy models will be described in the future work.

This paper is organized as follows. Section 2 is a briefly introduction to Sugeno fuzzy models and mechanism of clonal selection in the immune system. Section 3 describes extracting knowledge from the numerical data. Section 4 presets briefly identification of parameters for Sugeno fuzzy models. Section 5 contains the experimental verification results. Conclusions and future work are given in Section 6.

2 Methods for the Construction of Smart Models

Artificial intelligence methods include issues of fuzzy modeling, evolutionary algorithms, artificial neural networks and artificial immune system. Inspired by the mechanisms that occur in the human body they can imitate human behavior; also use certain mechanisms. Based on an input-output data, a smart model will be constructed. It should create rules of inference, generalize knowledge and robust to data errors.

2.1 Fuzzy Models

Sugeno-type fuzzy model is computationally very efficient and works properly in conjunction with the methods of adaptation and with algorithms of optimization. It is particularly attractive when applied to modeling and control of nonlinear systems.

A typical fuzzy rule in a Sugeno fuzzy model [11] has the form:

$$\text{If } (x \text{ is } A_j) \text{ and } (y \text{ is } B_j) \text{ then } z_j = f_j(x, y)$$

where A_j and B_j are fuzzy sets in the antecedent of j -th rule, while $z_j = f_j(x, y)$ is a crisp function in consequent of j -th rule. Usually $f_j(x, y)$ is a polynomial in the input variables x and y , but it can be any functions as long as it can appropriately describe the output of the model within the fuzzy region specified by antecedent of the rule. Most often $f_j(x, y)$ is a first-order polynomial or a constant (zero-order polynomial). Consequent of j -th rule can be expressed as:

$$z_j = f_j(x, y) = p_j x + q_j y + r_j \quad (1)$$

where p_j, q_j, r_j - polynomial coefficients, in zero-order model: $z_j = r_j$.

Operation type Sugeno fuzzy model is based on the performance of switching between several optimal linear models. In this way, strongly non-linear objects can be modeled.

2.2 The Mechanism of Clonal Selection in the Immune System

The immune system is based on two aspects. The first is detection of the pathogen in terms of verifying the molecules that belong to the body and could be harmful. Second aspect is to eliminate the risk that must effectively select the appropriate method depending on the type of pathogen. Once you have identified the item as the enemy, the system performs a series of steps aimed at removing the hazard.

The elements of the immune system are the so-called lymphocytes, who take part in the detection and elimination of the pathogen. They are divided into T-cells and B-cells [9]. T-lymphocytes are responsible for considering the "suspicious" cells. They give a signal to undertake defensive actions. B-lymphocytes are subordinated to the T-lymphocytes. They produce antibodies to eliminate these pathogens and are involved in memorizing these structures which leads to the formation of immunological memory.

Clonal selection theory is used to describe the fundamental characteristics of the immune response to antigenic incentive. It involves the construction of a large number of cell clones responding to an antigen that was found in the body [9]. In response to the first contact with the antigen and under the influence of T-lymphocytes, B-lymphocytes are activated and cloned. This process is proportional to the similarity of the selected antigen - higher the similarity is, more generated B-cells. Then, clones are subject to hypermutation. The process is inversely proportional to the affinity of the antigen cell receptors - higher the affinity, lower intensity of the mutation. The population of the mutant clones is evaluated for the degree of adjustment for the antigen: antigen-binding clones are poorly removed, and the well-binding are included in the immunological memory [9].

The whole process of mutation and selection is known as "maturation of the immune response", and is a continuous process. Information about effective response is encoded in immunological memory, used, when calling the secondary response. The content of the memory is increased as a result of external or internal repeated expose to a pathogen.

Clonal selection algorithm is divided into two stages:

1. *Clonal expansion* is responsible for the selection of best-fit antibody and multiplication of proportion (but not necessarily) to their degree of affinity.
2. *Hypermutation* of antibody maturation process is carried out by transforming clones (were expanded) in such a way that some of them achieved better overall adjustment than their predecessors.

The two operations (selection of the best individuals and their cloning) are necessary for correct operation with this algorithm. The literature describes several implementations with the use of clonal selection algorithms [13]. It is most commonly used for optimization tasks, clustering and pattern recognition.

3 Extracting Knowledge from the Measured Data

It is very useful if rules base (knowledge base) of fuzzy model can be automatically extracted from the measured data. The most common method of automatic extraction of rules is based on grouping of numerical input-output data, often called clustering. Effective clustering algorithms specify the center of clusters, focus areas of measurement data. This involves the determination of parameters of fuzzy model together with the organization of its structure (rule base).

3.1 Clustering Algorithms

Fuzzy clustering creates assignment in ambiguously. One element of the set of data may belong to several clusters, each of them to a certain extent. The methods of fuzzy clustering are one of the unsupervised learning techniques.

Generally, these methods can be divided into:

- Methods, in which the number of clusters is not known. The user must initially assume the number of clusters and enters this number as a parameter of the algorithm. Center of each cluster is searched simultaneously with the division of data. Examples of algorithms of this type can be: *fuzzy C-means* method (called FCM) and the method of *Gustafson-Kessel*.
- Number of clusters may be determined automatically based on a measure of data density in space. This type of algorithms may include *subtractive clustering* method or a method based on the flow chart *clonal selection* algorithm.

For all four of the above methods of clustering implementations, a MATLAB code was prepared. These methods were tested for numeric test data sets *Iris* and *Trip*. The effectiveness of various clustering algorithms was evaluated using a number of quality indexes such as: partition, entropy, Xie-Bieni and Fukuyamy–Sugeno [4].

The correct results of all tested clustering algorithms were obtained. The quality and speed of clustering studied algorithms were dependent on the proper selection of their parameters [5, 6].

3.2 Clustering Using Clonal Selection Mechanism

Clustering algorithm using clonal selection mechanism is based on the method used for pattern recognition tasks, described in [9, 10]. The algorithm is performing successive iterations. Each section starts assuming initial population representing the downloaded data. Of these, the data pattern is also selected. If the element of the data set representing a pattern is fixed in advance within the cluster, then the other element is selected. The next step is to compare the pattern with all elements of the population.

Calculated similarity is presented in the form of matching vector, which is then sorted according to the value of similarity. Ordered population is cloned in proportion to the degree of alignment and the number of dependent values cloning. This creates a temporary population, which matures in the process of hypermutation. The intensity depends on the ratio hypermutation.

Mature temporary population is re-compared with the pattern and calculates the matching vector. Then it is sorted according to the value of similarity. Best suited element of the temporary population is written to memory as a new center of the cluster, provided that they are not in the other cluster. Iteration ends up substituting the worst elements of the population with the new matched. The algorithm terminates after a predetermined number of iterations or in the case where all points lie within clusters and it is incapable to determine a new pattern. The advantage of this algorithm is the independent matching of the optimal clusters number based on the declared cluster radius.

4 Identification of Parameters for Sugeno Fuzzy Models

Chiu presents the methods of extracting fuzzy rules from data for function approximation and for pattern classification applications [1, 2]. Extracting fuzzy rules from data for pattern classification applications starts from separating data into groups according to their respective classes.

Subtractive or clonal selection clustering may be applied to the input space of each group of data individually to extract the rules for identifying each class of data. The clusters found in the data of a given group identify regions in the input space that map into the associated class. Hence, we can translate each cluster center into a fuzzy rule for identifying the class.

Extracting fuzzy rules for approximation of function from data, clustering is performed in the combined input/output space, i.e., each data point x_i is a vector that contains both input and output values. Each center of cluster is essence of prototypical data point that exemplifies an input/output behavior of the system.

The main task in constructing the model is to determine the fuzzy rule base and the number of fuzzy sets (membership function) assigned to individual inputs and outputs of the model. Subsequently, it is also appropriately selecting algorithms for aggregation of simple premises.

4.1 Fuzzy Rules Extraction Directly from Numerical Data

Each cluster center x_i^* is considered as a fuzzy rule that describes the system behavior [1, 2]. Intuitively, each cluster center represents the rule:

Rule i : **If** {input is near y_i^* } **then** output is near z_i^* .

Given an input vector y , the degree of fulfillment of rule i is defined as

$$\mu = e^{-\alpha} \|y - y_i^*\|^2 \quad (2)$$

where $\| \dots \|$ denotes the Euclidean distance. Coefficient α is the constant and equals: $\alpha = 4/r_a^2$. The constant r_a is effectively the radius defining a neighborhood, used in subtractive and clonal selection clustering methods.

This computational model can be used as a fuzzy inference system employing traditional fuzzy if-then rules [1, 2]. Each rule has the following form:

if Y_1 is A_{i1} & Y_2 is A_{i2} & ... **then** Z_1 is B_{i1} & Z_2 is B_{i2} ...

where Y_j is the j -th input variable and Z_j is the j -th output variable; A_{ij} is an exponential membership function in the i -th rule associated with the j -th input and B_{ij} is a membership function in the i -th rule associated with the j -th output.

For the i -th rule, which is represented by cluster center x_i^* , A_{ij} is given by

$$A_{ij}(Y_j) = \exp \left\{ -\frac{1}{2} \left(\frac{Y_j - y_{ij}^*}{\sigma_{ij}} \right)^2 \right\} \quad (3)$$

and B_{ij} can be any symmetric membership function centered around z_{ij}^* , where y_{ij}^* is the j -th element of y_i^* , z_{ij}^* is the j -th element of z_i^* , and $\sigma_{ij}^2 = 1/(2\alpha)$ [1, 2].

This computational scheme is equivalent to an inference method that uses multiplication as the AND (&) operator, weights the consequent of each rule by the rule's degree of fulfillment, and computes the final output value as a weighted average of all the consequents.

Another approach is to let the consequent parameter z_{ij}^* be a linear function of the input variables. That is, it let

$$z_{ij}^* = G_{ij} y + h_{ij} \quad (4)$$

where G_{ij} is an N -element vector of coefficients and h_{ij} is a scalar constant.

The *if-then* rules then become the Sugeno-type. For given a set of rules with fixed premise membership functions, optimization G_{ij} and h_{ij} in all consequent equations is a simple linear least-squares estimation problem [1].

In Section 5 fuzzy rules are extracted from the real data set for function approximation applications purposes. In our example we use these approximations for testing *Trip* data and to apply *DWP* measurements for creating the fuzzy model of a *fire hazard* index.

4.2 Optimization of Coefficients in Consequent and Premise the Fuzzy Rules

Fuzzy model Sugeno type identification method consist of two steps: (1) find cluster centers to establish number of fuzzy rules and parameters the rules premise and (2) calculate with optimizing the rules consequents parameters. Optimizing only the coefficients in consequent equations allows a significant degree of model optimization to be performed without adding much computational complexity.

This identification method was used to parameters of fuzzy model of a *fire hazard* index which describes the problem of fire detection. The accuracy of this fuzzy model was sufficient.

If better accuracy of estimation fuzzy model is needed it can be obtain by optimizing the parameters of membership functions with chosen evolutionary algorithm, for example clonal selection method. For new parameters of premise membership functions, the parameters all consequent equations should be recalculated.

5 Application Examples

To illustrate application the immune algorithm for fuzzy models generation was applied to fuzzy modeling problem based on the numerical data. Two testing examples are presented. They have few input data and one output and it is very difficult (impossible in practice) to make their mathematical models.

5.1 Model of Automobile Trip Prediction

Trip Generation Data was published in [2, 3]. *Trip* data set contains the relationship between the number of automobile trips generated from an area and the area's

demographic factors. Demographic and trips data are from 100 traffic zones in New Castle County, Delaware. Five demographic factors are considered: population, number of dwelling units, vehicle ownership, median household income and total employment.

Hence, the model has 5 input variables and 1 output variable. Fuzzy model make estimation of the number of automobile trips generated from an area based on 5 demographics factor from of this area. 75 data points (of 100 available) was selected for extracting fuzzy rules for Sugeno fuzzy model by using 4 clustering algorithms.

Trip data sets and results of clustering for 4 methods (fuzzy C-means, Gustafson–Kessel, subtractive and clonal selection methods) are presented in Fig. 1 [6].

All employed clustering methods give the best results for 3 determinate clusters. The location of determinate clusters is similar for all used clustering methods with the little variations. The values of all the clustering quality indexes [4] are also very similar. The partition index has the values between 0.573 and 0.618. The entropy index has the values between 0.663 and 0.735. Indexes Xie–Bieni and Fukuyamy–Sugeno has better values for fuzzy C-means and Gustafson–Kessel methods [5, 6].

The shortest computation time was for subtractive clustering method (only 0.03 sec) and clonal selection method (0.07 sec).The computation time for fuzzy C-means method (0,29 sec) and Gustafson–Kessel (0.31 sec have the same order of magnitude [5, 6].

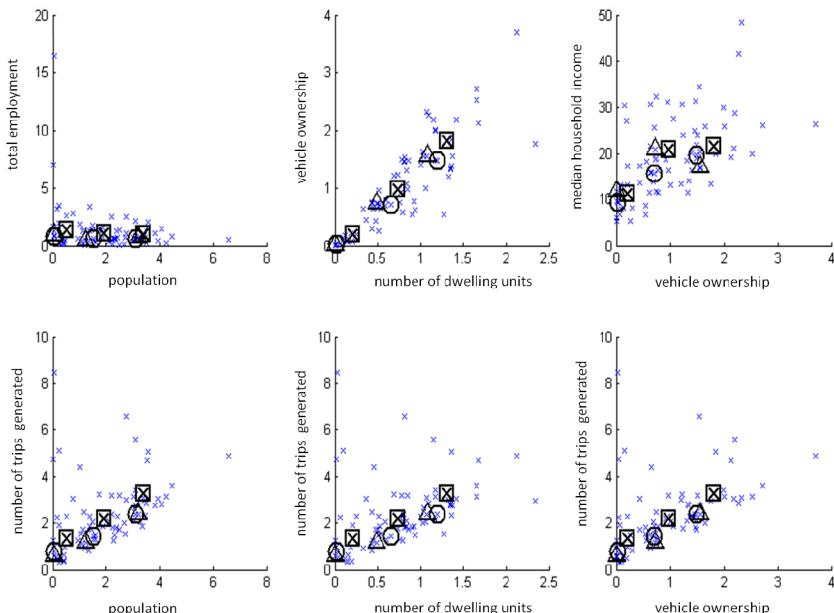


Fig. 1. Trip data set (all numbers are expressed in thousands) and its cluster centers calculated with different methods, \times – fuzzy C-means, \square – Gustafson–Kessel, \circ – subtractive clustering, \triangle – clonal selection

5.2 Determination of a Fire Hazard Index from Fuzzy Model

DWP measurement of data set is based on the measurement of data obtained from the multi-detector sensor DWP-1 performed at the Institute of Innovative Technologies EMAG within the European project, EDAFFIC. The sensor was working in the environment of high methane and coal dust explosion risk [8]. By using special detectors (carbon monoxide, hydrogen cyanide and smoke) various parameters are controlled and a fire may be detected in its early stage. The main idea is to create fuzzy model to determine the index of fire risk based on independent, pre-processed analog signals.

The proposed fuzzy model of a fire hazard index describes the problem of fire detection in the mine. The process of creating such a fuzzy model is exemplified by the DWP data set. Input measurement data – 3 input (580 measurements for each value) as follows:

- Detector CzU: smoke – range data: 413 ÷ 2036 mV,
- Detector KCO: carbon monoxide (CO) – range: 0 ÷ 14.2 ppm,
- Detector KHCN: hydrogen cyanide (HCN) – range: 0 ÷ 5.4 ppm.
- Output data – fire hazard index, range data: 29.94 ÷ 351.6.

In the laboratory tests, the air physical parameters such as: temperature, humidity and pressure do not change. Laboratory measurement data were processed by clustering algorithms [5, 6, 7]. In Tab. 1 are presented coordinates of the calculated cluster centers for 4 mentioned above clustering methods. Analyzing the clustering quality indexes for each method one can see that all the methods perform better for the 5 appointed clusters. For all clustering methods the location of clusters is similar. The obtained clustering quality indexes are good and it ranges from 0.838 to 0.850 (partition index) and from 0.338 to 0.357 (entropy index).

Table 1. DWP data set: coordinates of the calculated cluster centers

Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
fuzzy C-means ($c = 5$)	418.95 1.22 1.64 33.93	1264.29 8.58 4.19 45.41	1842.35 12.62 5.17 340.74	1774.11 1.83 1.42 166.32	1992.24 3.84 4.77 255.65
Gustafson– Kessel ($c = 5$)	418.94 1.22 1.64 33.93	1264.40 8.58 4.19 45.41	1842.36 12.62 5.17 340.74	1773.5 3 1.83 1.42 166.25	1992.20 3.84 4.77 255.63
Subtractive ($r_a = 0.5$)	413.00 1.20 1.70 32.22	1319.00 9.20 4.30 41.28	1856.00 13.00 5.20 344.69	1964.0 2.00 2.00 187.60	2035.0 3.90 5.10 251.60
Clonal selection ($r_k = 0.7$)	413.00 1.20 1.70 31.97	1143.90 7.44 4.20 39.44	1856.00 13.00 5.20 344.69	1928.0 1.60 1.60 160.43	2035.0 3.90 5.10 249.79

In case of Xie–Bieni and Fukuyama–Sugeno indexes one can see, as in the other test sets of numerical data, that the methods which require to declare the number of clusters (fuzzy C-means and Gustafson–Kessel) have better indexes values than indexes obtained via methods that independently determine the number of clusters.

Taking into account the computation time found that the fastest is subtractive method (0.82 sec), the second in sequence is the method of clonal selection (1.54 s). These algorithms perform the task of clustering process faster than fuzzy C-means and Gustafson–Kessel methods. On Fig. 2 are presented DWP data sets together with calculated cluster centers for different clustering methods such as fuzzy C-means, Gustafson–Kessel, subtractive and clonal selection [5, 6, 7] are presented.

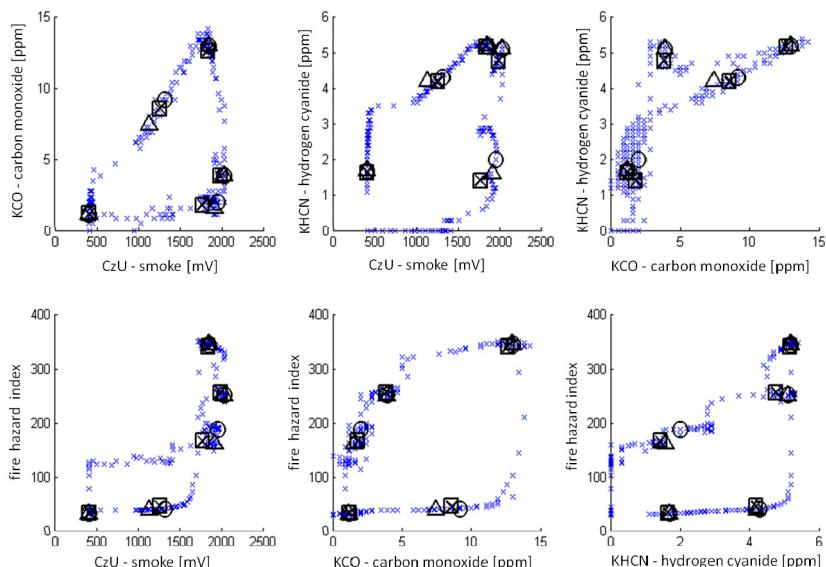


Fig. 2. DWP data set and its calculated cluster centers with methods: \times – fuzzy C-means, \square – Gustafson–Kessel, \circ – subtractive clustering, \triangle – clonal selection

Number of inputs and outputs of the fuzzy model corresponds to the number of data sets. The three inputs determine the values measured by the sensors, they are: carbon monoxide, hydrogen cyanide and smoke. Output values determine the index fire hazard. Laboratory tests fuzzy model for a fire hazard index, with pre-selected parameters, gave good results.

6 Conclusion and Future Work

Fuzzy models allow describing a very complex phenomenon that are difficult to be represented in the form of mathematical formulas. Therefore they may be a useful complement to conventional methods of acquiring knowledge. Efficient methods for extracting fuzzy rules from the high dimensional numerical data based on cluster estimation have been presented. Fuzzy rules can be generated to solve function approximation problems for a variety of technical applications.

Here the fuzzy model to determine the index of fire risk based on independent, pre-processed analog signals was presented. The proposed fuzzy model describes the problem of fire detection in the mine, in its early stages. The model structure and antecedent parameters was extracted by using 4 clustering methods.

Clustering method with applied clonal selection mechanism was implemented in MATLAB. This method was tested with *Trip* data and compare with subtractive, fuzzy C-means, Gustafson-Kessel clustering algorithms. Two experiments with *Trip* and *DWP* data indicate that clonal selection clustering algorithm is computationally fast and robust.

The future works will involve the applications clonal selection mechanism for optimization antecedent parameters of Sugeno type fuzzy models.

References

1. Chiu, S.: Fuzzy model identification based on cluster estimation. *J. of Intelligent and Fuzzy Systems* 2, 267–278 (1994)
2. Chiu, S.: Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification. In: Dubois, D., Prade, H., Yager, R. (eds.) *Fuzzy Information Engineering: A Guided Tour of Application*, ch. 9. John Wiley&Sons (1997)
3. Fuzzy Logic Toolbox User's Guide, Version 5.0 (R2010b), The MathWorks, Inc.
4. Łęski, J.: Systemy neuronowo – rozmyte. WNT, Warszawa (2008) (in polish)
5. Felka, D.: Metody budowy inteligentnych modeli na bazie danych numerycznych, Innowacyjne rozwiązania w obszarze automatyki, robotyki i pomiarów. In: Kacprzyk, J. (ed.) Konkurs Młodzi Innowacyjni 2012, pp. 75–88. PIAP Warszawa (2012) (in polish)
6. Felka D.: Metody budowy inteligentnych modeli na bazie danych numerycznych, diploma dissertation, supervisor dr inż. B. Mrozek, PK, Kraków (2011) (in polish)
7. Mrozek, B., Felka, D.: Inteligentny model wskaźnika zagrożenia pożarowego w kopalni węgla. In: Annual Conference Automation 2012, Pomiary Automatyka Robotyka 2/2012, Warszawa, pp. 540–545 (2012) (in polish)
8. Mróz, J., Broja, A., Małachowski, M., Szczygielska, M.: Środowisko EDAFFIC (Elary Detection And Fighting in belt Conveyor). Opracowanie czujnika wielodetektorowego oraz jego budowa i testy, Projekt badawczy EDAFFIC finansowany ze środków EU w ramach programu Coal & Steel nr RFCR-CT-2008-00002, Instytut Technik Innowacyjnych EMAG, Katowice (2010)
9. de Castro, L.N., von Zuben, F.J.: Artificial Immune Systems: Part I – Basic Theory and Applications, Technical Report – RT DCA 01/99 (1999)
10. de Castro, L.N., Von Zuben, F.J.: Learning and Optimization Using the Clonal Selection Principle. *IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems* 6(3) (2002)
11. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. on Systems, Man & Cybernetics* 15, 116–132 (1985)
12. Wang, H., Zhao, L., Du, W., Qian, F.: A hybrid method for identifying T-S fuzzy models. In: 8th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shanghai, vol. 1, pp. 11–15 (2011)
13. Wong, E., Lau, H.: Advancement in the twentieth century in artificial immune systems for optimization: review and future outlook. In: *IEEE International Conference on Systems, Man, and Cybernetics*, San Antonio, pp. 4195–4202 (2009)

Augmented Reality of Technological Environment in Correlation with Brain Computer Interfaces for Control Processes

Szczepan Paszkiel*

Opole University of Technology,
Faculty of Electrical Engineering, Automatic Control and Informatics,
Institute of Control and Computer Engineering
Gen. Kazimierza Sosnkowskiego 31, 45-271 Opole, Poland
s.paszkiel@po.opole.pl <http://www.s.paszkiel.po.opole.pl>

Abstract. The article describes the idea of linking solutions from Brain Computer Interfaces technology with augmented reality based on a Google device. A system composed of devices such as MindWave Mobile and Google Glass, which support the process of controlling mobile robots without a need of issuing voice commands, an identification of an eye movement or the use of upper and lower limbs, is presented. The article discusses parameters of Google Glass for control processes. A unit of MindWave Mobile by NeuroSky is characterised as well.

Keywords: Brain Computer Interfaces, augmented reality, mobile robots, control process, MindWave Mobile, Google Glass.

1 Introduction

The development of technical sciences and the correlation of conclusions from the conducted research, including the ever-increasing knowledge of the mechanisms of the human brain, implies the possibility of developing new solutions that can have a very wide field of applications. Augmented Reality is a relatively new field that has grown on the basis of the technological environment, which has its origins in the 60s of the twentieth century. Nowadays it is supported in a quite significant way by Google Glass. This device may become very common in the near future. Brain-computer interfaces, which develop rapidly for several years, are an excellent example of a technology that is in line with the immersion reality. This technology may be an interesting tool among other things for the implementation of control processes including mobile robots.

2 Brain Computer Interface Technology

Brain-computer interfaces (BCI) allow direct control of different applications installed on a workstation and the work of mobile robots using brain. BCI tech-

* Dr. eng. Szczepan Paszkiel, Assistant Professor.

nology is a clearly interdisciplinary issue combining sciences such as biomedical engineering, medicine, signal analyses, electronics and information technology [1]. BCI interfaces are created mainly in order to help people with disabilities and patients in severe stages of diseases such as multiple sclerosis, stroke, amyotrophic lateral sclerosis or Guillain-Barre syndrome. But this is not the only use of this type of interfaces. They are also implementations in the military system, entertainment and media industries. Emotiv System Inc. and NeuroSky implement the above interfaces to support the service of computer games. There are two main types of brain-computer interfaces because of the reception of information from the human brain: interfaces based on the invasive method lie in the implementation of measurement electrodes directly into the brain of a patient and interfaces based on the non-invasive method - they are often used due to the lack of interference in the human structure. Unfortunately, BCI based on EEG (the non-invasive method) is characterized by high complexity of the EEG signal resulting from disruptions that occur in this type of signal. This is connected with the fact that the non-invasive solution is based on a readout of a signal from the surface of a skull. Brain Computer Interfaces already developed are based on one of three paradigms: SCP - free cortical potentials characterized by a free communication; P300 - evoked potentials used to design synchronous BCI constructions; ERD/ERS a desynchronization/synchronization associated with a stimulus. This phenomenon is to change the amplitude of the oscillations occurring in some bands when planning a move. A major role in the creation of BCI interfaces played Polish neurophysiologist Adolf Beck. He was the explorer of EEG desynchronization in response to stimuli. The main aim of his study was the use of electrophysiology for the location of functions in the brain. P300 evoked potentials are, in turn, characteristic brain responses to external stimuli. Research in this area is carried out by repeating external stimuli in time. The test person is subjected to these stimuli. Changes in the activity of neurons repeated in time are observed in the record of the EEG signal [2]. Other signal components unrelated to the stimulus are stochastic. Habituation effect should be taken into account during the analysis of the EEG signal based on evoked potentials. It results from weakening of potentials evoked by subsequent repetitions of the stimulus. Therefore, when the EEG signal is modeled based on brain responses to various external stimuli a particular attention should be paid to the deterministic repeatability, which may not always be correct [3]. Research for a long time in this area seeks to determine the appropriate parameterization of single evoked potentials. In the EEG signal evoked potentials are observed as changes in the amplitude of the signal in a given period of time. A proper synchronization of individual neurons activities affects the output signal extracted by the electroencephalograph [4]. A structure of connections between particular network elements and synaptic processes forming local circuits affect directly the synchronization of neurons. The factors that directly affect the synchronous oscillations are among others the neural membrane properties, the influence

of neurotransmitters, the structure of connections between particular network elements and synaptic processes. Many systems used for controlling mobile robots and many applications supporting the processes of acquisition and the classification of EEG signals have originated on the basis of BCI technology among others in the Laboratory of Biomedical Measurements at Opole University of Technology [5],[6].

3 Immersion Environment and Augmented Reality

Immersion allows you to "absorb" the human being into the electronic reality. Then you have to deal with the so-called immersion of the senses. Immersion can be seen in many cases, including in the area of arts, culture and technological environment. In our case the latest range of immersion is the most important from the point of view of the technology [7]. Immersion in this range is directly connected with the electronics of advanced immersive technologies. An example of immersive technologies already developed is a system of Cave Automatic Virtual Environment and Head Mounted Displays. Cave Automatic Virtual Environment (CAVE) creates the environment by organizing an enclosed space in the area perceived by reality. Technology is the projection of digital images on the walls, at the same time it is based on the three-dimensionality. Head Mounted Displays (HMD) is a technology which allows a man to enter the world of virtual reality thanks to a device attached to a head in the shape of a helmet and a suit holding sensors. The main difference between the CAVE and HMD technology is that in the case of the CAVE technology a recipient is surrounded by environment, and the system recipient of the HMD technology is in its own space created in a digital way [8]. In science, there is the concept of reality known as Augmented reality (AR), which combines real and virtual world. Under this system, with the help of special glasses we are able to observe the world around us in the streets and the elements produced by virtual reality. The assumptions, on the basis of which Augmented reality is based on, are mentioned before a connection of both worlds, a real-time interaction and freedom of movement in three dimensions. It should be noticed that, among others, at the Massachusetts Institute of Technology (MIT) more and more lectures for the students are conducted on the basis of Augmented reality. Students use their smart phones and GPS devices gradually exploring previously enriched in essence campus sites for learning. Augmented reality can be used in the study to obtain information on subjects in which students work (as is the case at MIT), by immediate verification and electronic feedback in the form of information; in medicine an access to data on the structure of the internal organs of a patient; in marketing, and the most significant area in terms of this article in robotics by identifying objects among which the robot moves and supporting the generation of potentials evoked in the brain of a person who controls the robot motion by using BCI technology in the feedback loop.

4 A Device Which Operates on the Basis of BCI Technology

NeuroSky is a manufacturer of devices which are technologically based on the idea of brain-computer interfaces. The wide range of products includes among others MindWave Mobile which is a device shaped like headphones with a ground electrode attached to your ear. This device is based on the EEG. The electroencephalographic signal is measured during its work and thus activities of individual brain waves are observed. Based on the above fact we can make the process of control and influence in different ways to different objects in space including the movement of a mobile robot. When testing the device in the Laboratory of Biomedical Measurements at Opole University of Technology, it was possible to use it in order to evoke intentional activities in terms of the control process. Stimulation of the brain, by driving it in certain states of meditation and enhanced thinking, brought practical results. An identification of changes in the amplitude of the EEG signal in correlation with induced potentials enabled the use of the product to carry out simple tasks. Figure 1 shows a visualization of brain wave activities identified by the application available with MindWave Mobile.

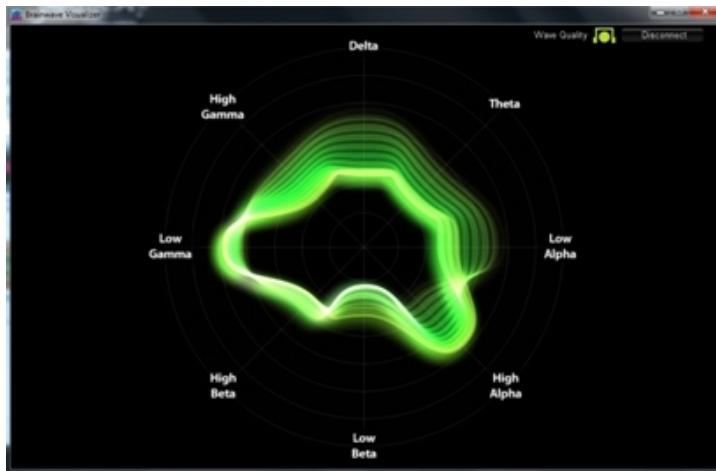


Fig. 1. A window showing the visualization of brain wave activities identified by the applications available with MindWave Mobile by NeuroSky

5 A Device Which Operates on the Basis of Augmented Reality

Google Glass is a typical example of a device based on partial immersion, namely on augmented reality. Eventually, as Google plans the glasses may serve as a

smartphone. However, due to some design solutions they can operate on the basis of the supplementary device of control processes based on the technology brain-computer interfaces. The Google Glass works on the Android operating system. It has combined its duties with prescription glasses which are used by people with vision defects. Therefore, it does not discriminate against this group of people in terms of inability to use the above-mentioned technologies in practice. Implementing Google's Android operating system on the device will allow free access to many free applications and facilitate the implementation of a connection between the BCI technology and solutions of immersed reality. Image resolution displayed by Google Glass, even if it is not high, enables to use the display as a part which supports the control process by using the BCI technology. The Google Glass is also equipped with 682MB of RAM and 16 GB of storage, which meets the requirements of the proposed system solution presented in the next section of this article. Built-in microUSB port allows you to charge the battery in the device. It will also be possible to use the developed system at night in poor visibility by implementing HDR technology by Google in their product. Google on your device implemented a mechanism for transmitting voice to the person using the glasses using bone conduction. A sound wave is passed through a built-in speaker rests on the human temporal bone directly into the middle ear. Unfortunately, it could negatively affect the EEG signal acquisition process, and therefore the functionality of the device with the proposed system will not be used. Another feature of Google Glass is that the device is controlled by voice and touch. In terms of the proposed system these two control methods will not be necessary. Using a brain-computer interface allows the use of the device for control processes without the need of controlling it. Supply of the device becomes an important issue from the point of view of control processes in long continuous time. Currently, there are already Google Glass models that have a capacious battery and that is why they can work longer.

6 Applications in Control Processes

Controlling by means of the human mind without the use of evoked potentials is currently difficult in terms of implementations in everyday conditions as evidenced by the authors neuromarketing studies. Nevertheless, the control based on evoked potentials is relatively easy to implement. With Google Glass paintings on glass can be emitted, thereby stimulating the brain in terms of inducing the desired changes in the amplitude of the brain waves to avoid manipulations of the limbs or eyes closed or using voice commands. In the next step on this basis the control process of an application on both the workstation and mobile robots which explore the site using the correlation of brain-computer interface with Google Glass glasses can be carried out. Of course, images on the glass would be directly linked to what the eye of the person making the process control sees and their corresponding formula would affect positively the stimulation of the brain in terms of the control process. The following figure provides an example of the correlation between the state of the human brain activity, different

State of the human brain	Brain wave activity	Shown image	Planned action
Meditation	Alpha waves	Forest, sea	Stopping the movement of the robot
Increased activity	Gamma waves	The problem to solve - the task	Moving the robot forward

Fig. 2. Examples of connections between the state of the human brain, brain wave activity, shown images and the planned action

types of brain waves, presented images, and the planned action connected with controlling the mobile robot.

Currently, tests have been conducted on controlling Quadro Helicopter with Google Glass. A camera in Google Glass identified the position of the head of the person controlling the Quadro Helicopter due to the reference point which in this case was the floor. Head movements affected the movement with the Quadro Helicopter. It should be born in mind that Google Glass technology also found its application in medicine. A case of a doctor who operates a patient using Google Glass in terms of videoconferencing through Google Glass with a camera and transferring information to their trainees about the details of the operation is known. This action was possible by using Google Hangouts. During the development of the system and the analysis for and against Google's use of the device the problem of wind which can cause noise in the speakers and the microphone is encountered. Nevertheless, the proposed system does not use the part of the auditory device that works without a microphone and a speaker. Therefore, this type of interference will not affect the operation of the proposed system. Figure 3 shows the overall architecture of the system based on Augmented Reality and Brain Computer Interfaces technology. Person's eyes and brain are the two main control units without which the proposed system will not

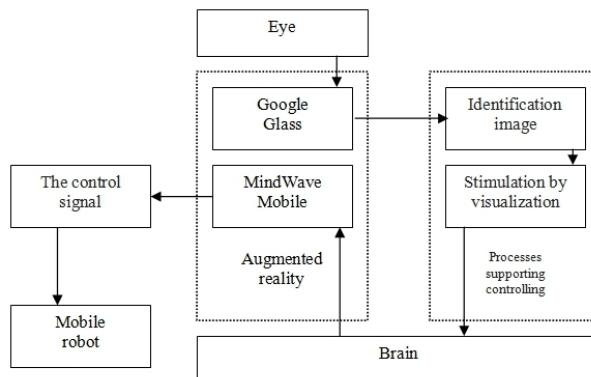


Fig. 3. Outline of the idea of using Augmented Reality with BCI technology

work. Its advantage is the fact that both voice signals and the need for the use of upper and lower limbs are not necessary. The system is based on Augmented Reality in which the main role is played by Google Glass. A block of the system known as processes which support controlling consists of the identification of an image observed by the user of the system and the stimulation based on the visualization of images directly related to the observed reality. A control signal is identified for the mobile robot control processes after producing the appropriate response in the brain by means of MindWave Mobile by NeuroSky.

7 Summary

Currently there are practical implementations of technology based on augmented reality among other things in the entertainment industry for the construction of urban games. Augmented reality can also be used among other things to scan products at shelves in a store, for an immediate comparison of their composition, caloric testing, prices, etc.. The development work is carried out in the context of the People+ or the face identification of people with whom interviews are conducted to obtain information about them. The field to use Augmented reality is wider and wider. BCI Technology can become in the future a great addition to the above-mentioned technologies. It may be a very useful tool especially for people with paralysis or speech impairment in the processes of controlling mobile robots.

References

1. Tadeusiewicz, R., et al.: Neurocybernetyka teoretyczna. Wydawnictwo Uniwersytetu Warszawskiego, Warszawa (2009)
2. Bear, M.F., Connors, B.W., Paradiso, M.A.: Neuroscience: Exploring the Brain. Lippincott, Philadelphia (2006)
3. Andreasen, N.C.: Brave New Brain: Conquering Mental Illness in the Era of the Genome. Oxford University Press (2004)
4. Paszkiel, S.: Modelling of the EEG signal in the aspect of disruptive artefacts. In: Materiały III International Interdisciplinary Technical Conference of Young Scientists InterTech, Poznan, May 19-21, pp. 218–220 (2010)
5. Blachowicz, A., Paszkiel, S.: A mobile system for measurements of incomplete discharges controlled by electroencephalographic waves. Journal of Automation, Mobile Robotics and Intelligent Systems, JAMRIS 4(3), 31–35 (2010)
6. Paszkiel, S., Blachowicz, A.: The application of electroencephalographic signals in the aspect of controlling a mobile robot for measurements of incomplete discharges. Przegląd Elektrotechniczny 86(8), 303–306 (2010)
7. Burdea, G.C., Coiffet, P.: Virtual Reality Technology. John Wiley and Sons, Inc. (2003)
8. Supan, P., Haller, M., Stuppacher, I.: Image based shadowing in real-time augmented reality. International Journal of Virtual Reality (2006)

Modeling and Dynamic Analysis of the Precise Electromechanical Systems Driven by the Stepping Motors

Agnieszka Pręgowska¹, Tomasz Szolc¹, Andrzej Pochanke², and Robert Konowrocki¹

¹ Institute of Fundamental Technological Research of the Polish Academy of Sciences,
ul. Pawińskiego 5B, 02-106 Warsaw,
aprego@ippt.gov.pl

² Faculty of Electrical Engineering of the Warsaw University of Technology,
Pl. Politechniki 1, 00-661 Warsaw
andrzej.pochanke@ee.pw.edu.pl

Abstract. In the paper there is investigated electromechanical dynamic interaction between the driving stepping motor and the driven laboratory belt-transporter system imitating an operation of the robotic device in the form of working tool-carrier under translational motion. The considerations are performed by means of the circuit model of the electric motor and the discrete, non-linear model of the mechanical system. In the computational examples various scenarios of the working tool-carrier motion and positioning by the belt-transporter are simulated.

Keywords: electromechanical interaction, stepping motor, drive system, belt-transporter, transient and steady-state vibrations, tool-carrier positioning.

1 Introduction

The stepping motors are commonly known sources of power usually applied for a possibly exact positioning of selected elements of driven precise mechanical systems. On the one hand, in such systems an accuracy of positioning essentially depends on electrical properties and proper control of the driving motor. But on the other hand, this accuracy follows from flexibility and dynamic properties of the driven object. Here, because the flexible mechanical systems usually indicate a natural ability of vibrations causing a fluctuation of the angular velocity of the stepping motor rotor, the flows of electric currents in the motor windings become affected by these mechanical oscillations, which results in additional variable components of the driving electromagnetic torque generated by the stepping motor. According to the above, the electrical current oscillations are coupled with mechanical vibrations of the driven object. In order to assure a possibly accurate positioning of elements of the mechanical systems driven by the stepping motors, the phenomenon of electromechanical interaction between the mechanical and electrical parts should be thoroughly investigated theoretically and experimentally.

This problem has been considered till present by many authors, but usually from the viewpoint of modeling of the stepping motor. Majority of them have applied the circuit models of the stepping motor and the driven mechanical system was reduced to only one rigid body, the mass moment of inertia of which is added to that of the motor rotor, as e.g. in [1–4]. In [1] one can find fundamentals of electrical modeling of the stepping motor. The main target of investigations performed in [2–4] was a possibly the most exact control of the stepping motor operation in order to assure a precise angular positioning of the rotor. For this purpose in [2] an influence of electrical parameters of the motor and its controller has been studied. In [3, 4] various stepping motor controllers were developed, where in [4] also higher harmonic components of the motor electromagnetic torque have been taken into consideration for an analysis of a possible resonance excitation. Some introductory attempts to investigations of dynamic interaction between the driving stepping motor and the driven mechanical system can be found in [5, 6]. In [5] there are presented quantitative simulation examples of torsional vibrations of the drive train in the form of few mutually visco-elastically connected rigid bodies excited by the stepping motor. One of the first approach to more qualitative analysis of the circuit model of the stepping motor interacting with the torsionally vibrating structural model of the geared drive system, where all necessary geometrical and material parameters are taken into consideration, has been realized in [6]. In that paper, in addition to simulation examples of transient and steady-state operating conditions, also a qualitative spectral analysis of the electrical-to-mechanical response has been performed.

In the paper presented here there is studied a dynamic interaction between the typical two-phase hybrid stepping motor and the mechanical laboratory belt-transporter system representing a robotic device characterized by the moving inertial working tool-carrier under translational motion. The main goal of this work is to develop a possibly efficient and realistic electromechanical model of this object, using which an exact positioning of the tool-carrier could be realized.

2 Modeling of the Electromechanical System

The object of considerations is the abovementioned belt-transporter system representing a robotic device characterized by the moving inertial working tool-carrier under translational motion, as shown in Fig. 1. This mechanism is directly driven by the stepping motor.

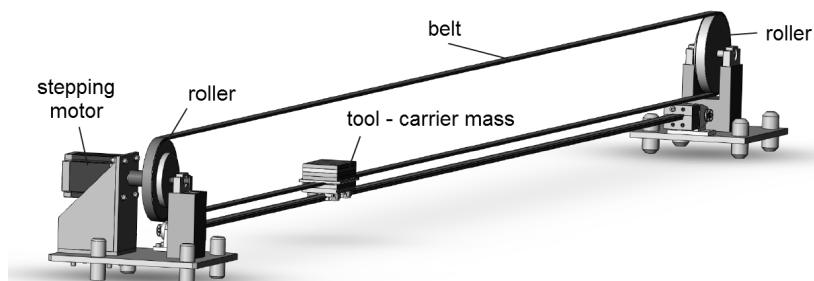


Fig. 1. The laboratory belt-transporter system driven by the stepping motor

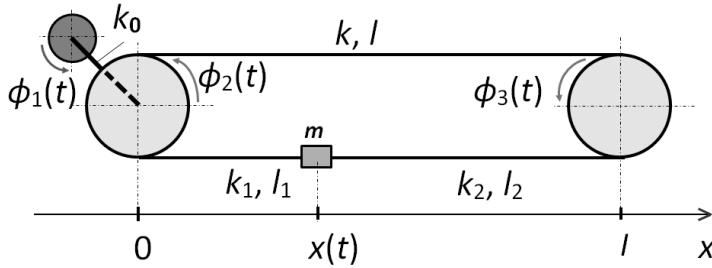


Fig. 2. Physical model of the laboratory belt-transporter system

Thus, a dynamic behavior of this mechanical system can be described by the non-linear discrete model of four degrees of freedom presented in Fig. 2. Its motion is governed by the following system of ordinary differential equations:

$$\begin{aligned}
 I_1\ddot{\phi}_1 + d_0\dot{\phi}_1 + c_0(\phi_1 - \phi_2) + k_0(\phi_1 - \phi_2) &= T_{el}(t), \\
 I_2\ddot{\phi}_2 + c_0(\phi_2 - \dot{\phi}_1) + r^2c(\phi_2 - \phi_3) + rc_1(t)(r\dot{\phi}_2 - \dot{x}) - k_0\phi_1 + [k_0 + r^2(k + k_1(t))] &\phi_2 - \\
 - r^2k\phi_3 - rk_1(t)x &= -[T_2 + F_2\dot{\phi}_2^2]\operatorname{sgn}(\dot{\phi}_2), \\
 I_3\ddot{\phi}_3 + r^2c(\phi_3 - \dot{\phi}_2) + rc_2(t)(r\dot{\phi}_3 - \dot{x}) - r^2k\phi_2 + r^2(k + k_2(t))\phi_3 - rk_2(t)x &= -[T_3 + F_3\dot{\phi}_3^2]\operatorname{sgn}(\dot{\phi}_3), \quad (1) \\
 m\ddot{x} + c_1(t)(\dot{x} - r\dot{\phi}_1) + c_2(t)(\dot{x} - r\dot{\phi}_3) - rk_1(t)\phi_2 - rk_2(t)\phi_3 + (k(t)_1 + k_2(t))x &= \\
 -[T_4 + F_4\dot{x}^2] &\operatorname{sgn}(\dot{x}),
 \end{aligned}$$

where ϕ_i , x , $i = 1, 2, 3$ are respectively the rotational and translational time-dependent generalized coordinates describing motion of the discrete mechanical model, I_i denotes the mass moment of inertia of the stepping motor rotor, I_2 , I_3 are the mass moments of inertia of the belt-transporter rollers of radius r , m is the mass of the tool-carrier, k_0 , c_0 denote respectively the torsional stiffness and material damping coefficient of the shaft segment connecting the motor rotor with the driving roller, k , c are the constant and $k_1(t)$, $c_1(t)$, $k_2(t)$, $c_2(t)$ are the variable, tool-carrier position dependent longitudinal stiffness and material damping coefficient of the pre-tensioned belt between the driving and driven roller, between the driving roller and the tool-carrier and between the tool-carrier and driven roller, respectively, see Figs. 1 and 2. The symbols $T_e(t)$ and d_0 denote the electromagnetic torque function and the absolute rotor-to-stator mechanical damping coefficient. In this model there is assumed that the retarding torques and the retarding force imposed respectively on the both rollers and

on the tool-carrier are expressed as sums of slowly varying components T_j representing dry friction effects in the drive system and of the square functions of the current rotational or translational speed $F_j \omega_j^2(t)$, where $j = 2, 3, 4$ and $\omega_j(t) = \dot{\phi}_j$ for $j = 2, 3$, $\omega_j(t) = \dot{x}$ for $j = 4$ and F_j denote the proper constant coefficients.

Here, for the assumed proportional material damping in the system the constant and variable belt longitudinal stiffness and damping coefficients can be determined by means of the following formulae:

$$c = \gamma k, c_i(t) = \gamma k_i(t), i = 1, 2, k = \frac{EA}{l}, k_1(t) = \frac{EA}{x(t)}, k_2(t) = \frac{EA}{l - x(t)},$$

where EA is the belt longitudinal cross-sectional stiffness, l denotes the distance between the roller axles, see Fig. 2, and γ is the constant material loss factor.

According e.g. to [1,5], the circuit model of the two-phase hybrid stepping motor can be described by two voltage ordinary differential equations of the first order:

$$\begin{aligned} \frac{di_1}{dt} & [L_0 + L_2 \cos(2Z_r \Theta)] + \frac{di_2}{dt} [L_2 \sin(2Z_r \Theta)] - \frac{d\Theta}{dt} [2L_2 Z_r i_1 \sin(2Z_r \Theta)] + \\ & + \frac{d\Theta}{dt} [2L_2 Z_r i_2 \cos(2Z_r \Theta)] - \frac{d\Theta}{dt} [K_U \sin(Z_r \gamma \Theta)] + R i_1 = -U(t) \operatorname{sgn}\{\sin(\Phi(t))\}, \\ \frac{di_2}{dt} & [L_0 - L_2 \cos(2Z_r \Theta)] + \frac{di_1}{dt} [L_2 \sin(2Z_r \Theta)] + \frac{d\Theta}{dt} [2L_2 Z_r i_2 \sin(2Z_r \Theta)] + \\ & + \frac{d\Theta}{dt} [2L_2 Z_r i_1 \cos(2Z_r \Theta)] + \frac{d\Theta}{dt} [K_U \cos(Z_r \Theta)] + R i_2 = U(t) \operatorname{sgn}\{\cos(\Phi(t))\}, \Phi(t) = \frac{\pi}{2} \int_0^t f_e(\tau) d\tau, \end{aligned} \quad (2)$$

where $i_1(t)$, $i_2(t)$ denote the electric currents in both motor phases, L_0 is the phase inductance, L_2 is the inductance following from reluctance fluctuation of the magnetic circuit, R denotes the resistance of the one phase, K_U is the motor voltage constant, $\Theta(t)$ denotes the instantaneous rotation angle of the rotor including the rigid body motion and the vibratory component, $U(t)$ is the slowly varying control voltage, Z_r denotes the rotor pole number and $f_e(t)$ is the voltage supply commutation frequency. Here, a sufficiently good commutation realized by means of a proper stepping motor control should result in the control voltage supply phase angle $\Phi(t) \equiv Z_r \Theta(t)$.

The electromagnetic torque generated by such a double-phase stepping motor is expressed by the following formula:

$$\begin{aligned} T_{el}(t) = K_T & [-i_1 \sin(Z_r \Theta) + i_2 \cos(Z_r \Theta)] + L_2 Z_r (-i_1^2 + i_2^2) \sin(2Z_r \Theta) + \\ & + 2L_2 Z_r i_1 i_2 \cos(2Z_r \Theta), \end{aligned} \quad (3)$$

where K_T denotes the stepping motor torque constant.

As it follows e.g. from [1,5], a fluctuation of the rotor-to-stator reluctance in the considered type of the stepping motor is small enough to neglect L_2 in Eqs. (2) and (3). Next, if the rotor rotation angle $\Theta(t)$ can be equated to the generalized coordinate φ_1 in (1), one obtains the following system of coupled non-linear ordinary differential equations of motion of the determined in this way electromechanical model of the considered object:

$$\mathbf{M} \cdot \ddot{\mathbf{s}}(t) + \mathbf{C}(\mathbf{s}(t), Z_r \Theta(t)) \cdot \dot{\mathbf{s}}(t) + \mathbf{K}(\mathbf{s}(t), Z_r \Theta(t)) \cdot \mathbf{s}(t) = \mathbf{F}(t, \dot{\mathbf{s}}(t)), \quad (4)$$

$$\text{where: } \mathbf{C}(\mathbf{s}(t), Z_r \Theta(t)) = \mathbf{C}_0 + \mathbf{C}(\mathbf{s}(t)) + \mathbf{C}_E(Z_r \Theta(t)),$$

$$\mathbf{K}(\mathbf{s}(t), Z_r \Theta(t)) = \mathbf{K}_0 + \mathbf{K}(\mathbf{s}(t)) + \mathbf{K}_E(Z_r \Theta(t)),$$

$$\mathbf{s}(t) = \text{col}[i_1(t), i_2(t), \varphi_1(t), \varphi_2(t), \varphi_3(t), x(t)],$$

$$\mathbf{F}(t, \dot{\mathbf{s}}(t)) = \begin{bmatrix} -U(t) \cdot \text{sgn}\{\sin(\Phi(t))\} \\ U(t) \cdot \text{sgn}\{\cos(\Phi(t))\} \\ 0 \\ -\left(T_2 + F_2 \dot{\varphi}_2^2(t)\right) \cdot \text{sgn}(\dot{\varphi}_2(t)) \\ -\left(T_3 + F_3 \dot{\varphi}_3^2(t)\right) \cdot \text{sgn}(\dot{\varphi}_3(t)) \\ -\left(T_4 + F_4 \dot{x}^2(t)\right) \cdot \text{sgn}(\dot{x}(t)) \end{bmatrix}.$$

The symbols \mathbf{M} , \mathbf{C}_0 and \mathbf{K}_0 denote, respectively, the constant mass, damping and stiffness matrices, $\mathbf{C}(\mathbf{s}(t))$, $\mathbf{K}(\mathbf{s}(t))$ are the symmetrical, tool-carrier position dependent damping and stiffness matrices, $\mathbf{C}_E(Z_r \Theta(t))$ is the band matrix of the inductive-electromagnetic effects and $\mathbf{K}_E(Z_r \Theta(t))$ denotes the band matrix of the resistive-electromagnetic effects, both of harmonically variable coefficients with the frequency following from the current electric rotation angle. The symbol $\mathbf{F}(t, \dot{\mathbf{s}}(t))$ denotes the external excitation vector due to the control input voltage and the retarding torques. The unknown co-ordinate vector $\mathbf{s}(t)$ consists of the electric currents in both motor phases and of the unknown generalized coordinates describing motion of the mechanical model.

In order to perform a qualitative analysis of the obtained in this way electromechanical mathematical model of the considered object, some linearization of the mechanical system described by Eqs. (1) seem to be rather difficult at this stage, since its dynamic properties are essentially dependent on the current translational positions of the inertial tool-carrier with respect of the both rollers, see Figs. 1 and 2. Namely, the

first natural frequencies obtained by means of eigenproblem solution formulated for Eqs. (1) vary of ca 6% only for the tool-carrier central and extreme left-hand and right-hand positions on the belt. However, these differences reach respectively 17.2%, 59.5% and 139.3% in the case of the second, third and the fourth natural frequency of the mechanical system. Thus, before some selected linearization assumptions will be made for a qualitative analysis for the considered object, at the first attempt, the nonlinear system of Eqs. (4) can be solved numerically by means of a direct integration using Newmark's method in order to obtain the system's dynamic response due to 'a priori' assumed modes of the stepping motor control. In this way, numerous scenarios of the electromechanical system start-ups, steady-state operations and run-downs can be realized.

3 Computational Examples

The fundamental target to be obtained in the computational examples is to position the tool-carrier into the given point between the rollers from its standstill at the left-hand side extreme belt location within the required time instant. For this purpose the tool-carrier should be properly accelerated and then decelerated back to its standstill. Such movement can be realized by means of several modes of the stepping motor voltage-frequency control. Each acceleration-deceleration mode can result in a given accuracy of positioning with an expected level of unavoidable vibrations of the flexible mechanical system as well as with corresponding electric current oscillations and electric energy consumption. Then, the control voltage $U(t)$ and the voltage supply commutation frequency $f_e(t)$ in (2) and (4) should be continuously being selected by the motor controller in order to realize an expected acceleration, deceleration and steady-state motion of the driven mechanical system. By the use of the simplest, constant average accelerations/deceleration modes a positioning of the tool-carrier can be achieved in various 'trapezoidal' forms consisting each of a linear speed increase, then of a motion with an average constant speed and finally of a linear speed decrease to a standstill. Here, in a particular case, the tool-carrier can be accelerated and decelerated only, i.e. without any motion phase with a constant average speed. Then, for a case of linear speed increase and decrease, one obtains a 'triangular' form of the tool-carrier positioning mode.

The exemplary computations have been carried out for parameters of the mentioned above hybrid, four-cycle, double-phase stepping motor FL110STH driving the real laboratory belt transporter system shown in Fig. 1. This motor is characterized by $L_0 = 11.5 \text{ mH}$, $R = 0.72 \Omega$, the maximal static torque 21 Nm at the corresponding current 6.5 A and by the fundamental step angle 1.8 deg = 0.0314 rad, which means

that its rotor has $Z_r = 50$ poles. The distance between the rollers with radius $r = 0.1$ m in the belt-transporter is equal to $l = 4.375$ m and the tool-carrier mass $m = 7.0$ kg.

In Fig. 3 there is shown the trajectory of the tool-carrier translational speed versus run distance from its starting left-hand side extreme location until the desired position on the belt. This result has been obtained for the case of the above defined ‘triangular’ form of the tool-carrier positioning mode characterized by the required distance $x_0 = 4.0$ m achieved within $t_0 = 2$ s. In Fig. 4 the corresponding time-histories of the system electromechanical response are presented. Fig. 4a demonstrates plots of the electromagnetic torque generated by the stepping motor (grey solid line) and of the retarding torque resulting from the motion of the driven mechanical system (black dashed line). Consequently, Fig. 4b presents the time-history plot of translational acceleration which the working tool-carrier experiences during its entire run till the final positioning. In Figs. 5a,b there are respectively depicted by the grey and black solid lines the corresponding time-histories of the control voltage and the electric currents in both motor phases.

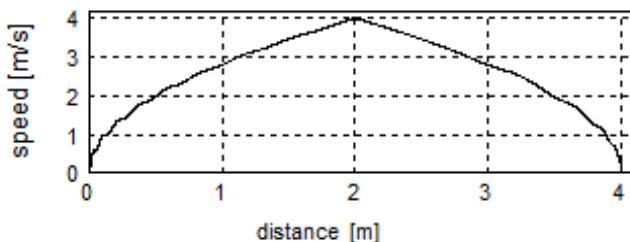


Fig. 3. The trajectory of the working tool-carrier motion for the ‘triangular’ control scenario

Figs. 6, 7a,b and 8a,b present respectively analogous results obtained for the ‘trapezoidal’ form of the tool-carrier positioning mode and for the identical as above distance x_0 achieved within the same time instant t_0 . Here, in order to run exactly the same distance within the identical time, the tool-carrier should experience appropriately greater rate of acceleration and deceleration to be able in the meantime to move with a constant average speed, which is remarkably smaller than the maximal speed obtained in the case of the ‘triangular’ positioning mode. But in the considered ‘trapezoidal’ case, the appropriately associated more rapid velocity changes result in the significantly greater mechanical and electrical vibration amplitudes which make the precise positioning effect more difficult. Namely, the greater acceleration and deceleration of the mechanical system induce much bigger transient fluctuations of the motor electromagnetic torque, translational acceleration of the tool-carrier and of the electric currents in the stepping motor windings. However, for a realization of the ‘trapezoidal’ positioning mode significantly smaller control voltage amplitudes are observed, as it follows from the respective plots in Figs. 5a and 8a.

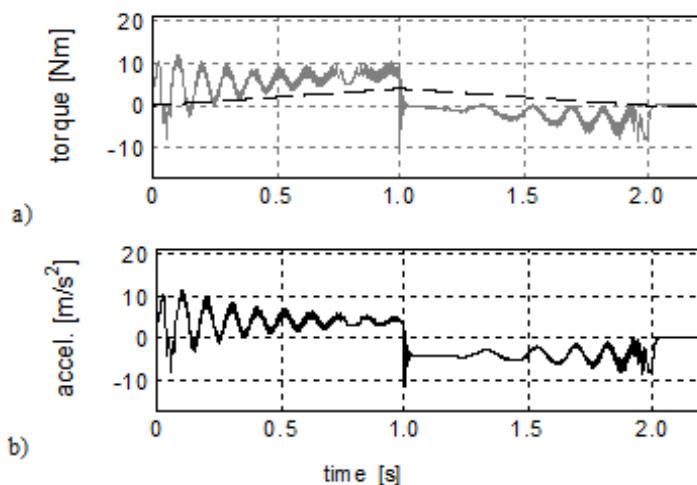


Fig. 4. Time-histories of the electromagnetic motor torque and driven system retarding torque (a) and of the tool-carrier translational acceleration (b) for the ‘triangular’ control scenario

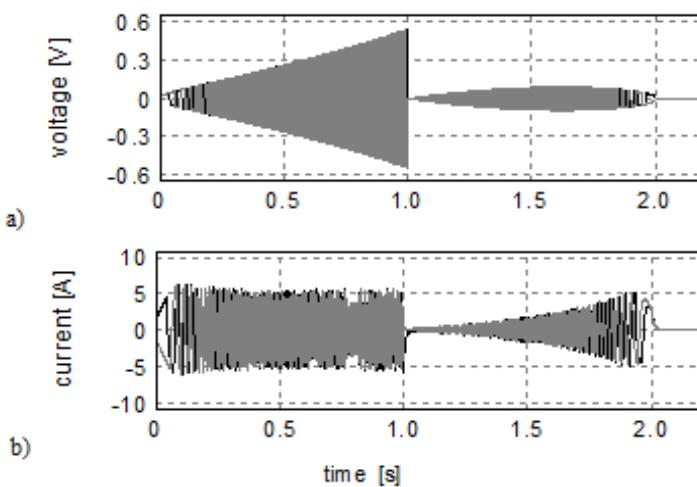


Fig. 5. Time-histories of the control voltages (a) and of the electric currents (b) in both stepping motor phases for the ‘triangular’ control scenario

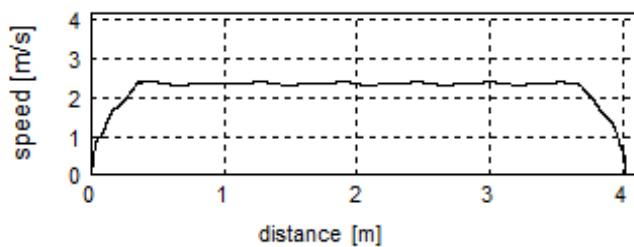


Fig. 6. The trajectory of the working tool-carrier motion for the ‘trapezoidal’ control scenario

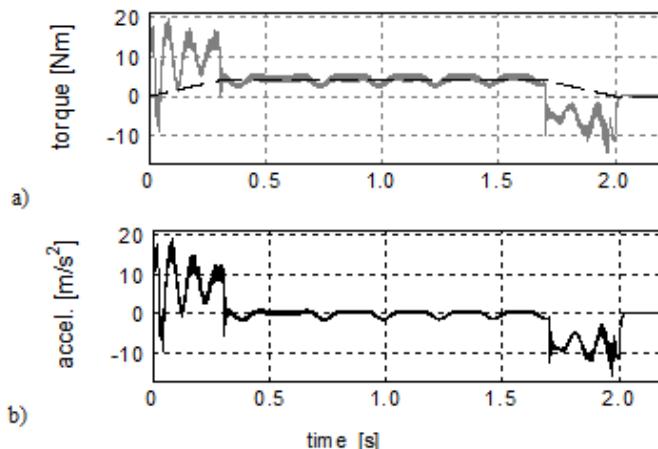


Fig. 7. Time-histories of the electromagnetic motor torque and driven system retarding torque (a) and of the tool-carrier translational acceleration (b) for the ‘trapezoidal’ control scenario

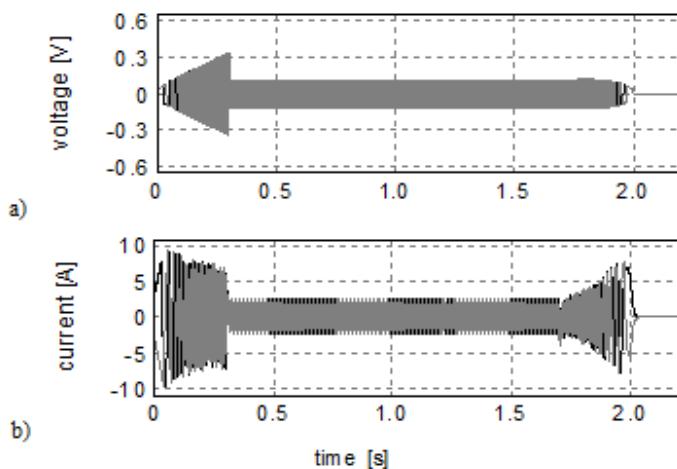


Fig. 8. Time-histories of the control voltages (a) and of the electric currents (b) in both stepping motor phases for the ‘trapezoidal’ control scenario

In analogous ways numerous positioning scenarios can be realized by means of the assumed electromechanical model of the considered laboratory belt-transporter system. Here, for various ‘trapezoidal’ positioning modes the obtained system electrodynamic responses gradually began to be similar to that for the ‘triangular’ mode when acceleration/deceleration decreased and the average tool-carrier translational speed respectively increased, i.e. where the trapezoidal shapes were being converted into the form of a triangle.

4 Final Remarks

In the paper the electromechanical model of the robotic tool-carrier moved by the belt-transporter system driven by the stepping motor has been developed. Using this model several scenarios of the tool-carrier positioning were simulated. The presented computational examples have demonstrated an influence of driven object accelerations and decelerations on transient and steady-state vibrations of the flexible mechanical system coupled with the electric current oscillations in the motor windings. Since the greater rates of accelerations/decelerations are associated, on the one hand, by smaller running speed values of the tool-carrier and smaller control voltage fluctuation amplitudes and, on the other hand, by bigger electric current and mechanical system oscillation amplitudes, optimal positioning scenarios should be searched. Thus, in the next steps of investigations in this field an optimum positioning mode is going to be determined, where possibly the greatest precision of the tool-carrier motion associated with the smallest electric energy consumption will be regarded as general objectives. Moreover, a proper experimental verification of the theoretical results is expected by the use of the test-rig being at present prepared for this purpose.

Acknowledgement. These investigations have been supported by the National Science Center: Research Project: UMO-2011/03/N/ST8/03904 titled: “Semi-active control of vibration of the drive systems of machines and vehicles using the actuators with the magneto-rheological fluid”.

References

1. Sochocki, R.: Electrical Micro-machines. Eds. of the Warsaw University of Technology, Warsaw (1996) (in Polish) ISBN 83-86569-84-0
2. Liu, X., Zhao, D.: The simulation of stepping motor under two kinds of condition. Procedia Engineering 23, 464–467 (2011)
3. Jahani, R., Chahkandi Nejad, H., Shayanfar, H.A., Zare, A.: Positioning control of PM stepper motor based on type-2 fuzzy robust control. International Journal on “Technical and Physical Problems of Engineering” 2((4)5), 19–26 (2010)

4. Wang-Hay Tsui, K., Chow Cheung, N., Chi-Wah Yuen, K.: Novel modeling and damping technique for hybrid stepper motor. *IEEE Transactions on Industrial Electronics* 56(1), 202–211 (2009)
5. Pochanke, A., Szolc, T.: Oddziaływanie elektromechaniczne w systemie napędowym z silnikiem skokowym. *Przegląd Elektrotechniczny (Electrical Review)* 87(11), 64–67 (2011) ISSN 0033-2097
6. Szolc, T., Pochanke, A.: Dynamic investigations of electromechanical coupling effects in the mechanism driven by the stepping motor. *Journal of Theoretical and Applied Mechanics* 50(2), 653–673 (2012)

Practical Stability and Asymptotic Stability of Interval Fractional Discrete-Time Linear State-Space System

Andrzej Ruszewski

Bialystok University of Technology, Faculty of Electrical Engineering,
Bialystok, Poland
andrusz@pb.edu.pl

Abstract. This paper presents the problems of robust practical stability and robust asymptotic stability of fractional-order discrete-time linear systems with uncertainty. It is supposed that the system matrix is the interval matrix and the fractional order α satisfies $0 < \alpha < 1$. Using Gershgorin's theorem for the interval matrices and the matrix measure the robust stability conditions are given. The considerations are illustrated by numerical examples.

Keywords: discrete-time linear system, fractional-order, interval, robust stability.

1 Introduction

In recent years a considerable attention has been paid to fractional calculus and its application in many areas in science and engineering (see, e.g. monographs [14, 15, 17, 20, 22] and reference therein).

The stability problem is the fundamental matter in the dynamical systems theory. This problem for linear continuous-time fractional systems has been considered in many publications (see, e.g. [4, 5, 14, 17, 19, 21]), whereas a stability problem of linear discrete-time fractional systems is more complicated and less advanced. In this case beside the asymptotic stability it is also considered the so-called practical stability defined for the length of practical implementation. The problem of practical stability of fractional discrete-time systems has been considered in [6, 7, 13, 14] for positive systems and in [6, 11, 12] for non-positive (standard) systems. Recently, the stability regions in the complex plane of fractional discrete-time linear systems were presented in [8, 18, 23]. The parametric descriptions of boundaries of these regions have been given.

The robust asymptotic stability problem of the fractional continuous-time interval systems has been studied among others in [1, 9, 16]. For the fractional discrete-time interval systems this problem has been analysed only in the case of positive systems in [3].

In this paper we focus on the robust practical stability and asymptotic stability of the fractional discrete-time interval systems (standard systems). To the best knowledge of the author this problem has not been considered yet.

2 Problem Formulation

Consider an uncertain discrete-time linear system of fractional order described by the homogeneous state equation

$$\Delta^\alpha x_{i+1} = A_I x_i, \quad \alpha \in (0, 1), \quad (1)$$

with the initial condition x_0 , where $x_i \in \mathbb{R}^n$. The system matrix $A_I \in \mathbb{R}^{n \times n}$ is the interval matrix in which all elements are known only to within a specific closed intervals defined as follows

$$A_I = \{A = [a_{ij}], b_{ij} \leq a_{ij} \leq c_{ij}, \quad i, j = 1, 2, \dots, n\}, \quad (2)$$

where b_{ij} and c_{ij} are elements of bounded matrices B and C . The system (1) is called the interval system.

The following definition of the fractional difference [13, 14] will be used

$$\Delta^\alpha x_i = \sum_{k=0}^i (-1)^k \binom{\alpha}{k} x_{i-k} \quad (3)$$

where $\alpha \in \mathbb{R}$ is the order of the fractional difference and

$$\binom{\alpha}{k} = \frac{\alpha!}{k!(\alpha-k)!}. \quad (4)$$

Using definition (3) we may write the equation (1) in the form

$$x_{i+1} = (A_I + I\alpha)x_i + \sum_{k=1}^i Ic_k(\alpha)x_{i-k}, \quad k = 1, 2, \dots \quad (5)$$

where I is the $n \times n$ identity matrix and

$$c_k(\alpha) = (-1)^k \binom{\alpha}{k+1}, \quad k = 1, 2, \dots. \quad (6)$$

The coefficients (6) can be easily calculated using the following formula [7]

$$c_{k+1}(\alpha) = c_k(\alpha) \frac{k+1-\alpha}{k+2}, \quad k = 1, 2, \dots \quad (7)$$

with $c_1(\alpha) = 0.5\alpha(1-\alpha)$.

Note that the equation (5) represents a linear discrete-time system with increasing number of delays in state. From (7) it follows that the coefficients $c_k(\alpha)$ are positive for $\alpha \in (0, 1)$ and decrease rapidly with an increase of k . Therefore, we can assume that the value of k in the equation (5) may be limited by some natural number L . This

number is called the length of the practical implementation [13]. In this case the equation (5) can be written in the form

$$x_{i+1} = (A_I + I\alpha)x_i + \sum_{k=1}^L Ic_k(\alpha)x_{i-k}, \quad k = 1, 2, \dots \quad (8)$$

The equation (8) represents the interval linear discrete-time system with L delays in state. Moreover, the system (8) is called the practical realization of the interval fractional system (1).

The definition of practical and asymptotic stability for fractional discrete-time systems has been introduced in the work [13]. With regard to the interval system (1) this definitions take the following forms.

Definition 1. The fractional interval system (1) is called robust practically stable if the system (8) is asymptotically stable for every matrix $A \in A_I \in \Re^{n \times n}$.

Definition 2. The fractional interval system (1) is called robust asymptotically stable if the system (8) is asymptotically stable with $L \rightarrow \infty$ for every matrix $A \in A_I \in \Re^{n \times n}$.

The aim of the paper is to give the robust stability conditions of the practical and asymptotic stability for the discrete-time linear interval system (1) of fractional order $\alpha \in (0, 1)$. To the best knowledge of the author such conditions have not been established yet.

3 Practical and Asymptotic Stability Regions

In paper [8] the practical stability and the asymptotic stability of system

$$\Delta^\alpha x_{i+1} = Ax_i, \quad \alpha \in (0, 1), \quad (9)$$

with precisely known matrix A has been considered. Necessary and sufficient conditions for the practical stability and for the asymptotic stability for system (9) have been established. The conditions have been given in terms of eigenvalues of the matrix $A_0 = A + I\alpha$ for the practical stability and in terms of eigenvalues of the matrix A for the asymptotic stability. In particular, it has been shown that location of all eigenvalues of the matrix A_0 or A in the stability regions is necessary and sufficient for practical and asymptotic stability, respectively. The parametric description of boundary of practical stability region denoted by $S(\alpha, L)$ in the complex plane of eigenvalues of the matrix A_0 has been given

$$\rho(\omega) = e^{j\omega} - \sum_{k=1}^L c_k(\alpha)e^{-jk\omega}, \quad \omega \in [0, 2\pi]. \quad (10)$$

Denote by $\lambda_i(A_0)$ i -th eigenvalue of $A_0 = A + I\alpha$ ($i = 1, 2, \dots, n$) and by $\lambda_i(A)$ i -th eigenvalue of A ($i = 1, 2, \dots, n$). From the paper [8] we have the following theorem.

Theorem 1. [8] The fractional system (9) with given length L of practical implementation is practically stable if and only if all eigenvalues $\lambda_i(A_0)$ ($i = 1, 2, \dots, n$) are located in the stability region $S(\alpha, L)$, i.e. $\lambda_i(A_0) \in S(\alpha, L)$ for all $i = 1, 2, \dots, n$.

Adopting a similar approach as in the work [8], we can define the following theorem.

Theorem 2. The fractional system (9) with given length L of practical implementation is practically stable if and only if all eigenvalues $\lambda_i(A)$ ($i = 1, 2, \dots, n$) are located in the stability region $S_A(\alpha, L)$, i.e. $\lambda_i(A) \in S_A(\alpha, L)$ for all $i = 1, 2, \dots, n$, where

$$\rho(\omega) = e^{j\omega} - \alpha - \sum_{k=1}^L c_k(\alpha) e^{-jk\omega}, \quad \omega \in [0, 2\pi]. \quad (11)$$

is the parametric description of boundary of stability region $S_A(\alpha, L)$, in the complex ρ -plane.

Practical stability regions $S_A(\alpha, L)$, for $L = 50$ and a few values of fractional order $\alpha \in (0, 1)$ are shown in Fig. 1 on the plane of eigenvalues of A .

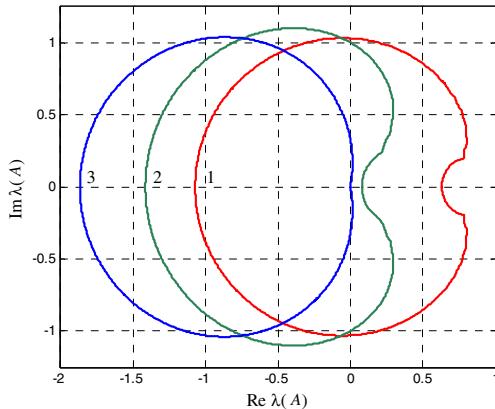


Fig. 1. Regions $S_A(\alpha, L)$ for $L = 50$ and $\alpha = 0.1$ (boundary 1), $\alpha = 0.5$ (boundary 2) and $\alpha = 0.9$ (boundary 3)

The practical stability condition of system (9) in case if all eigenvalues of the matrix A_0 are real has been given in [8].

Lemma 1. [8] If all eigenvalues $\lambda_i(A_0)$ are real, then the fractional system (9) with given length L of practical implementation is practically stable if and only if

$$-1 - \sum_{k=1}^L c_k(\alpha)(-1)^k < \lambda_i(A_0) < 1 - \sum_{k=1}^L c_k(\alpha), \quad i = 1, 2, \dots, n. \quad (12)$$

Similarly we can define the following lemma for the eigenvalues of matrix A .

Lemma 2. If all eigenvalues $\lambda_i(A)$ are real, then the fractional system (9) with given length L of practical implementation is practically stable if and only if

$$-1 - \alpha - \sum_{k=1}^L c_k(\alpha)(-1)^k < \lambda_i(A) < 1 - \alpha - \sum_{k=1}^L c_k(\alpha), \quad i = 1, 2, \dots, n. \quad (13)$$

According to the asymptotic stability of system (9) we have the following theorem and lemma.

Theorem 3. [8] The fractional system (9) is asymptotically stable if and only if all eigenvalues of the matrix A are located in the stability region, where

$$\eta(\omega) = (e^{j\omega} - 1)^\alpha (e^{j\omega})^{1-\alpha}, \quad \omega \in [0, 2\pi]. \quad (14)$$

is the parametric description of boundary of the stability region $S(\alpha)$ in the complex η -plane.

Lemma 3. [8] If all eigenvalues $\lambda_i(A)$ are real, then the fractional system (9) is asymptotically stable if and only if

$$-2^\alpha < \lambda_i(A) < 0, \quad i = 1, 2, \dots, n. \quad (15)$$

Asymptotic stability regions $S(\alpha)$ for a few values of α are shown in Fig. 2.

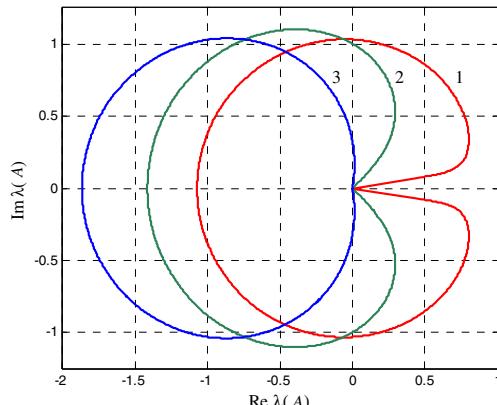


Fig. 2. Region $S(\alpha)$ for $\alpha = 0.1$ (boundary 1), $\alpha = 0.5$ (boundary 2) and $\alpha = 0.9$ (boundary 3)

4 Main Results

The eigenvalues regions of the interval matrix based on Gershgorin's theorem and the matrix measure has been presented in the paper [2]. Using the results of [2] and the parametric description of boundaries of stability regions (11) and (14) the conditions of robust practical stability and robust asymptotic stability of the fractional discrete-time interval systems (1) will be given.

Lemma 4. The interval system (1) is robust practically stable if and only if all the eigenvalues of the matrix A_I lie in the open region $S_A(\alpha, L)$.

Lemma 5. The interval system (1) is robust asymptotically stable if and only if all the eigenvalues of the matrix A_I lie in the open region $S(\alpha)$.

Proof. All eigenvalues of the matrix A_I are in the open region $S_A(\alpha, L)$ ($S(\alpha)$) if and only if all the eigenvalues of every matrix $A \in A_I$ are in this region. The proof follows from Theorem 2 (Theorem 3).

The matrix $A \in \Re^{n \times n}$ has exactly n eigenvalues, but the interval matrix A_I has an infinite number of eigenvalues. Therefore, it is impossible to check the condition given in Lemmas 4 and 5 completely.

Denote by $\lambda_i(B)$ and $\lambda_i(C)$ i -th eigenvalue of the matrices B and C ($i = 1, 2, \dots, n$), respectively.

Every element a_{ij} of the interval matrix (2) is bounded by the elements b_{ij} and c_{ij} of the matrices B and C . Hence, we have the following simple necessary conditions of robust practical and robust asymptotic stability of interval system (1).

Lemma 6. If all eigenvalues $\lambda_i(B)$ or $\lambda_i(C)$ do not lie in the open region $S_A(\alpha, L)$ then the interval system (1) is not robust practically stable.

Lemma 7. If all eigenvalues $\lambda_i(B)$ or $\lambda_i(C)$ do not lie in the open region $S(\alpha)$ then the interval system (1) is not robust asymptotically stable.

From the above and lemma 2 and 3 we have the following lemmas.

Lemma 8. If the condition (13) does not hold for the real eigenvalues $\lambda_i(B)$ or $\lambda_i(C)$ then the interval system (1) is not robust practically stable.

Lemma 9. If the real eigenvalues $\lambda_i(B)$ or $\lambda_i(C)$ are positive than the interval system (1) is not robust asymptotically stable.

For the interval matrix A_I we can determine the eigenvalues region, i.e. the region which consists the eigenvalues of a matrix A_I . This region can be determined by the generalized Gershgorin's theorem and the matrix measure [2].

For a matrix $X = [x_{ij}] \in C^{n \times n}$ (C – field of complex number), Gershgorin circle $K_i(x_{ii}, R_i)$ is defined as the circle with radius $R_i = \sum_{j \neq i}^n |x_{ij}|$ and centered at x_{ii} [24].

By the generalized Gershgorin's theorem to the interval matrices (2) we have [2]

$$R_i = \sum_{j=1, j \neq i}^n \max \{ |b_{ij}|, |c_{ij}| \} \quad i = 1, 2, \dots, n, \quad (16)$$

$$R'_i = \sum_{j=1, j \neq i}^n \max \{ |b_{ji}|, |c_{ji}| \} \quad i = 1, 2, \dots, n, \quad (17)$$

where b_{ij} and c_{ij} are the elements of the matrices B and C , respectively.

Let D be the union of the regions $K_i(b_{ii}, R_i)$ and $K_i(c_{ii}, R_i)$ ($i = 1, 2, \dots, n$) and let D' be the union of the regions $K'_i(b_{ii}, R'_i)$ and $K'_i(c_{ii}, R'_i)$ ($i = 1, 2, \dots, n$). Denote by DD' the intersection of the regions D and D' .

All the eigenvalues of the interval matrix (2) lie in the closed region DD' .

Theorem 4. If DD' lies in the open region $S_A(\alpha, L)$, then the interval system (1) is robust practically stable for a given α and L .

Theorem 5. If DD' lies in the open region $S(\alpha)$, then the interval system (1) is robust asymptotically stable for a given α .

The eigenvalues region of the interval matrix A_I can be also easily determined on the basis of the matrix measure [2].

For a matrix $X = [x_{ij}] \in C^{n \times n}$, the measures $\mu_k(X)$ ($k = 1, \infty$) are calculated as follows [10]

$$\mu_1(X) = \max_j [\operatorname{Re}(x_{jj}) + \sum_{i=1, i \neq j}^n |x_{ij}|], \quad (18)$$

$$\mu_\infty(X) = \max_i [\operatorname{Re}(x_{ii}) + \sum_{j=1, j \neq i}^n |x_{ij}|]. \quad (19)$$

Let $L = [l_{ij}]$ and $R = [r_{ij}]$ ($i, j = 1, 2, \dots, n$) be constant matrices, where

$$\left. \begin{aligned} l_{ii} &= b_{ii}, \quad i = 1, 2, \dots, n, \\ r_{ii} &= c_{ii}, \quad i = 1, 2, \dots, n, \\ l_{ij} &= r_{ij} = \max \{ |b_{ij}|, |c_{ij}| \}, \quad i, j = 1, 2, \dots, n, \quad i \neq j, \end{aligned} \right\} \quad (20)$$

where b_{ij} and c_{ij} are the elements of the known boundary matrices B and C of the interval matrix (2).

From the paper [2] we have the following lemma.

Lemma 10. [2] The eigenvalues of the interval matrix A_I , lie in the region determined by the following inequalities

$$u_l \leq \operatorname{Re} \lambda_i(A_I) \leq u_r, \quad (21)$$

$$-v \leq \operatorname{Im} \lambda_i(A_I) \leq v, \quad (22)$$

where

$$u_l = -\min\{\mu_1(-L), \mu_\infty(-L)\}, \quad (23)$$

$$u_r = \min\{\mu_1(-R), \mu_\infty(-R)\}, \quad (24)$$

$$v = \min\{\mu_1(jR), \mu_\infty(jR)\}, \quad j^2 = -1. \quad (25)$$

The smallest eigenvalues region of the interval matrix A_I is determined by the inequalities (21) and (22). Moreover, the region determined by (21) and (22) is the smallest rectangle in which lies the region DD'.

From above we have the following theorems.

Theorem 6. If rectangle determined by the inequalities (21) and (22) lies in the open region $S_A(\alpha, L)$, then the interval system (1) is robust practically stable for a given α and L .

Theorem 7. If rectangle determined by the inequalities (21) and (22) lies in the open region $S(\alpha)$, then the interval system (1) is robust asymptotically stable for a given α .

Theorems 5 and 6 are easier to apply than theorems 3 and 4.

5 Illustrative Examples

Example 1. Check practical stability of the interval system (1) with length $L = 50$ of practical implementation, fractional order $\alpha = 0.1$ and the boundary matrices B and C of the interval matrix A of the form

$$B = \begin{bmatrix} -0.1 & -0.04 & 0.4 \\ 0.03 & -0.05 & 0.1 \\ 0.02 & -0.15 & -0.2 \end{bmatrix}, \quad C = \begin{bmatrix} -0.01 & 0.03 & 0.55 \\ 0.05 & 0 & 0.2 \\ 0.04 & -0.1 & 0.02 \end{bmatrix}. \quad (26)$$

It is easy to check that the necessary conditions of the practical stability of interval system (1), (26) is satisfied (Lemma 6), i.e. all the eigenvalues of the matrices B and C lie in the open region $S_A(\alpha, L)$ for $L = 50$ and $\alpha = 0.1$ (Fig. 3).

From (20) and (26) we have

$$L = \begin{bmatrix} -0.15 & 0.04 & 0.3 \\ 0.05 & -0.05 & 0.2 \\ 0.04 & 0.15 & -0.20 \end{bmatrix}, \quad R = \begin{bmatrix} -0.01 & 0.04 & 0.3 \\ 0.05 & 0 & 0.2 \\ 0.04 & 0.15 & 0.02 \end{bmatrix}. \quad (27)$$

Calculating the measures of the suitable matrices from (18) and (19) and u_l , u_r and v from (23), (24) and (25), we obtain

$$u_l = -\min\{0.95, 0.69\} = -0.69, \quad (28)$$

$$u_r = \min\{0.77, 0.58\} = 0.58, \quad (29)$$

$$v = \max\{0.75, 0.59\} = 0.59. \quad (30)$$

Hence, from Lemma 10 we have

$$-0.49 \leq \operatorname{Re} \lambda_i(A_I) \leq 0.58, \quad (31)$$

$$-0.59 \leq \operatorname{Im} \lambda_i(A_I) \leq 0.59. \quad (32)$$

The eigenvalues region, determined by (31) and (32), the region $S_A(\alpha, L)$ for $L = 50$ and $\alpha = 0.1$ and the eigenvalues of matrices B and C are shown in Fig. 3. In this figure, the region DD' obtained by the generalized Gershgorin's theorem is also shown (union of two circles with the radius 0.59 and the centers in points $(-0.1, 0)$ and $(-0.01, 0)$).

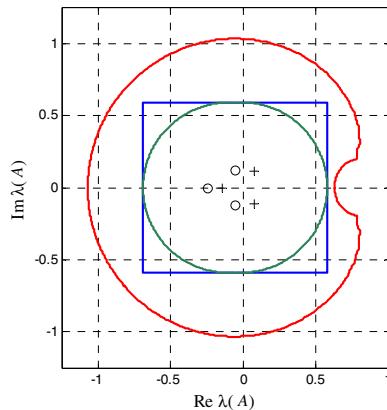


Fig. 3. Region $S_A(\alpha, L)$ for $L = 50$ and $\alpha = 0.1$, eigenvalues region of matrix A_I (rectangle), eigenvalues of B (\circ), eigenvalues of C (+) and region DD'

From Fig. 3 it follows that eigenvalues region of considered interval system lies in the practical stability region $S_A(\alpha, L)$. This means that system (1), (26) is robust practically stable for $L = 50$ and $\alpha = 0.1$.

Taking into account Fig. 1 and Fig. 3 we can notice that for system (1), (26) with fixed $L = 50$ the eigenvalues region of matrix A_l is not located in the regions $S_A(\alpha, L)$ for $\alpha = 0.5$ and $\alpha = 0.9$.

Example 2. Check asymptotic stability of the interval system (1) with fractional order $\alpha = 0.5$ and the boundary matrices B and C of the interval matrix A of the form

$$B = \begin{bmatrix} -0.1 & -0.2 & -0.8 \\ 0.01 & 0 & 0.3 \\ -0.3 & -0.15 & 0.01 \end{bmatrix}, \quad C = \begin{bmatrix} 0.08 & 0.1 & -0.5 \\ 0.04 & 0.02 & 0.7 \\ 0 & -0.1 & 0.04 \end{bmatrix}. \quad (33)$$

Calculating the eigenvalues of the matrix B we obtain $\lambda_1 = 0.434$, $\lambda_2 = -0.449$, $\lambda_3 = -0.075$. From above and Fig. 2 it follows that eigenvalue $\lambda_1 = 0.434$ does not lie in the stability region $S(\alpha)$ for $\alpha = 0.5$. According to Lemma 9 we have that the fractional interval system (1) with the matrices (33) is not robust asymptotically stable for $\alpha = 0.5$. Moreover, this system is not robust asymptotically stable for any $\alpha \in (0, 1)$.

6 Concluding Remarks

The problems of robust practical stability and robust asymptotic stability of discrete-time linear system (1) of fractional order $\alpha \in (0, 1)$ with the system matrix as the interval matrix have been addressed. It has been shown that location of all eigenvalues of the interval matrix A_l in the stability region is necessary and sufficient for the robust practical stability and robust asymptotic stability. Using Gershgorin's theorem for the interval matrices, the matrix measure and the parametric description of boundaries of stability regions necessary and sufficient conditions for the robust practical stability (Theorem 4, Theorem 6) and for the robust asymptotic stability (Theorem 5, Theorem 7) have been established. Also the simple necessary condition of the robust practical and robust asymptotic stability of the interval system (1) has been given in Lemmas 6 and 8, respectively.

Acknowledgement. The work was supported by the National Science Centre in Poland under grant N N514 638940.

References

1. Ahn, H.-S., Chen, Y.Q.: Necessary and sufficient stability condition of fractional-order interval linear systems. *Automatica* 44, 2985–2988 (2008)
2. Busłowicz, M.: Asymptotic stability of dynamical interval systems with pure delay. *Scientific Journal Białystok University of Technology, Technical Sciences* 83, Electricity 11, 61–77 (1992)

3. Busłowicz, M.: Robust stability of positive discrete-time linear systems of fractional order. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 58, 567–572 (2010)
4. Busłowicz, M.: Stability of state-space models of linear continuous-time fractional order systems. *Acta Mechanica et Automatica* 5, 15–22 (2011)
5. Busłowicz, M.: Stability analysis of continuous-time linear systems consisting of n subsystems with different fractional orders. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 60, 279–284 (2012)
6. Busłowicz, M.: Simple analytic conditions for stability of fractional discrete-time linear systems with diagonal state matrix. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 60, 809–814 (2012)
7. Busłowicz, M., Kaczorek, T.: Simple conditions for practical stability of linear positive fractional discrete-time linear systems. *International Journal of Applied Mathematics and Computer Science* 19, 263–269 (2009)
8. Busłowicz, M., Ruszewski, A.: Necessary and sufficient conditions for stability of fractional discrete-time linear state-space systems. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 61 (2013) (in press)
9. Chen, Y.Q., Ahn, H.-S., Podlubny, I.: Robust stability check of fractional order linear time invariant systems with interval uncertainties. *Signal Processing* 86, 2611–2618 (2006)
10. Desoer, C.A., Vidyasagar, M.: *Feedback Systems: Input-output properties*. Acad. Press, New York (1975)
11. Dzieliński, A., Sierociuk, D.: Stability of discrete fractional state-space systems. *Journal of Vibration and Control* 14, 1543–1556 (2008)
12. Guermah, S., Djennoune, S., Bettayeb, M.: A new approach for stability analysis of linear discrete-time fractional-order systems. In: Baleanu, D., et al. (eds.) *New Trends in Nanotechnology and Fractional Calculus Applications*, pp. 151–162. Springer (2010)
13. Kaczorek, T.: Practical stability of positive fractional discrete-time systems. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 56, 313–317 (2008)
14. Kaczorek, T.: *Selected Problems of Fractional Systems Theory*. Springer, Berlin (2011)
15. Kilbas, A.A., Srivastava, H.M., Trujillo, J.J.: *Theory and Applications of Fractional Differential Equations*. Elsevier, Amsterdam (2006)
16. Liao, Z., Peng, C., Li, W., Wang, Y.: Robust stability analysis for a class of fractional order systems with uncertain parameters. *Journal of the Franklin Institute* 348, 1101–1113 (2011)
17. Monje, C., Chen, Y., Vinagre, B., Xue, D., Feliu, V.: *Fractional-order Systems and Controls*. Springer, London (2010)
18. Ostalczyk, P.: Equivalent descriptions of a discrete-time fractional-order linear system and its stability domains. *Int. J. Applied Mathematics and Computer Science* 22, 533–538 (2012)
19. Petras, I.: Stability of fractional-order systems with rational orders: a survey. *Fractional Calculus & Applied Analysis. An International Journal for Theory and Applications* 12, 269–298 (2009)
20. Podlubny, I.: *Fractional Differential Equations*. Academic Press, San Diego (1999)
21. Ruszewski, A.: Stability regions of closed loop system with time delay inertial plant of fractional order and fractional order PI controller. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 56, 329–332 (2008)
22. Sabatier, J., Agrawal, O.P., Machado, J.A.T. (eds.): *Advances in Fractional Calculus, Theoretical Developments and Applications in Physics and Engineering*. Springer, London (2007)
23. Stanisławski, R., Latawiec, K.J.: Stability analysis for discrete-time fractional-order LTI state-space systems. Part I: New necessary and sufficient conditions for asymptotic stability. *Bulletin of the Polish Academy of Sciences* 61, 353–361 (2013)
24. Varga, R.S.: *Gershgorin and His Circles*. Springer, Berlin (2004)

Interfacing Inputs and Outputs with IEC 61131-3 Control Software

Dariusz Rzońca, Jan Sadolewski, and Bartosz Trybus

Rzeszow University of Technology, Department of Computer and Control Engineering,
al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
drzonca@prz-rzeszow.pl
<http://kia.prz.edu.pl/>

Abstract. The paper presents interfaces for handling inputs and outputs from IEC 61131-3 control software available in the CPDev engineering environment. The programmer can use code-embedded clauses which directly connect addresses of variables to particular I/Os. Another approach is to perform additional configuration step to map the variables to hardware channels, what is especially convenient when external modules are used. I/O function blocks are yet another possibility to connect variables to inputs and outputs. The blocks may be either generic or specialized for particular purposes such as fieldbus communications or to perform additional signal processing.

Keywords: distributed control system, IEC 61131-3, function block.

1 Introduction

Control system programming is often performed by means of IEC 61131-3 standard [2], which defines universal languages such as ST or FBD. However, reading signal inputs and writing outputs processed by the programs depend on hardware solutions. This introduces a necessity of additional mechanism to interface controller inputs and outputs with program values. Two approaches for mapping of IEC variables into controller inputs and outputs are possible: program-independent and program-dependent. Separation of control program from hardware characteristics is typical for multi-module PLCs and DCSs (Distributed Control Systems) and involves additional configuration step. The configuration assigns program variables to inputs and outputs, or to communication interfaces. The assignment is independent from programming, allowing to use the same program for various hardware solutions. Even greater portability can be achieved by using common formats (such as XML) for storing the configuration, as shown in [9]. The other approach involves direct referencing of I/Os in the program, which does not require the additional step, but makes the program hardware-dependent, i.e. it will run on a particular hardware.

The paper deals with interfacing of input and outputs with IEC program variables in CPDev programming environment [6]. CPDev is used to program and run IEC 61131-3 control software and can be adapted to various platforms, such

as PLCs, PACs or DCSSes. The controllers may have their inputs and outputs integrated on-board, or be connected to external I/O modules [7]. CPDev allows to choose various mechanisms for accessing them from IEC programs, accordingly to the system specifics or programmer's preference.

2 IEC 61131-3 and CPDev Enhancements

2.1 Direct I/O References

The IEC 61131-3 standard allows to connect variables with inputs and outputs directly in code using AT clause. For instance, the variable declaration denoted as:

```
VAR
  ENGINE AT %I1 : BOOL;
END_VAR
```

means that the Boolean variable **ENGINE** will be mapped to input I number 1 (binary). Due to various hardware architectures, the standard introduces a set of available prefixes (Table 1). It can be noticed, that the prefixes can be used to define a category of the variable (input, output or memory stored) and its size. Since variable types (such as BOOL or INT) also define the size (indirectly), a coherent variable declaration require the type to match the size modifier from Table 1. When they don't match, the compiler reports an error.

The IEC standard allows to use the AT %-notation instead of a variable name, so the following code is grammatically correct:

```
VAR
  AT %Q1 : BOOL;
  AT %IW4 : INT;
END_VAR

%Q1 := %IW4 >= 1_000;
```

Table 1. Variable prefixes available in IEC 61131-3

Prefix	Description
I	Input
Q	Output
M	Memory
X	Size: 1 bit
<i>none</i>	Size: 1 bit
B	Size: 1 byte (8 bits)
W	Size: 1 word (16 bits)
D	Size: 1 double word (32 bits)
L	Size: 1 long word (64 bits)

The code checks whether a hardware counter connected to 16-bit input number 4 has the value greater or equal than 1000 (underscores can be used in numeric literals for easier reading). Boolean result of the check is put into the binary output 1.

The IEC 61131-3 standard leaves handling of I/O variables to compiler providers, because real meaning of the given numbers depends on hardware. As the result, it may lead to different interpretation of the same code on different platforms. For example, the notation %QX1.1 can mean 'bit 1 of binary output 1' as well as 'first output bit of the device number 1 at the controller rail' depending on the manufacturer. In addition, the manufacturers introduce their own variable prefixes aside of the standarized ones.

2.2 CPDev Extensions

The CPDev compiler keeps all program variables, including I/O ones, in operational memory, so additional mechanism is needed to map the variables into physical inputs and outputs. The following rules are obeyed for maintaining the variables:

1. Variables are stored in a common memory space, each at its own address.
2. The addresses are automatically generated by the compiler.
3. Variables with AT clause are located at the given fixed addresses (overlaps are notified as warnings).
4. Numeric addresses are multiplied by size prefixes.
5. Outputs (Q) are shifted by 16 bytes.

Memory variables without the AT clause are located at the lowest address where they fit. A compiler option can be set to generate addresses dividable by the variable size. Since I/O variables (with AT clause) use fixed addresses, their memory location will not change between compilations of the program, so it is possible to statically map particular CPDev addresses into hardware input and outputs (so-called memory mapped I/O).

The CPDev introduces some extra compilation directives for I/O handling. They are seen as comments by third-party tools, so this complies with IEC 61131-3 standard. One of the common directives is (*\$CHANNEL*). It is placed after the keyword VAR_GLOBAL and denotes a group of variables with values taken from a hardware input channel or stored in a hardware output channel. Since such variables are connected to the channels, the program does not need to initialize them. As the result, the program code is shorter and system initialization is faster.

Two other CPDev directives are (*\$READ*) and (*\$WRITE*), useful when multiple control tasks are executed in parallel. The directives indicate that program code will only read or write some global variables (given in VAR_EXTERNAL). Since the globals can be shared among control tasks, local copies are used to avoid conflicts. The local copies are synchronized with the globals, and the two directives indicate required synchronization (read for input, write for output).

```

VAR_GLOBAL (*$CHANNEL*)
  START : BOOL;
  STOP : BOOL;
END_VAR

VAR_GLOBAL
  ENGINE : BOOL;
END_VAR

PROGRAM MOTOR
  VAR_EXTERNAL
    START (*$READ*) : BOOL;
    STOP (*$READ*) : BOOL;
    ENGINE : BOOL;
  END_VAR

  ENGINE := (ENGINE OR START) AND (NOT STOP);
END_PROGRAM

```

In the above example, two global variables `START` and `STOP` form an input channel and are read from binary inputs. Following the directives `(*$READ*)` in the program `MOTOR`, the compiler does not perform unnecessary synchronization of the global variables at the end of the program. If the directive `(*$WRITE*)` was used, obtaining a value for a local copy of a variable at the beginning of a program would be skipped. However, this does not apply to the variable `ENGINE` in the program above, because its previous value is read in the logical expression.

3 Variable Mapping with a Configuration Tool

Defining direct connections between variables and I/O using `AT` keyword is cumbersome and may lead to errors. Software engineer has to determine hardware details in early stages of development instead of focusing on control algorithm. Programs written in that way are oriented on specific hardware configuration and generally not portable. Much more flexibility could be achieved by separation of logic and hardware layers. Such an approach has been introduced to CPDev [6]. Universal executable code generated by the compiler in logic layer is hardware-independent. Specification of communication and I/O interfaces in a form of hardware allocation map connects the code with the target platform. In such a case, the hardware layer can be defined in the final phase, when control programs are already developed and tested on simulator. Hardware configuration can be easily modified if necessary, without recompilation of control software.

Typically an industrial controller can be considered as a finite state machine with state diagram shown in Fig. 1. Controller software works in an infinite loop. At the beginning of every cycle some initial routines are called, e.g. reading the inputs. Then, a single cycle of the control task is executed. Finishing the cycle

may require some additional activity, e.g. updating the outputs. Finally the controller is waiting for the next cycle to keep the cycle time constant. At this point, communications with other devices may be performed.

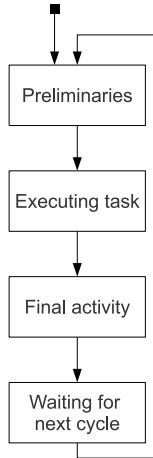


Fig. 1. State diagram of a controller

Basing on the diagram from Fig. 1 further models can be developed to specify more details of the system. Such a model constructed with the use Hierarchical Timed Coloured Petri Nets (HTCPN) represents the CPDev virtual machine (VM) has been presented in [10]. It represents the runtime part of CPDev which executes the universal code generated by the compiler. The general model is shown in Fig. 2, whereas specific activities can be further modeled as lower-level subnets introduced by substitution transitions and fusion places (e.g. "Preliminaries" or "Executing task").

Analysis and simulation of HTCPN models help to identify and resolve potential performance problems e.g. bottleneck or blockage. Such models may be even integrated with larger applications [3] e.g. engineering environment to improve quality of software. Interfacing the CPDev VM with I/O boards has been modeled in [7]. Two modes of interaction between the VM and external I/O boards have been considered there, i.e. synchronous and asynchronous one. In synchronous mode, I/O reading and writing are performed in "Preliminaries" and "Final activity" substitution transitions accordingly (Fig. 2). Such approach may result in extending the program cycle time if large amount of date needs to be transferred. Asynchronous mode requires a separate process to handle I/O boards and synchronize their internal buffers with VM variables periodically. This mode will short program cycle time, however it may also lead to calculations performed on values being out-of-date.

HTCPN models of the communication with remote I/O modules has been described in [8]. Frequently a controller is connected with external modules forming

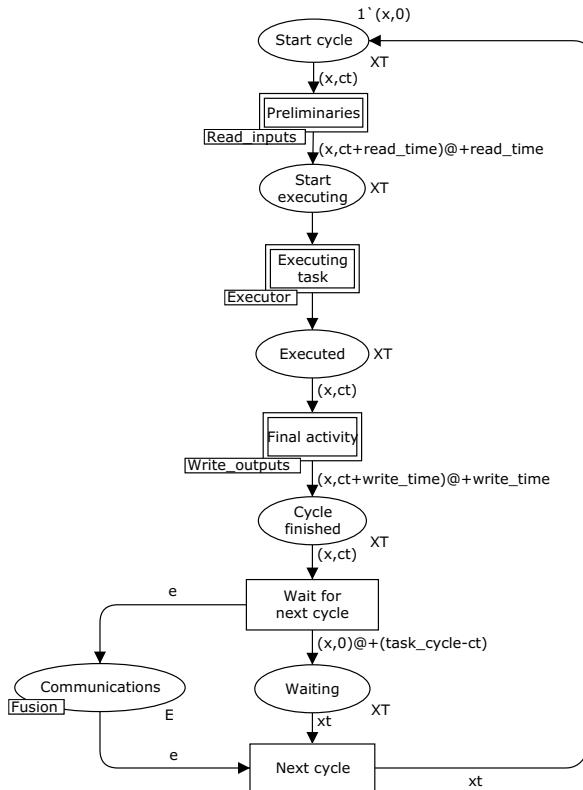


Fig. 2. HTCPN model of CPDev virtual machine

a distributed control system (DCS). Different paradigms of data transfers scheme can be considered, depending on communication bus and field protocol [1]. Usually models of access to the communication link and scenario of data exchanges are fixed, however concept of dynamic change to improve communication flexibility and reliability has also been described [5]. For example, in master-slave communication typically a controller (master) will periodically send commands to I/O modules (slaves). Data are exchanged in transactions, each one consist of two data frames – request and response. Transactions repeated periodically and related to particular device registers form a communication task. Typically a distributed system involves numerous communications tasks, which may be executed simultaneously on different transmission lines [4] or sequentially on single data bus. Communication tasks which connects local variables with remote physical I/O have to be defined by design engineer, however such relationship can be made in final development phase, which improves flexibility and portability of the system.

Both local I/O boards and remote I/O modules handled in the described way need an external tool to map control program variables with inputs and outputs.

Such program will link logical and hardware layer binding the software to target platform. CPDev includes the CPCon application (Fig. 3) allowing the user to configure the mapping during transferring a control program to a controller.

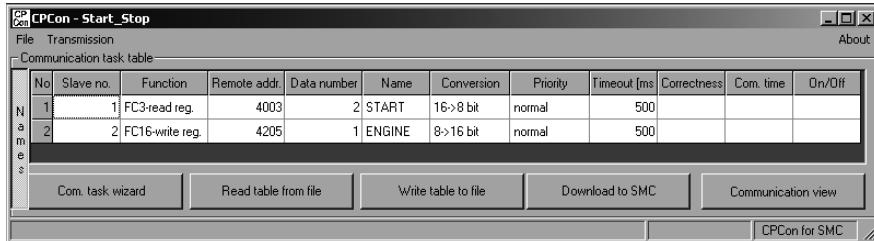


Fig. 3. CPCon application

Communication task table configured in the example (Fig. 3) consists of two tasks. The first provides communication with slave no. 1 to read two consecutive inputs as registers from address 4003 using Modbus function FC3. Those values will be converted to 8-bit and stored in the START and STOP variables respectively. This task has normal priority and the timeout has been set to 500 ms. Similarly the second task will set the output line in the remote I/O module according to value of the ENGINE variable.

4 Input and output function blocks

4.1 Generic I/O Blocks

Another method of assigning program variables to inputs and outputs involves dedicated I/O function blocks. Contrary to the configuration tool described above, I/O blocks are components of the user program, i.e. POU, providing hardware dependent functions internally. The external signals are either outputs or inputs of such blocks and can be connected to program variables. The I/O blocks may represent:

- inputs and outputs integrated in the controller
- external slave input or output modules
- signal processing algorithm.

Generic I/O function blocks can be constructed to handle most of the cases when on-board binary or analog inputs and outputs are considered. Such blocks will interface the hardware in a way that is transparent to IEC 61131-3 programmer. The Function Block Diagram (FBD) from Fig. 4 contains two types of I/O function blocks, namely **BINARY_INPUT** and **BINARY_OUTPUT**. The blocks are thought to be easy to use, so they contain only a single input or output of Boolean type (**BOOL**) and a status indicator (not used in the example). In the

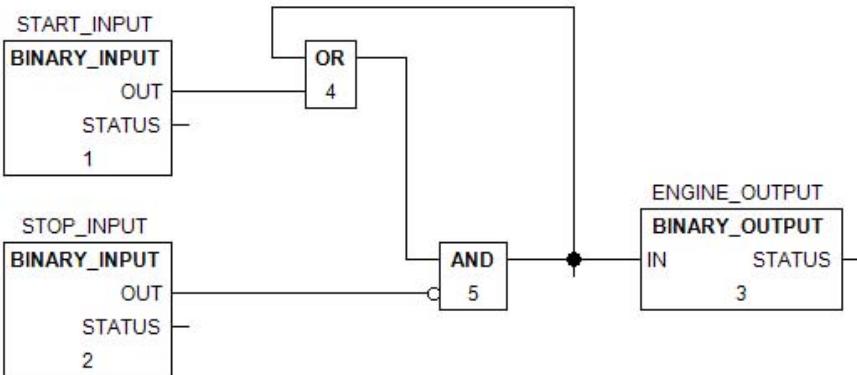


Fig. 4. FBD diagram with generic binary I/O function blocks

FBD, which corresponds to the ST program described in Sec. 2.2 above, the two binary input blocks provide START and STOP signals, while the output block passes the control signal to the engine.

4.2 Signal Processing Blocks

Signal processing blocks represent an internal algorithm which is required to process incoming signals or to convert the calculated value to hardware output characteristics. Since IEC 61131-3 languages are not always sufficient for covering some sophisticated algorithms, the blocks are usually prepared at the lower programming level. CPDev runtime provides support for so-called native blocks, which are usually written in C/C++ and linked by a library. They interface the low-level signal processing with IEC programs and are used in standard way, i.e. as other function blocks. CPDev native blocks can, for instance, provide read/write of program variables into non-volatile memory (flash), communicate over serial and CAN, handle LCD displays.

Fig. 5 shows declaration and usage of a signal processing function block which reads data from a GPS module using NMEA 0183 protocol, introduced by National Marine Electronics Association. The block **NMEA_GGA** (GGA is a command in NMEA) involves two inputs, serial port number (**PORT**) and **DATA_RATE**. The outputs provide UTC time (absolute), LATitude, LONgitude, ALTitude of actual position, together with **QUALITY** of readings. A low-level C code extracts those data from a stream of characters supplied by the GPS module. The example ST program at right assigns outputs of the block instance to program variables.

Similar approach has been used in a mini-DCS control system with SMC controller and external input/output modules [6]. The controller communicates with the I/O modules using Modbus protocol. Due to limited performance and communication characteristics, dedicated communication function blocks were developed. Fig. 6 shows two of them, **COM_SM2** (4-channels module of analogue inputs) and **COM_SM4** (module of 8 logic outputs). Boolean input **EN** will enable

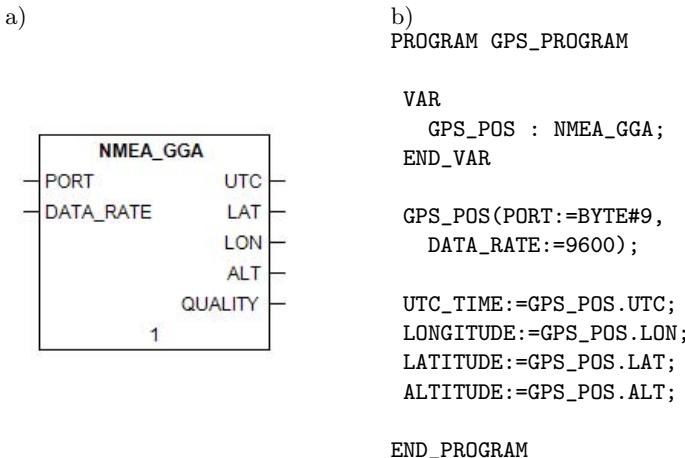


Fig. 5. NMEA function block a) FBD declaration, b) sample ST program

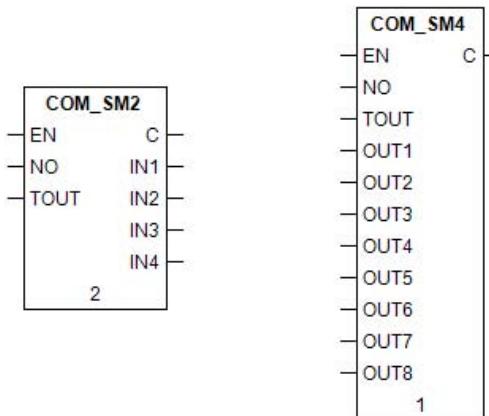


Fig. 6. Function blocks for communications with external I/O modules

the communication, with value at input **NO** as the Modbus identifier of particular module. **TOUT** is used to set the required timeout. The output **C** will raise false if the communication fails, so it may be used to trigger the control into a safe state. The communication blocks automatically handle communication tasks (see Sect. 3). Transfer of input or output data for all channels (i.e. analog or logical) is carried out in a single Modbus request to reduce communication transactions.

5 Summary

There are several options for IEC 61131-3 programmers to handle inputs and outputs in control software. CPDev engineering environment provides IEC 61131-3

%AT clause as well as CPDev-specific extensions to control memory allocation of variables. Another scenario involves an external configuration tool to map program variables into I/O channels. Such tool may be used to parametrize communication tasks which are required to exchange data between modules of a distributed control system. Function blocks can also be used to logically connect variables and I/O ports. Generic I/O function blocks are thought to be easy to use, while signal processing blocks allow to handle specialized data transfers.

References

1. Gaj, P., Jasperneite, J., Felser, M.: Computer Communication Within Industrial Distributed Environment – a Survey. *IEEE Transactions on Industrial Informatics* 9(1), 182–189 (2013)
2. IEC 61131-3 – programmable controllers – part 3: Programming languages (2013)
3. Kowalski, M., Rząsa, W.: Object-oriented approach to Timed Colored Petri Net simulation. In: 2013 Federated Conference on Computer Science and Information Systems (FedCSIS), September 8-11, pp. 1401–1404 (2013)
4. Kwiecień, A., Sidzina, M.: Dual Bus as a Method for Data Interchange Transaction Acceleration in Distributed Real Time Systems. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) CN 2009. CCIS, vol. 39, pp. 252–263. Springer, Heidelberg (2009)
5. Kwiecień, A., Sidzina, M., Maćkowski, M.: The Concept of Using Multi-protocol Nodes in Real-Time Distributed Systems for Increasing Communication Reliability. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) CN 2013. CCIS, vol. 370, pp. 177–188. Springer, Heidelberg (2013)
6. Rzońca, D., Sadolewski, J., Stec, A., Świdler, Z., Trybus, B., Trybus, L.: Programming controllers in Structured Text language of IEC 61131-3 standard. *Journal of Applied Computer Science* 16(1), 49–67 (2008)
7. Rzońca, D., Sadolewski, J., Trybus, B.: Coloured Petri-nets models of CPDev soft controller with I/O boards. *Przegląd Elektrotechniczny* (Electrical Review) 9, 170–173 (2010)
8. Rzońca, D., Trybus, B.: Hierarchical Petri Net for the CPDev Virtual Machine with Communications. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) CN 2009. CCIS, vol. 39, pp. 264–271. Springer, Heidelberg (2009)
9. Simros, M., Theurich, S., Wollschlaeger, M.: Programming Embedded Devices in IEC 61131-Languages With Industrial PLC Tools Using PLCOpen XML. In: 10th Portuguese Conference on Automatic Control, CONTROLO 2012, Funchal, Portugal, July 16-18 (2012)
10. Trybus, B.: Development and Implementation of IEC 61131-3 Virtual Machine. *Theoretical and Applied Informatics* 23(1), 21–35 (2011)

Reachability of Fractional Positive Continuous-Time Linear Systems with Two Different Fractional Orders

Łukasz Sajewski

Bialystok University of Technology, Faculty of Electrical Engineering,
Wiejska 45D, 15-351 Bialystok
l.sajewski@pb.edu.pl

Abstract. The reachability problem for the fractional positive continuous-time linear systems with two different fractional orders is formulated and solved. Sufficient conditions for the reachability are established. Applications of the proposed conditions is demonstrated on example of electrical circuit.

Keywords: fractional, different order, reachability, positive, continuous-time.

1 Introduction

In positive systems inputs, state variables and outputs take only non-negative values. Examples of positive systems are industrial processes involving chemical reactors, heat exchangers and distillation columns, storage systems, compartmental systems, water and atmospheric pollution models. A variety of models having positive linear systems behavior can be found in engineering, management science, economics, social sciences, biology and medicine, etc. Positive linear systems are defined on cones and not on linear spaces. Therefore, the theory of positive systems is more complicated and less advanced. An overview of state of art in positive systems theory is given in the monographs [4, 10].

The first definition of the fractional derivative was introduced by Liouville and Riemann at the end of the 19th century [22, 23] and another one was proposed in 20th century by Caputo [26]. This idea has been used by engineers for modeling different process [3, 5]. Mathematical fundamentals of fractional calculus are given in the monographs [14, 21–23, 26]. The fractional order controllers have been developed in [24, 25]. The positive fractional linear systems have been investigated in [7, 8, 11, 12, 14]. Stability of fractional linear 1D discrete-time and continuous-time systems has been investigated in the papers [1, 14, 16] of 2D fractional positive linear systems in [6] and of continuous-time linear systems consisting of n subsystem with different fractional orders [2]. The minimum energy control problem for standard linear systems has been formulated and solved in [16, 18, 20] and for 2D linear systems with variable coefficients in [15, 17]. The controllability and minimum energy control problem of fractional discrete-time linear systems has been investigated in [19]. The minimum energy control of fractional positive continuous-time linear systems has been addressed in [9]. Positive realization of 2D different orders fractional discrete-time linear systems has been solved in [27].

In this paper reachability problem for the fractional positive continuous-time linear systems with two different fractional orders will be formulated and solved.

The paper is organized as follows. In section 2 the basic definitions and theorems of the fractional continuous-time linear systems with two different fractional orders are recalled and the necessary and sufficient conditions for the positivity of the systems are given. The reachability problem for this class of fractional systems is formulated and solved in section 3, where sufficient conditions are established and illustrated by the example of electrical circuit with ultracapacitors. Concluding remarks are given in section 4.

The following notation will be used: \mathfrak{R} – the set of real numbers, $\mathfrak{R}^{n \times m}$ – the set of $n \times m$ real matrices, $\mathfrak{R}_+^{n \times m}$ – the set of $n \times m$ matrices with nonnegative entries and $\mathfrak{R}_+^n = \mathfrak{R}_+^{n \times 1}$, M_n – the set of $n \times n$ Metzler matrices (real matrices with nonnegative off-diagonal entries), I_n – the $n \times n$ identity matrix, A^T – the transpose matrix A . A real square matrix is called monomial if each its row and each its column contains only one positive entry and the remaining entries are zero.

2 Linear Differential Equations with Two Different Fractional Orders

In this paper the following Caputo definition of the fractional derivative will be used [26]

$$\frac{d^\alpha f(t)}{dt^\alpha} = \frac{1}{\Gamma(n-\alpha)} \int_0^\infty \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha+1-n}} d\tau, \quad (1)$$

where $n-1 < \alpha < n$, $n \in N = \{1, 2, \dots\}$,

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \quad (2)$$

is the gamma Euler function and

$$f^{(n)}(\tau) = \frac{d^n f(\tau)}{d\tau^n}. \quad (3)$$

It is well known [14] that the Laplace transform (\mathcal{L}) of (1) is given by the formula

$$\mathcal{L}\left[\frac{d^\alpha f(t)}{dt^\alpha}\right] = \int_0^\infty \frac{d^\alpha f(t)}{dt^\alpha} e^{-st} dt = s^\alpha F(s) - \sum_{k=1}^n s^{\alpha-k} f^{(k-1)}(0+), \quad (4)$$

where $F(s) = \mathcal{L}[f(t)]$ and $n-1 < \alpha < n$, $n \in N$.

Consider a fractional linear system described by the equation [12]

$$\begin{bmatrix} \frac{d^\alpha x_1(t)}{dt^\alpha} \\ \frac{d^\beta x_2(t)}{dt^\beta} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(t), \quad (5)$$

where $p-1 < \alpha < p$; $q-1 < \beta < q$; $p, q \in N$, $x_1(t) \in \Re^{n_1}$ and $x_2(t) \in \Re^{n_2}$ are the state vectors, $A_{ij} \in \Re^{n_i \times n_j}$, $B_i \in \Re^{n_i \times m}$; $i, j = 1, 2$, and $u \in \Re^m$ is the input vector. Initial conditions for (5) have the form

$$x_1(0) = x_{10}, \quad x_2(0) = x_{20} \quad \text{and} \quad x_0 = \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix}. \quad (6)$$

Remark 1. The state equation (5) of fractional continuous-time linear systems with two different fractional orders has similar notion to the 2D Roeesser type models [10, 14].

Theorem 1. The solution of the equation (5) for $0 < \alpha < 1$; $0 < \beta < 1$ with initial conditions (6) has the form

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \Phi_0(t)x_0 + \int_0^t M(t-\tau)u(\tau)d\tau, \quad (7a)$$

where

$$\begin{aligned} M(t) &= \Phi_1(t)B_{10} + \Phi_2(t)B_{01} = \begin{bmatrix} \Phi_{11}^1(t) & \Phi_{12}^1(t) \\ \Phi_{21}^1(t) & \Phi_{22}^1(t) \end{bmatrix} \begin{bmatrix} B_1 \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_{11}^2(t) & \Phi_{12}^2(t) \\ \Phi_{21}^2(t) & \Phi_{22}^2(t) \end{bmatrix} \begin{bmatrix} 0 \\ B_2 \end{bmatrix} \\ &= \begin{bmatrix} \Phi_{11}^1(t)B_1 + \Phi_{12}^2(t)B_2 \\ \Phi_{21}^1(t)B_1 + \Phi_{22}^2(t)B_2 \end{bmatrix} = \begin{bmatrix} \Phi_{11}^1(t) & \Phi_{12}^2(t) \\ \Phi_{21}^1(t) & \Phi_{22}^2(t) \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \end{aligned} \quad (7b)$$

and

$$\Phi_0(t) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} T_{kl} \frac{t^{k\alpha+l\beta}}{\Gamma(k\alpha+l\beta+1)}, \quad (8a)$$

$$\Phi_1(t) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} T_{kl} \frac{t^{(k+1)\alpha+l\beta-1}}{\Gamma[(k+1)\alpha+l\beta]}, \quad (8b)$$

$$\Phi_2(t) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} T_{kl} \frac{t^{k\alpha+(l+1)\beta-1}}{\Gamma[k\alpha+(l+1)\beta]}, \quad (8c)$$

$$T_{kl} = \begin{cases} I_n & \text{for } k = l = 0 \\ \begin{bmatrix} A_{11} & A_{12} \\ 0 & 0 \end{bmatrix} & \text{for } k = 1, l = 0 \\ \begin{bmatrix} 0 & 0 \\ A_{21} & A_{22} \end{bmatrix} & \text{for } k = 0, l = 1 \\ T_{10}T_{k-1,l} + T_{01}T_{k,l-1} & \text{for } k + l > 1 \end{cases} \quad (8d)$$

Proof is given in [11].

Note that if $\alpha = \beta$ then from (8a) we have [6, 11]

$$\Phi_0|_{\alpha=\beta}(t) = \sum_{k=0}^{\infty} \frac{A^k t^{k\alpha}}{\Gamma(k\alpha+1)}. \quad (9a)$$

From comparison of (8a) and (9a) and using (8d) it is easy to show that

$$\left. \sum_{i=0}^k \sum_{j=0}^k T_{ij} \frac{t^{i\alpha+j\beta}}{\Gamma(i\alpha+j\beta+1)} \right|_{\alpha=\beta} = \frac{A^k t^{k\alpha}}{\Gamma(k\alpha+1)}. \quad (9b)$$

Definition 1. The fractional system (5) is called positive if $x_1(t) \in \Re_+^{n_1}$ and $x_2(t) \in \Re_+^{n_2}$, $t \geq 0$ for any initial conditions $x_{10} \in \Re_+^{n_1}$, $x_{20} \in \Re_+^{n_2}$ and all input vectors $u \in \Re_+^m$, $t \geq 0$.

Theorem 2. The fractional system (5) for $0 < \alpha < 1$; $0 < \beta < 1$ is positive if and only if

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in M_N, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \in \Re_+^{N \times m}, \quad (N = n_1 + n_2). \quad (10)$$

Proof is given in [9, 10].

3 Reachability

Definition 2. [13] The state $x_f \in \Re_+^N$ of the fractional system (5) is called reachable in time t_f if there exist an input $u(t) \in \Re_+^m$, $t \in [0, t_f]$ which steers the state of system (5) from zero initial state $x_0 = 0$ to the state x_f . If every state $x_f \in \Re_+^N$ is reachable in time t_f the system is called reachable in time t_f . If for every state $x_f \in \Re_+^N$ there exist a time t_f such that the state is reachable in time t_f then the system (2) is called reachable.

Theorem 3. The fractional continuous-time system linear system with different fractional orders (5) is reachable in time t_f if the matrix

$$R(t_f) = \int_0^{t_f} M(\tau)M^T(\tau)d\tau \quad (11)$$

is a monomial matrix and the input which steers the state of the system (5) from $x_0 = 0$ to x_f is given by

$$u(t) = M^T(t_f - t)R^{-1}(t_f)x_f. \quad (12)$$

Proof. If the matrix (11) is a monomial matrix then $R^{-1}(t_f) \in \Re_+^{N \times N}$ and the input defined by (12) is nonnegative vector, i.e. $u(t) \in \Re_+^m$, $t \geq 0$. Using (7a), (12) and (11)

for $x_0 = \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix} = 0$, $t = t_f$ we obtain

$$\begin{aligned} x(t_f) &= \int_0^{t_f} M(t_f - \tau)M^T(t_f - \tau)R^{-1}(t_f)d\tau x_f \\ &= \int_0^{t_f} M(\tau)M^T(\tau)d\tau R^{-1}(t_f)x_f = R(t_f)R^{-1}(t_f)x_f = x_f. \end{aligned} \quad (13)$$

Therefore, the input (12) steers the state of the system (5) from $x_0 = \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix} = 0$ to

$$x_f = \begin{bmatrix} x_{f,1} \\ x_{f,2} \end{bmatrix}. \quad \blacksquare$$

Theorem 4. If $A \in M_N$ is diagonal and $B \in \Re_+^{N \times N}$ is a monomial matrix then the fractional continuous-time linear system with two different fractional orders (5) is reachable.

Proof. Sufficiency. From (7) and (8) it follows that if the matrix $A \in M_N$ is diagonal then the matrix $\begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix} \in \Re_+^{N \times N}$ is also diagonal and since by assumption

$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \in \Re_+^{N \times N}$ is monomial then the matrix $M(t) = \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$ and $BB^T \in \Re_+^{N \times N}$ are monomial. From (11) written in the form

$$R(t_f) = \int_0^{t_f} \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}^T \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix}^T d\tau \quad (14)$$

it follows that the matrix (4) is monomial. Thus by Theorem 3 the fractional system is reachable.

Necessity. Let

$$p(s^\alpha, s^\beta) = s^{\alpha n_1} + a_{n_1-1}^1 s^{\alpha(n_1-1)} + \dots + a_1^1 s^\alpha + a_0^1 + s^{\beta n_2} + a_{n_2-1}^2 s^{\beta(n_2-1)} + \dots + a_1^2 s^\beta + a_0^2 \quad (15)$$

be the characteristic polynomial of the diagonal matrix $A = \text{blkdiag}[A_1, A_2] = \text{diag}[a_0^1, \dots, a_{n_1}^1, a_0^2, \dots, a_{n_2}^2] \in M_N$. Then by the well-known [1, 19] Cayley-Hamilton theorem we have

$$\begin{aligned} p(A_1, A_2) &= A_1^{\alpha n_1} + a_{n_1-1}^1 A_1^{\alpha(n_1-1)} + \dots + a_1^1 A_1^\alpha + a_0^1 I_{n_1} \\ &\quad + A_2^{\beta n_2} + a_{n_2-1}^2 A_2^{\beta(n_2-1)} + \dots + a_1^2 A_2^\beta + a_0^2 I_{n_2} = 0. \end{aligned} \quad (16)$$

Using (16) we may eliminate from (8b) and (8c) T_{kl} for $k = n_1, n_1+1, \dots; l = n_2, n_2+1, \dots$ and we obtain

$$\Phi_1(t) = \sum_{k=0}^{n_1-1} \sum_{l=0}^{n_2-1} T_{kl} c_{kl}^1(t) = \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^1(t) \end{bmatrix}, \quad \Phi_2(t) = \sum_{k=0}^{n_1-1} \sum_{l=0}^{n_2-1} T_{kl} c_{kl}^2(t) = \begin{bmatrix} \Phi_{11}^2(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix}, \quad (17)$$

where $c_{kl}^1(t)$, $c_{kl}^2(t)$, $k = 0, 1, \dots, n_1-1$; $l = 0, 1, \dots, n_2-1$ are some nonzero functions of time depending on the diagonal matrix T and

$$T_{kl} = \begin{cases} I_n & \text{for } k = l = 0 \\ \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix} & \text{for } k = 1, l = 0 \\ \begin{bmatrix} 0 & 0 \\ 0 & A_2 \end{bmatrix} & \text{for } k = 0, l = 1 \\ T_{10} T_{k-1,l} + T_{01} T_{k,l-1} & \text{for } k + l > 1 \end{cases} \quad (18)$$

Substitution of (17) into

$$\int_0^{t_f} [\Phi_1(t_f - \tau) B_{10} + \Phi_2(t_f - \tau) B_{01}] u(\tau) d\tau = \int_0^{t_f} \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(\tau) d\tau \quad (19)$$

yields

$$x_f = \begin{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} & T_{01} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} & T_{10} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} & \dots & T_{n_1-1, n_2-1} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \end{bmatrix} \begin{bmatrix} v_{00}(t_f) \\ v_{01}(t_f) \\ v_{10}(t_f) \\ \vdots \\ v_{n_1-1, n_2-1}(t_f) \end{bmatrix}, \quad (20)$$

where

$$v_{kl}(t_f) = \int_0^{t_f} \begin{bmatrix} c_{kl}^1(\tau) & 0 \\ 0 & c_{kl}^2(\tau) \end{bmatrix} u(t_f - \tau) d\tau, \quad k = 0, 1, \dots, n_1 - 1; \quad l = 0, 1, \dots, n_2 - 1. \quad (21)$$

For given $x_f \in \mathfrak{R}_+^N$ it is possible to find nonnegative $v_{kl}(t_f)$ for $k = 0, 1, \dots, n_1 - 1$; $l = 0, 1, \dots, n_2 - 1$ if and only if the matrix

$$[B \quad T_{01}B \quad T_{10}B \quad \dots \quad T_{n_1-1, n_2-1}B] \quad (22)$$

has N linearly independent monomial columns and this takes place only if the matrix $[B, A]$ contains N linearly independent columns [14]. Note that for the nonnegative $v_{kl}(t_f)$, $k = 0, 1, \dots, n_1 - 1$; $l = 0, 1, \dots, n_2 - 1$ it is possible to find a nonnegative input $u(t) \in \mathfrak{R}_+^N$, $t \in [0, t_f]$ only if the matrix $B \in \mathfrak{R}_+^{N \times N}$ is monomial and the matrix $A \in M_N$ is diagonal. ■

Example 1. [13] Consider the fractional electrical circuit (Fig. 1) with given source voltages e_1, e_2 , ultracapacitor C_1 of the fractional order $\alpha = 0.7$, ultracapacitor C_2 of the fractional order $\beta = 0.6$, conductances G_1, G_2, G'_1, G'_2 and $G_{12} = 0$.

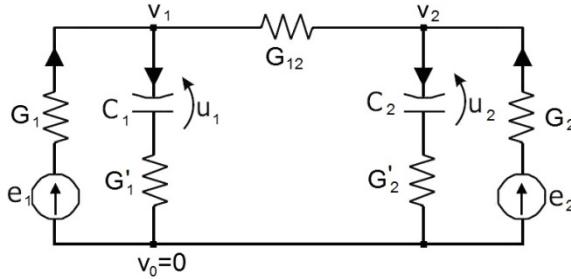


Fig. 1. Fractional electrical circuit

Using the Kirchhoff's laws we can write the equations

$$\begin{aligned} C_1 \frac{d^\alpha u_1}{dt^\alpha} &= G'_1(v_1 - u_1), \\ C_2 \frac{d^\beta u_2}{dt^\beta} &= G'_2(v_2 - u_2) \end{aligned} \quad (23)$$

and

$$\begin{bmatrix} (G_1 + G'_1) & 0 \\ 0 & (G_2 + G'_2) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} G'_1 & 0 \\ 0 & G'_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}. \quad (24)$$

From (24) we obtain

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} (G_1 + G'_1) & 0 \\ 0 & (G_2 + G'_2) \end{bmatrix}^{-1} \begin{bmatrix} G'_1 & 0 \\ 0 & G'_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$+ \begin{bmatrix} (G_1 + G'_1) & 0 \\ 0 & (G_2 + G'_2) \end{bmatrix}^{-1} \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}. \quad (25)$$

Substitution of (25) into

$$\begin{bmatrix} \frac{d^\alpha u_1}{dt^\alpha} \\ \frac{d^\beta u_2}{dt^\beta} \end{bmatrix} = \begin{bmatrix} -\frac{G'_1}{C_1} & 0 \\ 0 & -\frac{G'_2}{C_2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} \frac{G'_1}{C_1} & 0 \\ 0 & \frac{G'_2}{C_2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (26)$$

we obtain

$$\begin{bmatrix} \frac{d^\alpha u_1}{dt^\alpha} \\ \frac{d^\beta u_2}{dt^\beta} \end{bmatrix} = A \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + B \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}, \quad (27)$$

where

$$A = \begin{bmatrix} \frac{G'_1}{C_1} \left(\frac{G'_1}{(G_1 + G'_1)} - 1 \right) & 0 \\ 0 & \frac{G'_2}{C_2} \left(\frac{G'_2}{(G_2 + G'_2)} - 1 \right) \end{bmatrix}, \quad (28)$$

$$B = \begin{bmatrix} \frac{G'_1 G_1}{C_1 (G_1 + G'_1)} & 0 \\ 0 & \frac{G'_2 G_2}{C_2 (G_2 + G'_2)} \end{bmatrix}.$$

From (28) it follows that A is a diagonal Metzler matrix and the matrix B is monomial matrix with positive diagonal entries. Therefore, the fractional electrical circuit is positive for all values of the conductances and capacitances.

Without lost of generality, to simplify the notation, let take for example values of the electrical circuit: $C_1 = 1$, $\alpha = 0.7$, $C_2 = 2$, $\beta = 0.6$, $G_1 = 4$, $G'_1 = 4$, $G_2 = 3$, $G'_2 = 6$, $G_{12} = 0$ and $N = n_1 + n_2 = 2$. The matrices (28) takes the form

$$A = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \quad (29)$$

Check the reachability of the system and find the input (source voltage) $e(t) \in \Re_+^2$, $t \geq 0$ which steers the system from initial state (voltage drop on capacitances) $u_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ to the finite state $u_f = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. By Theorem 4 the fractional positive system is reachable in time $[0, t_f] = [0, 1]$ since the matrix $A \in M_2$ is a Metzler matrix and the matrix $B \in \Re_+^{2 \times 2}$ is a monomial matrix.

Using (8), (11) and (29) we obtain

$$R(1) = \int_0^1 \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix}^T d\tau, \quad (30)$$

where

$$\begin{aligned} \Phi_{11}^1(t) &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{t^{(k+1)0.7+l0.6-1}}{\Gamma[(k+1)0.7+l0.6]} (-2)^k, \\ \Phi_{22}^2(t) &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{t^{k0.7+(l+1)0.6-1}}{\Gamma[k0.7+(l+1)0.6]} (-1)^l \end{aligned} \quad (31)$$

and the matrix

$$R(1) = \int_0^1 M(\tau) M^T(\tau) d\tau \quad (32)$$

is a diagonal.

From (12) we have

$$\begin{aligned} e(t) &= \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} = M^T(1-t) R^{-1}(1) u_f = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^T \begin{bmatrix} \Phi_{11}^1(t) & 0 \\ 0 & \Phi_{22}^2(t) \end{bmatrix}^T R^{-1}(1) \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 4\Phi_{11}^{1 T}(t) \int_0^1 [4\Phi_{11}^1(\tau)\Phi_{11}^{1 T}(\tau)]^{-1} d\tau \\ 3\Phi_{22}^{2 T}(t) \int_0^1 [\Phi_{22}^2(\tau)\Phi_{22}^{2 T}(\tau)]^{-1} d\tau \end{bmatrix} \end{aligned} \quad (33)$$

the desired input which steers the state of the system from $u_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ to the final state

$$u_f = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

4 Concluding Remarks

The reachability problem for the fractional positive continuous-time linear systems with two different fractional orders have been formulated and solved. Sufficient conditions for the reachability has been established (Theorem 4). Effectiveness of the proposed conditions has been demonstrated on example of electrical circuit (Example 1). Next step is extension of this considerations on systems with state matrix A in arbitrary form and for minimum energy control [9, 11].

Acknowledgment. This work was supported by National Science Centre in Poland under work No. N N514 6389 40.

References

1. Busłowicz, M.: Stability of linear continuous time fractional order systems with delays of the retarded type. *Bull. Pol. Acad. Sci. Tech.* 56(4), 319–324 (2008)
2. Busłowicz, M.: Stability analysis of continuous-time linear systems consisting of n subsystem with different fractional orders. *Bull. Pol. Acad. Sci. Tech.* 60(2), 279–284 (2012)
3. Dzieliński, A., Sierociuk, D., Sarwas, G.: Ultracapacitor parameters identification based on fractional order model. In: Proc. ECC 2009, Budapest (2009)
4. Farina, L., Rinaldi, S.: Positive Linear Systems, Theory and Applications. J. Wiley, New York (2000)
5. Ferreira, N.M.F., Machado, J.A.T.: Fractional-order hybrid control of robotic manipulators. In: Proc. 11th Int. Conf. Advanced Robotics, ICAR 2003, Coimbra, Portugal, pp. 393–398 (2003)
6. Kaczorek, T.: Asymptotic stability of positive fractional 2D linear systems. *Bull. Pol. Acad. Sci. Tech.* 57(3), 289–292 (2009)
7. Kaczorek, T.: Fractional positive continuous-time systems and their Reachability. *Int. J. Appl. Math. Comput. Sci.* 18(2), 223–228 (2008)
8. Kaczorek, T.: Fractional positive linear systems. *Kybernetes: The International Journal of Systems & Cybernetics* 38(7/8), 1059–1078 (2009)
9. Kaczorek, T.: Minimum energy control of fractional positive continuous-time linear systems. In: Proc. of Conf. MMAR, Miedzyzdroje, Poland (2013)
10. Kaczorek, T.: Positive 1D and 2D Systems. Springer, London (2002)
11. Kaczorek, T.: Positive linear systems consisting of n subsystems with different fractional orders. *IEEE Trans. Circuits and Systems* 58(6), 1203–1210 (2011)
12. Kaczorek, T.: Positive linear systems with different fractional orders. *Bull. Pol. Acad. Sci. Tech.* 58(3), 453–458 (2010)
13. Kaczorek, T.: Positivity and reachability of fractional electrical circuits. *Acta Mechanica et Automatica* 5(2), 42–51 (2011)
14. Kaczorek, T.: Selected Problems in Fractional Systems Theory. Springer, Heidelberg (2011)
15. Kaczorek, T., Klamka, J.: Minimum energy control of 2D linear systems with variable coefficients. *Int. J. of Control* 44(3), 645–650 (1986)

16. Klamka, J.: Controllability of Dynamical Systems. Kluwer Academic Press, Dordrecht (1991)
17. Klamka, J.: Minimum energy control of 2D systems in Hilbert spaces. System Sciences 9(1-2), 33–42 (1983)
18. Klamka, J.: Relative controllability and minimum energy control of linear systems with distributed delays in control. IEEE Trans. Autom. Contr. 21(4), 594–595 (1976)
19. Klamka, J.: New Trends in Nanotechnology and Fractional Calculus. In: Baleanu, D., Guvenc, Z.B., Tenreiro Machado, J.A. (eds.) Controllability and minimum energy control problem of fractional discrete-time systems, pp. 503–509. Springer, New York (2010)
20. Klamka, J.: Controllability of dynamical systems-a survey. Archives of Control Sciences 2(3-4), 281–307 (1993)
21. Miller, K.S., Ross, B.: An Introduction to the Fractional Calculus and Fractional Differential Equations. Wiley, New York (1993)
22. Nishimoto, K.: Fractional Calculus. Koriama Decartess Press (1984)
23. Oldham, K.B., Spanier, J.: The Fractional Calculus. Academic Press, New York (1974)
24. Ostalczyk, P.: The non-integer difference of the discrete-time function and its application to the control system synthesis. Int. J. Sys. Sci. 31(12), 1551–1561 (2000)
25. Podlubny, I., Dorcak, L., Kostial, I.: On fractional derivatives, fractional order systems and $P\lambda D^\mu$ -controllers. In: Proc. 36th IEEE Conf. Decision and Control, San Diego, CA, pp. 4985–4990 (1997)
26. Podlubny, I.: Fractional Differential Equations. Academic Press, San Diego (1999)
27. Sajewski, Ł.: Positive realization of SISO 2D different orders fractional discrete-time linear systems. Acta Mechanica et Automatica 5(2), 122–127 (2011)

A Hybrid Approach to the Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP)

Paweł Sitek

Kielce University of Technology
sitek@tu.kielce.pl

Abstract. The paper presents a concept and application of a hybrid approach to modeling and optimization the Two-Echelon Capacitated Vehicle Routing Problem. Two environments of mathematical programming (MP) and constraint logic programming (CLP) were integrated. The strengths of MP and CLP, in which constraints are treated in a different way and different methods are implemented, were combined to use the strengths of both. The proposed approach is particularly important for the decision models with an objective function and many discrete decision variables added up in multiple constraints. The Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP) is an extension of the classical Capacitated Vehicle Routing Problem (CVRP) where the delivery depot-customers pass through intermediate depots (called satellites). Multi-echelon distribution systems are quite common in supply-chain and logistic systems. The presented approach will be compared with classical mathematical programming on the same data sets.

Keywords: Vehicle Routing, Multi-echelon systems, Constraint Logic Programming, Mathematical Programming, Optimization.

1 Introduction

The freight transportation and logistics industry are one of the sources of employment and supports the country's economic improvement. However, freight transportation with logistic operations is also a disturbing activity, due to congestion and environmental aspects, that negatively affects the quality of life and natural environment, in particular in urban areas points and along the transport routes. In freight transportation there are two main distribution strategies: direct shipping and multi-echelon distribution. In the direct shipping, vehicles, starting from a depot, bring their freight directly to the destination, while in the multi-echelon systems, freight is delivered from the depot to the customers through intermediate.

In two-echelon distribution systems, freight is delivered to an intermediate depot and, from this depot, to the customers.

The majority of multi-echelon systems presented in the literature usually explicitly consider the routing problem at the last level of the transportation system, while a simplified routing problem is considered at higher levels [1].

In practice, and uses in recent years multi-echelon systems have been introduced in different areas:

- Logistics enterprises and express delivery service companies.
- Hypermarkets products distribution.
- Multimodal freight transportation and supply chains.
- E-commerce and home delivery services.
- City and public logistics.

The vast majority of models of decision support and/or optimization in freight transportation and logistics industry have been formulated as the mixed integer programming (MIP) or mixed integer linear programming (MILP) problems and solved using the operations research (OR) methods [3]. Their structures are similar and proceed from the principles and requirements of mathematical programming [2, 9]. It seems that better results will be obtained by the use of the constraint programming environments (CP/CLP) especially in modeling. The CP-based environments have the advantage over traditional methods of mathematical modeling in that they work with a much broader variety of interrelated constraints and allow producing “natural” solutions for highly combinatorial problems. The CP/CLP environments have declarative nature [8, 11].

The main contribution of this paper is hybrid approach (mixed CP with MILP) to modeling and optimization Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP) or a similar problem. In addition, some extensions and modifications to the standard 2E-CVRP are presented.

The paper is organized as follows. In Section 2 the literature related to Multi-Echelon Vehicle Routing Problems has been reviewed. Next section is about motivation and contribution. In Section 4 the concept of hybrid approach to modeling and solving, and the hybrid solution environment have been presented. Then, the general description of Multi-Echelon Vehicle Routing Problems and mathematical model of 2E-CVRP has been discussed. Finally test instances for 2E-CVRP and some computational results were discussed in Section 6.

2 Literature Review

The Vehicle Routing Problem (VRP) is used to design an optimal route for a fleet of vehicles to service a set of customers' orders (known in advance), given a set of constraints. The VRP is used in supply chain management in the physical delivery of goods and services. The VRP is of the NP-hard type.

Nowadays, the VRP literature offers a wealth of heuristic and metaheuristic approaches, which are surveyed in the papers of [3–5] because exact VRP methods have a size limit of 50–100 orders depending on the VRP variant and the time-response requirements.

There are several variants and classes of VRP like the capacitated VRP (CVRP), VRP with Time Windows (VRPTW) and Dynamic Vehicle Routing Problems (DVRP), sometimes referred to as On-line Vehicle Routing Problems, etc. [3].

Different distribution strategies are used in freight transportation. The most developed strategy is based on the direct shipping: freight starts from a depot and arrives directly to customers. In many applications and real situations, this strategy is not the best one and the usage of a multi-echelon and particular two-echelon distribution system can optimize several features as the number of the vehicles, the transportation costs, loading factor and timing.

In the literature the multi-echelon system, and the two-echelon system in particular, refer mainly to supply chain and inventory problems [1]. These problems do not use an explicit routing approach for the different levels, focusing more on the production and supply chain management issues. The first real application of a two-tier distribution network optimizing the global transportation costs is due to [13] and is related to the city logistics area. They developed a two-tier freight distribution system for congested urban areas, using small intermediate platforms, called satellites (intermediate points for the freight distribution). This system is developed for a specific situation and a generalization of such a system has not already been formulated. The complete mathematic model of 2E-CVRP with the solution for sample test data in the classical approach has been proposed by [4], complemented with the method for boosting the computing efficiency (see Chapter/Section 5).

3 Motivation and Contribution

Based on [3–6] and our previous work [7–10] we observed some advantages and disadvantages of both (CLP/MILP) environments.

An integrated approach of constraint programming (CP/CLP) and mixed integer programming (MIP/MILP) can help to solve optimization problems that are intractable with either of the two methods alone [13–15]. Although Operations Research (OR) and Constraint Programming (CP) have different roots, the links between the two environments have grown stronger in recent years.

Both MIP/MILP and finite domain CP/CLP involve variables and constraints. However, the types of the variables and constraints that are used, and the way the constraints are solved, are different in the two approaches [15].

MIP/MILP relies completely on linear equations and inequalities in integer variables, i.e., there are only two types of constraints: linear arithmetic (linear equations or inequalities) and integrity (stating that the variables have to take their values in the integer numbers). In finite domain CP/CLP, the constraint language is richer. In addition to linear equations and inequalities, there are various other constraints: disequalities, nonlinear, symbolic (*alldifferent*, *disjunctive*, *cumulative*, etc.). In both MIP/MILP and CP/CLP, there is a group of constraints that can be solved with ease and a group of constraints that are difficult to solve. The easily solved constraints in MIP/MILP are linear equations and inequalities over rational numbers.

Integrity constraints are difficult to solve using mathematical programming methods and often the real problems of MIP/MILP make them NP-hard.

In CP/CLP, domain constraints with integers and equations between two variables are easy to solve. The system of such constraints can be solved over integer variables

in polynomial time. The inequalities between two variables, general linear constraints (more than two variables), and symbolic constraints are difficult to solve, which makes real problems in CP/CLP NP-hard. This type of constraints reduces the strength of constraint propagation. As a result, CP/CLP is incapable of finding even the first feasible solution.

Both approaches use various layers of the problem (methods, the structure of the problem, data) in different ways. The approach based on mathematical programming (MIP/MILP) focuses mainly on the methods of optimization and, to a lesser degree, on the structure of the problem. However, the data is completely outside the model. The same model without any changes can be solved for multiple instances of data. In the approach based on constraint programming (CP/CLP), due to its declarative nature, the methods are already built-in. The data and structure of the problem are used for its modelling in a significantly greater extent.

The motivation and contribution behind this work was to create a hybrid method for constrained decision problems modelling and optimization instead of using mathematical programming or constraint programming separately.

It follows from the above that what is difficult to solve in one environment can be easy to solve in the other.

Moreover, such a hybrid approach allows the use of all layers of the problem to solve it.

The hybrid method is not inferior to its component elements applied separately. This is due to the fact that the number of decision variables and the search area are reduced. The extent of the reduction directly affects the effectiveness of the method.

In our approach to modelling and optimization these problems we proposed the solution environment, where:

- knowledge related to the problem can be expressed as linear, logical and symbolic constraints;
- the decision models solved using the proposed framework can be formulated as a pure model of MIP/MILP or of CP/CLP, or it can also be a hybrid model;
- the problem is modelled in the constraint programming environment, which is far more flexible than the mathematical programming environment;
- transforming the decision model to explore its structure has been introduced;
- constrained domains of decision variables, new constraints and values for some variables are transferred from CP/CLP into MILP/MIP/IP;
- the efficiency of finding solutions to larger size problems is increased.

As a result, a more effective solution environment for a certain class of decision and optimization problems (2E-CVRP or similar) was obtained.

4 Hybrid Solution Framework (HSF)

Both environments have advantages and disadvantages. Environments based on the constraints such as CLPs are declarative and ensure a very simple modeling of decision problems, even those with poor structures if any. The problem is described by a

set of logical predicates. The constraints can be of different types (linear, non-linear, logical, binary, etc.). The CLP does not require any search algorithms. This feature is characteristic of all declarative backgrounds, in which modeling of the problem is also a solution, just as it is in Prolog, SQL, etc. The CLP seems perfect for modeling any decision problem.

Numerous MP models of decision-making have been developed and tested, particularly in the area of decision optimization. Constantly improved methods and mathematical programming algorithms, such as the simplex algorithm, branch and bound, branch-and-cost, etc., have become classics now.

The proposed method's strength lies in high efficiency of optimization algorithms and a substantial number of tested models.

Traditional methods when used alone to solve complex problems provide unsatisfactory results. This is related directly to different treatment of variables and constraints in those approaches (Chapter 3).

This schema of the Hybrid Solution Framework (HSE) and the concept of this framework with its phases (P1 .. P5, G1 .. G3) is presented in Fig. 1. The names and descriptions of the phases and the implementation environment are shown in Tab. 1.

A detailed description of the phases and their implementation aspects has been described in [16].

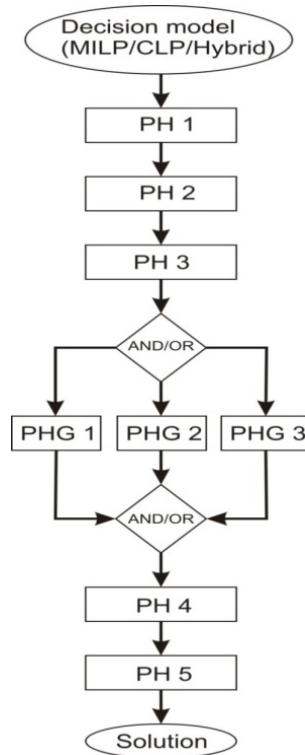


Fig. 1. Detailed scheme of the hybrid solution framework (HSF)

Table 1. Description of Phases

Phase	PH1
Name	Implementation of decision model – [environment: CLP]
Description	The implementation of the model in CLP, the term representation of the problem in the form of predicates.
Phase	PH2
Name	Transformation of implemented model for better constraint propagation (optional) – [environment: CLP]
Description	The transformation of the original problem aimed at extending the scope of constraint propagation. The transformation uses the structure of the problem. The most common effect is a change in the representation of the problem by reducing the number of decision variables, and the introduction of additional constraints and variables, changing the nature of the variables, etc.
Phase	PH3
Name	Constraint propagation – [environment: CLP]
Description	Constraint propagation for the model. Constraint propagation is one of the basic methods of CLP. As a result, the variable domains are narrowed, and in some cases, the values of variables are set, or even the solution can be found.
Phase	G1
Name	Generation of MILP/MIP/IP model – [environment: CLP]
Description	Generation of the model for mathematical programming. Generation performed automatically using CLP predicate. The resulting model is in a format accepted by the MILP system.
Phase	PHG2
Name	Generation of additional constraints (optional) – [environment: CLP]
Description	Generation of additional constraints on the basis of the results obtained in step PH3
Phase	PHG3
Name	Generation domains of decision variables and other values – [environment: CLP]
Description	Generation of domains for different decision variables and other parameters based on the propagation of constraints. Transmission of this information in the form of fixed value of certain variables and/or additional constraints to the MP.
Phase	PH4
Name	Merging MILP/MIP/IP model – [environment: MILP]
Description	Merging files generated during the phases PHG1, PHG2, PHG3 into one file. It is a model file format in MILP system.
Phase	PH5
Name	Solving MILP/MIP/IP model – [environment: MILP]
Description	The solution of the model from the previous stage by MILP solver. Generation of the report with the results and parameters of the solution.

5 Two-Echelon Capacitated Vehicle Routing Problem

The Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP) is an extension of the classical Capacitated Vehicle Routing Problem (CVRP) where the delivery depot-customers pass through intermediate depots (called satellites). As in CVRP, the goal is to deliver goods to customers with known demands, minimizing the total delivery cost in the respect of vehicle capacity constraints. Multi-echelon systems presented in the literature usually explicitly consider the routing problem at the last level of the transportation system, while a simplified routing problem is considered at higher levels [4, 12].

In 2E-CVRP, the freight delivery from the depot to the customers is managed by shipping the freight through intermediate depots. Thus, the transportation network is decomposed into two levels (Fig. 2): the 1st level connecting the depot (d) to intermediate depots (s) and the 2nd one connecting the intermediate depots (s) to the customers (c). The objective is to minimize the total transportation cost of the vehicles involved in both levels. Constraints on the maximum capacity of the vehicles and the intermediate depots are considered, while the timing of the deliveries is ignored.

From a practical point of view, a 2E-CVRP system operates as follows (Fig. 2):

- freight arrives at an external zone, the depot, where it is consolidated into the 1st-level vehicles, unless it is already carried into a fully-loaded 1st-level vehicles;
- each 1st-level vehicle travels to a subset of satellites that will be determined by the model and then it will return to the depot;
- at a satellite, freight is transferred from 1st-level vehicles to 2nd-level vehicles;

The mathematical model (MILP) was taken from [4]. Table 4 shows the parameters and decision variables of 2E-CVRP. Figure 2 shows an example of the 2E-CVRP – transportation network.

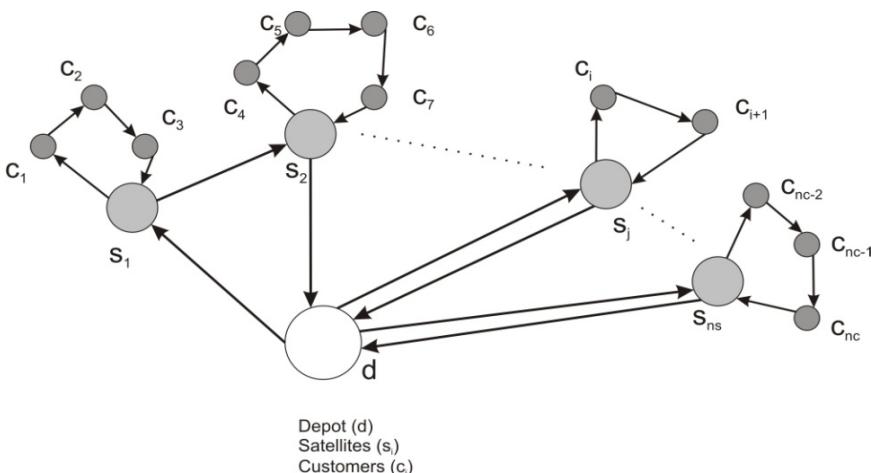


Fig. 2. Example of 2E-CVRP transportation network

Table 2. Summary indices, parameters and decision variables

Symbol	Description
<i>Indices</i>	
n_s	Number of satellites
n_c	Number of customers
$V_0 = \{v_o\}$	Deport
$V_s = \{v_{s1}, v_{s2}, v_{s3}, \dots, v_{sn_s}\}$	Set of satellites
$V_c = \{v_{c1}, v_{c2}, v_{c3}, \dots, v_{cn_c}\}$	Set of customers
<i>Parameters</i>	
M_1	Number of the 1st-level satellites
M_2	Number of the 2nd-level satellites
K_1	Capacity of the vehicles for the 1st level
K_2	Capacity of the vehicles for the 2nd level
d_i	Demand required by customer i
$c_{i,j}$	Cost of the arc (i,j)
s_k	Cost of loading/unloading operations of a unit of freight in satellite k
<i>Decision variables</i>	
$X_{i,j}$	An integer variable of the 1st-level routing is equal to the number of 1st-level vehicles using arc (i,j)
$Y_{k,i,j}$	A binary variable of the 2nd-level routing is equal to 1 if a 2nd-level vehicle makes a route starting from satellite k and goes from node i to node j and 0 otherwise
$Q_{i,j}^1$	The freight flow arc (i,j) for the 1st-level
$Q_{k,i,j}^2$	The freight arc (i,j) where k represents the satellite where the freight is passing through.
$Z_{k,j}$	A binary variable that is equal to 1 if the freight to be delivered to customer j is consolidated in satellite k and 0 otherwise

$$\min \sum_{i,j \in V_0 \cup V_s} (c_{i,j} \cdot X_{i,j}) + \sum_{k \in V_s \cup V_0 \cup V_c} (c_{i,j} \cdot Y_{k,i,j}) + \sum_{k \in V_s} (s_k \cdot Ds_k) \quad (1)$$

$$\sum_{i \in V_s} X_{0,i} \leq M_1 \quad (2)$$

$$\sum_{j \in V_s \cup V_0, j \neq k} X_{j,k} = \sum_{i \in V_s \cup V_0, i \neq k} X_{k,i} \text{ for } k \in V_s \cup V_0 \quad (3)$$

$$\sum_{k \in V_s} \sum_{j \in V_c} Y_{k,k,j} \leq M_2 \quad (4)$$

$$\sum_{i \in V_c, j \in V_c} Y_{k,i,j} = \sum_{i \in V_c, j \in V_c} Y_{k,j,i} \text{ for } k \in V_s \quad (5)$$

$$\sum_{i \in V_0 \cup V_s, i \neq j} Q_{i,j}^1 - \sum_{i \in V_s, i \neq j} Q_{j,i}^1 = \begin{cases} Ds_j & j \text{ is not the deport} \\ \sum_{i \in V_c} - d_i & \text{otherwise} \end{cases} \text{ for } j \in V_s \cup V_0 \quad (6)$$

$$Q_{i,j}^1 \leq K_1 \cdot X_{i,j} \text{ for } i, j \in V_s \cup V_0, i \neq j \quad (7)$$

$$\sum_{i \in V_s \cup V_c, i \neq j} Q_{k,i,j}^2 - \sum_{i \in V_c, i \neq j} Q_{k,j,i}^2 = \begin{cases} Z_{k,j} d_j & j \text{ is not a satellite} \\ -D_j & \text{otherwise} \end{cases} \text{ for } j \in V_c \cup V_s, k \in V_s \quad (8)$$

$$Q_{k,i,j}^2 \leq k_2 \cdot Y_{k,i,j} \text{ for } i, j \in V_s \cup V_c, i \neq j, k \in V_s \quad (9)$$

$$\sum_{i \in V_s} Q_{i,V_0}^l = 0 \quad (10)$$

$$\sum_{j \in V_c} Q_{k,j,k}^2 = 0 \text{ for } k \in V_s \quad (11)$$

$$Y_{k,i,j} \leq Z_{k,j} \text{ for } i \in V_s \cup V_c, j \in V_c, k \in V_s \quad (12)$$

$$Y_{k,j,i} \leq Z_{k,j} \text{ for } i \in V_s, j \in V_c, k \in V_s \quad (13)$$

$$\sum_{i \in V_s \cup V_c} Y_{k,i,j} = Z_{k,j} \text{ for } k \in V_s, j \in V_c, i \neq k \quad (14)$$

$$\sum_{i \in V_s} Y_{k,j,k} = Z_{k,j} \text{ for } k \in V_s, j \in V_c, i \neq k \quad (15)$$

$$\sum_{i \in V_s} Z_{i,j} = 1 \text{ for } j \in V_c \quad (16)$$

$$Y_{k,i,j} \leq \sum_{l \in V_s \cup V_0} X_{k,l} \text{ for } k \in V_s, i, j \in V_c \quad (17)$$

$$Y_{k,i,j} \in \{0,1\}, Z_{k,l} \in \{0,1\} \text{ for } k \in V_s, i, j \in V_s \cup V_c, l \in V_c \quad (18)$$

$$X_{k,j} \in \mathbb{Z}^+ \text{ for } k, j \in V_s \cup V_0 \quad (19)$$

$$Q_{i,j}^l \geq 0 \text{ for } i, j \in V_s \cup V_0; Q_{k,i,j}^2 \geq 0 \text{ for } i, j \in V_s \cup V_c, k \in V_s \quad (20)$$

$$Ds_k = \sum_{i \in V_c} (d_j \cdot Z_{k,j}) \text{ for } k \in V_s$$

$$\sum_{i, j \in S_c} Y_{k,i,j} \leq |S_c| - 1 \text{ for } S_c \subset V_c, 2 \leq |S_c| \leq |V_c| - 2 \quad (21)$$

$$Q_{k,i,j}^2 \leq (k_2 - d_j) \cdot Y_{k,i,j} \text{ for } i, j \in V_c, k \in V_s \quad (22)$$

$$Q_{k,i,j}^2 - \sum_{l \in V_s} Q_{k,j,l}^2 \leq (k_2 - d_j) \cdot Y_{k,i,j} \text{ for } i, j \in V_c, k \in V_s \quad (23)$$

The objective function minimizes the sum of the routing and handling operations costs. Constraints (3) ensure, for $k = v_0$, that each 1st-level route begins and ends at the depot, while when k is a satellite, impose the balance of vehicles entering and leaving that satellite. Constraints (5) force each 2nd-level route to begin and end to one satellite and the balance of vehicles entering and leaving each customer. The number of the routes in each level must not exceed the number of vehicles for that level, as imposed by constraints (2) and (4). The flows balance on each network node is equal to the demand of this node, except for the depot, where the exit flow is equal to the total demand of the customers, and for the satellites at the 2nd-level, where the flow is equal to the demand (unknown) assigned to the satellites which provide constraints (6) and (8). Moreover, constraints (6) and (8) forbid the presence of sub-tours not containing the depot or a satellite, respectively. In fact, each node receives an amount of flow equal to its demand, preventing the presence of sub-tours. Consider, for example, that a sub-tour is present between the nodes i, j and k at the 1st level. It is easy to check that, in such a case, does not exist any value for the variables $Q_{1,ij}$, $Q_{1,jk}$ and $Q_{1,ki}$ satisfying the constraints (6) and (8). The capacity constraints are

formulated in (7) and (9), for the 1st-level and the 2nd-level, respectively. Constraints (10) and (11) do not allow residual flows in the routes, making the returning flow of each route to the depot (1st-level) and to each satellite (2nd-level) equal to 0. Constraints (12) and (13) indicate that a customer j is served by a satellite k ($z_{kj} = 1$) only if it receives freight from that satellite ($y_{kij} = 1$). Constraint (16) assigns each customer to one and only one satellite, while constraints (14) and (15) indicate that there is only one 2nd-level route passing through each customer and connect the two levels. Constraints (17) allow to start a 2nd-level route from a satellite k only if a 1st-level route has served it. Constraints from 17 to 20 result from the character of the MILP-formulated problem. Additional constraints were introduced by [4] to increase the solution search efficiency. They strengthen the continuous relaxation of the flow model. In particular, authors in [4] used two families of cuts, one applied to the assignment variables derived from the sub-tour elimination constraints (edge cuts) and the other based on the flows. The edge cuts explicitly introduce the well-known sub-tours elimination constraints derived from the TSP (Traveling Sales Problem). They can be expressed as constraints (21). The inequalities explicitly forbid the presence in the solution of sub-tours not containing the depot, already forbidden by the constraints (8). The number of potential valid inequalities are exponential, so that each customer reduces the flow of an amount equal to its demand d_i – constraints (22) and (23).

A possibility to transform the hybrid approach model in the CLP environment (phase PH2) is an important aspect of that approach.

The our transformation of this model in the hybrid approach focused on the resizing of $Y_{k,i,j}$ decision variable by introducing additional imaginary volume of freight shipped from the satellite and re-delivered to it. Such transformation resulted in two facts. First of all, it forced the vehicle to return to the satellite from which it started its trip. Secondly, it reduced decision variable $Y_{k,i,j}$ to variable $Y_{i,j}$ which decreased the size of the combinatorial problem.

6 Computational Tests - Two-Echelon Capacitated Vehicle Routing Problem

For the final validation of the proposed hybrid approach, the benchmark data for 2E-CVRP was selected. 2E-CVRP, a well described and widely discussed problem, corresponded to the issues to which our hybrid approach was applied.

The instances for computational examples were built from the existing instances for CVRP [17] denoted as E-n13-k4. All the instance sets can be downloaded from the website [18]. The instance set was composed of 5 small-sized instances with 1 depot, 12 customers and 2 satellites. The full instance consisted of 66 small-sized instances because the two satellites were placed over twelve customers in all 66 possible ways (number of combinations: 2 out of 12).

All the instances had the same position for depot and customers, whose coordinates were the same as those of instance E-n13-k4. Small-sized instances differed in the choice of two customers who were also satellites (En13-k4-1, En13-k4-5, En13-k4-9, En13-k4-12, etc.).

Numerical experiments were conducted for the same data in three runs. The first run was a classical implementation of model (1)..(20) and its solution in the MILP environment. The second run used the same environment for model (1)..(23) with additional edge-cuts. In the final run the model (1)..(20) and its solution were implemented in the proposed framework. The calculations were performed using a computer with the following specifications: Intel(R) Core(TM) 2 QQuad CPU Q6600 @ 2 × 2,40 GHz 2,4GHz RAM 1,98 GB.

The analysis of the results for the benchmark instances demonstrates that the hybrid approach may be a superior approach to the classical mathematical programming. For all examples, the solutions were found 4–16 times faster than they are in the classical approach (Tab.3). In many cases the calculations ended after 600 s as they failed to indicate that the solution was optimal.

Table 3. The results of numerical examples for 2E-CVRP

E-n13-k4	MILP				MILP+ edge cuts constraints							
	F _c	T	C	V(int V)	F _c	T	C	V(int V)				
E-n13-k4-05	218	108	1262	744(368)	218	48	1982	744(368)				
E-n13-k4-09	244	93	1262	744(368)	244	67	1982	744(368)				
E-n13-k4-12	290	600*	1262	744(368)	290	600*	1982	744(368)				
E-n13-k4-22	312	600*	1262	744(368)	312	600*	1982	744(368)				
E-n13-k4-01	280	600*	1262	744(368)	280	600*	1982	744(368)				
E-n13-k4-13	288	600*	1262	744(368)	288	600*	1982	744(368)				
E-n13-k4-61	338	600*	1262	744(368)	338	600*	1982	744(368)				
E-n13-k4-66	400	600*	1262	744(368)	400	600*	1982	744(368)				
Hybrid												
E-n13-k4-05	218	9,58	21	1082(1079)								
E-n13-k4-09	244	13,22	21	1082(1079)								
E-n13-k4-12	290	24,38	21	1082(1079)								
E-n13-k4-22	312	9,97	21	1082(1079)								
E-n13-k4-01	280	20,36	21	1082(1079)								
E-n13-k4-13	288	18,14	21	1082(1079)								
E-n13-k4-61	338	12,20	21	1082(1079)								
E-n13-k4-66	400	14,28	21	1082(1079)								
F_c	the optimal value of the objective function											
T	time of finding solution											
V(int V)/C	the number of variables (integer variables) /constraints											
*	calculations stopped after 600 s, the feasible value of the objective function											

As the presented example was formulated as a MILP problem, the HSF was tested for the solution efficiency. Owing to the hybrid approach the 2E-CVRP models can be extended over logical, nonlinear, and other constraints. At the next stage logical constraints were introduced into the model. The logical relationship between mutually exclusive variables was taken into account, which in real-world distribution systems means that the same vehicle cannot transport two types of selected goods or two

points cannot be handled at the same time. Those constraints result from technological, marketing, sales or safety reasons. Only declarative application environments based on constraint satisfaction problem (CSP) make it possible to implement constraints such as. Table 4 presents the results of the numerical experiments conducted for 2E-CVRPs with logical constraints relating to the situation where two delivery points (customers) can be handled separately but not together in one route.

Table 4. The results of numerical examples for 2E-CVRP with logical constraints

E-n13-k4	F _c	T	C	V	exCustomer*
E-n13-k4-05	232	7,55	21	788 (785)	1,3; 1,4; 1,6; 2,8
E-n13-k4-09	252	8,42	21	788 (785)	1,3; 1,4; 1,6; 2,8
E-n13-k4-12	290	12,43	21	788 (785)	1,3; 1,4; 1,6; 2,8
E-n13-k4-22	314	12,91	21	788 (785)	1,3; 1,4; 1,6; 2,8

* pairs of customers that cannot be served on one route

7 Conclusion and Discussion on Possible Extension

The efficiency of the proposed approach is based on the reduction of the combinatorial problem and using the best properties of both environments. The hybrid approach (Table 3, Table 4) makes it possible to find solutions in the shorter time.

In addition to solving larger problems faster, the proposed approach provides virtually unlimited modeling options with many types of constraints. Therefore, the proposed solution is recommended for decision-making problems that have a structure similar to the presented models (Section 5). This structure is characterized by the constraints and objective function in which the decision variables are added together. Further work will focus on running the optimization models with non-linear and other logical constraints, multi-objective, uncertainty etc. in the hybrid optimization framework. The planned experiments will employ HSF for Two-Echelon Capacitated VRP with Time Windows (2E-CVRP-TW), Two-Echelon Capacitated VRP with Satellites Synchronization (2E-CVRPSS) and 2E-CVRP with Pickup and Deliveries (2E-CVRP-PD).

References

1. Verrijdt, J., de Kok, A.: Distribution planning for a divergent n-echelon network without intermediate stocks under service restrictions. International Journal of Production Economics 38, 225–243 (1995)
2. Schrijver, A.: Theory of Linear and Integer Programming. John Wiley & Sons (1998) ISBN 0-471-98232-6
3. Kumar, S.N., Panneerselvam, R.: A Survey on the Vehicle Routing Problem and Its Variants. Intelligent Information Management 4, 66–74 (2012)
4. Perboli, G., Tadei, R., Vigo, D.: The Two-Echelon Capacitated Vehicle Routing Problem: Models and Math-Based Heuristics. Transportation Science 45, 364–380 (2011)

5. Bocewicz, G., Banaszak, Z.: Declarative approach to cyclic steady states space refinement: periodic processes scheduling. *International Journal of Advanced Manufacturing Technology* 67(1-4), 137–155 (2013)
6. Relich, M.: Project prototyping with application of CP-based approach. *Management* 15(2), 364–377 (2011)
7. Sitek, P., Wikarek, J.: Supply chain optimization based on a MILP model from the perspective of a logistics provider. *Management and Production Engineering Review*, 49–61 (2012)
8. Sitek, P., Wikarek, J.: A Declarative Framework for Constrained Search Problems. In: Nguyen, N.T., Borzemski, L., Grzech, A., Ali, M. (eds.) IEA/AIE 2008. LNCS (LNAI), vol. 5027, pp. 728–737. Springer, Heidelberg (2008)
9. Sitek, P., Wikarek, J.: Cost optimization of supply chain with multimodal transport. In: Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1111–1118 (2012)
10. Sitek, P., Wikarek, J.: A hybrid approach to supply chain modeling and optimization. In: Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1223–1230 (2013)
11. Apt, K., Wallace, M.: Constraint Logic Programming using Eclipse. Cambridge University Press (2006)
12. Crainic, T., Ricciardi, N., Storchi, G.: Advanced freight transportation systems for congested urban areas. *Transportation Research Part C* 12, 119–137 (2004)
13. Jain, V., Grossmann, I.E.: Algorithms for hybrid MILP/CP models for a class of optimization problems. *INFORMS Journal on Computing* 13(4), 258–276 (2001)
14. Milano, M., Wallace, M.: Integrating Operations Research in Constraint Programming. *Annals of Operations Research* 175(1), 37–76 (2010)
15. Achterberg, T., Berthold, T., Koch, T., Wolter, K.: Constraint integer programming: A new approach to integrate CP and MIP. In: Trick, M.A. (ed.) CPAIOR 2008. LNCS, vol. 5015, pp. 6–20. Springer, Heidelberg (2008)
16. Wikarek, J.: Implementation aspects of hybrid solution framework. In: Automation 2014. AISC, vol. 267, pp. 309–320. Springer, Heidelberg (2014)
17. Christofides, N., Elion, S.: An algorithm for the vehicle dispatching problem. *Operational Research Quarterly* 20, 309–318 (1969)
18. <http://www.orggroup.polito.it/>

Identification of Thermal Response, of Plasmatron Plasma Reactor

Jakub Szałatkiewicz, Roman Szewczyk, Eugeniusz Budny,
Tadeusz Missala, and Wojciech Winiarski

Industrial Research Institute for Automation and Measurements PIAP
jszalatkiewicz@piap.pl, jakub.szalatkiewicz@gmail.com

Abstract. This paper presents identification of heat characteristic of plasmatron powered plasma reactor designed and build in Industrial Research Institute for Automation and Measurements. Plasma reactor is the key element of the test setup, designed to research processing and recovery of metals from waste of electric and electronic equipment, focusing on electronic printed circuit boards. Identification was based on step response of the reactor, to step input power of the plasmatrons. Two experiments were carried out. First experiment with power of 36 kW, and second experiment using 29 kW. In each experiment two temperatures were measured in two points of the reactor chamber, allowing identification of four thermal responses. Object transfer function $G(s)$ was assumed as second order inertial. This assumption was confirmed during identification. The model will be utilized for determination of PID controller with additional functionality.

Keywords: inertial object identification, recycling, electronic waste utilization, plasma technology.

1 Introduction

Increasing digitalization of appliances, and machines equip them with electronic circuit boards. Those electronic circuit boards after end of life of appliance, becomes hazardous waste that needs to be treated properly. Mass production of electric and electronic equipment requires huge amounts of nonrenewable resources, like metals including precious, and rare earth. It is important to develop new effective ways of treating WEEE waste like electronic printed circuit boards (PCB), because they are also new “renewable” resource that can supply recycled metals for new production. By processing the waste of electronic circuit boards, it is possible to recover metals, energy, and decrease its hazardous effect on environment.

Waste of electrical and electronic equipment (WEEE) is a global concern. In the 27 EU countries it is estimated that the weight of produced waste WEEE in 2005 was 8.3 - 9.1 million Mg (tones), 25% of which is collected and processed, while remaining 75% is not registered and does not occur in collection points [1,2]. Such state of waste management system can be caused by lack of processing capacities and suitable technologies which can utilize WEEE effectively. The amount of WEEE rises continuously [3,4]

in 2008 Sweden collects 16.7 kg/capita of WEEE, Britain 8.2 kg/capita, Austria 6,5 kg/capita [5]. Moreover European Commission proposes rising collection targets from 4 kg/capita to 65% of average mass of electrical and electronic equipment placed on market (WEEE directive 2002/96/EC) [6]. WEEE has to be utilized, but it also can become a source of valuable resources. Need for technology allowing recovery and neutralization of this waste is strong in Poland, due to huge technological and organizational gap between Poland and west European countries.

Traditional simple WEEE processing technologies i.e. manual dismantling, milling, allow recovery of most of the waste mass. Also new robotic technologies offer new approach to WEEE dismantling decreasing human labor and energy consumption in waste treatment [7]. However every waste processing technology also generates waste, and does not allow full neutralization and recovery. One of such waste that requires specialist processes are the printed circuit boards, and the second one is the “under sieve” fraction from milling of WEEE waste. Currently in Europe only few plants process electronic printed circuit boards, and they use pyrometallurgical processes. There is no such installation in Poland. Important is also that the waste of electronic printed circuit boards, is only a part of the total input in those technologies. Moreover apart of those, there are no complete processes for neutralization of waste of printed circuit boards and recovery of metals forms them. That is why research project was undertaken, financed by Polish National Centre for Research and Development, to investigate and design plasma process allowing processing of waste of printed circuit boards and recovery of metals they contain.



Fig. 1. Overview of laboratory setup: 1) Plasma reactor, 2) Plasmatron, 3) Molten product collection, 4) Fumes Exhaust – chimney, 5) Waste package transporter, 6) Plasmatron power supply, 7) PLC – automation and data collection apparatus cabinet, 8) Automatic waste package feeder

In Industrial Research Institute for Automation and Measurements PIAP the test setup was designed and constructed to investigate plasma processing of waste of electronic and electric equipment for recovering of metals and its neutralization. The stand is

presented on fig. 1. The key component of the test setup is the plasma reactor, equipped with three plasmatrons - plasma sources that are located 120° around the reactor chamber. The test position is equipped with peripheral systems, measurement and control apparatus for data acquisition and control of the process, during research.

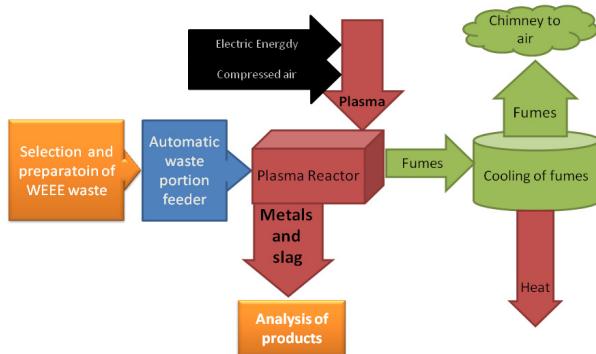


Fig. 2. Block diagram of the designed process for research over high temperature plasma technology for metals recovery and electronic waste utilization

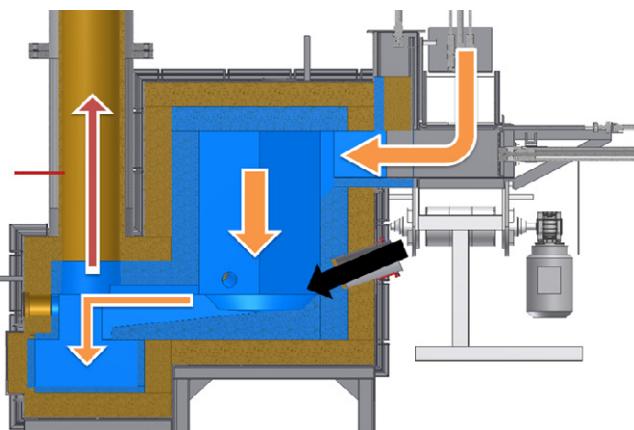


Fig. 3. Cross section through plasma reactor with presented material flow. Waste, and molten product – orange arrows, plasma stream – black arrow, fumes – red arrow

The high temperature plasmatron plasma reactor is the key component of laboratory setup for research over high temperature utilization of waste of printed circuit boards, for metals recovery. Block diagram of this laboratory setup is presented on figure 2. Designed test setup allows wide range of possible experiments and data acquisition during research over waste processing and metals recovery. Designed plasma proces is being carried out by the fallowing steps presented on figure 2. Prepared waste portion is transported through automatic feeder to the plasma reactor chamber. In the reactor chamber waste is being incinerated and molten by three plasma streams, next the incineration fumes are being transported to the scrubber where they are

neutralized, cooled and then released to the atmosphere. As to the metals and slag, in molten form metals and slag flow out from the reactor and set in casts, from which it can be recovered and recycled. Figure 3 presents cross section of the reactor, arrows marks the waste and metals route (orange arrow), plasma stream (black arrow), and the fumes exhaust direction (red arrow).

2 Reactor Chamber Construction

Reactor chamber construction consist three layers: first from inside is fire proof concrete, next is the thermal insulation, and last is the external metal construction shell. Reactor chamber is hexagonal and its construction is presented on the figures 3 and 4. Such construction allows containing in its volume temperatures ranging for 1500 °C up to 1650 °C. However, in the area where the plasma streams has direct effect on waste, the temperatures exceeds the temperature measured above, however due to difficulties of measurement of temperatures above 2000 °C, this temperature currently is not measured.

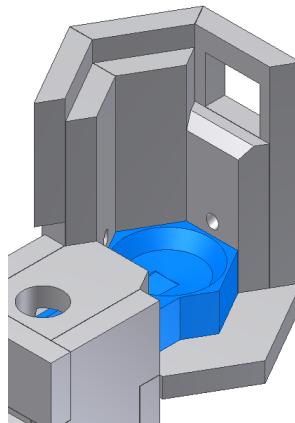


Fig. 4. Internal construction of plasma reactor – CAM model

The plasmatron plasma reactor has three sources of heat, which are 20 kW arc plasmatrons. Plasmatrons efficiency reaches 80% of energy to plasma heat efficiency. However calculating the plasmatron efficiency including efficiency of the power source, overall efficiency decreases to 70%.

Each plasmatron generates stream of plasma that flow out at the bottom of the reactor chamber. Plasma is produced from compressed air that is used as plasmatron working gas. Three plasmatrons consume 11 Nm³/h of air during normal operation.

2.1 Measurement of Temperature in the Reactor

Temperature inside the reactor chamber is measured in two points located 30 mm away from reactor wall, and placed in theirs centre. Temperature probe A is located

350 mm above reactors surface, and probe B is located 30 mm above reactor surface. Both sensors are thermocouples type B in ceramic cover. Sensors placement is shown on figure 5.

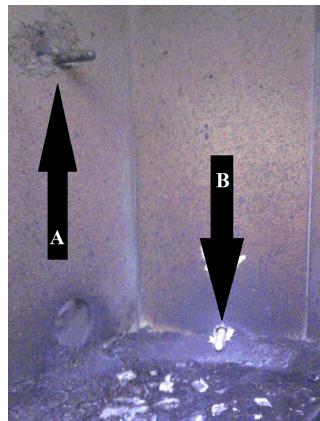


Fig. 5. Placement of the temperature measurement probes: A – 350 mm above reactor surface, B – 30 mm above reactor surface

3 Identification of Thermal Response

Thermal parameters identification of the plasma reactor was carried out during its initial heat up with use of 3 plasmatrons without any input material. Two experiments were carried out: First using 36 kW, and second – using 29 kW of thermal power, provided to the reactor chamber by 3 plasmatrons. For both experiments 2 temperatures were measured in two points of the reactor chamber (A and B), this allowed identification of four thermal responses.

Identification was based on step response of the reactor, with given step input power of the plasmatrons for each experiment. Object transfer function $G(s)$ was assumed as inertial second order given by equation (1). [8]

$$G(s) = \frac{K_p}{(1 + T_{p1}s)(1 + T_{p2}s)} \quad (1)$$

K_p – gain,
 T_{p1}, T_{p2} – time constants of inertial elements

This assumption has physical justification, and was confirmed during identification. In the reactor, there are two areas where the heat is being accumulated. First of it is the volume of the chamber (presented in the middle of figures 3 and 4), and the second one is the concrete wall of the reactor chamber. It is obvious that heat accumulation

constants will be different for those two masses. The third area of heat accumulation (heat insulation) is not taken into account.

Identification of transfer function was carried out assuming that this function is linear. This assumption is approximation of the real response, because heat conductance of concrete and other materials rises with temperature. However inertia of the object is high enough to narrow the identification error.

To identify reactor parameters according to transfer function presented in equation (1) gradient oriented optimization method was applied. During identification the delay of the object was confirmed to be negligible. As a result it is not represented in equation (1). Results of identification for four solutions are presented in the fig 6. Also the modeling results, plotted with measured temperature and input power level, are shown on figures 7, 8, 9, 10.

Experiment I (36kW)		Experiment II (29kW)	
Measurement point A	Measurement point B	Measurement point A	Measurement point B
$K_p = 94,8 \pm 0,7$	$K_p = 47,45 \pm 0,09$	$K_p = 106 \pm 0,3$	$K_p = 33,55 \pm 0,05$
$T_{p1} = 84235 \pm 934$	$T_{p1} = 21455 \pm 105,1$	$T_{p1} = 183700 \pm 0$	$T_{p1} = 7971 \pm 47$
$T_{p2} = 201,8 \pm 1,6$	$T_{p2} = 224,55 \pm 5,1$	$T_{p2} = 103 \pm 0$	$T_{p2} = 167 \pm 3,8$
$T_{max} = 3384 \text{ } ^\circ\text{C}$	$T_{max} = 1692 \text{ } ^\circ\text{C}$	$T_{max} = 3074 \text{ } ^\circ\text{C}$	$T_{max} = 957 \text{ } ^\circ\text{C}$

Fig. 6. Modeling results, and maximum temperature calculated by equation (2)

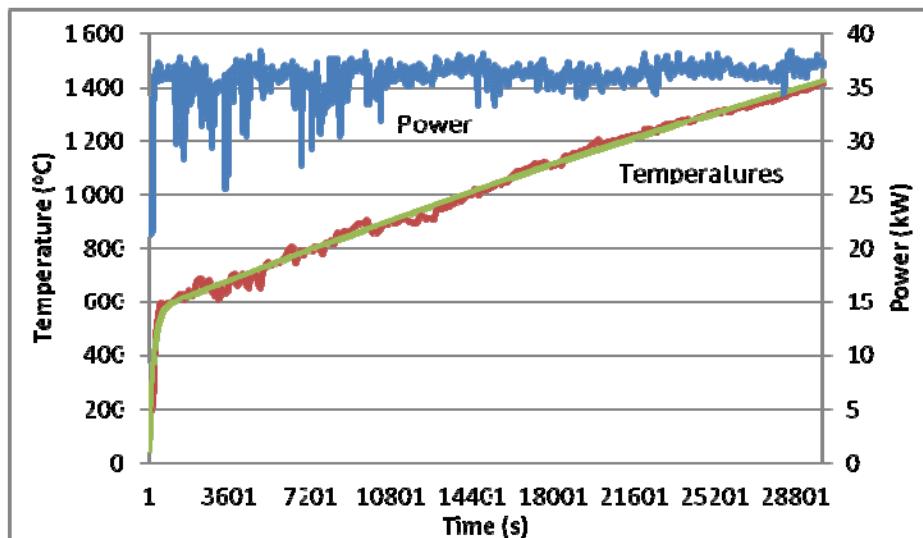


Fig. 7. Experiment I – Temperature measurement point “A” in plasma reactor (red), result of modeling (green), and input power (blue)

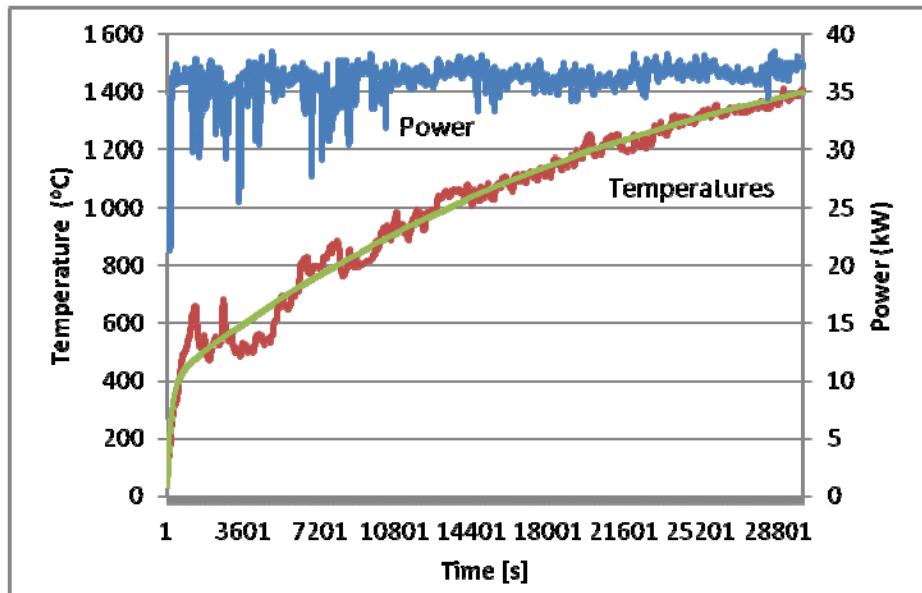


Fig. 8. Experiment I – Temperature measurement point “B” in plasma reactor (red), result of modeling (green), and input power (blue)

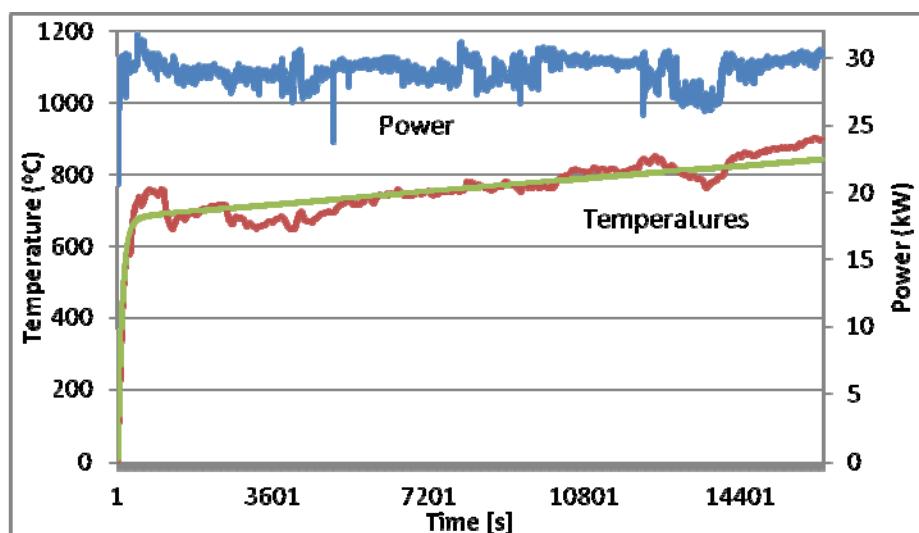


Fig. 9. Experiment II – Temperature measurement point “A” in plasma reactor (red), result of modeling (green), and input power (blue)

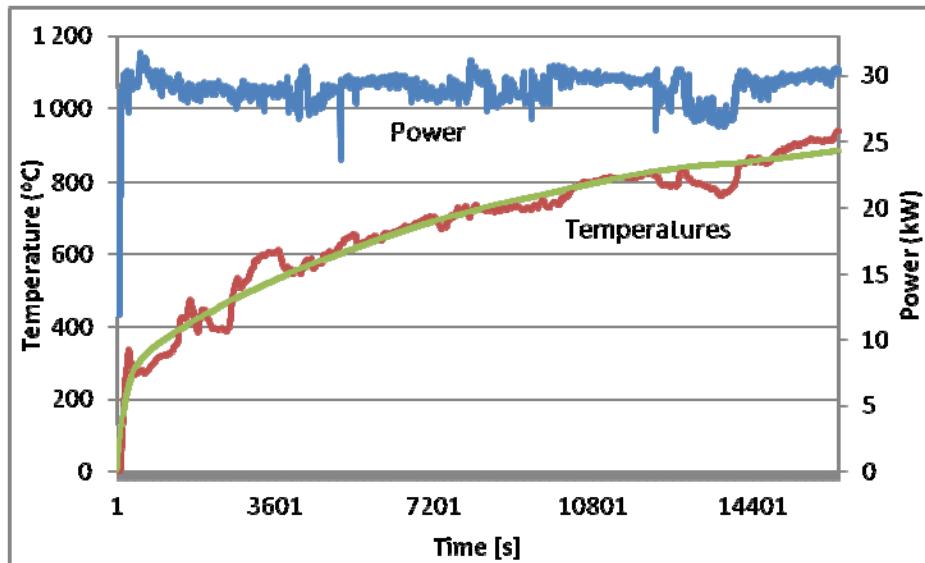


Fig. 10. Experiment II – Temperature measurement point “B” in plasma reactor (red), result of modeling (green), and input power (blue)

As it is presented on figures 7 to 10, and in the results figure 6, the plasmatron plasma reactor chamber has different heating characteristics in each point. The difference between temperature measurement point A and point B results of location of both points. Point A is located on flat vertical surface, allowing gas circulation. Point B is located close to two flat surfaces – walls, one horizontal and second vertical – bottom of the reactor. That is why its response is slower due to high capacity of surrounding walls and smaller convection. Location of both thermocouples can be described as measuring temperature of the gases and radiation inside the chamber in point A, and the temperature of the concrete in point B directly in the melting area.

Presented results show that the time constant of heating up of the reactor chamber, is about 3 minutes, and the time constant of heating of the concrete walls of the reactor is over 22 hours.

Transmittance (1) allows calculation of reactor temperature that can be reached during long heating time. This temperature can be calculated from the equation (2).

$$T_{\max} = K_p \cdot P \quad (2)$$

where T is maximum temperature ($^{\circ}\text{C}$), and P is the average sum of heat delivered to the reactor by plasmatrons (kW).

Using equation (2) the maximum temperature of the reactor was calculated for each experiment and measurement point, the results are presented in the figure 6. It should be indicated that the reactor have temperature limit due to use of certain materials. That is why above $1800\text{ }^{\circ}\text{C}$ some of reactor components will be destroyed.

Method of identification of plasmatron plasma reactor heat transfer parameters was presented on chosen power levels used in continuous experiments. Obtained object transmittance can be used to develop the control algorithm allowing stabilization of temperature inside the reactor on required level in range of 1200 °C to 1700 °C is presented in previous paper [9].

4 Summary

Determination of thermal characteristics, of plasmatron plasma reactor confirmed, that its behavior is similar to inertial second order object. Accepted for modeling of temperature response object transmittance, allows calculation of maximum temperature it will reach after given time. Also this transmittance was used to develop automatic temperature control algorithm to stabilize reactors parameters. Such automatic control is very important in undergoing research, allowing examination of energy saving by reducing of plasmatrons power, and stabilization of process parameters. Presented identification will be also used in mathematical modeling of plasmatron plasma reactor equipped with three plasma sources.

Developed plasmatron plasma reactor will allow printed circuit board waste neutralization and metals recovery in decentralized scale reducing waste transportation [10].

Acknowledgements. The work was financed from the means of the research and development project no. N R03 0083 10, founded by the Ministry of Science and Higher Education of Poland and realized through The National Centre for Research and Development (NCBiR).

References

1. Huisman, J.: Review of Directive 2002/96 on WEEE, Final Report, United Nations University (2007)
2. Cobbing, M.: Toxic Tech: Not In Our Backyard, Greenpeace.org (2008)
3. Lee, J., Song, H., Yoo, J.: Present status of the recycling of waste electrical and electronic equipment in Korea. Resources Conservation & Recycling (50), 380–397 (2007)
4. Kang, H., Schoenung, J.: Electronic waste recycling: A review of U.S. infrastructure and technology options. Resources Conservation & Recycling (45), 368–400 (2005)
5. Wawrzonek, R.: "Praktyczne aspekty funkcjonowania systemu gospodarowania zużytym sprzętem elektrycznym i elektronicznym", Elektro Eco (2009)
6. Environment Commission proposes revised laws on recycling and use of hazardous substances in electrical and electronic equipment,
<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/08/1878&format=HTML&aged=1&language=EN&guiLanguage=en>
7. Szałatkiewicz, J., Szewczyk, R.: Intelligent Utilization of Waste of Electrical and Electronic Equipment (Weee) with Robotized Tool. Journal of Automation, Mobile Robotics & Intelligent Systems 6(4) (2012)

8. Szewczyk, R., Szałatkiewicz, J., Budny, E., Missala, T., Winiarski, W.: Identyfikacja wybranych parametrów plazmotronowego reaktora plazmowego. *Pomiary Automatyka Robotyka* (11) (2012)
9. Szałatkiewicz, J., Szewczyk, R., Budny, E., Missala, T., Winiarski, W.: Determination of PID control parameters of plasmatron plasma reactor. *Journal of Applied Computer Science Methods* 4(2) (2012)
10. Szałatkiewicz, J., Szewczyk, R., Budny, E., Missala, T., Winiarski, W.: Construction aspects of plasma based technology for waste of electrical and electronic equipment (WEEE) management in urban areas. *Procedia Engineering* 57, 1101–1108 (2013)

Computational Problems Connected with Jiles-Atherton Model of Magnetic Hysteresis

Roman Szewczyk

Industrial Research Institute for Automation and Measurements,
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
rszewczyk@piap.pl

Abstract. Paper presents the most important problems connected with Jiles-Atherton model of magnetic B(H) hysteresis. These problems are mainly caused by accuracy of numerical integration as well as methods of solving the ordinary differential equations. Paper presents comparison of accuracy of calculation with MATLAB and OCTAVE for both Windows 7 and Scientific Linux 6.3. Moreover, the analyse of time efficiency is presented. On the base of numerical errors analyses and benchmarking, the guidelines for calculation of Jiles-Atherton model are given.

Keywords: ordinary differential equation, numerical error analyse, accuracy, magnetization.

1 Introduction

Recently Jiles-Atherton model of magnetization process [1, 2] is intensively developed [3–5] for both crystalline [6, 7] and amorphous alloys [8, 9]. However, development is focused on both physical [10–12] and engineering [13–15] aspects of this model, whereas accuracy assessment of numerical calculation seems to be neglected. This accuracy assessment is especially important, due to the fact, that calculation of Jiles-Atherton model requires solving of ordinary differential equations. Moreover, in the case of anisotropic materials, solving of Jiles-Atherton model's equations require numerical integration of sophisticated functions.

In addition, the software for calculation of Jiles-Atherton model is not presented together with results of calculation. As a result it is very difficult to determine the numerical errors of calculation. This lead to inconsistent results of calculations carried out by different researchers caused by different methods of calculation.

Presented paper is filling this gap. Different methods of calculation of Jiles-Atherton model are compared from the point of view of both its efficiency as well as numerical errors. Moreover, these errors are compared from the point of view of used matrix-oriented software (MATLAB and OCTAVE) as well as from the point of view of operating systems.

2 Jiles-Atherton Model of Magnetic Hysteresis

Modelling the magnetic hysteresis with Jiles-Atherton model covers two main steps [2]. In the first step, anhysteretic magnetization M_{ah} is calculated. Then, in the second step, hysteresis is modelled by differential equation considering the sign of changes of magnetizing field H .

Anhysteretic magnetization is value of magnetization M_{ah} in the material after electrical demagnetization process under constant value of magnetizing field H [1]. Demagnetization is carried out by exponentially damped sine wave current. However, measurement of anhysteretic magnetization is sophisticated from practical point of view, due to the fact, that it requires reliable magnetic flux density B measurement during dynamic changes of magnetizing field due to the demagnetization. As a result anhysteretic magnetization curve is rarely presented.

In Jiles-Atherton model anhysteretic magnetization in the ferromagnetic materials is modelled similarly to the model of magnetization of paramagnetic material [1, 2]. In the case of paramagnetic materials, value of magnetization M_{para} may be determined considering the Boltzman distribution of magnetic domain directions [16] given by following equation:

$$M_{para} = M_s \frac{\int_0^{\pi} e^{-\frac{E_m(\theta)}{k_B T}} \sin \theta \cdot \cos \theta \cdot d\theta}{\int_0^{\pi} e^{-\frac{E_m(\theta)}{k_B T}} \sin \theta \cdot d\theta} \quad (1)$$

where M_s is saturation magnetization of paramagnetic material, θ is angle between the atomic magnetic moment m_{at} and direction of magnetizing field H and k_B is Boltzman constant. Energy of the magnetic moment $E_m(\theta)$ is given as:

$$E_m(\theta) = -\mu_0 \cdot m_{at} \cdot H \cdot \cos \theta \quad (2)$$

In the case of ferromagnetic materials, in modelling of anhysteretic magnetization of ferromagnetic, Boltzman distribution is also used. However, atomic magnetic moment m_{at} is changed by average magnetization of domain m_d given as [1]:

$$m_d = \frac{M_s}{N} \quad (3)$$

where N is average domain density in the material. Moreover, in equation determining energy of domain m_d , the effective magnetizing field H_e appears, given by the equation (4) [2]:

$$H_e = H + \alpha \cdot M \quad (4)$$

where α determines interdomain coupling.

Considering equations (3) and (4), the Boltzman distribution based equation for anhysteretic magnetization (1) (and its antiderivative) leads to the Langevin equation for anhysteretic magnetization M_{ah_iso} for isotropic materials [2]:

$$M_{ah_iso} = M_s \left[\coth\left(\frac{H_e}{a}\right) - \left(\frac{a}{H_e}\right) \right] \quad (5)$$

where a is given as:

$$a = \frac{N \cdot k_B \cdot T}{\mu_0 \cdot M_s} \quad (6)$$

It should be stressed, that simplification to equation (5) is possible only due to the fact, that functions in equation (1) have antiderivative. Moreover, equation (5) is valid only for isotropic materials.

According to Ramesh extension of Jiles-Atherton model for anisotropic, ferromagnetic materials, equation (1) should be converted to the following form, giving value of anhysteretic magnetization in anisotropic magnetic materials M_{ah_aniso} [17, 18]:

$$M_{ah_aniso} = M_s \left[\frac{\int_0^{\pi} e^{E(1)+E(2)} \sin \theta \cdot \cos \theta \cdot d\theta}{\int_0^{\pi} e^{E(1)+E(2)} \sin \theta \cdot d\theta} \right] \quad (7)$$

where

$$E(1) = \frac{H_e}{a} \cos \theta - \frac{K_{an}}{M_s \cdot \mu_0 \cdot a} \sin^2(\psi - \theta) \quad (8)$$

$$E(2) = \frac{H_e}{a} \cos \theta - \frac{K_{an}}{M_s \cdot \mu_0 \cdot a} \sin^2(\psi + \theta) \quad (9)$$

In such a case K_{an} is the average energy density connected with uniaxial anisotropy in magnetic material and ψ is the angle between direction of magnetizing field and the easy axis of magnetization due to the anisotropy. It should be stressed, that functions in equation (7) have no known antiderivatives. As a result, equation (7) can be only solved using numerical integration.

In Jiles-Atherton model, the hysteresis loop is determined by the irreversible magnetization M_{irr} , which is given by the following equation [2, 19, 20]:

$$\frac{dM_{irr}}{dH} = \delta_M \frac{M_{ah} - M_{irr}}{\delta \cdot k} \quad (10)$$

where the parameter k quantifies average energy required to break pining site. In this equation parameter $\delta = +1$ for $\frac{dH}{dt} \geq 0$ and $\delta = -1$ for $\frac{dH}{dt} < 0$. Additional parameter

$\delta_M = 0$ when $\frac{dH}{dt} < 0$ and $M_{an} - M > 0$ as well as when $\frac{dH}{dt} \geq 0$ and $M_{ah} - M < 0$. In other cases $\delta_M = 1$. Parameter δ_M guarantees that incremental susceptibility is always positive, what is physically judged.

In the Jiles-Atherton model, the reversible magnetization M_{rev} is given by the equation [2]:

$$M_{rev} = c \cdot (M_{an} - M_{irr}) \quad (11)$$

where c is parameter describing magnetization reversibility. Considering the fact, that total magnetization M in magnetic materials is given by the sum of reversible M_{rev} and irreversible M_{irr} magnetization, this total magnetization M may be calculated from following ordinary differential equation (ODE) [2, 19, 20]:

$$\frac{dM}{dH} = \frac{\delta_M}{1+c} \frac{M_{ah} - M}{\delta \cdot k - \alpha(M_{ah} - M)} + \frac{c}{1+c} \frac{dM_{ah}}{dH} \quad (12)$$

It should be highlighted, that calculation of the magnetic hysteresis loop by solving equation (12), requires the initial states conditions. These conditions are determined as state of demagnetized material, where $H = 0$ and $M = 0$.

Example of parameters of Jiles-Atherton model of $B(H)$ hysteresis loop for the magnetic materials with uniaxial anisotropy was given by Ramesh et al. [17]. However, previously this solution was presented only for major hysteresis loop (up to saturation) of material magnetized in the easy axis. Figure 1 presents the results of these calculations for both easy and hard axes as well as considering the minor hysteresis loops.

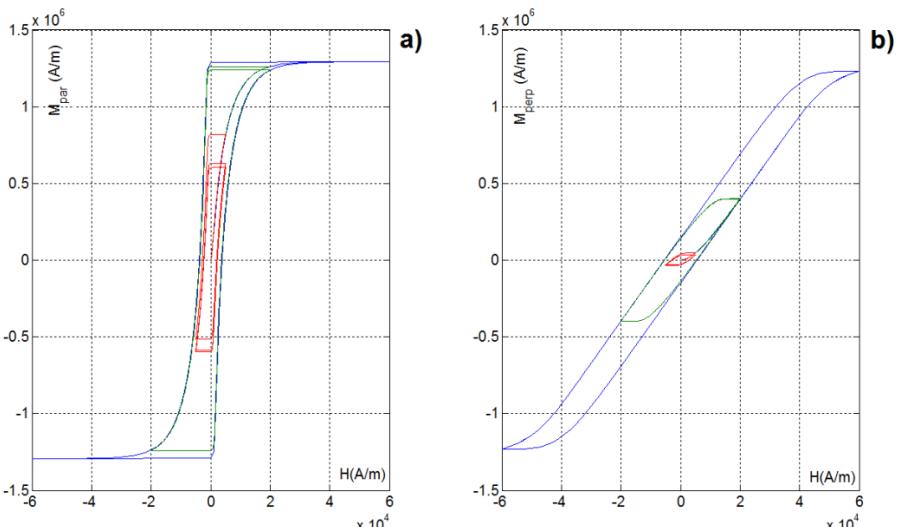


Fig. 1. The results of calculations of Jiles-Atherton model of hysteresis loops for easy [17] (a) and hard (b) axes of material with parameters proposed by proposed by Ramesh et al.: $M_s = 1.3 \cdot 10^6$ A/m, $a=1000$ A/m, $c=0.1$, $k=5000$ A/m, $\alpha=10^{-3}$, $K_{an}=4 \cdot 10^4$ J/m³

Major magnetic hysteresis loop of materials presented in figure 1a is in good agreement with the results given by Ramesh et al. [17]. However, for minor hysteresis loops, especially for lower values of amplitude of magnetizing field, significant values of accommodation were observed. [21] From qualitative point of view, such effect is physically judged [22]. On the other hand, quantitative values of magnetic accommodation in the series of hysteresis loops were not experimentally verified.

3 Computational Problems Connected with Jiles-Aherton Model

Analyses of equations (1)–(12) indicates, that there are two most important computational problems connected with calculation of the Jiles-Atherton model: numerical integration of the function for anhysteretic magnetisation in anisotropic material (given by the equations 7–9) as well as solving of ODE given by equation 12.

Analyse of the function given by equations 7–9 indicate, that this function is rather flat. Moreover, there is limited numbers of extreme, what judge trapezoidal approximation during the integration. On the other hand, number of points of approximation in $(0, \pi)$ integration limits was not specified in the literature. For this reason adaptive sampling with numerical error control should be used for this purpose. Due to oscillatory character of the integrated function, the most suitable method of such integration is Gauss-Kronrod approximation [23]. This method of numerical integration with adaptive sampling is implemented both in MATLAB and in OCTAVE as quadgk() function. It should be highlighted, that using numerical error control in quadgk() function, number of points for trapezoidal approximation with acceptable numerical error may be determined experimentally. Verification of the model's accuracy indicates, that negligible numerical errors may be achieved even for 90 points of trapezoidal approximation in the $(0, \pi)$ integration limits.

The most important source of numerical error during the calculation of Jiles-Atherton model of hysteresis loop is connected with solving of ODE given by equation (12). Typically fixed step solution with not specified method and step value is used [7, 24]. It should be highlighted, that for anisotropic materials, even for small value of integration step, significant numerical errors may occur if Riemann integration is applied. For this reason for solving equation (12) Runge-Kutta algorithm based solvers should be used. Runge-Kutta algorithm with numerical error control oriented on adaptive step is implemented in both MATLAB and OCTAVE in the group of functions diversified from the point of view of order of approximation, such as ode45(), ode23() or ode113(). However, application of adaptive step, Runge-Kutta algorithm based solvers is quite time consuming. Moreover, in some cases, these solvers may give unstable results, what lead to radical increase of calculation time.

For this reason application of the fixed-step, fourth order Runge-Kutta algorithm (rk4) is promising solution leading to increasing the speed of calculation of hysteresis loop with Jiles-Atherton model. However, number of steps in such algorithm have to be determined. This may be done experimentally, where numerical error controlled 2–3th order Runge-Kutta algorithm with adaptive steps may be used as a reference.

It should be highlighted, that practical experience leads to the conclusion, that in the case of Jiles-Atherton model, numerical error controlled 2-3th order Runge-Kutta algorithm (implemented as `ode23()` function) is more effective than 4-5th order Runge-Kutta algorithm. However, different implementations of this algorithm are made in MATLAB and OCTAVE, what may lead to incoherent results of calculation with use of these software.

4 Benchmarking and Verification of Numerical Accuracy

Benchmarking and verification of numerical accuracy was made on PC with Intel i5-650 @ 3.19 GHz 64 bits processor. Tests were carried out in Windows 7 system for both MATLAB2009b as well as for Octave 3.6.2/3.4.3 with `odepkg` (version 0.8.2 for Windows and 0.8.4 for Scientific Linux) for solving the ordinary differential equations.

Four methods for calculation of Jiles-Atherton model were implemented:

RK4, fixed 150 steps – Runge-Kutta 4-th order method with fixed step number 150 steps per each hysteresis loop, integration with 90 points of trapezoidal approximation,

RK4, fixed 450 steps – Runge-Kutta 4-th order method with fixed step number 450 steps per each hysteresis loop, integration with 90 points of trapezoidal approximation,

ODE45 – Runge-Kutta 4-th and 5-th order method with adaptive step size (10-3 relative accuracy), integration with adaptive Gauss-Kronrod adaptive approximation,

ODE23 – Runge-Kutta 2-th and 3-th order method with adaptive step size (10-4 relative accuracy), integration with adaptive Gauss-Kronrod adaptive approximation.

Calculation time for different operating systems, software and methods of solving the Jiles-Atherton equations is given in the table 1.

Table 1. Calculation time of 6 hysteresis loops (figure 1) for different operating systems, software and methods of solving the Jiles-Atherton equations

Operating system		Windows 7, 64-bit		Scientific Linux 6.3
Software		MATLAB2009b	Octave 3.6.2	Octave 3.4.3
ODE solver	Calculation time (s)			
	RK4, fixed 150 steps	7.5	39.7	31.5
	RK4, fixed 450 steps	26.9	120.5	98.5
	ODE45	11.1	78.1	60.7
	ODE23	16.3	76.4	57.7

During the assessment of relative, average numerical error, results of calculation by ODE23 method in MATLAB was treated as the reference. Results of this assessment are given in the table 2.

Table 2. Relative, average error of calculation of 6 hysteresis loops (figure 1) for different operating systems, software and methods of solving the Jiles-Atherton equations (MATLAB ODE23 method as reference)

Operating system		Windows 7, 64-bit		Scientific Linux 6.3
Software		MATLAB2009b	Octave 3.6.2	Octave 3.4.3
ODE solver	Relative, average calculation error MATLAB ODE23 method as reference			
	RK4, fixed 150 steps	0.0015	0.0007*	0.0006*
	RK4, fixed 450 steps	0.0003	0.0003	0.0003
	ODE45	0.0029	0.0004	0.0035
	ODE23	n.a.	0.0002	0.0027

* errors occur for the lowest value of amplitude of magnetizing field

The results indicate that numerical errors for Octave 3.6.2/3.4.3 odepakage vary significant for Windows 7 and Scientific Linux. This phenomenon is probably caused by differences in floating point operation error handling among these implementations of OCTAVE. Moreover, in both cases the results for Octave odepakage are significantly different than achieved for MATLAB2009b.

However, achieved results indicates, that the most reliable results of calculation of Jiles-Atherton model are achieved for fixed step, fourth order Runge-Kutta method with trapezoidal approximation for integration in calculation of anhysteretic magnetization. However, this method is the most time consuming.

5 Conclusion

Analyse presented in the paper indicates, that there are two most important computational problems connected with calculation of the Jiles-Atherton model: numerical integration of the function for anhysteretic magnetisation in anisotropic material as well as solving of ODE connected with modelling of the magnetic hysteresis.

Numerical errors connected with integration are on quite low level, even in the case of using trapezoidal approximation integration limited to 90 points. However, as solving of ODE connected with modelling of the magnetic hysteresis causes more serious problem. To limit possibility of instability during solving of ODE constituting the Jiles-Atherton model, fixed step, 4th order Runge-Kutta method may be used with 450 steps per hysteresis loop. However, to increase accuracy and sped of calculation, the 2-3th order Runge-Kutta method with adaptive sampling should be applied with relative calculation error limited to 10^{-4} .

Benchmarking confirms, that calculations with MATLAB2009 are about five times faster than with OCTAVE 3.6.2. However, Octave works faster in Scientific Linux environment. In such a case it is approximately four times slower than MATLAB2009b.

Experimental results clearly indicate, that packages for solving ordinary differential equations with OCTAVE require further development and validation to provide comparable results of calculation under different operation systems. Moreover, to provide possibility of verification of the results of calculation, the software source codes for calculation should be publicated on-line. Source code for calculations presented in this paper is available at <http://zisp.mchtr.pw.edu.pl/JASmodel/benchmark>.

References

1. Jiles, D.C., Atherton, D.: Theory of ferromagnetic hysteresis. *Journal of Applied Physics* 55, 2115 (1984)
2. Jiles, D.C., Atherton, D.: Theory of ferromagnetic hysteresis. *Journal of Magnetism and Magnetic Materials* 61, 48 (1986)
3. Venkataraman, R., Krishnaprasad, P.: Qualitative analysis of a bulk ferromagnetic hysteresis model. In: Proceedings of the 37th IEEE Conference on Decision and Control, p. 2443 (1998)
4. Szewczyk, R.: Modelling of the magnetic and magnetostrictive properties of high permeability Mn-Zn ferrites. *PRAMANA-Journal of Physics* 67, 1165–1171 (2006)
5. Chwastek, K., Szczyglowski, J.: Estimation methods for the Jiles-Atherton model parameters – a review. *Electrical Review (Przeglad Elektrotechniczny)* 84, 145 (2008)
6. Pop, N., Caltun, O.: Jiles-Atherton model used in the magnetization process study for the composite magnetoelectric materials based on cobalt ferrite and barium titanate. *Canadian Journal of Physics* 89, 787 (2011)
7. Baghel, A., Kulkarni, S.: Hysteresis modeling of the grain-oriented laminations with inclusion of crystalline and textured structure in a modified Jiles-Atherton model. *Journal of Applied Physics* 113, 043908 (2013)
8. Stoklosa, Z., Rasek, J., Kwapulinski, P.: Magnetic, electrical and plastic properties of Fe76Nb2Si13B9, Fe75Ag1Nb2Si13B9 and Fe75Cu1Nb2Si13B9 amorphous alloys. *Journal of Alloys and Compounds* 509, 9050 (2011)
9. Bienkowski, A., Szewczyk, R., Kolano, R.: Influence of thermal treatment on magnetoelastic Villari effect in Fe78Si13B9 amorphous alloy. *Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing* 375, 1024–1026 (2004)
10. Hamimid, M., Mimoune, S., Feliachi, M.: Evaluation of minor hysteresis loops using Langevin transforms in modified inverse Jiles-Atherton model. *Physica B-Condensed Matter* 429, 115 (2013)
11. Jackiewicz, D., Szewczyk, R., Salach, J.: Modelowanie charakterystyk magnesowania stali konstrukcyjnych. *Pomiary Automatyka Robotyka* 2, 552 (2012)
12. Andrei, P., Dimian, M.: Clockwise Jiles-Atherton Hysteresis Model. *IEEE Transactions on Magnetics* 49, 3183 (2013)
13. Pop, N.C., Caltun, O.F.: Using the Jiles Atherton model to analyze the magnetic properties of magnetoelectric materials (BaTiO₃)_x (CoFe₂O₄)_(1-x). *Indian Journal of Physics* 86, 283–289 (2012)
14. Gorecki, K., Detka, K.: Electrothermal model of choking-coils for the analysis of dc-dc converters. *Materials Science and Engineering B-Advanced Functional Solid-State Materials* 177, 1248 (2012)

15. Raghunathan, A., Klimczyk, P., Melikhov, Y.: Application of Jiles-Atherton Model to Stress Induced Magnetic Two-Phase Hysteresis. *IEEE Transactions on Magnetics* 49, 3187 (2013)
16. Jiles, D.C.: Introduction to Magnetism and Magnetic Materials. Chapman and Hall, London (1998)
17. Ramesh, A., Jiles, D.C., Bi, Y.: Generalization of hysteresis modeling to anisotropic materials. *Journal of Applied Physics* 81, 5585 (1997)
18. Ramesh, A., Jiles, D., Roderik, J.: A model of anisotropic anhysteretic magnetization. *IEEE Transactions on Magnetics* 32, 4234 (1996)
19. Chwastek, K., Szczygłowski, J.: Identification of a hysteresis model parameters with genetic algorithms. *Mathematics and Computers in Simulation* 71, 206 (2006)
20. Davidson, R., Charap, S.: Combined vector hysteresis models and applications. *IEEE Transactions on Magnetics* 32, 4198 (1996)
21. Tellini, B., Giannetti, R., Lizon-Martinez, S., Marracci, M.: Characterization of the Accommodation Effect in Soft Hysteretic Materials via Sensorless Measurement Technique. *IEEE Transactions on Instrumentation and Measurement* 58, 2807 (2009)
22. Della Torre, E.: A Preisach model for accommodation. *IEEE Transactions on Magnetics* 30, 2701 (1994)
23. Shampine, L.F.: Vectorized Adaptive Quadrature in MATLAB. *Journal of Computational and Applied Mathematics* 211, 131 (2008)
24. Calkins, F., Smith, R., Flatau, A.: Energy-based hysteresis model for magnetostrictive transducers. *IEEE Transactions on Magnetics* 36, 429 (2000)
25. Szewczyk, R.: Modelling of the magnetic and magnetostrictive properties of high permeability Mn-Zn ferrites. *J. of Physics* 67(6), 1165–1171 (2006)
26. Szewczyk, R., Bienkowski, A.: Magnetoelastic Villari effect in high-permeability Mn-Zn ferrites and modeling of this effect. In: Conference: 15th International Symposium on Soft Magnetic Materials (2001); *J. of Magnetism and Magnetic Materials* 254, SI 284– SI 286 (2003)

The Tester of the Actuator with ARINC 429 Data Bus

Ewelina Szpakowska-Peas

Institute of Aviation, Warsaw, Poland

Abstract. The article presents a device for testing avionics systems, that steering flight control surfaces on the aircraft. The tester uses ARINC 429 data bus to communicate with controller of the actuator. In this article ARINC 429 data bus standard was briefly described. The construction and operation of the tester were discussed also sample test executed by the device was presented. Propositions for further tester improvements were presented in this paper.

Keywords: tester, ARINC 429, data bus, flight control surfaces, controller of a actuator.

While working on the system controller of the actuator of flight control surfaces on the plane in the Institute of Aviation, there was a need to create an additional instrument to validate its proper operation. It was created a piece of equipment that is a cheaper alternative to the accredited instruments. The device is designed to carry out a preliminary tests of a controller, however the results should be considered as auxiliary data in the evaluation of proper operation of the tested system. The specifics are the use of ARINC 429 communications between devices. The commonness of this type of data bus in the aerospace industry gives the tester universal nature of the instrument. In case of need to examine the other type of controller, it would be possibly make some modifications to the device software, which may indicate a versatile applications of the tester for aviation equipment.

1 Introduction

A schematic diagram of the test controller is shown in figure 1. Operation of the device is based on a feedback loop. Incoming external measurements of the ARINC 429 data bus form the basis of the estimated angle α_e , which is subsequently calculated in the next block. Further, the feedback loop comes to the real angle α_{rz} executed by the mechanism and these data also have the form based on the standard ARINC 429. At junction the node error δ_α is calculated based on these two angles and is forwarded to the control block. Moreover, it controls the gear-motor to adjust the angle of the aircraft flight control surfaces. Thus both the measurement of the actual angle executed by the mechanism and the estimated angle calculation will be one of the main tasks of the tester. The

simulated parameters will be transmitted by the tester using the ARINC 429 bus to the controller. Furthermore the device will check if the measured angle is within the error. Sample results of this study were presented at the point.

The tester main task is to conduct selected tests chosen by the user. The results of each experiment are saved to the μ SD card for further analysis. Device will perform miscellaneous functions according to the selected tests. Among other things, measure the position of the actuator, transmit specific parameter values via the ARINC 429 data bus, will ask various external brightness of LED via built-in PWM module, also examine selected logic signals.

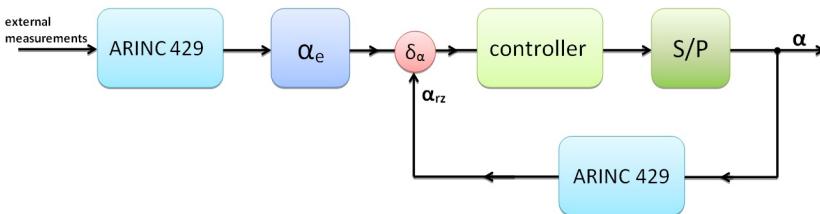


Fig. 1. The general block diagram of controller of actuator

2 Standard of the ARINC 429 Data Bus

ARINC 429 is the most common standard data bus, which is widely used in aircraft. It connects directly the on-board equipment, which transmit and receive informations. This standard defines the electrical characteristics, data and protocols equipped with on-board bus 429 Arica [1].

The physical form of ARINC is a shielded twisted-pair (symmetrical signal) and systems of input and output devices. Unidirectional data bus standard known as Mark 33 Digital Information Transfer System (DITS) is used. In practice, this means that the signal reception (Rx) and transmit (Tx) are located on separate ports. The message format is 32-bit words, which are transmitted at a frequency of 100 kHz or slower - 12.5 kHz. In the digital transmitted signals are extracted three states: high (+5 V), low (-5V) and the value of zero (NULL - 0 V). The transmitter always gives the 32-bit words or NULL state. The transmission of the words must be separated by at least 4 bits of NULL, which means that the two lines will be 0V. Figure 1 is an example of the data transmission. The data bus can not be tethered more than 20 receivers, but not less than one.

Typical data format used in the ARINC 429 is shown in table 1. It consists of five main segments: Parity, SSM, data, SDI and label [2].

- *Parity [P]*

Parity flag is always set to the most significant bit. This means that a 32-bit word must be an odd number of ones (including parity), beside to certain exceptions. This helps to find errors in data transmission.

Table 1. Generalized ARINC 429 word format

32	31	30	29		11	10	9	8	1
P	SSM			data		SDI		label	

– *SSM*

Matrix character / status contains different information depending on the type of encoding, such as condition of the equipment hardware, operating mode, and the validity of the data contents. Exemplary table 2 contains the BCD encoding used.

Table 2. Codes for BCD

Bit	Meaning	
31	30	
0	0	Plus, North, East, Right, To, Above
0	1	No Computed Data (NCD)
1	0	Functional Test (FT)
1	1	Minus, South, West, Left, From, Below

– *data*

Bits from 11 to 29 contain the encoded data, which strictly defines the label. They can take the form of BCD coding, or the representation of decimal digits or two's complement code BNR.

– *SDI*

A field consisting of 10 and 9 bit is the ID of sender and recipient. It aims to identify the receiver amid a variety of appliances, to which the message is addressed. It can also be used for finding the source of transmission among multiple systems.

– *label*

Fields bits from 1 to 8 contain information on the label that specifies the type of data in the message, and the parameters associated directly with them. The label is a very important part of the transmission. It is always sent before the data and it starts from the MSB, which means that the most significant bit is transmitted first and the least significant bit last. What is important is the fact that the same label value may comprise two different information for different systems according to the type of system installed on the airplane.

3 A Description of the Tester Construction

Tester is designed to be easy to use for the user. All elements such as buttons, connectors and the display is placed on the front panel so that the recipient had

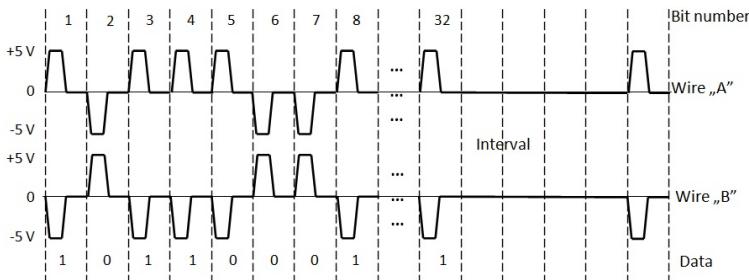


Fig. 2. Standard ARINC 429

easy access to them. Stickers were put in important places to make device more accessible to use. Also it's been thought of handles for the transport of device. They have been placed on opposite sides of the tester.

The tester front plate is comprised of four modules. The first one contains the power and signal connectors. On the second module two buttons are mounted, which control running of the motor built in the interior of the tester either directly or by the tested controller. It is an electric motor with gear and feedback potentiometer. It is used to control an aircraft flight control surfaces. Above the buttons is a red LED to indicate a failure of the actuator. Another module has two rows of three buttons with highlight. The assumed functions of the tester module performs the function of the buttons in the cockpit of the pilot. The last module contains a graphical display, 5 buttons and three LED and μ SD card slot. White buttons are used to select the menu items, and green is for their approval. The red button is used to restart the software. Orange LED announces that the tester is ready for operation and it does not take any test. Lastly, the red light notifies the duration of the test.

The graphic display used in the tester has a resolution of 128×64 pixels. This allows to display 8 lines of 21 characters. If the font is used 5×7 5x7 pixels it displays 168 characters. The display also has the ability to display some graphics. Applied display is based on the KS0108 LCD driver. In addition, on the panel, below the buttons, is fitted μ SD card slot that allows to gather information on the μ SD card while tester is performing various tests. Information stored on the μ SD are the FAT file system („File Allocation Table”). SD card support („Secure Digital Cards”) is made possible by the fatfs library. Description of menu displayed on the graphic LCD is shown in figure 5.

All tasks performed by the tester are controlled by two microcontrollers. Inside the chassis there are two processor modules. The first is a STM32F4 Discovery board. It controls front plate module, which contains the graphic display, and perform some tests. The second board includes SM32F2812 signal processor from Texas Instrument. Its presence is essential, because the circuit has been designed in such a way that it has an electronic circuit HI-3585, that is responsible for the

activation of data signals in accordance with standard ARINC 429 Both boards are communicating with each other via the serial COM by a standard RS232. In figure 4 is shown a generalized diagram of the tester circuits with signals. The following sections the various modules of processors and their functions in the tester will be discuss.



Fig. 3. The appearance of front panel of the Tester

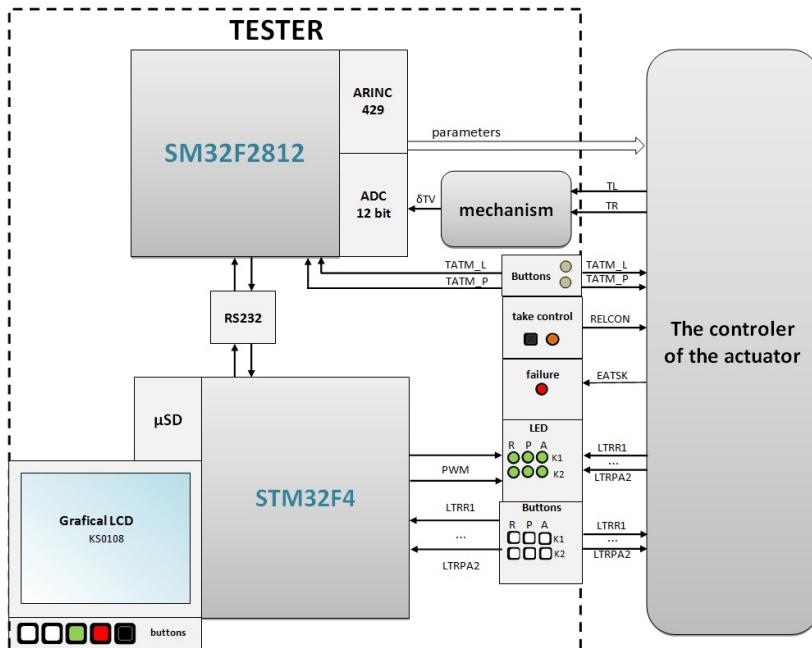


Fig. 4. Schematic diagram of the Tester

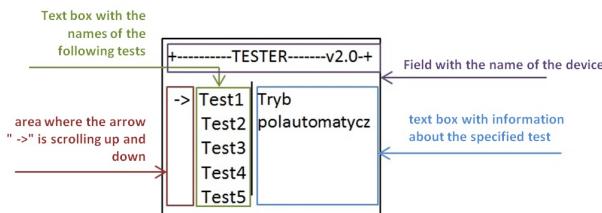


Fig. 5. Tester menu displayed on the GLCD during Tester operation

3.1 STM32F4 Discovery

Evaluation Kit based on STM32F407VGT microcontroller is equipped with the latest core Cortex-M4F of ARM. The main tasks of the module processor type Cortex-M4F to be undertaken are:

- support for the graphic display,
- support for buttons,
- support for LED,
- support for SD card reader,
- communication via RS232 data bus,
- carry out of selected tests.

3.2 SM32F2812

The main tasks to be undertaken by an electronic system with a processor SM32F2812 are:

- support for ARINC 429 module via SPI,
- voltage measurement using the 12-bit analog-to-digital converter,
- communication via the data bus RS232,
- carry out and evaluate tests.

4 Device Operation

This device tests proper operation of the actuator in different situations. User of the tester selects test he/she wants to perform from the menu that appears on the graphic display.

In the standby mode, when any test is being carried out, on the ARINC 429 bus is issued a set of nominal values of parameters, which are necessary for the proper operation of the actuator and which are specific to the aircraft. This task is performed by the system processor SM32F2812. Furthermore, in this mode, the menu is displayed on the LCD, as shown in Figure 3, below the screen should be light orange LED „READY” and highlighted in green buttons „MANUAL” in „CAB 1” and „CAB 2”. If these conditions are fulfilled, Tester is ready for testing.

4.1 Short Description of the Algorithms and Purpose of Their Action for STM32F4

Each electronic module mounted inside the tester fulfils the specific features that have been mentioned. The main algorithm, that is implemented on a processor STM32F4, is designed to support the graphical display and buttons. The program such is constructed that the first step after power is carried initialization of the processor, external interrupts, SPI and SD card support. Subsequently, the graphic display and peripherals USART module for communication with the processor SM32F2812 are initialized.

The main loop of the program - the main task of this section is to examine whether the data is transmitted from the second set of electronics and the μ SD card is inserted into the slot. In case of receiving data for a particular test, which is being performed, information is stored on the μ SD card to the file named „testn.txt”, where „n” is the number of ongoing test, eg for test No. 1, the file is named „test01.txt”. A schematic of this algorithm is shown in figure.

Buttons are read in the interrupts. In addition interrupt service is protected from the effect of contact bounce, because buttons incline to vibrate during the closing and / or open contacts. Bypass of this negative occurrence, which by nature is purely mechanical, is made by using the available counter in the system. The diagram of the algorithm used in the program is shown in figure 7.

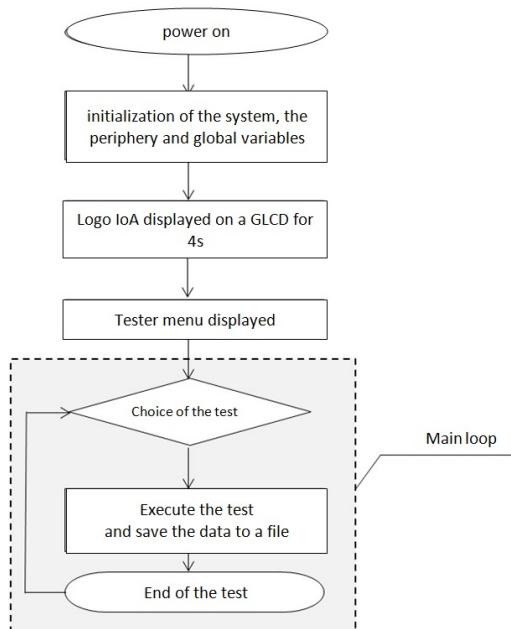


Fig. 6. The main loop of the program to the processor STM32F4

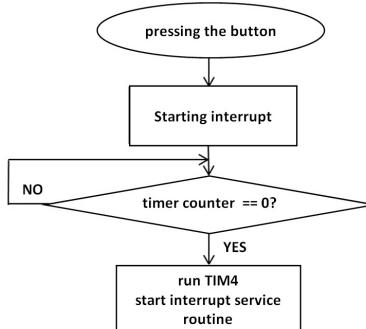


Fig. 7. Protecting algorithm against bouncing effect of buttons

4.2 Short Description of the Algorithms and Purpose of Their Action for SM32F2812

The second set of electronics used in the Tester has the main task to conduct the tests. In most studies, The ARINC 429 data bus is used to transmit message of the specific values of the parameters to the actuator. The first step after powering the device is to initialize the CPU, external interrupts, event handling, SPI and initialize ARINC 429 support and USART peripheral module for communication with the processor STM32F4.

The main loop of the program - both the examination of which test should be run and nominal parameter set transmission via ARINC 429 data bus are the main tasks of this part of the program.

During the interrupt, which occurs at the time of the incoming character via RS232, it is checked if a sequence of incoming characters is „% n%”, where n stands for a decimal number. Due to the received number via serial bus corresponding test is started. Moreover, at some stage in the test data can be transmitted to another system, which subsequently are stored on the μ SD card. However, for this to happen processor sends a character „^” in ASCII code, which indicates the second electronic set that transmission of data to write will begin. Sending a character „\$” ends the transmission.

The angular position of the actuator, which is determined through the measurement of the voltage value on the potentiometer of a gear-motor, is a part of the message transmitted via the ARINC 429. The measurement of voltage is made by using a 12-bit analog to digital converter located on the board with digital signal processor. For each measure will be called the same algorithm for voltage value determination.

To determine the current voltage value it will be sampled sequentially with 10 samples, which will be lowpass filtered. After collecting sufficient number of samples, sampling is being stopped until the next prompt.

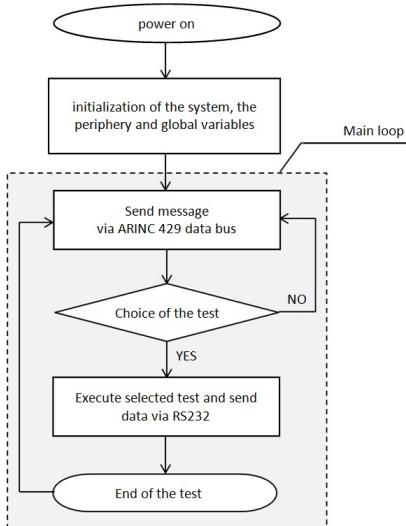


Fig. 8. The main loop of the program to the processor SM32F2812

5 Sample Tests and Their Results

All tests were conducted in laboratory conditions. The tester is connected to the power supply and to the actuator via the control signal connectors.

A series of tests was carried out in order to validate the operation of the controller of the actuator in different modes and its interaction with the aircraft cockpit. The interplay of both devices is good, and examinations conducted by the tester showed the correct operation of the controller.

The results of particular test are presented in this section as an example of the tester operation. The examination verifies the correctness of a actuator controller to received parameters via the ARINC 429 data bus. An angle with which the flight control surface is being rotated in the plane can be fixed by measuring the voltage value at the potentiometer mechanism. In one of the tests the effectiveness of the controller is evaluated on a basis of the value of the difference angle (real and estimated), which should be within the defined limits of error.

Both the real and actual and estimated value of the angle, that was worked out by the mechanism due to currently transmitted parameters are shown in figure 9. The chart highlights the difference of the two values δ_α , that according to the design criteria should not exceed 3° . The program also evaluate the test. During the test, measurement and calculation were carried out for 27 combinations of different sets of parameters transmitted via ARINC 429 data bus to the system controller. The test was successful. All parameters and results are up to date, during the test, recorded on the μ SD card for further revision and possible analysis.

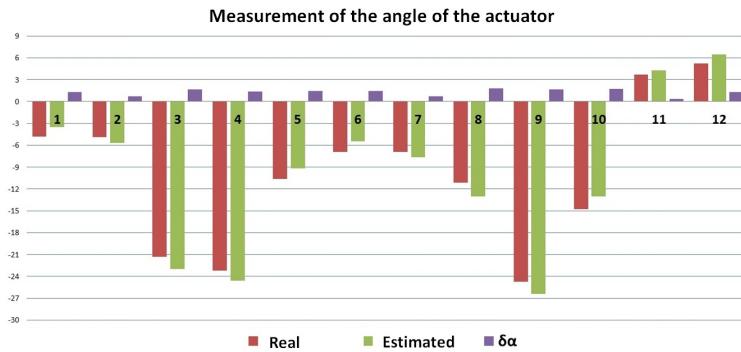


Fig. 9. Measuring the angle of the actuator

6 Summary

Tester is an easy-to-use and cost-effective alternative device to examine the mechanisms of implementing controllers of aircraft flight control surfaces with digital ARINC 429 bus. The presented device met the basic assumptions.

Further work on the device should be focused on convergence of electronics. It would be to design a board that would connect all the necessary modules of currently used processor. This modification would reduce the space currently used by the electronics and would reduce the complexity of connections between components. Additionally, due to standardization of the software, only one would suffice, and not two separate, each dedicated to a particular processor. Software would be unified and possibly making things easier to adapt software to the system controller.

Moreover, the flexibility of the software is worth to notice, hence Tester has the ability to change the program in order to adjust to the demands of the user.

References

1. http://en.wikipedia.org/wiki/ARINC_429
2. http://www.actel.com/ipdocs/CoreARINC429_DS.pdf

Relay Self-tuning of Industrial PID Temperature Controller with Set-Point Weighting

Leszek Trybus, Zbigniew Świder, and Andrzej Stec

Rzeszow University of Technology,
al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
{ltrybus, swiderzb, astec}@kia.prz.edu.pl

Abstract. Relay self-tuning of PID loops with on-off PWM control and no-overshoot plus short settling time requirements is considered. Temperature loops in electric heating and cooling are typical examples. No-overshoot can be provided by I-PD structure of PID controller and short settling time by set-point weighting. No user parameters are allowed. The approach is restricted to lag plants, whose dynamics is evaluated by the shift between relay control and plant output during ultimate oscillation. Gain and phase margins for calculation of PID [1-3], as well as set-point weight, are selected in terms of this shift. Asymmetry of relay control is compensated by modification of the oscillation period. The self-tuner has been implemented in low-cost temperature controllers.

Keywords: relay self-tuning, set-point weighting, PID control, temperature control.

1 Introduction

Automatic tuning of PID loops, another words self-tuning, is now common in PLCs, PALs, DCS systems and instrument controllers. General purpose self-tuners require some user input to set parameters of tuning experiment and requirements for closed-loop response. In case of self-tuners dedicated for particular plants, installations or machines, experiment parameters and response requirements are usually fixed, so user input may be limited to pressing a single “tune” button. Such solutions are found in low-cost instrument controllers and some PLC/PAC libraries for specific applications.

Single button self-tuner considered here is dedicated for lag plants with on-off PWM control (Pulse Width Modulation), and no-overshoot plus short settling time requirements. Lag plants with on-off control are typical for electric heating, drying and cooling, so in fact we deal with temperature controllers.

Recall that in case of no-overshoot requirement, I-PD structure of PID controller is preferred. Here the integral action operates on the error $E = Y_{SP} - Y$ and proportional action on the plant output Y (Y_{SP} denotes set-point). Unfortunately, I-PD responses are sluggish. They may be shortened, however, by a set-point weighting, where proportional action operates on modified error $bY_{SP} - Y$ with some weight $b < 1$ [3-4].

Here we show how b can be selected to extend frequency bandwidth of I-PD loop, i.e. to decrease settling time.

Relay self-tuning introduced by Åström and Hägglund [1–3] is natural choice for plants with on-off control. PID settings are calculated using amplitude and period of ultimate oscillation, with some gain and phase margins. In our case the margins should provide no-overshoot and short settling time. Moreover, despite restricting applications to lag plants, the margins must still be adjusted to dynamics of particular plant. Hence, besides amplitude and period, another parameter of ultimate oscillation is needed. Here we choose a shift between relay control and plant output as this third parameter, so gain and phase margins, as well as set-point weight b , are selected from appropriate nomograms in terms of the shift.

Asymmetry of switch-on and switch-off times during relay control violates original assumption on relay tuning. It is taken into account here by adjusting ultimate oscillation period taken for calculation of the settings according to degree of asymmetry [5–7].

Relay self-tuners involving set-point weighting and compensation of asymmetry have been implemented in temperature controllers manufactured by Lumel, PL.

2 Set-Point Weighting and Bandwidth

Control U generated by PID controller with set-point weighting and derivative action operating on the plant output Y is given by

$$U = k_p((bY_{SP} - Y) + \frac{1}{T_i s} E - \frac{T_d s}{T_s s / N + 1} Y) \quad (1)$$

This can be implemented in parallel form as in Fig. 1a. The SPW is a compromise between PI-D structure for $b=1$ and I-PD for $b=0$. The SPW, called also two-degrees of freedom, was first explored by Japanese researches [2] (see also [3–4]). Current research on the subject is concerned with dynamic set-point weight $b(t)$ [8–9].

By simple rearrangement of Fig. 1a we obtain the diagram of Fig. 1b, i.e. the I-PD loop preceded by the SPW block $bT_i + 1$. Bode plot of $bT_i + 1$ is shown in Fig. 2a. Suppose the I-PD loop has been tuned for no-overshoot and reasonable settling time requirements, so its Bode plot lies below frequency axis. Bandwidth BD of the loop is determined for the magnitude -3dB (Fig. 2a). Let the weight b in $bT_i + 1$ provide somewhat smaller magnitude than $+3\text{dB}$ at the frequency BD , i.e.

$$\frac{1}{bT_i} > BD \quad \text{or} \quad b < \frac{1}{T_i BD} \quad (2)$$

Then the combined plot of $(bT_i s + 1) \& I\text{-PD}$ looks as in Fig. 2a, so it has wider bandwidth BD_1 . This is reflected in shorter settling time of set-point response (Fig. 2b).

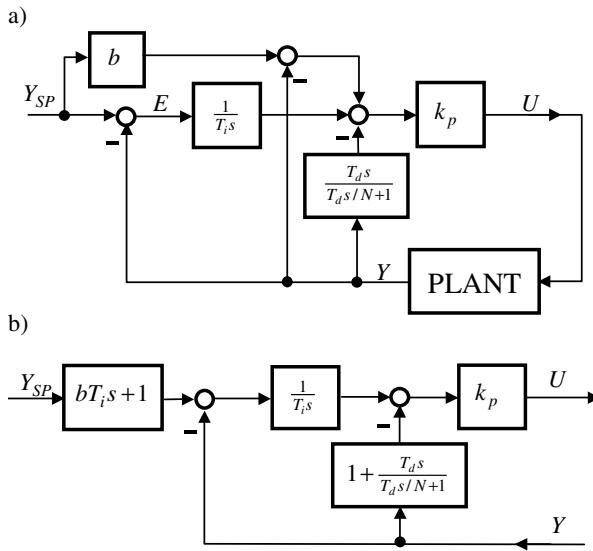


Fig. 1. PID control loop with set-point weighting: (a) implementation, (b) rearranged diagram

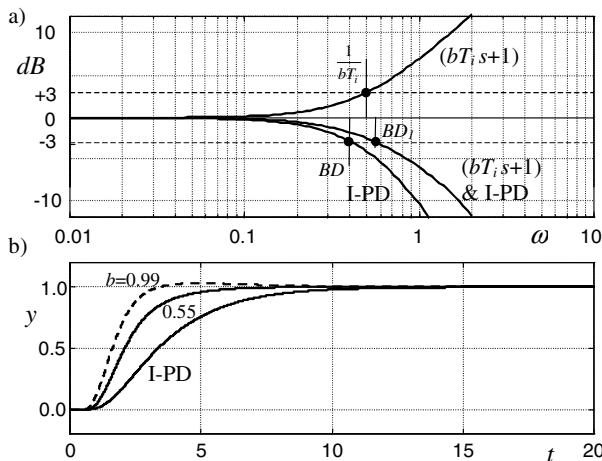


Fig. 2. Bode plots (a) and set-point responses (b) for I-PD loop with and without set-point weighting

Note that taking \$b\$ exactly at \$1/(T_i BD)\$ to cancel \$-3\text{dB}\$ of I-PD loop by \$+3\text{dB}\$ of \$b T_i + 1\$ is risky, because we cannot guarantee that the combined plot will not exceed frequency axis in some range below the bandwidth \$BD\$. This may result in some overshoot.

Example. The plots and responses presented above (and further) correspond to the sample plant

$$G(s) = \frac{1}{(s+1)^2} e^{-0.5s} \quad (3)$$

and PID settings $k_p = 1.66$, $T_i = 2.43$, $T_d = 0.61$ (Sec. 3). Since $BD = 0.41$ in Fig. 2a, so the condition (2) becomes $b < 0.99$. The nomogram of Sec. 4 provides $b = 0.55$ used in Fig. 2. The limit value $b = 0.99$ would give small overshoot of 2.5%.

3 Relay Self-tuning with Plant Depended Margins

3.1 PID Settings

Relay self-tuning introduced 30 years ago by Åström and Hägglund [1, 3] is now widely used in process automation. The approach is described in many texts, so here we quote only the expressions needed to calculate PID settings.

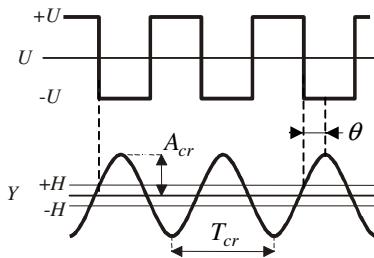


Fig. 3. Ultimate oscillation plots for relay tuning

Given the amplitude A_{cr} and period T_{cr} of ultimate oscillation generated by relay control with magnitude U and hysteresis H (Fig. 3), single value $G(j\omega_{cr})$ of the plant frequency characteristic at $\omega_{cr} = 2\pi/T_{cr}$ is calculated from

$$\begin{aligned} G(j\omega_{cr}) &= -R - jI = M e^{j\phi} \\ M &= \sqrt{R^2 + I^2}, \quad \phi = \pi - \arctan \frac{I}{R} \\ R &= \frac{\pi}{4U} \sqrt{A_{cr}^2 - H^2}, \quad I = \frac{\pi H}{4U} \end{aligned} \quad (4)$$

To determine the settings Åström and Hägglund applied the condition [2, 3]

$$G(j\omega_{cr})G_{PID}(j\omega_{cr}) = \frac{1}{GM} e^{-j(\pi - PM)} \quad (5)$$

with some gain GM and phase PM margins. By using (4), PID transfer function $k_p(1+1/T_i s + T_d s)$ and Ziegler-Nichols condition $T_d = T_i/4$ in (5), one obtains the settings

$$\begin{aligned} k_p &= \frac{GM}{M} \frac{\omega_{cr} T_i}{(\omega_{cr} T_i)^2 + 4}, \\ T_i &= \frac{2}{\omega_{cr}} \tan \frac{1}{2} (\phi - \frac{\pi}{2} + PM) \end{aligned} \quad (6)$$

Our problem is to select such margins GM , PM , so as to get set-point response with no-overshoot and short settling time. Recall that a "single button" controller is required, so the magnitude U and hysteresis H are fixed.

3.2 Shift between Relay and Plant

The on-off control admitted here assumes 0 and 100% values, so $U = 50\%$. The hysteresis H must cover measurement noise but cannot be too large for practical reasons. Hence H is small, so $H/U \ll 1$ ($H = 5\%$ in our application, i.e. $H/U = 0.1$). The imaginary part I in (4) becomes small and the angle ϕ is only slightly less than π .

Lag plants considered here are assumed to have S-shape reaction curve, so they may be reasonably well approximated by a model with double time constant and delay, i.e.

$$G(s) = \frac{k_o}{(Ts+1)^2} e^{-\tau s} \quad (7)$$

(see (3)). This is also fairly good approximation for plants with multiple time constants such as $1/(Ts+1)^n$. Models of this type represent electric furnaces, dryers, coolers, etc. We stress however that the self-tuner will leave some overshoot if the plant is described by single time constant with delay, i.e. $k_o e^{-\tau s}/(Ts+1)$.

To justify the model (7), in the Table 1 we present approximations of two benchmark plants used originally by Kraus and Myron [10] to demonstrate convergence of Foxboro EXACT adaptive control. Such plants are considered bracketing extremes for most lag processes encountered in practice. Plant A is "easy",

Table 1. Benchmark plants and approximations [5]

	Plant	Approximation
A	$\frac{1}{(20s+1)(0.5s+1)^4}$	$\frac{1}{(9.3s+1)^2} e^{-2.1s}$
B	$\frac{1}{(1.2s+1)^5} e^{-2s}$	$\frac{1}{(2.2s+1)^2} e^{-3.7s}$

plant B "difficult". Despite that the plant A is better approximated by the model $e^{-2.1s}/(19.5s+1)$ with single time constant, it is still handled by our self-tuner (no-overshoot).

Plant dynamics in (7) is characterized by relative delay τ/T which depends on physical construction of the plant, equipment and load. Naturally, the margins GM , PM in the settings (6) must depend on τ/T , or on some other equivalent measure. Therefore we need another parameter of ultimate oscillation, besides amplitude and period. Let us look in Fig.3 at the shift θ between relay control and plant output. If τ/T increases, the shift θ also increases. For the model (7), the ultimate oscillation is characterized by the phase equation $\tau\omega_{cr} + 2\arctan(T\omega_{cr}) = \phi$, whose left side involves τ/T_{cr} and T/T_{cr} . Since θ depends on τ/T , so the left side can be viewed as a function of θ/T_{cr} . Therefore we may use the relative shift θ/T_{cr} as a measure of plant dynamics.

3.3 Margins GM , PM vs. Shift θ/T_{cr}

Given θ/T_{cr} the margins GM , PM are selected from the nomograms in Fig. 4. They have been developed for the model (7) by executing two steps for each of a number of relative delays τ/T :

1. Simulation of relay control; readings of A_{cr} , T_{cr} and θ from ultimate oscillation (Fig. 3).
2. By numerous simulations of I-PD loop with the settings (6) obtained for different gain and phase margins, selection of such pair GM , PM which provides set-point response without overshoot and shortest settling time.

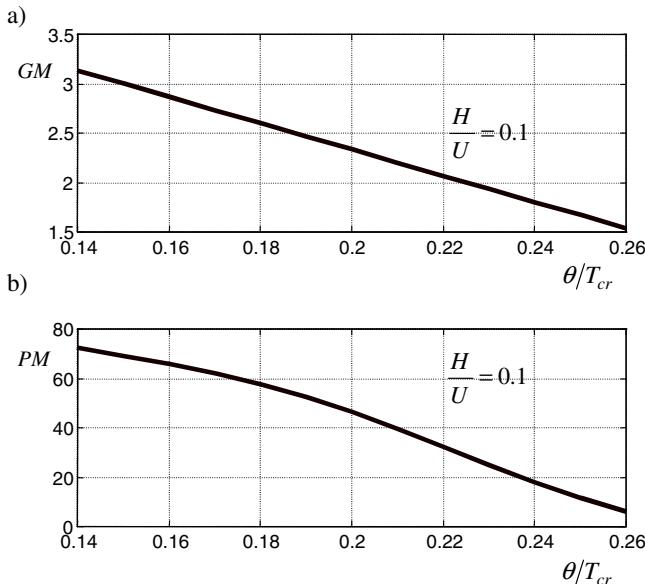


Fig. 4. Nomograms of design margins in terms of relative shift θ/T_{cr} : (a) GM , (b) PM

For specific τ/T , the values of θ/T_{cr} and GM , PM determine a single point on respective nomograms.

Example – ctd. Relay control of the plant (3) with $U = 50\%$ and $H/U = 0.1$ yields $A_{cr} = 0.191$, $T_{cr} = 3.96$, $\theta = 0.76$. So $\theta/T_{cr} = 0.192$, and $GM = 2.44$, $PM = 50.4^\circ$ from Fig. 4. Using this in (6) gives $k_p = 1.66$, $T_i = 2.43$, $T_d = 0.61$ (see Sec. 2).

As seen in Fig. 4, if the shift θ/T_{cr} increases (larger delay), the two margins gradually decrease. Recall that the first ECA 400 controller with relay self-tuning applied the fixed margins $GM = 2$, $PM = 45^\circ$ [2–3]. Whereas $PM = 45^\circ$ is in the middle of our second nomogram, $GM = 2$ is somewhat below. So here the gain margins are larger due to no-overshoot requirement, not imposed on ECA 400. Other recent relevant results are concerned with analytic expressions for PID settings [11] and stability regions for gain and phase margins [12].

4 Set-Point Weighting in the Self-tuner

4.1 Weight b vs. Shift θ/T_{cr}

Naturally, the weight b of the SPW bT_is+1 must also depend on the shift θ/T_{cr} . Corresponding nomogram of b in terms of θ/T_{cr} is shown in Fig. 5. The nomogram gives the largest b satisfying the condition (2), for which the (bT_is+1) &I-PD loop tuned as above still does not exhibit overshoot. However, in addition to the basic model (7), a few other typical lag plants with different parameters have been examined, such as $k_o/(Ts+1)^n$, $k_o/((T_1s+1)(T_2s+1))$, $k_o e^{-\tau s}/(Ts+1)^n$, etc., including the bracketing extremes A, B of Table 1. For given θ/T_{cr} the smallest weight b has been selected from a number of weights obtained for different plants, to be sure that the overshoot never appears. Hence the nomogram of Fig. 5 absorbs the difference between the model (7) and other lag plants.

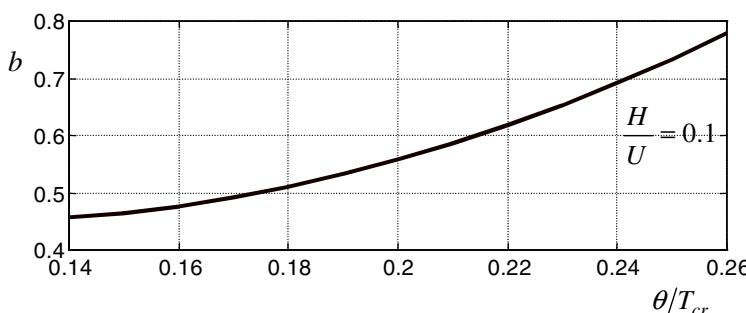


Fig. 5. Set-point b as a function of relative shift θ/T_{cr}

4.2 Equivalence of the SPW and a Set-Point Filter

Before turning attention to set-point weighting the authors examined application of set-point lag filter $(T_i s/2 + 1)$ for elimination of the overshoot. Recall that for the Ziegler-Nichols condition $T_d = T_i/4$, the PID transfer function becomes $k_p(T_i s/2 + 1)^2 / (T_i s)$, so one of the closed loop zeroes can be cancelled by such filter.

We discovered that the overshoot can be eliminated and at the same time the settling time kept short when the set-point filter affects integral action only, as shown in Fig. 6a [5–6]. However, the nomograms of corresponding self-tuner, although basically similar to those in Fig. 4, must be somewhat modified.

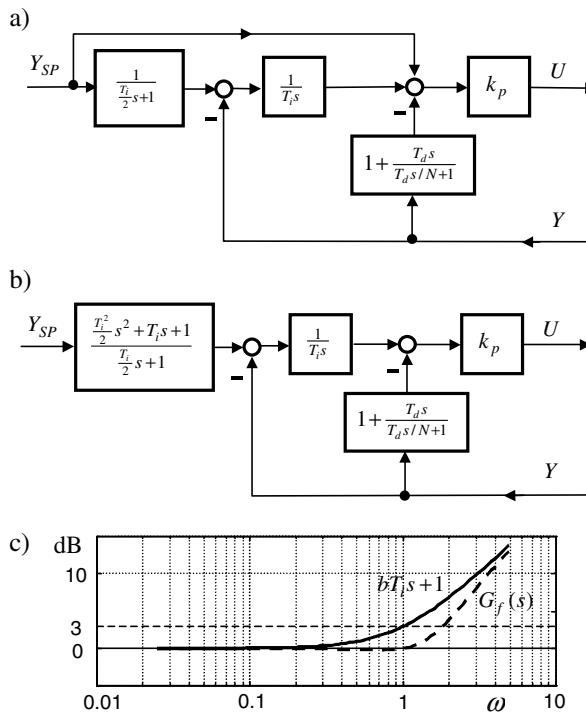


Fig. 6. (a) PID loop with set-point filter for integral action, (b) rearranged diagram, (c) Bode plots of $bT_i s + 1$ and $G_f(s)$ for sample data

The structure with the set-point filter for integral action shows interesting correspondence to the standard SPW $bT_i s + 1$. It can be discovered after rearrangement to Fig. 6b, where the transfer function

$$G_f(s) = \frac{\frac{T_i^2}{2}s^2 + T_i s + 1}{\frac{T_i}{2}s + 1} \quad (8)$$

proceeds the I-PD loop (instead of $bT_i s + 1$ in Fig. 1b). Bode plots of $bT_i s + 1$ for $bT_i = 1$ and $G_f(s)$ for $T_i = 1$ are shown in Fig. 6c. The two plots overlap for small and large frequencies, and differ only in the middle range. So by appropriate selection of breakpoints of the two plots we are able to extend bandwidth of the corresponding loops (compare Fig. 2a). Thus the two solutions are more or less equivalent as far as practical applications are concerned.

Note also that (8) can be generalized into

$$G_F(s) = \frac{T_1 s^2 + 2\zeta T_1 s + 1}{T_2 s + 1} \quad (9)$$

and incorporated into I-PD structure. By selecting appropriate $\zeta < 1$ and T_1, T_2 one could eliminate the overshoot even if the loop is tuned for oscillatory responses.

5 Implementation

5.1 Relay Asymmetry

The ultimate oscillation method assumes symmetry of relay control, i.e. the same switch-on and switch-off times. Here control can be either 0 or 100%, so if plant gain is unity, symmetry is provided only for the set-point $Y_{SP} = 50\%$. Any deviation from this value or non-unity plant gain causes asymmetry extending the oscillation period T_{cr} and deteriorating closed-loop responses.

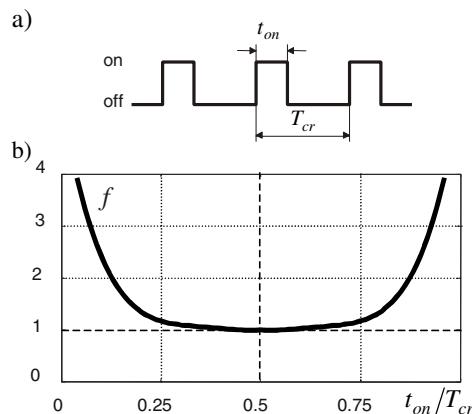


Fig. 7. Asymmetry of relay tuning: (a) control plot, (b) compensating divisor f of the period T_{cr} as a function of degree of asymmetry t_{on}/T_{cr}

The asymmetry can be dealt with by decreasing oscillation period used for calculation of the settings, as proposed by the authors in [5]. It involves the following steps:

1. Measurement of switch-on time t_{on} (Fig. 7a) and calculation of the ratio t_{on}/T_{cr} to express degree of asymmetry.
3. Determination of compensating divisor $f(t_{on}/T_{cr})$ from the nomogram of Fig. 7b.
4. Calculation of the corrected period

$$T_{cr}^* = \frac{T_{cr}}{f(t_{on}/T_{cr})} \quad (10)$$

used in the nomograms of Figs. 4 and 5 instead of the original T_{cr} .

The nomogram of Fig. 7b has been developed under assumption that the settings obtained at any set-point for the plant (7) with the gain k_o not necessarily unity will be the same as the ones obtained for $Y_{SP} = 50\%$ and $k_o = 1$, i.e. for fully symmetric relay control (where $f(t_{on}/T_{cr}) = 0.5 = 1$ and $T_{cr}^* = T_{cr}$). In other cases we have $f > 1$ and $T_{cr}^* < T_{cr}$, so extension of the period due to asymmetry is compensated. Naturally, the compensation becomes poorer for nonlinear plants.

5.2 Temperature Controllers

The relay self-tuner with set-point filter for integral action or set-point weighting and compensation of relay asymmetry have been implemented in low-cost temperature controllers manufactured by Lumel, PL. Two members of RE-series, simple RE-71 and more advanced RE-92 involving gain scheduling, are shown in Fig. 8. The devices are equipped with 8-bit AVR microcontrollers.

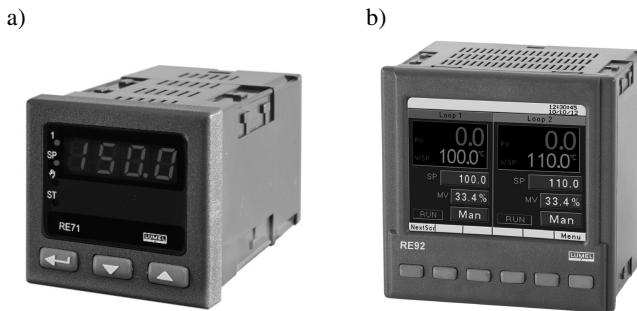


Fig. 8. Temperature controllers with relay self-tuning and set-point weighting: (a) RE-71, (b) RE-92 with gain-scheduling (courtesy of Lumel <www.lumel.com.pl>)

This algorithm has been also implemented in SMC controller as an function block written in ST (Structured Text) language (or native block written in C). This controller (manufactured also by Lumel) with test panel and small laboratory heater is shown in Fig. 9.



Fig. 9. The SMC controller with test panel and laboratory heater

6 Conclusions

Solutions to problems encountered while development of self-tuning for low-cost PID temperature controllers with PWM output have been presented in the paper. Low-cost implies minimum involvement of the staff in self-tuning, i.e. single button operation. It is expected that the controller will provide set-point responses without overshoot. Short settling time indicates reasonable self-tuning quality.

I-PD structure is natural choice for no-overshoot requirement. Sluggish responses can be made faster by application of set-point weighting which extends frequency bandwidth. The self-tuner of such SPW&I-PD loop must determine both PID settings and set-point weight.

Relay self-tuning is preferred for on-off PWM control. However, fairly precise selection of the settings and weight is possible only if more information on plant dynamics is available. Our self-tuner evaluates the dynamics as the shift between relay control and plant output. Design margins for calculation of the settings are selected from appropriate nomograms in terms of this shift. The nomograms cover lag plants approximated by the model with double time constant and delay. Set-point weight is determined from another nomogram. Settling time can be also made shorter by set-point filter affecting integral action. Asymmetry of switch-on/switch-off relay control is compensated by appropriate modification of oscillation period taken for selection of design margins.

Acknowledgments. The authors appreciate cooperation with R&D group of Lumel, Zielona Góra, PL.

References

1. Åström, K.J., Hägglund, T.: Automatic tuning of simple regulators with specifications on phase and gain margins. *Automatica* 20(5), 645–651 (1984)
2. Åström, K.J., Hägglund, T., Hang, C.C., Ho, W.K.: Automatic tuning and adaptation for PID controllers - a survey. *Control Engineering Practice* 1(4), 699–714 (1993)

3. Åström, K.J., Hägglund, T.: Advanced PID control. ISA, Research Triangle Park, NC, USA (2006)
4. Shigemasa, T., Iino, Y., Kanda, M.: Two degrees of freedom PID auto-tuning controller. In: Proceedings of ISA Annual Conference, pp. 703–711 (1987)
5. Korbicz, J., Kościelny, J.M. (eds.): Modeling, diagnostics and process control. Implementation in the DIaster system. Springer, Heidelberg (2010)
6. Stec, A., Świder, Z., Trybus, L.: Samostrojenie przekaźnikowe w systemie DIASTER. In: Systemy wykrywające, analizujące i tolerujące usterki (red. Z. Kowalcuk), Pomorskie Wydawnictwo Naukowo-Techniczne, pp. s.31–s.38 (2009)
7. Świder, Z., Trybus, L., Stec, A.: Automatyczne strojenie przekaźnikowe mikroregulatora temperatury. In: Przemysłowy Instytut Automatyki i Pomiarów, Warszawa, PAR, z.2, pp. s.588–s.597 (2010)
8. Mudji, R.K., Dey, C.: Performance improvement of PI controllers through dynamic set-point weighting. ISA Transactions 50, 220–230 (2011)
9. Mantz, K.J.: A PI controller with dynamic set-point weighting for nonlinear processes. In: IFAC Conference on Advances in PID Control, Brescia, March 28-30 (2012)
10. Kraus, T.W., Myron, T.J.: Self-tuning PID controller uses pattern recognition approach. Control Engineering 31, 106–111 (1984)
11. Hu, W., Xiao, G., Li, X.: An analytical method for PID controller tuning with specified gain and phase margins for integral plus time delay processes. ISA Transactions 50, 267–276 (2011)
12. Wang, D.J.: A PID controller set of guaranteeing stability and gain and phase margins for time-delay systems. Journal of Process Control 22, 1298–1306 (2012)

Pointwise Completeness and Pointwise Degeneracy of Linear Continuous-Time Systems with Different Fractional Orders

Wojciech Trzasko

Faculty of Electrical Engineering, Bialystok University of Technology, Bialystok, Poland
w.trzasko@pb.edu.pl

Abstract. In the paper the positive linear continuous-time systems with different fractional orders described by homogeneous state equations are considered. Definition and necessary and sufficient conditions for the pointwise completeness and the pointwise degeneracy are given. The considerations are illustrated by examples.

Keywords: linear system, fractional order, continuous-time, positive, pointwise completeness, degeneracy.

1 Introduction

A dynamical system without input signals (described by homogeneous equation) is called pointwise complete if every given final state of the system can be reached by a suitable choice of the initial state. The system, which is not pointwise complete, is called pointwise degenerated.

The problems of pointwise completeness and pointwise degeneracy of linear continuous-time systems with delays have been investigated in [3, 12, 16], of discrete-time and continuous-time systems of fractional orders – in [1, 7, 13], and of positive discrete-time systems in [2, 15].

The new class of positive linear systems with different fractional orders has been introduced in [4, 5]. In positive systems inputs, state variables and outputs take only non-negative values for non-negative initial conditions and non-negative controls. Positive linear systems are defined on cones and not on linear spaces. Therefore, theory of positive systems is more complicated and less advanced. An overview of state of the art in positive fractional systems is given in the monograph [8].

In this paper using recent results, given in [4–8] a problem of pointwise completeness and pointwise degeneracy of positive continuous-time systems with different fractional orders will be considered. The paper is organized as follows. In section 2 using the Caputo definition of the fractional order derivative the definition of the positive continuous-time systems with different fractional orders is introduced and basic system properties are given as well. For such a system the necessary and sufficient conditions for pointwise completeness and pointwise degeneracy are established in section 3. Some examples are given in sections 2 and 3.

2 Linear Continuous-Time Systems with Different Fractional Orders

Let $\Re^{n \times m}$ be the set of $n \times m$ real matrices and $\Re^n = \Re^{n \times 1}$. The set of $n \times m$ real matrices with nonnegative entries will be denoted by $\Re_+^{n \times m}$, and $\Re_+^n = \Re_+^{n \times 1}$. The set of nonnegative integers will be denoted by Z_+ , and $n \times n$ identity matrix by I_n .

The following Caputo definition of the fractional derivative will be used [8]

$$\frac{d^\alpha f(t)}{dt^\alpha} = \frac{1}{\Gamma(p-\alpha)} \int_0^\infty \frac{f^{(p)}(\tau)}{(t-\tau)^{\alpha+1-p}} d\tau, \quad (1)$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, $\operatorname{Re}(x) > 0$, is the Euler gamma function and $p-1 < \alpha < p$,

$$p \in N = \{1, 2, \dots\} \text{ and } f^{(n)}(\tau) = \frac{d^n f(\tau)}{d\tau^n}.$$

Consider the linear continuous-time system with different fractional orders described by the homogeneous equation [4]

$$\begin{bmatrix} \frac{d^{\alpha_1} x_1(t)}{dt^{\alpha_1}} \\ \vdots \\ \frac{d^{\alpha_n} x_n(t)}{dt^{\alpha_n}} \end{bmatrix} = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \dots & \vdots \\ A_{n1} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad 0 < \alpha_k < 1, \quad k = 1, \dots, n \quad (2)$$

where $x_k(t) \in \Re^{n_k}$, $k = 1, \dots, n$ are components of the state vector $x(t) \in \Re^N$, $N = n_1 + \dots + n_n$, $A_{kj} \in \Re^{n_k \times n_j}$, $k, j = 1, \dots, n$.

Initial conditions for (2) have the form

$$x_k(0) = x_{k0} \in \Re^{n_k}, \quad k, j = 1, \dots, n. \quad (3)$$

Theorem 1. The solution of the equation (2) for $0 < \alpha_k < 1$, $k = 1, \dots, n$ with initial conditions (3) has the form [4]

$$x(t) = \Phi_0(t)x_0 \quad (4)$$

where

$$x(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \Re^N, \quad N = n_1 + \dots + n_n, \quad x_0 = \begin{bmatrix} x_{10} \\ \vdots \\ x_{n0} \end{bmatrix}, \quad (5)$$

$$\Phi_0(t) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} T_{k_1 \dots k_n} \frac{t^{k_1 \alpha_1 + \dots + k_n \alpha_n}}{\Gamma(k_1 \alpha_1 + \dots + k_n \alpha_n + 1)} \quad (6)$$

$$T_{k_1 \dots k_n} = \begin{cases} I_N & \text{for } k_1 = \dots = k_n = 0 \\ \begin{bmatrix} A_{11} & \dots & A_{1n} \\ 0 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{bmatrix} & \text{for } k_1 = 1, k_2 = \dots = k_n = 0 \\ \vdots & \\ \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \\ A_{n1} & \dots & A_{nn} \end{bmatrix} & \text{for } k_1 = \dots = k_{n-1} = 0, k_n = 1 \\ T_{10 \dots 0} T_{01 \dots 1} + \dots + T_{0 \dots 01} T_{1 \dots 10} & \text{for } k_1 = \dots = k_n = 1 \\ \vdots & \\ T_{10 \dots 0} T_{k_1-1, k_2, \dots, k_n} + \dots + T_{0 \dots 01} T_{k_1, \dots, k_{n-1}, k_n-1} & \text{for } k_1 + \dots + k_n > 0 \\ 0 & \text{for at least one } k_i < 0, i = 1, \dots, n \end{cases} \quad (7)$$

Proof. The proof using the Laplace transform is similar as is given in [4] in the general case, i.e. $p_k - 1 < \alpha_k < p_k$, $p_k \in \{1, 2, \dots\}$ and non-zero input vectors $u \in \Re^m$.

Remark 1. Note that if $\alpha_1 = \alpha_2 = \dots = \alpha_n = \alpha$ then we have fractional order continuous-time systems described by equation [8]

$$\frac{d^\alpha x(t)}{dt^\alpha} = Ax(t). \quad (8)$$

The solution of equation (8) is given by (4), where

$$\Phi_0(t)|_{\alpha_1 = \dots = \alpha_n = \alpha} = E_\alpha(At^\alpha) = \sum_{k=0}^{\infty} \frac{A^k t^{k\alpha}}{\Gamma(k\alpha + 1)} \quad (9)$$

is the fundamental matrix and $E_\alpha(At^\alpha)$ is the Mittag-Leffler matrix function.

From comparison of (6) and (9) and using (7) it is easy to show that

$$\left. \sum_{k_1=0}^k \cdots \sum_{k_n=0}^k T_{k_1 \dots k_n} \frac{t^{k_1 \alpha_1 + \dots + k_n \alpha_n}}{\Gamma(k_1 \alpha_1 + \dots + k_n \alpha_n + 1)} \right|_{\substack{\alpha_1 = \dots = \alpha_n = \alpha \\ k_1 + \dots + k_n = k}} = \frac{A^k t^{k\alpha}}{\Gamma(k\alpha + 1)}. \quad (10)$$

Definition 1. The fractional system (2) is called positive if $x_k(t) \in \Re_+^{n_k}$, $k = 1, \dots, n$, $t \geq 0$ for any nonnegative initial conditions $x_{k0} \in \Re_+^{n_k}$, $k = 1, \dots, n$.

Let M_n be the set of $n \times n$ Metzler matrices, i.e. real matrices with nonnegative off-diagonal entries.

In the paper [4] the following theorem has been proved.

Theorem 2. The fractional system (2) for $0 < \alpha_k < 1$, $k = 1, \dots, n$ is positive if and only if

$$A = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \dots & \vdots \\ A_{n1} & \dots & A_{nn} \end{bmatrix} \in M_N. \quad (11)$$

Proof. Necessity. From the expansions (7) we have

$$\Phi_0(t) = \begin{bmatrix} I_{n_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & I_{n_n} \end{bmatrix} + \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \frac{t^{\alpha_1}}{\Gamma(\alpha_1 + 1)} + \dots + \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{bmatrix} \frac{t^{\alpha_n}}{\Gamma(\alpha_n + 1)} + \dots \quad (12)$$

From above it follows that $\Phi_0(t) \in \Re_+^{N \times N}$ for small value of $t > 0$ only if the condition (11) is satisfied.

Sufficiency. In the monograph [8] it has been proved that

$$\frac{A^k t^{k\alpha}}{\Gamma(k\alpha + 1)} \geq e^{At} \geq 0 \quad (13)$$

for $t \geq 0$ and $0 < \alpha_k < 1$, since $e^{At} \in \Re_+^{N \times N}$ if and only if A is a Metzler matrix.

Thus, from above and using (10), it can be shown that if (11) holds then

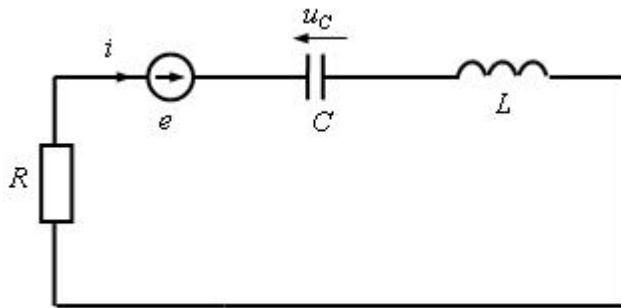
$$\Phi_0(t) \in \Re_+^{N \times N} \quad t \geq 0. \quad (14)$$

In this case from (4) we have $x(t) \in \Re_+^N$, $t \geq 0$ since by definition $x_0 \in \Re_+^N$. ■

2.1 Example 1

Consider the fractional linear circuit, shown in Fig. 1, with given resistance R , capacitance of supercapacitor C , inductance L and source of voltage e with the state-feedback

$$e = Kx. \quad (15)$$

**Fig. 1.** Electrical circuit

Using the second Kirchhoff's law we obtain for the circuit the homogeneous equation [5]

$$\begin{bmatrix} \frac{d^{\alpha_1}x_1}{dt^{\alpha_1}} \\ \frac{d^{\alpha_2}x_2}{dt^{\alpha_2}} \end{bmatrix} = Ax = \begin{bmatrix} 0 & \frac{1}{C} \\ \frac{a}{L} & \frac{b-R}{L} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad 0 < \alpha_1 < 1; \quad 0 < \alpha_2 < 1 \quad (16)$$

where $x_1 = u_c$, $x_2 = i$ and

$$A = A_0 + BK, \quad A_0 = \begin{bmatrix} 0 & \frac{1}{C} \\ -\frac{1}{L} & -\frac{R}{L} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \frac{1}{L} \end{bmatrix}, \quad K = [k_1 \quad k_2] = [a+1 \quad b]. \quad (17)$$

From (17) it follows that A_0 is not a Metzler matrix. It is easy to check that there exists a gain matrix $K \in \mathfrak{R}_+^2$ such that the closed-loop system matrix A is a Metzler matrix if and only if $a, b \geq 0$. The linear electrical circuits with the state-feedback (15) are positive systems with different fractional orders. ■

3 Main Results

Taking into account papers [7-8] we may formulate the following definitions of pointwise completeness and pointwise degeneracy of positive continuous-time systems with different fractional orders.

Definition 2. The positive fractional system (2) is called pointwise complete at $t = t_f$ if for every final state $x_f \in \mathfrak{R}_+^N$ there exists an initial state $x_0 \in \mathfrak{R}_+^N$ such that $x(t_f) = x_f$.

Definition 3. The positive fractional system (2) is called pointwise degenerated if it is not pointwise complete, that is there exists at least one final state $x_f \in \Re_+^N$ which can not be reached from any initial state $x_0 \in \Re_+^N$, i.e. there does not exist $t = t_f$ $x_0 \in \Re_+^N$, such that $x(t_f) = x_f$.

Theorem 2. The positive fractional system (2) for $0 < \alpha_k < 1$, $k = 1, \dots, n$ is pointwise complete at $t = t_f$ if and only if

1. all block diagonal matrices $A_{kk} \in \Re^{n_k \times n_k}$, $k = 1, \dots, n$ are diagonal and
2. all off-diagonal blocks are zero matrices, i.e. $A_{kj} = 0$, $k \neq j$.

Proof. In positive systems, according to Definition 1, $x_0 \in \Re_+^N$ and $x_f \in \Re_+^N$. From (4) it follows that for any final state $x_f \in \Re_+^N$ there exists positive initial state

$$x_0 = \Phi_0^{-1}(t_f)x_f \in \Re_+^N \quad (18)$$

if and only if $\Phi_0^{-1}(t_f) \in \Re_+^{NxN}$. It is well-known [9] that $\Phi_0^{-1}(t_f) \in \Re_+^{NxN}$ if and only if $\Phi_0(t_f)$ is monomial matrix (in each row and in each column only one entry is positive and remaining entries are zero). From expansion (12) it follows that $\Phi_0(t_f)$ is monomial matrix if and only if all matrices $A_{kk} \in \Re^{n_k \times n_k}$, $k = 1, \dots, n$ are diagonal matrices and the off-diagonal blocks are zero matrices, i.e. $A_{kj} = 0$, $k \neq j$. ■

Remark 2. If the conditions of Theorem 2 are satisfied, the fundamental matrix (6) can be written in the form

$$\Phi_0(t) = \begin{bmatrix} \Phi_1(t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Phi_n(t) \end{bmatrix}, \quad (19)$$

where

$$\Phi_k(t) = I_{n_k} + \sum_{j=1}^{\infty} A_{kk}^j \frac{t^{j\alpha_k}}{\Gamma(j\alpha_k + 1)}, \quad k = 1, \dots, n. \quad (20)$$

In this case, from comparison of (6) and (20) and using (7), it is easy to show that

$$T_{k_1 \dots k_n} = \begin{bmatrix} A_{11}^{k_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}, \dots, T_{00 \dots k_n} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{nn}^{k_n} \end{bmatrix} \quad (21)$$

and other matrices $T_{k_1 \dots k_n} = 0$, $k_1 + \dots + k_n > 0$.

Note that the suitable initial state (18) can be computed from the formula

$$x_0 = \Phi_0^{-1}(t_f)x_f = \begin{bmatrix} \Phi_1^{-1}(t_f) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Phi_n^{-1}(t_f) \end{bmatrix} x_f \in \Re_+^N. \quad (22)$$

The matrix $\Phi_k(t)$, $k = 1, \dots, n$ can be computed by using the Sylvester formula. In the case of distinct eigenvalues of matrix $A_{kk} \in \Re^{n_k \times n_k}$ we have the following lemma [7].

Lemma 1. If $A_{kk} \in \Re^{n_k \times n_k}$ has only distinct eigenvalues then the matrix (20) can be computed from the formula

$$\Phi_k(A_{kk}, t) = \sum_{i=1}^{n_k} \prod_{\substack{j=1 \\ j \neq i}}^{n_k} \frac{A_{kk} - I_{n_k} \lambda_j}{\lambda_i - \lambda_j} \Phi_k(\lambda_j, t). \quad (23)$$

From Definition 3 and Theorem 2 we have following theorem.

Theorem 3. The positive fractional system (2) for $0 < \alpha_k < 1$, $k = 1, \dots, n$ is pointwise degenerated if and only if one of the block diagonal matrices $A_{kk} \in \Re^{n_k \times n_k}$, $k = 1, \dots, n$ is not a diagonal matrix or one of the off-diagonal blocks is non-zero matrix.

3.1 Example 2 (Continuation of Example 1)

Let $K = [k_1 \ k_2] = [a+1 \ b] = [1 \ 0]$. In this case the matrix A (16) is not a diagonal matrix and conditions of Theorem 2 are not satisfied. By Theorem 3 the positive fractional electrical circuit with the state-feedback is pointwise degenerated for any $a, b \geq 0$. It is easy to see that the fundamental matrix $\Phi_0(t)$ is not a diagonal matrix. From the expansions (7) we have

$$\Phi_0(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{C} \\ 0 & 0 \end{bmatrix} \frac{t^{\alpha_1}}{\Gamma(\alpha_1 + 1)} + \begin{bmatrix} 0 & 0 \\ 0 & -R \\ 0 & L \end{bmatrix} \frac{t^{\alpha_2}}{\Gamma(\alpha_2 + 1)} + \begin{bmatrix} 0 & -R \\ 0 & LC \\ 0 & 0 \end{bmatrix} \frac{t^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2 + 1)} + \dots \quad (24)$$

3.2 Example 3

Consider the fractional linear circuit shown in Fig. 2 with given resistances R_1 , R_2 , R_3 , capacitance of supercapacitors C_1 , C_2 and voltage source e . The voltages on the capacitors $u_1 = u_1(t)$ and $u_2 = u_2(t)$ are the state variables $x_1 = u_1$ and $x_2 = u_2$ and the voltage source $u = e$ is input. Using the Kirchhoff's laws we obtain for the circuit the fractional state equation [8]

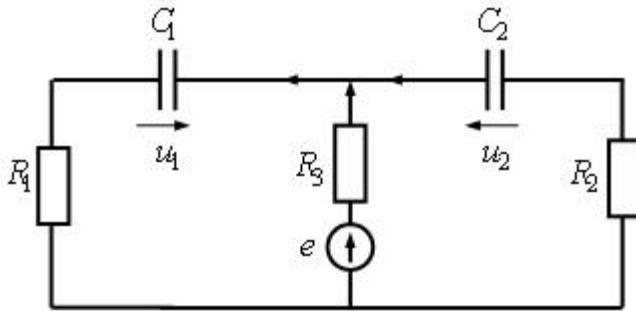


Fig. 2. Electrical circuit

$$\begin{bmatrix} \frac{d^{\alpha_1} x_1}{dt^{\alpha_1}} \\ \frac{d^{\alpha_2} x_2}{dt^{\alpha_2}} \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + Be, \quad 0 < \alpha_1 < 1; \quad 0 < \alpha_2 < 1 \quad (25)$$

where

$$A = \begin{bmatrix} -\frac{R_2 + R_3}{C_1[R_1(R_2 + R_3) + R_2R_3]} & \frac{R_3}{C_1[R_1(R_2 + R_3) + R_2R_3]} \\ \frac{R_3}{C_2[R_1(R_2 + R_3) + R_2R_3]} & -\frac{R_1 + R_3}{C_2[R_1(R_2 + R_3) + R_2R_3]} \end{bmatrix}, \quad (26)$$

$$B = \begin{bmatrix} \frac{R_2}{C_1[R_1(R_2 + R_3) + R_2R_3]} \\ \frac{R_1}{C_2[R_1(R_2 + R_3) + R_2R_3]} \end{bmatrix}. \quad (27)$$

The matrix \$A\$ is a Metzler matrix and the matrix \$B\$ has nonnegative entries. The electrical circuit is an example of positive fractional continuous-time linear system [8].

By Theorem 2 the electrical circuit is pointwise complete for \$t = t_f\$ if and only if its \$A\$ (26) matrix is diagonal, which implies \$R_3 = 0\$.

By Theorem 3 the electrical circuit is pointwise degenerated for every \$t = t_f\$ if and only if its \$A\$ (26) matrix is not a diagonal, which implies \$R_3 > 0\$.

Consider pointwise completeness of the positive system (25) without input, i.e. \$u = e = 0\$ and \$R_3 = 0\$. In this case the matrix (26) has form

$$A = \begin{bmatrix} -\frac{1}{C_1 R_1} & 0 \\ 0 & -\frac{1}{C_2 R_2} \end{bmatrix}. \quad (28)$$

Note that the electrical circuit (28) is asymptotically stable [8].

Using relations (22) and (20) we may find $x_0 \in \Re^2_+$ for any given $x_f \in \Re^2_+$. If $x_f = [x_{f1} \ x_{f2}]^T \in \Re^2_+$ then from (19)–(22) we have

$$x_0 = \begin{bmatrix} 1/\varphi_1(t_f) & 0 \\ 0 & 1/\varphi_2(t_f) \end{bmatrix} \begin{bmatrix} x_{f1} \\ x_{f2} \end{bmatrix} \in \Re^2_+, \quad (29)$$

where

$$\varphi_1(t_f) = \sum_{k=0}^{\infty} \frac{(-t_f^{\alpha_1})^k}{C_1 R_1 \Gamma(j\alpha_1 + 1)} \in \Re_+, \quad \varphi_2(t_f) = \sum_{k=0}^{\infty} \frac{(-t_f^{\alpha_2})^k}{C_2 R_2 \Gamma(j\alpha_2 + 1)} \in \Re_+. \quad (30)$$

Expressions (30) for a fixed values can be computed by using the MATLAB function MLF which calculates the Mittag-Leffler function with desired accuracy (2001–2012 Igor Podlubny, Martin Kacanak). ■

4 Concluding Remarks

The paper considers the problem of pointwise completeness and pointwise degeneracy of positive linear continuous-time systems with different fractional orders, described by the homogeneous equation (2), where order of the Caputo derivative α_k satisfies the following condition $0 < \alpha_k \leq 1$, $k = 1, \dots, n$, $n \leq N$.

Definitions as well as necessary and sufficient conditions of the pointwise completeness (Theorem 2) and pointwise degeneracy (Theorem 3) of the positive systems have been given. In case of pointwise completeness, a simple formula (22) for computing a nonnegative initial state $x_0 \in \Re^N_+$ for any given final state $x_f \in \Re^N_+$ has also been given. The considerations have been illustrated by examples of linear electrical circuits.

The considerations can be further extended to cover the positive descriptor (singular) fractional continuous-time linear systems.

Acknowledgment. This work was supported by National Science Centre of Poland under work No. G/WE/1/2011.

References

1. Busłowicz, M.: Punktowa zupełność i punktowa degeneracja liniowych układów dyskretnych ułamkowego rzędu. *Zesz. Nauk. Politechniki Śląskiej, ser. Automatyka* 151, 17–29 (2008)
2. Busłowicz, M., Kociszewski, R., Trzasko, W.: Punktowa degeneracja i punktowa zupełność liniowych dodatnich układów dyskretnych z opóźnieniami. *Zesz. Nauk. Politechniki Śląskiej, ser. Automatyka* 145, 51–56 (2006)
3. Choudhury, A.K.: Necessary and sufficient conditions of pointwise completeness of linear time-invariant delay-differential systems. *Int. J. Control* 16(6), 1083–1100 (1972)
4. Kaczorek, T.: Positive linear systems consisting of n subsystems with different fractional orders. *IEEE Trans. Circ. Sys. I* 58(6), 1203–1210 (2011)
5. Kaczorek, T.: Positive linear systems with different fractional orders. *Bull. Pol. Ac.: Tech.* 58(3), 453–458 (2010)
6. Kaczorek, T.: Pointwise completeness and pointwise degeneracy of standard and positive hybrid systems described by the general model. *Archives of Control Sciences* 20(2), 121–131 (2010)
7. Kaczorek, T., Busłowicz, M.: Pointwise completeness and pointwise degeneracy of linear continuous-time fractional order systems. *JAMRIS* 3(1), 8–11 (2009)
8. Kaczorek, T.: Wybrane zagadnienia teorii układów niecałkowitego rzędu, Oficyna Wydawnicza Politechniki Białostockiej, Białystok (2009)
9. Kaczorek, T.: Positive 1D and 2D Systems. Springer, London (2002)
10. Miller, K.S., Ross, B.: An Introduction to the fractional calculus and fractional differential equations. Wiley, New York (1993)
11. Podlubny, I.: Matrix approach to discrete fractional calculus. *An International Journal for Theory and Applications* 3(4), 359–386 (2000)
12. Popov, V.M.: Pointwise degeneracy of linear time-invariant delay-differential equations. *J. Diff. Equation* 11, 541–561 (1972)
13. Trzasko, W.: Punktowa zupełność i punktowa degeneracja układów dyskretnych niecałkowitego rzędu. In: *Automation 2012, Pomiary Automatyka Robotyka*, vol. 2, pp. 332–337 (2012)
14. Trzasko, W.: Względna punktowa zupełność dodatnich układów ciągłe-dyskretnych niecałkowitego rzędu. In: *Automation 2011, Pomiary Automatyka Robotyka*, vol. 2, pp. 528–537 (2011)
15. Trzasko, W., Busłowicz, M., Kaczorek, T.: Pointwise completeness of discrete-time cone systems with delays. In: *Proc. of EUROCON 2007*, pp. 606–611. IEEE Xplore (2007)
16. Weiss, L.: Controllability for various linear and nonlinear systems models. In: *Seminar on Differential Equations and Dynamic System. Lecture Notes in Mathematics*, vol. 144, pp. 250–262. Springer, Heidelberg (1970)
17. Vinagre, B.M.: Fractional Calculus Fundamentals. Tutorial Workshop #2. Fractional Calculus Applications in Automatic Control and Robotics. In: *41st IEEE Conference on Decision and Control*, Las Vegas (2002)

Implementation Aspects of Hybrid Solution Framework

Jarosław Wikarek

Kielce University of Technology
j.wikarek@tu.kielce.pl

Abstract. The paper presents application and implementation aspects of a hybrid approach to modeling and optimization for constrained problems. Two environments of mathematical programming (MP) and constraint programming (CP) were integrated into Hybrid Solution Framework (HSF). The strengths of MP and CP, in which constraints are treated in a different way and different methods are implemented, were combined to use the strengths of both. The proposed approach is particularly important for the decision models in manufacturing, logistic and supply chain management, where an objective function and many discrete decision variables added up in multiple constraints. Implementation details of the proposed framework were presented on example of the supply chain optimization.

Keywords: Constraint Logic Programming, Mathematical Programming, Optimization, Supply Chain, Hybrid Methods.

1 Introduction

The concept of constraint was used originally in physics and combinatorial optimization. It was first applied in computer science for describing interactive drawing system Sketchpad in 1963. In the next decades several experimental languages were proposed that used the notion of constraints and concept of constraint solving. At this time in the field of artificial intelligence (AI), the concept of constraint satisfaction problem (CSP) was formulated and used to describe computer vision. In the nineties, there was a rapid development of constraint-based environments. Commercial systems such as CHIP, ILOG, as well as freeware ones like ECLiPSe [1], were created.

The Constraint Satisfaction Problem (CSP) [1] offers a very good framework for representing the knowledge and information needed for decision constrained problems. A CSP consists of a set of variables and a set of constraints that must be satisfied. In the production, logistics and supply chain domains, many business rules can be easily represented as constraints. CSPs are used often in constraint programming and constraint logic programming (CLP). Constraint programming (CP) is the use of constraints as a programming language to encode and solve problems. Most problems that the constraint programming concerns belong to the group that conventional programming techniques find the hardest. Time needed to solve such problems using unconstrained search increases exponentially with the problem size. Unfortunately, CP-based environments poorly deal with the optimization problems.

The main contribution of this paper is design and description of implementation aspects of Hybrid Solution Framework (HSE). HSE is based on a combination of constraint programming (CP) and mathematical programming (MP).

2 Constraint Logic Programming

Constraint Logic Programming (CLP) is a paradigm that represents a successful attempt to merge the best features of logic programming (LP) and constraint solving.

CLP is also a tool for solving constraint satisfaction problems (CSP) [1, 3]. For the important combinatorial case CSP is characterized by following features:

- a finite set S of integer variables X_1, \dots, X_n , with values from finite domains D_1, \dots, D_n ;
- a set of constraints between variables. The i -th constraint $C_i(X_{i1}, \dots, X_{ik})$ between k variables from S is given by a relation defined as subset of the Cartesian product $D_{i1} \times, \dots, \times D_{ik}$ that determines variable values corresponding to each other in a sense defined by the problem considered;
- a CSP solution is given by any assignment of domain values to variables that satisfies all constraints.

The semantics of constraint logic programs can be defined in terms of a virtual interpreter that maintains a pair $\langle G, S \rangle$ during execution. The first element of this pair is called current goal; the second element is called constraint store.

The current goal contains the literals the interpreter is trying to prove and may also contain some constraints it is trying to satisfy; the constraint store contains all constraints the interpreter has assumed satisfiable so far. At the beginning, the current goal is the goal and the constraint store is empty. The interpreter proceeds by removing the first element from the current goal and analyzing it. In the end this analysis should produce a successful termination or a failure. This analysis could involve recursive calls and addition of new literals to the current goal and new constraint to the constraint store. The interpreter backtracks if a failure is generated. A successful termination is generated when the current goal is empty and the constraint store is satisfiable. CLP can use Artificial Intelligence (AI) techniques to improve the search: constraint propagation, data-driven computation, “forward checking” and “look-ahead” [1]. CLP is a form of CP, in which LP is extended to include concepts from constraint satisfaction. A constraint logic program is a logic program that contains constraints in the body of clauses. Constraints can also be present in the goal. These environments are declarative.

3 Motivation and Contribution

The declarative approach and the use of logic programming provide incomparably greater possibilities for decision problems modeling than the pervasive approach

based on mathematical programming. Unfortunately, those environments deal less with the optimization.

Based on [1–6], and previous works [7–11] we observed and checked some advantages and disadvantages of CP-based and MP-based environments. An integrated approach of constraint logic programming (CLP) and mixed integer programming / mixed integer linear programming (MIP/MILP) can help to solve optimization problems that are intractable with either of the two methods alone [5, 6].

Both MIP/MILP and finite domain CP/CLP involve variables and constraints. However, the types of the variables and constraints that are used, and the way the constraints are solved, are different in the two approaches. Thus, some types of constraints are easier to solve in the CLP and the other in the MP, and vice versa [6, 12].

The motivation and contribution behind this work was to create a hybrid framework for constrained decision problems modeling, solving and optimization instead of using CP-based and MP-based environments separately.

In our approach to modeling and optimization of these problems, we proposed the solution environment, where:

- knowledge related to the problem can be expressed as linear and logical constraints (implementing all types of constraints of the previous MILP/MIP/IP models [7, 9] and introducing new types of constraints (logical, nonlinear, symbolic, etc.));
- the decision models solved using the proposed framework can be formulated as a mathematical programming model or constraint programming model, or it can also be a hybrid model;
- the problem is modeled in CP-based environment, which is far more flexible than MP-based environment;
- the novel method of constraint propagation is introduced (obtained by transforming the decision model to explore its structure);
- constrained domains of decision variables, new constraints and values for some variables are transferred from CP-based environment into MP-based environment;
- the efficiency of finding solutions to larger size problems is increased.

4 Hybrid Solution Environment (HSE)

Both environments have advantages and disadvantages. CP-based environments are declarative and ensure a very simple modeling of decision problems, even those with poor structures if any. The problem is described by a set of logical predicates. The constraints can be of different types (linear, non-linear, logical, binary, etc.). The CLP does not require any search algorithms. This feature is characteristic of all declarative backgrounds, in which modeling of the problem is also a solution, just as it is in Prolog, SQL, etc. The CLP seems perfect for modeling any decision problem.

Numerous MP-based models of decision-making optimization have been developed and tested. Constantly improved methods and mathematical programming

algorithms, such as the simplex algorithm, branch and bound, branch-and-cost [2] etc., have become classics now. Traditional methods when used alone to solve complex problems provide unsatisfactory results. This is related directly to different treatment of variables and constraints in those approaches [6, 10]. The proposed hybrid approach, a composition of methods as described in [12], offers the optimal system for specific contexts.

4.1 Architecture and Implementation of Hybrid Solution Framework (HSF)

The concept of this framework with its phases (P1, ..., P5, G1, ..., G3) is presented in Fig. 1. The names and descriptions of the phases and the implementation environment is shown in Table 1.

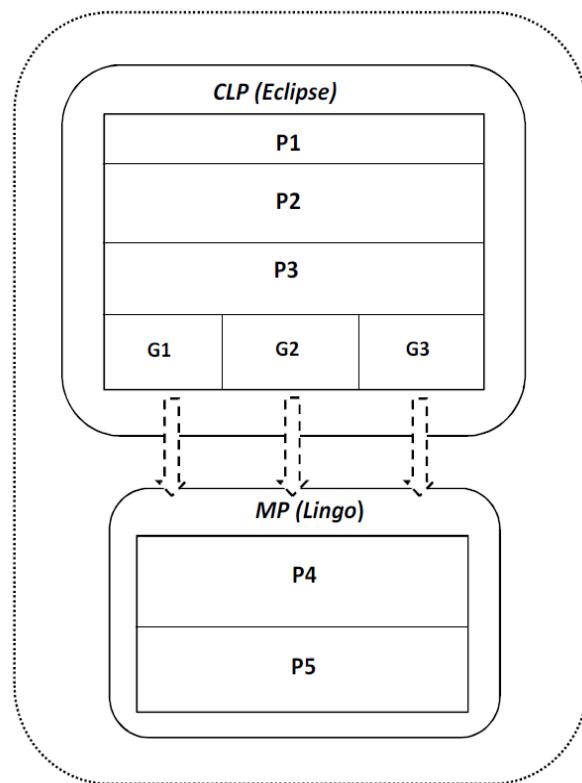


Fig. 1. Scheme of the hybrid solution framework (HSF)

Table 1. Description of Phases

Phase	P1
Name	Implementation of decision model – [environment: CLP].
Description	Implementation of the model in CLP, the term representation of the problem in the form of predicates.
Phase	P2
Name	Transformation of implemented model for better constraint propagation (optional) – [environment: CLP].
Description	Transformation of the original problem aimed at extending the scope of constraint propagation. The transformation uses the structure of the problem. The most common effect is change in the representation of the problem by reducing the number of decision variables, and introduction of additional constraints and variables, changing the nature of the variables, etc.
Phase	P3
Name	Constraint propagation – [environment: CLP]
Description	Constraint propagation for the model. Constraint propagation is one of the basic methods of CLP. As a result, the variable domains are narrowed, and in some cases, the values of variables are set, or even the solution can be found.
Phase	G1
Name	Generation of MILP/MIP/IP model – [environment: CLP].
Description	Generation of the model for mathematical programming. Generation performed automatically using CLP predicate. The resulting model is in a format accepted by the MILP system.
Phase	PG2
Name	Generation of additional constraints (optional) – [environment: CLP].
Description	Generation of additional constraints on the basis of the results obtained in step P3.
Phase	PG3
Name	Generation domains of decision variables and other values – [environment: CLP]
Description	Generation of domains for different decision variables and other parameters based on the propagation of constraints. Transmission of this information in the form of fixed value of certain variables and/or additional constraints to the MP.
Phase	P4
Name	Merging MILP/MIP/IP model – [environment: MILP].
Description	Merging files generated during the phases PG1, PG2, PG3 into one file. It is a model file format in MILP system.
Phase	P5
Name	Solving MILP/MIP/IP model – [environment: MILP].
Description	The solution of the model from the previous stage by a MILP solver. Generation of a report with the results and parameters of the solution.

From a variety of tools for the implementation of the CP-based environment in HSF, ECLiPSe software [13] was selected. ECLiPSe is an open-source software system for the cost-effective development and deployment of constraint programming applications. MP-based environment in HSF was LINGO by LINDO Systems [14]. LINGO Optimization Modeling Software is a powerful tool for building and solving mathematical optimization models [14]. ECLiPSe software is the environmental leader in HSE. ECLiPSe was used to implement the following phases of the framework: P1, P2, P3, G1, G2, G3 (Fig. 1, Table 1). The transformed files of the model were transferred from ECLiPSe to LINGO where they were merged (P4). Then the complete model was solved using LINGO efficient solvers (P5). Constraint propagation (phase P3) significantly affected efficiency of the solution. Therefore phase P2 was introduced. During this phase, the transformation was performed using the structure and properties of the model. This is an optional phase that depends on the modeled

problem. Details of this phase will be presented in one of the illustrative examples in section 5 (cost optimization of supply chain).

5 Illustrative Example – Implementation of the Model

The proposed framework was verified and tested on illustrative example. The example is the authors' original model of cost optimization of supply chain with multimodal transport [9]. During the first stage, the model was formulated as a MILP problem [2] in order to test the HSF (Fig. 1) against the classical integer-programming environment [2, 14]. Indices, parameters and decision variables in the models together with their descriptions are provided in Table 2. The simplified structure of the supply chain network for this model, composed of factories, distributors and customers is presented in Fig. 2.

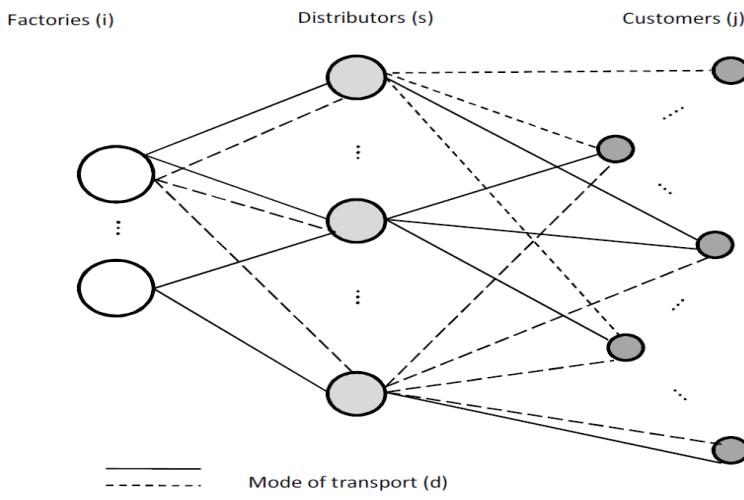


Fig. 2. The simplified structure of the supply chain network for illustrative example

In the construction of this model the following assumptions are valid:

- the shared information process in the supply chain consists of resources (capacity, versatility, costs), inventory (capacity, versatility, costs, time), production (capacity, versatility, costs), product (volume), transport (cost, mode, time), demand, etc.;
- the transport is multimodal. (Several modes of transport. A limited number of means of transport for each mode);
- the environmental aspects of use of transport modes;
- different products are combined in one batch of transport;
- the cost of supplies is presented in the form of a function (in this approach linear function of fixed and variable costs);
- constraints on the simultaneous distribution of certain products can occur.

Table 2. Summary indices, parameters and decision variables

Symbol	Description
<i>Indices</i>	
N	number of manufacturers/factories
M	number of delivery points/customers
E	number of distributors
O	number of product types
L	number of modes of transport
k	product type ($k=1..O$)
j	delivery point/customer/city ($j=1..M$)
i	manufacturer/factory ($i=1..N$)
s	distributor /distribution center ($s=1..E$)
d	mode of transport ($d=1..L$)
<i>Input parameters</i>	
F_s	the fixed cost of distributor/distribution center s
P_k	the area/volume occupied by product k
V_s	distributor s maximum capacity/volume
$W_{i,k}$	production capacity at factory i for product k
$C_{i,k}$	the cost of product k at factory i
$R_{s,k}$	if distributor s can deliver product k then $R_{s,k}=1$, otherwise $R_{s,k}=0$
$T_{p,s,k}$	the time needed for distributor s to prepare the shipment of product k
$T_{c,j,k}$	the cut-off time of delivery to the delivery point/customer j of product k
$Z_{i,k}$	customer demand/order j for product k
Z_t_d	the number of transport units using mode of transport d
Pt_d	the capacity of transport unit using mode of transport d
$Tf_{i,s,d}$	the time of delivery from manufacturer i to distributor s using mode of transport d
$Ka_{i,s,k,d}$	the variable cost of delivery of product k from manufacturer i o distributor s using mode of transport d
$Ra_{i,s,d}$	if manufacturer i can deliver to distributor s using mode of transport d then $Ra_{i,s,d}=1$, otherwise $Ra_{i,s,d}=0$
$A_{i,s,d}$	the fixed cost of delivery from manufacturer i to distributor s using mode of transport d
$Tm_{s,j,d}$	the time of delivery from distributor s to customer j using mode of transport d
$Kb_{s,j,k,d}$	the variable cost of delivery of product k from distributor s to customer j using mode of transport d
$Rb_{s,j,d}$	if distributor s can deliver to customer j using mode of transport d then $Rb_{s,j,d}=1$ otherwise $Rb_{s,j,d}=0$
$G_{s,j,d}$	the fixed cost of delivery from distributor s to customer j using mode of transport d
Od_d	the environmental cost of using mode of transport d
<i>Decision variables</i>	
$X_{i,s,k,d}$	delivery quantity of product k from manufacturer i to distributor s using mode of transport d
$Y_{s,j,k,d}$	delivery quantity of product k from distributor s to customer j using mode of transport d
$Xa_{i,s,d}$	if delivery is from manufacturer i to distributor s using mode of transport d then $Xa_{i,s,d}=1$, otherwise $Xa_{i,s,d}=0$
$Xb_{i,s,d}$	the number of courses from manufacturer i to distributor s using mode of transport d
$Ya_{s,j,d}$	if delivery is from distributor s to customer j using mode of transport d then $Ya_{s,j,d}=1$, otherwise $Ya_{s,j,d}=0$
$Yb_{s,j,d}$	the number of courses from distributor s to customer j using mode of transport d
Tc_s	if distributor s participates in deliveries, then $Tc_s=1$, otherwise $Tc_s=0$
<i>Values calculated</i>	
$Koa_{i,s,d}$	the total cost of delivery from distributor s to customer j using mode of transport d
$Kog_{s,j,d}$	the total cost of delivery from distributor s to customer j using mode of transport d
Cw	Arbitrarily large constant, e.g. the sum of all orders

In the classical method of implementation that model in CP-based environment on the basis of the facts contained in the files *orders.ecl* and *configuration.ecl* generated

adequate representation of the problem and, together with the facts are used in the file *op.ecl*. The file *op.ecl* contains a set of predicates implementing the decision model under constraints [9].

The proposed novel implementations of the problem introduced additional step generation marked with dashed line in Fig. 5. Generation is based on the facts of the files *configuration.ecl* and *orders.ecl*, all feasible routes as well as other feasible facts and placing them sequentially in files *routes.ecl* and *others.ecl*.

In this approach also different is the representation of the problem. Because it contains only one value is not set while in the classical approach, there are five. Each CLP predicate has a corresponding multi-dimensional vector representation. While modeling both problems, quantities i, s, k, d and decision variable $X_{i,s,k,d}$ were vector parameters (Fig. 3a). As shown in Fig. 3b, for each vector instance there are 5 values to be determined. They define the size of the delivery, factories, distributors involved in the delivery and the mode of transport.

[O_n,P,M,D,F,Tu,Tu,Oq,X,T]

Fig. 3a. Representation of the problem in the classical approach – vector definition

[[o_1,p1,m1,_,_,_,10,_,8],
[o_2,p1,m2,_,_,_,20,_,6],...]

Fig. 3b. Representation of the problem in the classical approach – a set of vector instances in the process of finding a solution

Symbols necessary to understand both the representation of the problem and their descriptions are presented in Table 3.

Table 3. Symbols used in the representation of the problem

Symbol	Description
O_n	order number
P	products, $P \in \{p_1, p_2, \dots, p_o\}$
M	customers, $M \in \{m_1, m_2, \dots, m_m\}$
D	distributors, $D \in \{c_1, c_2, \dots, c_e\}$
F	factories, $F \in \{f_1, f_2, \dots, f_n\}$
Tu	transport unit, $Tu \in \{s_1, s_2, \dots, s_l\}$
T	delivery time/period
Oq	order quantity
X	delivery quantity
Name_No	routes name-number

The process of finding the solution may consist in using the constraint propagation methods, variable labeling and the backtracking mechanism. The numbers of parameters that must be specified/labeled in the given predicate/vector critically affect the quality of constraint propagation and the number of backtracks. In both models presented above, the classical problem representation included five parameters: i, s, k, d and $X_{i,s,k,d}$. Considering the domain size of each parameter, the process was complex and time-consuming. In addition, the above representation (Fig. 3a, Fig. 3b) arising from the structure of the problem is the cause of many backtracks.

Our idea involved the transformation of the problem by changing its representation without changing the very problem. All permissible routes were first generated based on the fixed data and a set of orders, then the specific values of parameters i, s, k, d were assigned to each of the routes. In this way, only decision variables $X_{i,s,k,d}$ (deliveries) had to be specified (Fig. 4). This transformation fundamentally improved the efficiency of the constraint propagation and reduced the number of backtracks. A route model is a name adopted for the models that underwent the transformation.

```
[[name_1,f1,p1,c1,m1,s1,s1,5,12,100,_],
 [name_2,f1,p1,c1,m1,s1,s2,6,14,100,_],
 [name_3,f1,p1,c1,m1,s2,s1,6,22,100,_],...]
```

Fig. 4. Representation of the problem in the novel approach – a set of feasible routes

Then all feasible facts, and the facts of *orders.ecl* file is transferred to the main file *opn.ecl* (Fig. 5). Then, it executes the last predicates in CLP-generating components of elements of the model and the data in the format of the LINGO system (gen_mod_1, gen_dan_1, gen_pro).

Intermediate step associated with the generation of acceptable facts and knowledge of the structure of the problem increases in a fundamental way the scope of propagation of constraints and narrowing of domains.

The structure of the main predicates developed in the process of implementation and their descriptions are shown in Table 4.

Table 4. Predicate descriptions

Predicate name	Description	Phase
prod(name, capacity)	The predicate of facts describing the volume of individual products	P1
cust(name_c)	The predicate of facts describing the customers.	P1
dist(name_d, capacity, cost)	The predicate of facts describing the capacity and cost of individual distributors.	P1
fact(name_f)	The predicate of facts describing the factories.	P1
tran_unit(name_tu, capacity, quantity, cost)	The predicate of facts describing the capacity, quantity, and cost of individual transport units.	P1
fact_Dist_Tran_Unit(name_f, name_d, name_tu, cost, time)	Predicates describing the costs and time possible connections between the factory and the distributor carried out the selected means of transport	P1
dist_prod(name_d, product, time)	Predicates specify whether the product and service time by a given distributor.	P1
fact_prod(name_f, product, capacity, cost)	Predicates defining the capacity and cost of the product	P1
exclusion_d(name_d, product, product)	Exclude simultaneous distribution of two selected products by the distributor	P1
order(name_o, product, name_c, time, quantity)	Orders specifying customer demand for a product and delivery date.	PH1
gen_model	Generates predicate of facts in a new form in accordance with the transformed model – a file called p_01	P2
gen_constraint	Constraint propagation - generate possible X created p_02 file (predicate tr (route))	P3
gen_mod_1	Generates LINGO model – a file called l_model.lng	PG1
gen_dan_1	Generates data for the model – create l_date.ldt	PG2
gen_pro	Additional constraint propagation and narrow domains - a file called l_03_dane.lng, l_propagation.lng	PG3

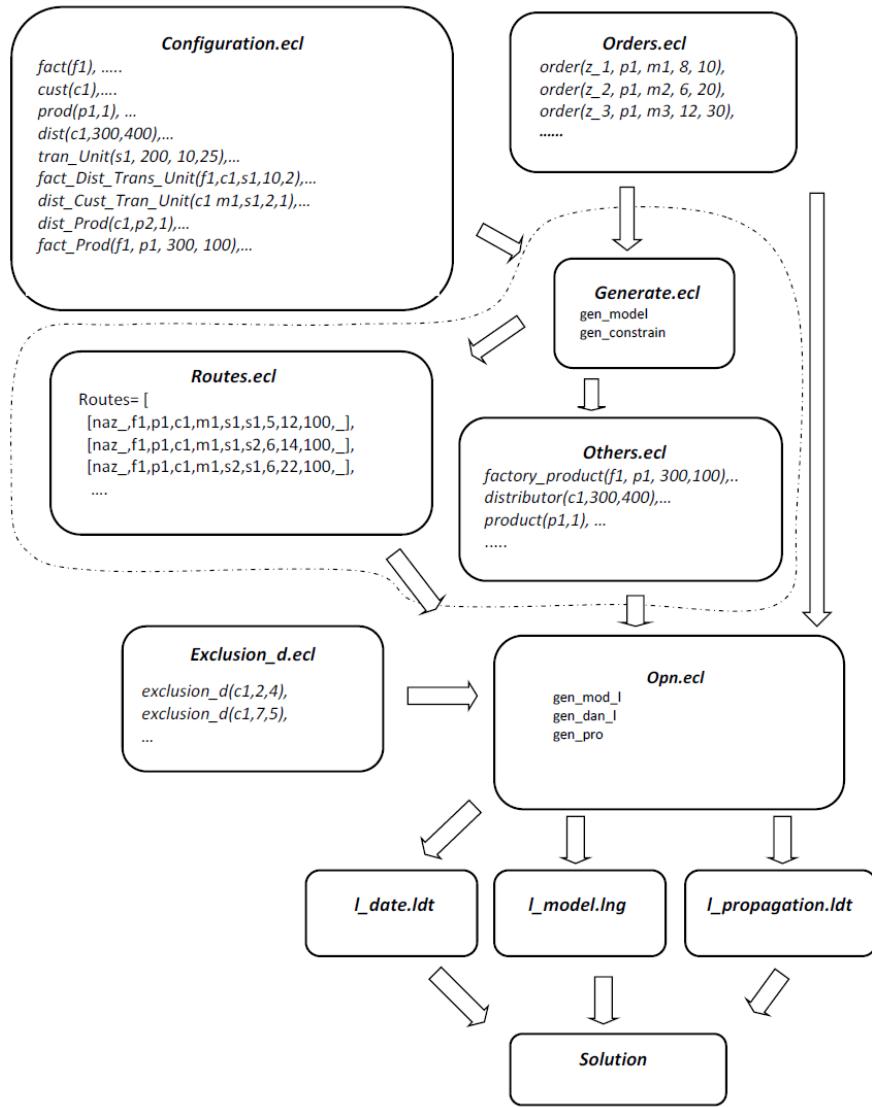


Fig. 5. Block diagram of the implementation of the model in the framework HSF – the extra step marked with a dashed line

6 Numerical Experiments

In order to verify and evaluate the proposed approach, many numerical experiments were performed for illustrative examples. All the examples relate to the supply chain

with two manufacturers ($i=1..2$), three distributors ($s=1..3$), five customers ($j=1..5$), three modes of transport ($d=1..3$), and ten types of products ($k=1..10$).

The series of experiments was designed to show the advantages of the hybrid approach and framework used. The experiments began with five examples: E1, E2, E3, E4, E5 for the problem formulated in MILP [9]. Two approaches were used to implement the proposed model: mathematical programming (LINGO) and the hybrid framework HSF (LINGO, ECLIPSe, transformation). The examples E1 .. E5 varied in terms of the number of orders (No).

Table 5. Results of numerical experiments

<i>E(No)</i>	<i>MILP-LINGO</i>				<i>MILP-HSF</i>			
	<i>F_c</i>	<i>T</i>	<i>V</i>	<i>C</i>	<i>F_c</i>	<i>T</i>	<i>V</i>	<i>C</i>
E1(10)	19 699	353	1541(1389)	1631	19 699	152	186(164)	174
E2(15)	27 345*	400**	1541(1389)	1891	27 306	23	246(224)	174
E3(20)	42 286*	400**	1541(1389)	2161	41 993	123	280(259)	174
E4(25)	44 124*	400**	1541(1389)	2431	43 347	235	340(319)	174
E5(30)	45 945*	400**	1541(1389)	2701	44 632	258	378(357)	174
F_c	the optimal value of the objective function							
T	Solution finding time							
V/C	the number of integer variables/constraints							
*	the feasible value of the objective function after the time T							
**	calculation was stopped after T=400s							

The analysis of the outcome indicates that the hybrid framework provided better results in terms of the time needed to find the solution in each case, and to obtain the optimal solution in some cases, which was impossible to do within the acceptable time limits using the traditional approaches.

7 Conclusion and Future Works

The efficiency of the proposed approach is based on the reduction of the combinatorial problem and use of the best properties of both environments. The hybrid approach (HSF) enables finding better solutions in shorter time (Table 5).

In addition to solving larger problems faster, the proposed approach provides virtually unlimited modeling options.

Therefore, the proposed solution is recommended for decision-making problems that have a structure similar to the presented models [7, 8]. This structure is characterized by the constraints and objective function in which the decision variables are added together. Further work will focus on running the optimization models with non-linear and logical constraints, multi-objective, uncertainty, etc. in the hybrid optimization framework.

References

1. Apt, K., Wallace, M.: Constraint Logic Programming using Eclipse. Cambridge University Press (2006)
2. Schrijver, A.: Theory of Linear and Integer Programming. John Wiley & Sons (1998)
ISBN 0-471-98232-6

3. Rossi, F., Van Beek, P., Walsh, T.: *Handbook of Constraint Programming (Foundations of Artificial Intelligence)*. Elsevier Science Inc., New York (2006)
4. Williams, H.P.: *Logic and Integer Programming*. Springer, Berlin (2009)
5. Achterberg, T., Berthold, T., Koch, T., Wolter, K.: Constraint Integer Programming: A New Approach to Integrate CP and MIP. In: Trick, M.A. (ed.) CPAIOR 2008. LNCS, vol. 5015, pp. 6–20. Springer, Heidelberg (2008)
6. Bockmayr, A., Kasper, T.: A Framework for Combining CP and IP, Branch-and-Infer. *Constraint and Integer Programming Operations Research/Computer Science Interfaces Series* 27, 59–87 (2004)
7. Sitek, P., Wikarek, J.: Supply chain optimization based on a MILP model from the perspective of a logistics provider. *Management and Production Engineering Review*, 49–61 (2012)
8. Sitek, P., Wikarek, J.: A Declarative Framework for Constrained Search Problems. In: Nguyen, N.T., Borzemski, L., Grzech, A., Ali, M. (eds.) IEA/AIE 2008. LNCS (LNAI), vol. 5027, pp. 728–737. Springer, Heidelberg (2008)
9. Sitek, P., Wikarek, J.: Cost optimization of supply chain with multimodal transport. In: Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1111–1118 (2012)
10. Sitek, P., Wikarek, J.: A hybrid approach to supply chain modeling and optimization. In: Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1223–1230 (2013)
11. Sitek, P.: Application of constraint logic programming to decision support for the supply chain management. *Pomiary Automatyka Robotyka* (2), 129–134 (2013)
12. Sitek, P.: A hybrid approach to the Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP). In: Automation 2014, Advances in Intelligent Systems and Computing (printing, 2014)
13. <http://www.eclipse.org>
14. <http://www.lindo.com>

On Choice of the Sampling Period and the Horizons in Generalized Predictive Control

Antoni Wysocki and Maciej Lawryńczuk

Institute of Control and Computation Engineering, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
A.T.Wysocki@stud.elka.pw.edu.pl

Abstract. This paper describes an efficient tuning procedure of the Generalized Predictive Control (GPC) algorithm. It consists of two stages: at first the sampling period is chosen and next the prediction and control horizons are selected which result in the best control quality. Tuning of the GPC algorithm applied to a simulated multi-input multi-output depropaniser distillation column is considered.

Keywords: Process control, Model Predictive Control (MPC), Generalized Predictive Control (GPC), Tuning.

1 Introduction

The core idea of Model Predictive Control (MPC) algorithms is to use on-line a dynamic model of the process to calculate some predicted control errors and to minimise a predefined cost-function which defines the future control quality [1,4,8]. Different variants of MPC algorithms are nowadays successfully used in thousands of industrial applications [5]. It is due to their advantages:

- a) the ability to take into account constraints imposed on input and output variables (or state variables) in a systematic way,
- b) the ability to control multi-input multi-output processes,
- c) very good control quality.

As a result, the MPC technique is a sound alternative to the classical Proportional-Integral-Derivative (PID) controller, which may turn out to be useful in the simplest case of single-input single-output not delayed processes with no constraints.

Simple step-response or impulse-response non-parametric models are used in the first versions of the MPC algorithms [5,8]. A characteristic feature of the Generalized Predictive Control (GPC) strategy [2,3] is the fact that it uses for prediction a discrete-time transfer function (a difference equation). In consequence, the GPC algorithm, unlike some of its predecessors, is able to control efficiently a very broad range of processes, including the systems with integration as well as nonminimum-phase and open-loop unstable systems.

This paper describes a relatively simple, but efficient tuning procedure of the GPC algorithm, i.e. selection of the sampling period and of the prediction

and control horizons. Although some general remarks on tuning are given elsewhere [3,7], the discussed procedure makes it possible to select the horizons which result in the best control quality. To illustrate the tuning procedure, an application of GPC algorithm to a simulated multi-input multi-output depropaniser distillation column is considered.

2 Generalized Predictive Control for Multi-input Multi-output Processes

2.1 The Model

The multi-input multi-output process under consideration has n_u inputs (manipulated variables) and n_y outputs (controlled variables). It means that $u = [u_1 \dots u_{n_u}]^T$, $y = [y_1 \dots y_{n_y}]^T$. The linear discrete-time model of the process is defined by the equation

$$\mathbf{A}(q^{-1})y(k) = \mathbf{B}(q^{-1})u(k) \quad (1)$$

where $\mathbf{A}(q^{-1}) = \text{diag}(A_1(q^{-1}), \dots, A_{n_y}(q^{-1}))$ and

$$\mathbf{B}(q^{-1}) = \begin{bmatrix} B_{1,1}(q^{-1}) & \dots & B_{1,n_u}(q^{-1}) \\ \vdots & \ddots & \vdots \\ B_{n_y,1}(q^{-1}) & \dots & B_{n_y,n_u}(q^{-1}) \end{bmatrix} \quad (2)$$

are polynomial matrices. The polynomials are: $A_i(q^{-1}) = 1 + a_1^i q^{-1} + \dots + a_{n_A}^i q^{-1}$ and $B_{i,j}(q^{-1}) = 1 + b_{1,j}^{i,j} q^{-1} + \dots + b_{n_B}^{i,j} q^{-1}$.

2.2 The GPC Algorithm

At each consecutive sampling instant of the GPC algorithm the following set of future control increments

$$\Delta \mathbf{u}(k) = \begin{bmatrix} \Delta u(k|k) \\ \vdots \\ \Delta u(k + N_u - 1|k) \end{bmatrix} \in \mathbb{R}^{n_u N_u} \quad (3)$$

is calculated, where $\Delta u(k + p|k) = u(k + p|k) - u(k + p - 1|k)$, N_u is the control horizon, in such a way that the predicted control errors are minimised over the prediction horizon N . Typically, the quadratic cost-function is used for optimisation

$$\begin{aligned}
J(k) &= \sum_{p=1}^N \sum_{m=1}^{n_y} \mu_{p,m} (y_m^{\text{sp}}(k+p|k) - \hat{y}_m(k+p|k))^2 \\
&\quad + \sum_{p=0}^{N_u-1} \sum_{n=1}^{n_u} \lambda_{p,n} (\Delta u_n(k+p|k))^2 \\
&= \sum_{p=1}^N \|y^{\text{sp}}(k+p|k) - \hat{y}(k+p|k)\|_{\mathbf{M}_p}^2 + \sum_{p=0}^{N_u-1} \|\Delta u(k+p|k)\|_{\mathbf{A}_p}^2 \quad (4)
\end{aligned}$$

where $y^{\text{sp}}(k+p|k)$ denotes the set-point vector for the sampling instant $k+p$ known in the current instant k , $\hat{y}(k+p|k)$ is the output vector for the sampling instant $k+p$ predicted in the current instant k , \mathbf{M}_p and \mathbf{A}_p are weighting matrices of dimensionality $n_y \times n_y$ and $n_u \times n_u$, respectively.

In general, it is possible to take into account some constraints imposed on manipulated and controlled variables, but in the majority of practical applications it is usually sufficient to consider only the first of them. Let the vectors $u^{\min} = [u_1^{\min} \dots u_{n_u}^{\min}]^T$, $u^{\max} = [u_1^{\max} \dots u_{n_u}^{\max}]^T$, $\Delta u^{\max} = [\Delta u_1^{\max} \dots \Delta u_{n_u}^{\max}]^T$ define the constraints imposed on the minimal value, the maximal value and the rate of change of the manipulated variables. Using the cost-function (4) and the vector-matrix notation, the GPC optimisation problem is

$$\begin{aligned}
\min_{\Delta \mathbf{u}(k)} \{ J(k) = \|y^{\text{sp}}(k) - \hat{y}(k)\|_{\mathbf{M}}^2 + \|\Delta \mathbf{u}(k)\|_{\mathbf{A}}^2 \} \\
\text{subject to} \\
\mathbf{u}^{\min} \leq \mathbf{u}(k) \leq \mathbf{u}^{\max} \\
-\Delta \mathbf{u}^{\max} \leq \Delta \mathbf{u}(k) \leq \Delta \mathbf{u}^{\max} \quad (5)
\end{aligned}$$

where

$$\mathbf{y}^{\text{sp}}(k) = \begin{bmatrix} y^{\text{sp}}(k) \\ \vdots \\ y^{\text{sp}}(k) \end{bmatrix}, \quad \hat{\mathbf{y}}(k) = \begin{bmatrix} \hat{y}(k+1|k)(k) \\ \vdots \\ \hat{y}(k+N|k)(k) \end{bmatrix} \in \mathbb{R}^{n_y N} \quad (6)$$

$$\mathbf{u}^{\min} = \begin{bmatrix} u^{\min} \\ \vdots \\ u^{\min} \end{bmatrix}, \quad \mathbf{u}^{\max} = \begin{bmatrix} u^{\max} \\ \vdots \\ u^{\max} \end{bmatrix}, \quad \Delta \mathbf{u}^{\max} = \begin{bmatrix} \Delta u^{\max} \\ \vdots \\ \Delta u^{\max} \end{bmatrix} \in \mathbb{R}^{n_u N_u} \quad (7)$$

and $\mathbf{M} = \text{diag}(\mathbf{M}_1, \dots, \mathbf{M}_N)$, $\mathbf{A} = \text{diag}(\mathbf{A}_0, \dots, \mathbf{A}_{N_u-1})$ are matrices of dimensionality $n_y N \times n_y N$ and $n_u N_u \times n_u N_u$, respectively.

Thanks to the linear nature of the model, the prediction vector can be expressed as the following sum [4,8]

$$\hat{y}(k) = \mathbf{G} \Delta \mathbf{u}(k) + \mathbf{y}^0(k) \quad (8)$$

where the dynamic step-response matrix (of dimensionality $n_y N \times n_u N_u$) and the free trajectory vector are

$$\mathbf{G} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0}_{n_y \times n_u} & \dots & \mathbf{0}_{n_y \times n_u} \\ \mathbf{S}_2 & \mathbf{S}_1 & \dots & \mathbf{0}_{n_y \times n_u} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_N & \mathbf{S}_{N-1} & \dots & \mathbf{S}_{N-N_u+1} \end{bmatrix}, \quad \mathbf{y}^0(k) = \begin{bmatrix} y^0(k+1|k) \\ \vdots \\ y^0(k+N|k) \end{bmatrix} \in \mathbb{R}^{n_y N} \quad (9)$$

The free trajectory is calculated on-line assuming only the influence of the past.

Taking into account Eq. (8), the general GPC optimisation problem (5) becomes the following quadratic programming task

$$\begin{aligned} \min_{\Delta \mathbf{u}(k)} \{ J(k) = & \| \mathbf{y}^{\text{sp}}(k) - \mathbf{G} \Delta \mathbf{u}(k) - \mathbf{y}^0(k) \|_M^2 + \| \Delta \mathbf{u}(k) \|_{\Lambda}^2 \} \\ \text{subject to} \\ \mathbf{u}^{\min} \leq \mathbf{u}(k) \leq \mathbf{u}^{\max} \\ -\Delta \mathbf{u}^{\max} \leq \Delta \mathbf{u}(k) \leq \Delta \mathbf{u}^{\max} \end{aligned} \quad (10)$$

2.3 Algorithm Tuning

Two objectives must be met:

- a) control quality must be good,
- b) computational burden should be as low as possible.

Three tuning parameters of the GPC algorithm are considered in this paper: the control horizon N_u , the prediction horizon N and the sampling period T_s . The penalty coefficients $\mu_{p,m}$, $\lambda_{p,n}$ are assumed to be fixed and the constraints result from some technological requirements. From the point of view of control quality:

- a) The horizons should be as long as possible, ideally infinite. The infinite control horizon is not possible in practice because it leads to an infinite number of decisions variables in the GPC optimisation problem (while infinite prediction horizon is possible [8]).
- b) The sampling period should be short, but the algorithm must be implementable, i.e. the maximal expected time necessary for completing calculations in one iteration must be shorter than the sampling period.

In general, computational complexity of the optimisation problem (10) depends on all the three considered tuning parameters and the number of process inputs n_u and outputs n_y . The shorter the sampling period, the longer must be the horizons.

The proposed tuning procedure consists of two stages:

- a) At first, relatively long prediction and control horizons are used. The sampling period is chosen as a compromise between good control quality (to be short enough) and implementation possibility (to be long enough).

- b) Next, for the selected sampling period, process trajectories for a set of different prediction and control horizons are considered. The horizons are chosen which give the best control quality.

To compare control quality, the following index is defined

$$J_{\text{sim}} = \frac{1}{N_k} \sum_{k=1}^{N_k} \|y^{\text{sp}}(k) - y(k)\|^2 \quad (11)$$

where N_k is the number of iterations of the algorithm, $y^{\text{sp}}(k)$ and $y(k)$ are the desired set-point vector and the real process output vector, respectively. All simulations (or experiments with a real process) are carried out for some assumed set-point trajectories, fixed penalty coefficients $\mu_{p,m}$, $\lambda_{p,n}$ and constraints.

3 Simulation Results

3.1 The Process and Its Model

The considered process is a depropaniser distillation column [6]. The process has three inputs: u_1 —column top reflux flow, u_2 —column bottom steam flow, u_3 —column overhead pressure, and three outputs: y_1 —top butane concentration, y_2 —bottom propane concentration, y_3 —column flooding. The continuous-time model is defined by the following transfer function matrix

$$G(s) = \begin{bmatrix} \frac{-0.26978e^{-27.5s}}{97.5s + 1} & \frac{1.978e^{-53.5s}}{118.5s + 1} & \frac{0.07724e^{-56s}}{96s + 1} \\ \frac{0.4881e^{-117s}}{56s + 1} & \frac{5.26e^{-26.5s}}{58.5s + 1} & \frac{0.19996e^{-35s}}{51s + 1} \\ \frac{-0.6e^{-16.5s}}{40.5s + 1} & \frac{5.5e^{-15.5s}}{19.5s + 1} & \frac{-0.5e^{-17s}}{18s + 1} \end{bmatrix} \quad (12)$$

Table 1 shows the influence of the sampling time T_s on dynamic order of the discrete-time model, i.e. the values of $\tau^{i,j}$, n_A^i , $n_B^{i,j}$ are given. The longer the sampling time, the lower the order of dynamics.

3.2 Algorithm Tuning

The weighting parameters of the GPC algorithm are: $\mathbf{M}_p = \text{diag}(10, 2, 5)$, $\mathbf{A}_p = \text{diag}(1, 1, 1)$, the constraints are defined by: $u^{\min} = [-2.5 \ -1 \ -0.5]^T$, $u^{\max} = [1 \ 1 \ 10]^T$, $\Delta u^{\max} = [1 \ 0.2 \ 0.2]^T$.

Table 1. The influence of the sampling time T_s on dynamic order of the model

Model parameters	$T_s = 1$ sec.	$T_s = 6$ sec.	$T_s = 15$ sec.
$\begin{bmatrix} \tau^{1,1} & \tau^{1,2} & \tau^{1,3} \\ \tau^{2,1} & \tau^{2,2} & \tau^{2,3} \\ \tau^{3,1} & \tau^{3,2} & \tau^{3,3} \end{bmatrix}$	$\begin{bmatrix} 27 & 53 & 56 \\ 117 & 26 & 35 \\ 16 & 15 & 17 \end{bmatrix}$	$\begin{bmatrix} 4 & 8 & 9 \\ 19 & 4 & 5 \\ 2 & 2 & 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 3 & 3 \\ 7 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$
$\begin{bmatrix} n_A^1 \\ n_A^2 \\ n_A^3 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$
$\begin{bmatrix} n_B^{1,1} & n_B^{1,2} & n_B^{1,3} \\ n_B^{2,1} & n_B^{2,2} & n_B^{2,3} \\ n_B^{3,1} & n_B^{3,2} & n_B^{3,3} \end{bmatrix}$	$\begin{bmatrix} 30 & 56 & 59 \\ 120 & 29 & 38 \\ 19 & 18 & 20 \end{bmatrix}$	$\begin{bmatrix} 7 & 11 & 12 \\ 22 & 7 & 8 \\ 5 & 5 & 5 \end{bmatrix}$	$\begin{bmatrix} 4 & 6 & 6 \\ 10 & 4 & 5 \\ 4 & 4 & 4 \end{bmatrix}$

Choice of the sampling period. At first, relatively long prediction and control horizons are used: $N=200$, $N_u=30$. The horizons are not infinite, but long enough, because further increase of the horizons does not lead to any significant change of process behaviour.

Fig. 1 depicts simulation results for $T_s = 1$ sec. and $T_s = 6$ sec., Fig. 2 depicts simulation results for $T_s = 6$ sec. and $T_s = 15$ sec. for some assumed set-point trajectories. For $T_s = 1$ sec. control signals change very rapidly, as a result the output trajectories are fast. For $T_s = 6$ sec. all changes are slower, the overshoot is bigger, but still acceptable. Finally, for $T_s = 15$ sec. all the trajectories are much slower, regulation time of the first and the second outputs is very long. Fig. 3 shows how the control quality index J_{sim} depends on the sampling period. Taking into account two contradictory objectives, i.e. good control quality and the fact that short sampling periods may be not possible in practice, the sampling period $T_s = 6$ sec. is chosen as a reasonable compromise, for which $J_{\text{sim}} = 0.2198$ (for $T_s = 1$ sec. $J_{\text{sim}} = 0.1764$, for $T_s = 15$ sec. $J_{\text{sim}} = 0.3035$).

Choice of the horizons. All the simulations presented so far assume that the horizons are long enough (ideally infinite). It is interesting to analyse how shortening of the horizons influences the quality of control. Table 2 shows the influence of the prediction and control horizons on the index J_{sim} for the selected sampling period and the assumed set-point trajectories. Its best (lowest) value 0.2168 is obtained for the horizons $N = 9$ and $N_u = 3$. One may notice, however, that for shorter horizons $N = 8$ and $N_u = 2$ it is possible to obtain a little bit worse value $J_{\text{sim}} = 0.2193$, which is only by some 1.15% greater than the best one. Moreover, further shortening of the horizons does not lead to noticeable deterioration of control quality and long horizons do not improve the quality significantly. That is why the horizons $N = 8$ and $N_u = 2$ are finally chosen.

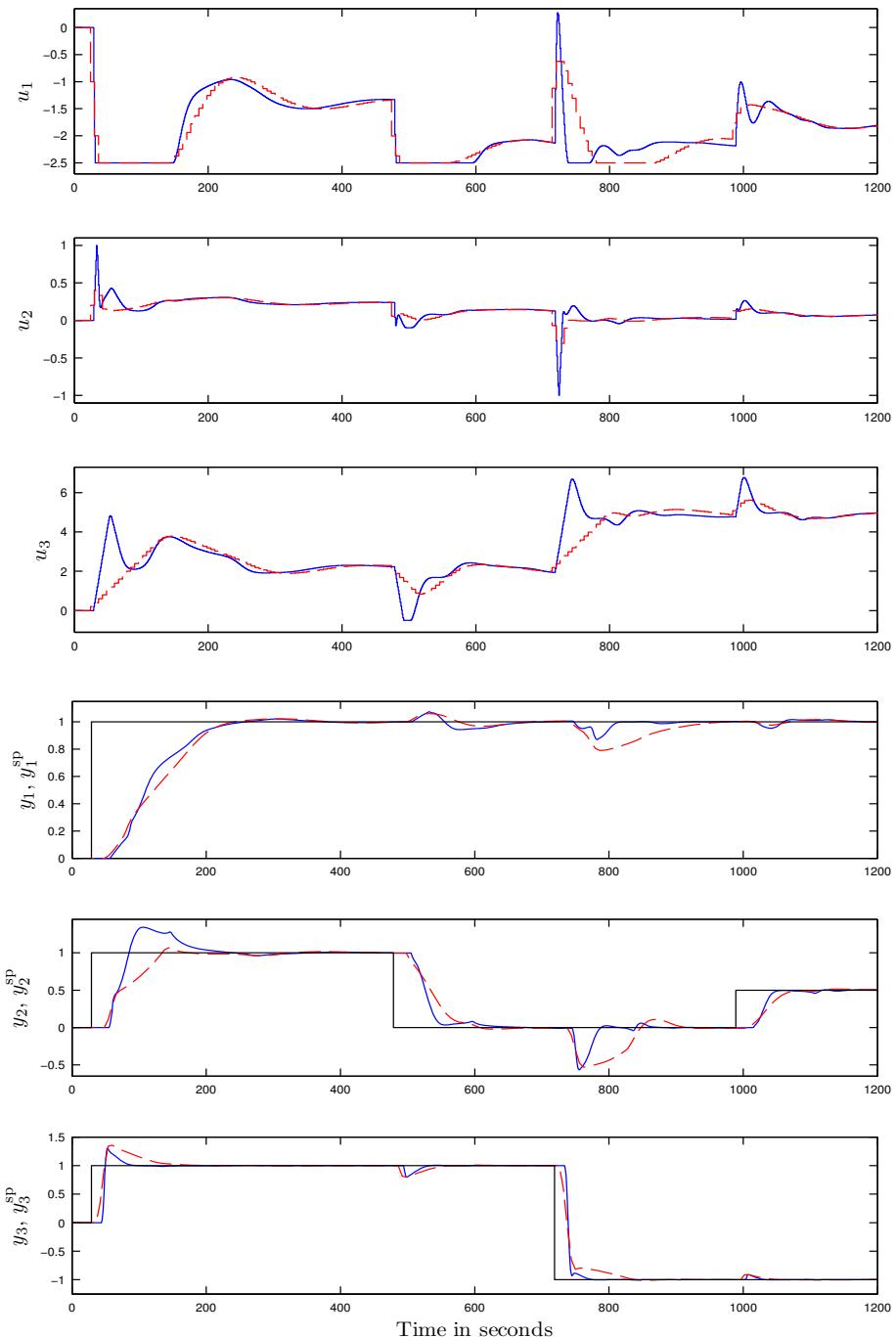


Fig. 1. Simulation results for long horizons $N = 200$, $N_u = 30$: the trajectories for $T_s = 1 \text{ sec.}$ (*solid line*) and $T_s = 6 \text{ sec.}$ (*dashed line*)

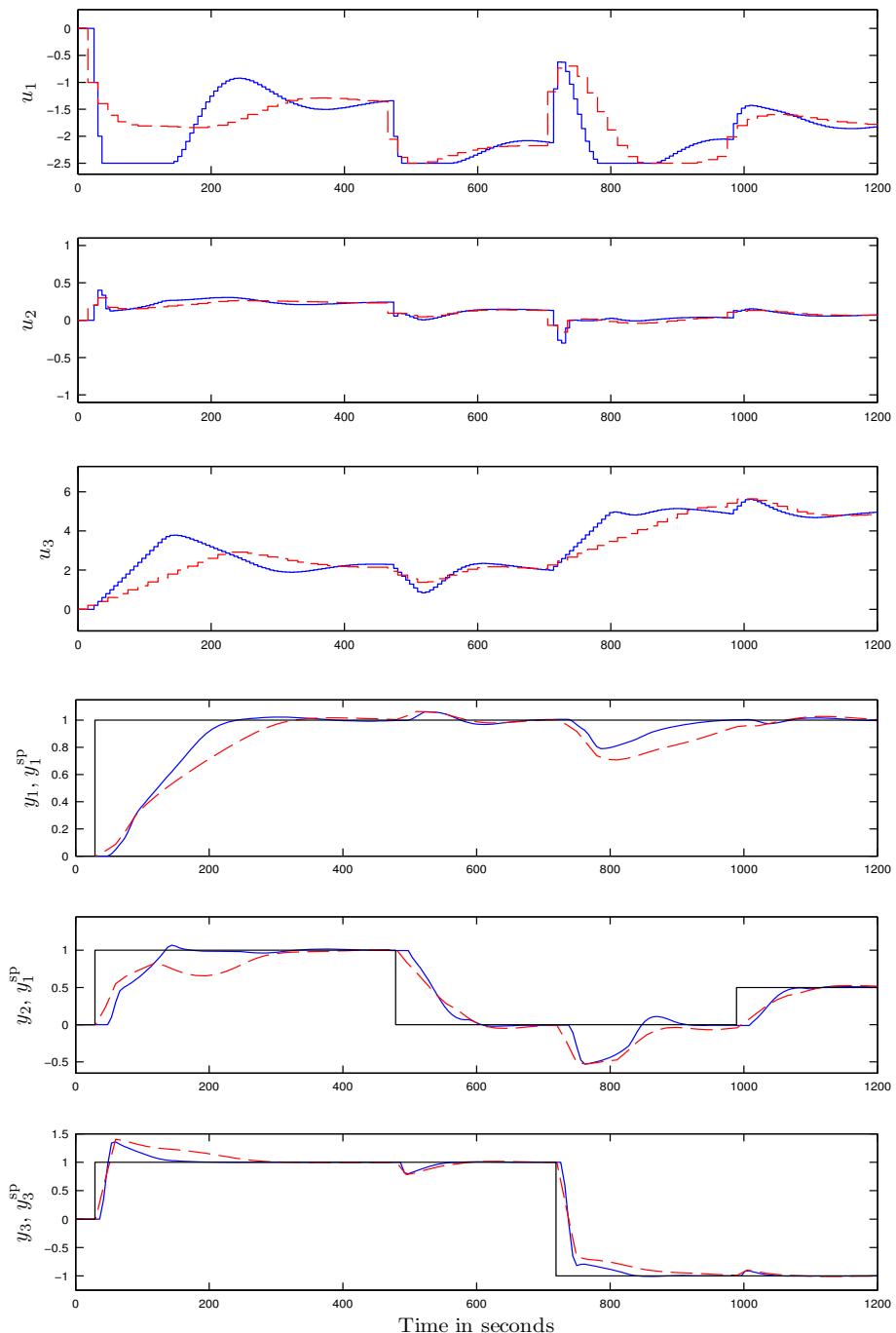


Fig. 2. Simulation results for long horizons $N = 200$, $N_u = 30$: the trajectories for $T_s = 6$ sec. (solid line) and $T_s = 15$ sec. (dashed line)

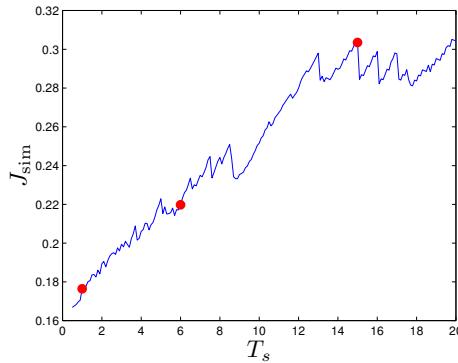


Fig. 3. The influence of the sampling period T_s on the control quality index J_{sim} for long horizons $N = 200$, $N_u = 30$

Table 2. The influence of the prediction and control horizons (N and N_u) on the control quality index J_{sim} , $T_s = 6$ sec.

N	N_u									
	1	2	3	4	5	10	15	20	25	30
3	1.0164	1.0164	1.0164							
4	1.0220	1.0206	1.0206	1.0206						
5	0.6165	0.5898	0.5909	0.5909	0.5909					
6	0.3199	0.3130	0.3163	0.3162	0.3162					
7	0.2472	0.2362	0.2360	0.2353	0.2353					
8	0.2316	0.2193	0.2178	0.2177	0.2175					
9	0.2325	0.2178	0.2168	0.2170	0.2170					
10	0.2344	0.2180	0.2173	0.2175	0.2175	0.2175				
15	0.2422	0.2224	0.2191	0.2186	0.2182	0.2180	0.2180			
20	0.2485	0.2261	0.2209	0.2192	0.2183	0.2179	0.2180	0.2180		
25	0.2501	0.2283	0.2222	0.2196	0.2182	0.2182	0.2185	0.2185	0.2185	
30	0.2529	0.2302	0.2229	0.2196	0.2180	0.2186	0.2192	0.2191	0.2190	0.2190

Finally, Fig. 4 compares process trajectories for long horizons ($N_u = 200$, $N_u = 30$) and for short horizons ($N = 8$, $N_u = 2$) for the chosen sampling period $T_s = 6$. Although some differences are present, it is interesting to notice that from the point of view of the quality index obtained results are very similar: in the first case $J_{\text{sim}} = 0.2198$, in the second case $J_{\text{sim}} = 0.2193$.

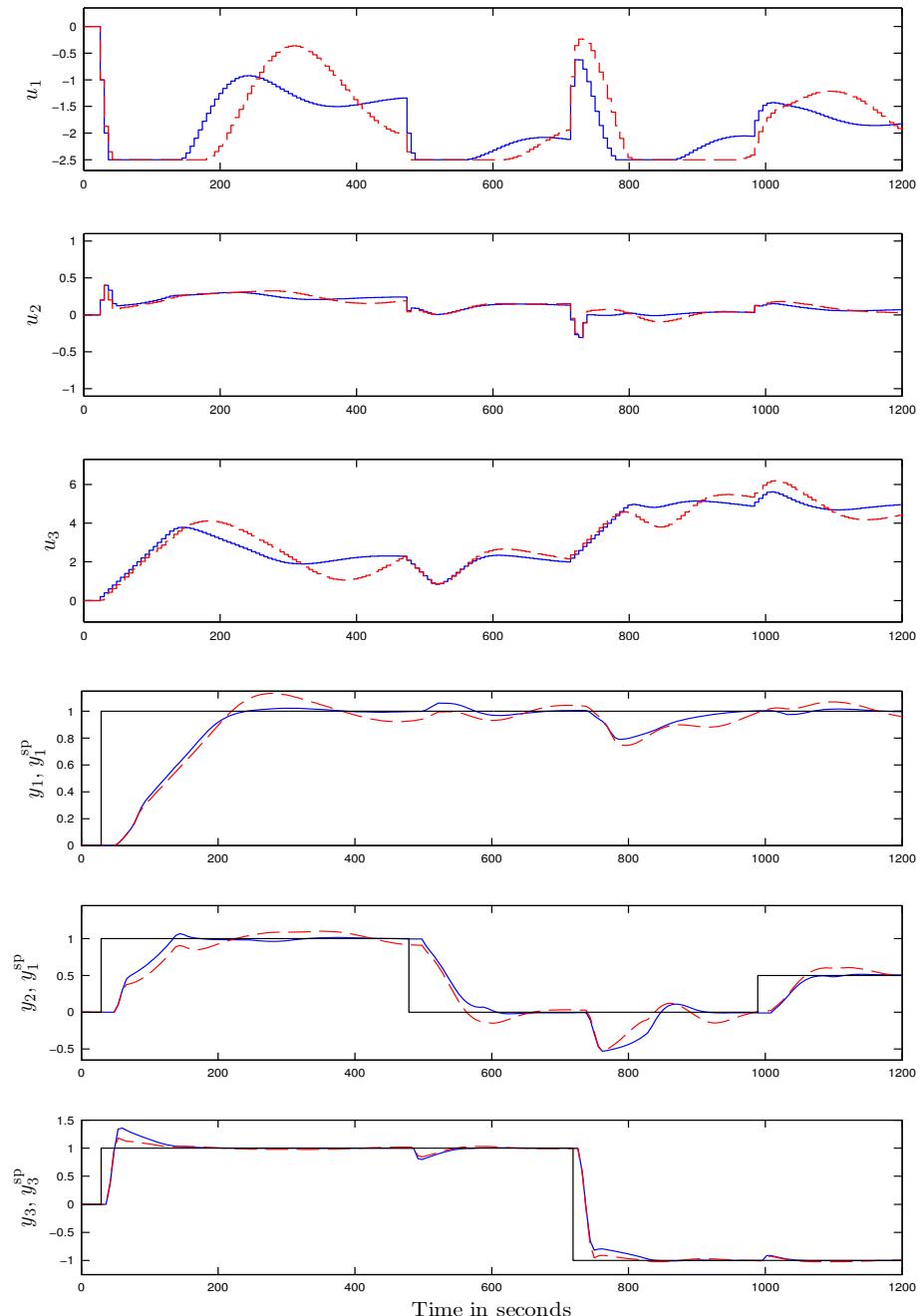


Fig. 4. Simulation results for the chosen sampling period $T_s = 6$ sec.: the trajectories for long horizons $N = 200, N_u = 30$ (solid line) and for short horizons $N = 8, N_u = 2$ (dashed line)

4 Summary

On the one hand, good control quality is the objective of any MPC algorithm, which needs short sampling periods and long horizons. On the other hand, the algorithm must be implemented using the available hardware and software, so the sampling period should be long enough and the horizons must be moderate. This paper shows a practical two-stage method of selecting the tuning parameters to reach a compromise between the contradictory objectives. The described method is quite intuitive and universal, i.e it is not limited to the GPC algorithm.

Acknowledgement. The work presented in this paper was supported by Polish national budget funds for science.

References

1. Camacho, E.F., Bordons, C.: *Model Predictive Control*. Springer, London (1999)
2. Clarke, D.W., Mohtadi, C.: Properties of generalized predictive control. *Automatica* 25, 859–875 (1989)
3. Clarke, D.W., Mohtadi, C., Tuffs, P.S.: Generalized predictive control. *Automatica* 23, 137–160 (1987)
4. Maciejowski, J.M.: *Predictive control with constraints*. Prentice Hall, Harlow (2002)
5. Qin, S.J., Badgwell, T.A.: A survey of industrial model predictive control technology. *Control Engineering Practice* 11, 733–764 (2003)
6. Qing-Guo, W., Yu, Z., Min-Sen, C.: Non-interacting control design for multivariable industrial processes. *Journal of Process Control* 13, 253–265 (2003)
7. Scattolini, R., Bittanti, S.: On the choice of the horizon in long-range predictive control—some simple criteria. *Automatica* 26, 915–917 (1990)
8. Tatjewski, P.: *Advanced control of industrial processes, Structures and algorithms*. Springer, London (2007)

Part II

Robotics

A Compact Walking Robot – Flexible Research and Development Platform

Dominik Belter and Krzysztof Walas

Institute of Control and Information Engineering, Poznan University of Technology,
ul. Piotrowo 3a, 60-965 Poznan, Poland

{dominik.belter,krzysztof.walas}@put.poznan.pl

Abstract. In the paper new six-legged robot Messor II is described. The new machine is the improved version of the previous robot Messor. The current design has better *power to mass* ratio. Additionally new servos, which power the joint of the robot, allows for better control and motion execution. The paper contains three main parts. In the first section mechanical design is presented. Then, the electronic part of the robot is described. Next the control system of the robot is outlined.

Keywords: walking robot, hexapod, design.

1 Introduction

The scientific research on machines that mimic animals locomotion has a long story and according to [1] is dated back to the 19th century. Despite the widely available wheeled locomotion systems people are still interested in building machines that walk for several reasons. Many parts of earth land mass are inaccessible for the wheeled machines. Legged robots are more agile in the obstacles negotiation and above all study of the legged locomotion is a very interesting research topic. Mainly driven by the reasons mentioned above the new generation of the six-legged robot was built.

In the development process the experience gathered during our earlier research on walking robots was exploited. As early as in 2007 the Ragno robot was presented [2]. The machine was used for gait algorithms development [3] and in preliminary research on obstacles negotiation. However, size of the Ragno robot was the limiting factor in its use in the real world scenarios.

The next generation robot called Messor [4] was twice as big as Ragno. It was able to carry more sophisticated sensors and perform more complex tasks. The first research was conducted on supporting locomotive functions of the robot [5]. Next, the algorithms for stair climbing were developed [6], followed by the extension of the approach to negotiation of other urban obstacles [7]. Furthermore, the study of the posture optimization [8] was performed. Using the strong connection between sensing and control of the robot foothold selection method was proposed [9]. Additionally, some specific methods for walking

robot self-localization [10] were investigated. Latest research is focused on terrain recognition [11] and the negotiation of the ground under various traction conditions. All the research allowed to gain sufficient knowledge to build the improved walking machine.

1.1 Related Work

The thorough study of initial development and control of the walking machines was presented in [12]. One of the early six-legged walking machines which was expected to work as a planetary rover was Ambler [13]. The six-legged machines may seem to be outdated. But according to authors of [14] six legged robots are not to be replaced by the quadrupeds due to two reasons: its stability and large margin of safety when walking on uneven ground. Even when two legs are loosing the contact with the ground the robot is still able to perform its gait.

However, it is unjust to omit the new achievements for the quadruped robots. One of the most impressive machines is the BigDog [15]. Powered with hydraulic motors it is able to perform robust gaits in the occurrence of the unperceived obstacles. Its smaller version LittleDog [16] based on electrical motors is not yet so powerful and hence not so impressive in its abilities. One of the latests research robots is HyQ [17] build in Italy is as in case of the BigDog also hydraulically powered robot. The researchers are aiming to compete with BigDog. Other example of the quadruped robot is the one developed in Switzerland. Electrically powered StarIETH [18] due to its compliance is able to fall safely from significant height (comparing to its size).

Moving back to the topic of hexapods. The research on the six-legged machines is still a current topic. To support this point of view the short list of the six legged machines built in last three years is given. This year the new Lauron V [19] was presented. In the paper from previous year the space rover named SpaceClimber [20] was described. Additionally, the biological studies on six-legged robot are performed in the research project conducted at the University of Bielefeld. The prototype of the insect-like robot Hector [21] is being built. Looking back into the past, in the last ten years some other interesting examples of the six-legged robots could be found. One of them is the robot which is able to climb on the vertical walls called RiSE [14]. The other one is the robot based on fingers of the robotics hand DLRHand [22]. When looking at the application of the robots one of the possible task could be demining mission. Examples of such robots are: SILO6 [23], and the robot presented in [24].

1.2 Our Approach

The requirements for the design of the new robot were the following. To make it lighter and more compact then the previous version of Messor robot. Additionally the designed robot should be a flexible research platform. The requirements posed are: good power to weight ratio, reduced electronic design and reliability.

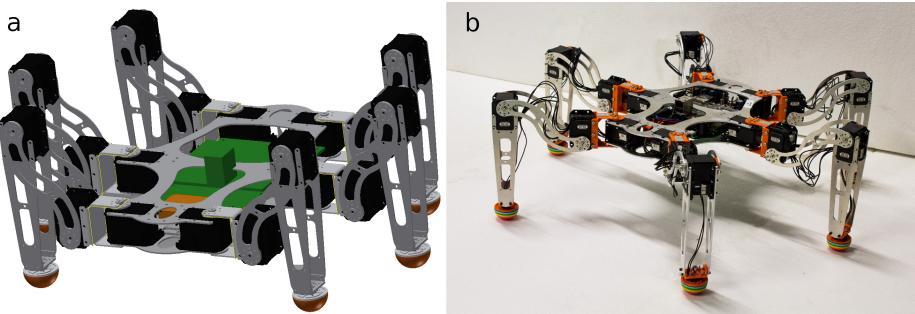


Fig. 1. Messor II: CAD design(a), real robot(b)

The paper consist of three main parts. First the mechanical design is presented. Then the electronic part of the robot is described. Next the control system of the robot is outlined. At the end concluding remarks are given.

2 Mechanical Design

In the mechanical part new robot has the similar shape to the previous robot Messor. However, the similarities in the shape are not constraining the better design of the robot in terms power to weight ratio.

2.1 Project Phase

Messor II was designed in CAD environment. The work was focused on adjusting the dimensions of the trunk and the legs of the robot. The machine should be able to work in the man-made environment e.g. climb stairs, but in the same time to have the reduced power consumption i.e. shorter parts of the legs – lower torques. Additionally, the range of movements of the robot legs was supposed to be as high as possible. Other mechanical improvement was the ease of exchanging parts of the robot. The leg of the robot could be removed by loosing several screws and taking out two electrical plugs. The robot in the project phase is shown in Figure 1a.

2.2 Real Machine

The real robot was manufactured using aluminum parts and several 3D printed tiles. The robot is shown in Figure 1b. Its joint are powered by 18 servomotors Robots Dynamical RX-28. The available torque is 2.5 Nm at 1.8CM but it is possible to power the motor with higher voltage and obtain higher torque. No load speed is 67 RPM. The weight of the single servomotor is 72 g. Multiplying it by the total number of joints one obtains $18 \times 72 \text{ g} = 1296 \text{ g}$. Adding other mechanical and electronics parts mounted on board the total weight of the robot is 2.5 kg.

The dimensions of the robot trunk and robot legs are shown in Figure 2. The trunk of the robot is 299 mm long and 205 mm wide (Figure 2a). Whereas, the leg of the robot (Figure 2b) has the following dimensions: 49 mm, 120 mm, 174 mm (beginning from the trunk). The relation of the dimensions is 1.4:3.45:5. The proportions of the leg parts proposed in the design is close to the proportions of the insects legs present in nature. For the insects leg the relation is 1:4:5.

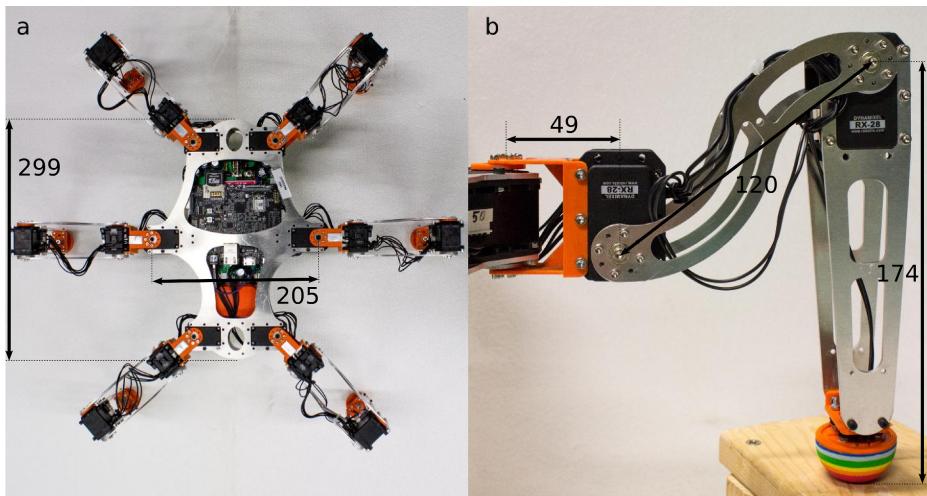


Fig. 2. Dimensions of the robot's trunk(a) and leg(b)

3 Electronic Design

The control system of the robot is designed to provide high flexibility of the development and to support research activities. The goal is to minimize an effort required to develop new algorithms. For this purpose most of the control tasks are solved by software. The electronic design is reduced to minimum which is required to establish communication of the control software with the robot's servomotors and on-board sensors. In this approach the changes to the control system require only modifications of the software architecture.

3.1 Sensors and Equipment

The sensory equipment of the Messor II robot is presented in Fig. 3. The robot is designed to operate with low cost, energy efficient RGB-D cameras. The exteroceptive sensor, which measure the state of the environment, is the PrimeSense sensor-based camera. Currently the Kinect sensor is used. In the future we are going to exchange the Kinect sensor with the smaller one – Carmine short range sensor from PrimeSense. Optionally we can also use the SwissRanger SR4000 Mesa Imaging sensor for outdoor experiments.

Most of the proprioceptive measurements are performed by Dynamical RX-28 servomotors. Servomotors are mounted directly in the joints of the robot. Thus, no additional encoder is needed to measure the joint position. The embedded measurement set consists of rotational potentiometer and analog-to-digital converter (ADC). The measurement results can be read using Dynamical interface protocol. The embedded functionality allows also to read current rotational speed of the joint and current load which is proportional to the joint torque. The information about current joint position and the joint torque allows to estimate the reaction force (direction and module of the force vector) in each foot of the robot. The information about the force of the interaction with the environment is useful in reactive controller to preserve stability of the robot [5]. Moreover, the force signal can be used to classify the terrain during walking on compliant terrain [11].

The measurements of the leg state are supported by two additional sensors. The hall sensors (ACS712) are used to measure total leg current. Each sensor is attached to the single leg of the robot. The analog output from the sensor is converted to digital form and is sent to the main microprocessor. The measurements are used to compute the total current load of the robot. The output can be used to compensate unequal load distribution of robot's legs. Moreover, the robot can turn off the leg when the load reaches predefined limit. Further increase of the load might damage the leg.

The inclination of the robot's platform is measured using Attitude and Heading Reference System (AHRS) Xsens MTi. The sensor contains micro electro-mechanical inertial sensors to measure rotational velocity and linear acceleration (three gyroscopes and three accelerometers). The measurements are supported by three-axis magnetometer – the absolute rotational sensors which prevents measurement drift. Measurements from all sensors are integrated by Kalman filter. The device returns accurate information about inclination of the robot's body.

The contact between robot's foot and the ground is additionally detected by using micro-switches mounted at the each leg's tip. We found this sensors the most robust and reliable among all sensors used for this purpose [4]. It works perfectly after thousands of working cycles. More complex mechanical sensors or tactile force sensors starts to loose reliability after hundreds of working cycles. The disadvantage of the micro-switches is the binary information. In our case the information about contact force can be obtained from the torque sensors mounted in robot's joints. The information about the torque and the inverse Jacobian of each leg is sufficient for this purpose.

3.2 Electronic Structure

The core of the electronic structure of the robot is the PandaBoard ES development platform with dual-core ARM Cortex A9 microprocessor. The CPU clock speed is 1.2 GHz. The board is equipped with 1 GB DDR2 RAM and SD card working as a hard disk. The installed operating system is Ubuntu Linux for OMAP4 microprocessors. Beyond standard sockets (USB, Ethernet, RS-232) the

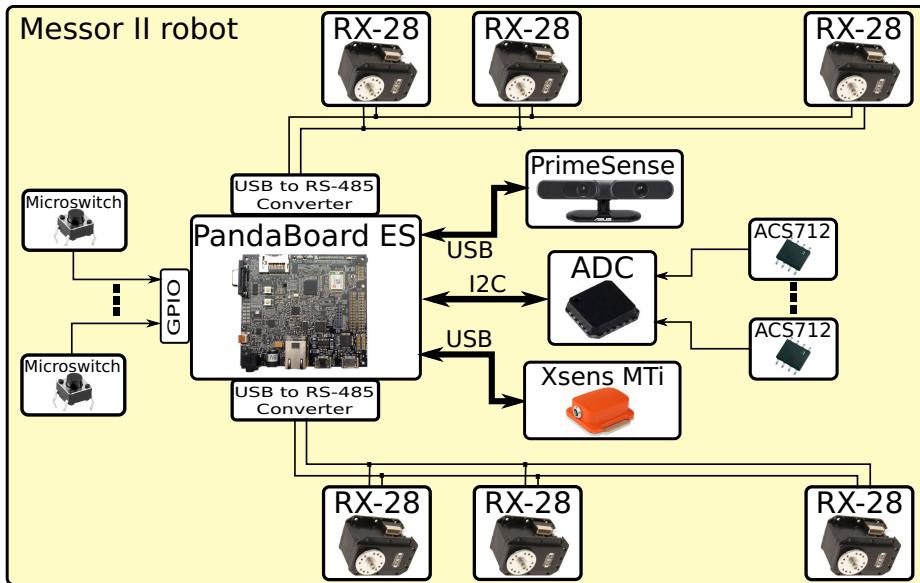


Fig. 3. Electronic design of the Messor II robot

board is equipped with extension port with additional USB, I2C, GPIO (General Purpose Input-Output) and RS-232 interfaces.

The servomotors are connected with RS-485 interface. To send messages to servomotor controllers USB socket is used. The interface between USB and RS-485 bus uses FT232RL converter. USB standard is converted to RS-232 standard. Then, the RS-232 signal is converted to RS-485 using MAX3443 integrated circuit. FT232RL sets automatically the direction of the RS-485 transmission.

PandaBoard development platform has only two channels of analog to digital converter (audio input). Thus, ADC which is connected to the board using I2C interface was added. Each channel measures total current of the corresponding leg. Micro-switches which measure contact with the ground are connected directly to the GPIO port. It allows to measure the binary state of the contact sensor.

4 Control System

The control strategy used on the Messor II robot is Sense-Plan-Act. The sequence of consecutive actions is presented in Fig. 4. At the beginning the robot uses its exteroceptive sensor to acquire information about the environment. The output from each exteroceptive sensor, which can be used on the robot, is a point cloud. The data are filtered and used to update the model of the terrain. The Messor II robot uses elevation map as a model of the surroundings [25]. The elevation map is used then to plan full body motion of the robot. We use RRT-based

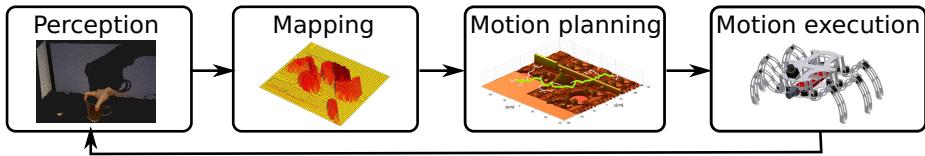


Fig. 4. Sequence of control procedures

planner for this purpose [8]. Finally, the robot executes the planned path. If the PrimeSense sensor is used the robot makes measurements after few steps. If the 2D active laser scanner is used the robot makes scans and updates the map during path execution at each point of the executed trajectory.

4.1 Software Architecture

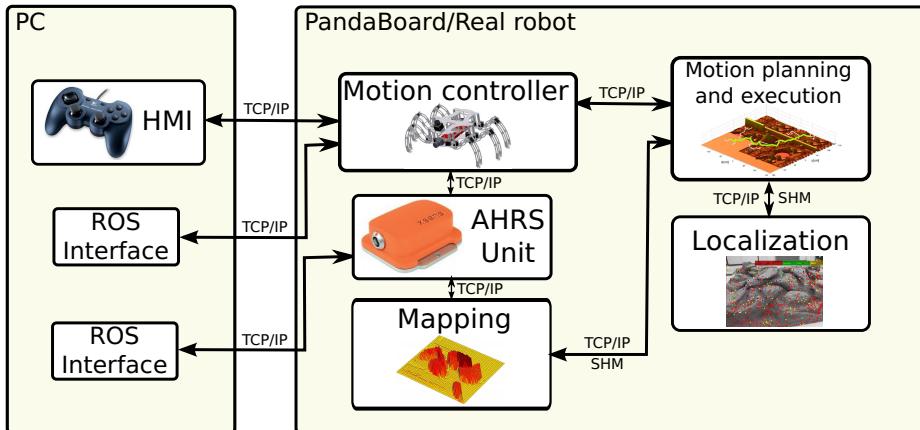


Fig. 5. Software architecture of the Messor II robot

The software for Messor II is written in C++. It is divided into several modules. The communication between modules is mainly implemented using TCP/IP protocol. Despite the fact that this type of communication is slow it allows to run various modules on various computers. This property is very useful during software development. However, we are going to compile each module as static library with standard C++ interface. The aim is to make communication faster on the same computer. In the case when the modules exchange big number of data (maps or point clouds) the modules use SHared Memory (SHM). This method is very fast but can be used to exchange data between modules which operate on the same physical platform.

The motion controller is responsible for communication with the devices. It sends proper orders to servomotors and receives information about joint's state (current position, speed and load). The module operates as a TCP/IP server. It waits for motion commands. It accepts low-level orders like motion of the single joint. It is also possible to send high-level orders regarding the motion of the robot's body or the type of gait. In this case motion controller computes all the reference values for the servomotors using kinematic model of the robot. Then, the controller sends the reference values to obtain proper motion trajectory.

The separate module is written for the AHRS unit. The server waits for a request from the client application. When the order is received it sends request to the sensor and re-send the response to the client application (mapping and motion control modules).

The localization module uses PTAM algorithm [10]. It provides information about pose of the robot (only x , y and z is used), uncertainty of the estimated pose and the set of key-points used for localization [26]. The pose of the robot is utilized by motion planning algorithm to correct elevation map.

The robot can also communicate with a teleoperation station. The human operator can use gamepad to control the motion of the robot. The computer can also run modules which interface robot controller and Robot Operating System. It gives an access to enormous number of libraries available on ROS.

5 Conclusions

In this paper the new improved version of the Messor robot was presented. The key features of the new design are: better *power to mass* ratio and flexible development of the control system due to the minimal electronic design. In the current work not only design process was shown but also the actual real robot.

The development of the research platform is the beginning of the possible work on legged locomotion. As a future work the study on the compliant locomotion is envisioned. Additionally the improvements of the Sense-Plan-Act control strategy are required i.e. new sensory data acquisition and motion planning algorithms. The authors are hoping to obtain autonomous robot locomotion in the outdoor scenario.

References

- [1] Raibert, M.H.: Legged robots that balance. Massachusetts Institute of Technology, Cambridge (1986)
- [2] Walas, K., Belter, D., Kasiński, A.: Control and environment sensing system for a six-legged robot. Journal of Automation, Mobile Robotics & Intelligent Systems 2, 26–31 (2008)
- [3] Belter, D., Skrzypczynski, P.: A biologically inspired approach to feasible gait learning for a hexapod robot. Applied Mathematics and Computer Science 20, 69–84 (2010)

- [4] Walas, K., Belter, D.: Messor – Versatile Walking Robot for Search and Rescue Missions. *Journal of Automation, Mobile Robotics & Intelligent Systems* 5, 28–34 (2011)
- [5] Walas, K., Belter, D.: Supporting locomotive functions of a six-legged walking robot. *Int. J. Appl. Math. Comput. Sci.* 21, 363–377 (2011)
- [6] Labecki, P., Walas, K., Kasinski, A.: Autonomous stair climbing with multisensor feedback. In: Proc. of the 18th World Congress, The International Federation of Automatic Control, Milano, Italy, pp. 8159–8164 (2011)
- [7] Walas, K., Kasinski, A.J.: Discrete event controller for urban obstacles negotiation with walking robot. In: IEEE IROS, pp. 181–186 (2012)
- [8] Belter, D., Skrzypczynski, P.: Posture optimization strategy for a statically stable robot traversing rough terrain. In: IEEE IROS, pp. 2204–2209 (2012)
- [9] Belter, D., Skrzypczynski, P.: Rough terrain mapping and classification for foothold selection in a walking robot. *J. Field Robotics* 28, 497–528 (2011)
- [10] Belter, D., Skrzypczynski, P.: Precise self-localization of a walking robot on rough terrain using parallel tracking and mapping. *Industrial Robot: An International Journal* 40, 229–237 (2013)
- [11] Walas, K.: Terrain Classification Using Vision, Depth and Tactile Perception. In: RSS Workshop RGB-D: Advanced Reasoning with Depth Cameras (2013); archived on the website of the workshop
- [12] Song, S.M., Waldron, K.: *Machines that Walk*. MIT Press, Cambridge (1989)
- [13] Krotkov, E., Bares, J., Kanade, T., Mitchell, T., Simmons, R., Whittaker, W.: Ambler: a six-legged planetary rover. In: International Conference on Advanced Robotics, Robots in Unstructured Environments (ICAR 1991), Pisa, Italy, pp. 712–722 (1991)
- [14] Spenko, M., Haynes, G.C., Saunders, J.A., Cutkosky, M.R., Rizzi, A.A., Full, R.J., Koditschek, D.E.: Biologically inspired climbing with a hexapedal robot. *J. Field Robotics* 25, 223–242 (2008)
- [15] Raibert, M., Blankespoor, K., Nelson, G., Playter, R.: Team, t.B.D.: Bigdog, the rough-terrain quadruped robot. In: Proc. of the 17th World Congress, The International Federation of Automatic Control, Seoul, Korea, pp. 10822–10825 (2008)
- [16] Kalakrishnan, M., Buchli, J., Pastor, P., Mistry, M., Schaal, S.: Fast, robust quadruped locomotion over challenging terrain. In: IEEE ICRA, pp. 2665–2670 (2010)
- [17] Barasuol, V., Buchli, J., Semini, C., Frigerio, M., de Pieri, E.R., Caldwell, D.G.: A reactive controller framework for quadrupedal locomotion on challenging terrain. In: IEEE ICRA, pp. 2554–2561 (2013)
- [18] Hutter, M., Gehring, C., Bloesch, M., Hoepflinger, M.A., Remy, C.D., Siegwart, R.: StarlETH: A compliant quadrupedal robot for fast, efficient, and versatile locomotion. In: Azad, A.K.M., Cowan, N.J., Tokhi, M.O., Virk, G.S. (eds.) *Proceedings of the 15th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, pp. 483–490. World Scientific, Singapore (2012)
- [19] Rönnau, A., Heppner, G., Pfotzer, L., Dillmann, R.: LAURON V: Optimized Leg Configuration for the Design of a Bio-Inspired Walking Robot. In: Waldron, K.J., Tokhi, M.O., Virk, G.S. (eds.) *Nature-Inspired Mobile Robotics, Proceedings of the 16th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, pp. 563–570. World Scientific, Singapore (2013)

- [20] Bartsch, S., Birnschein, T., Römmermann, M., Hilljegerdes, J., Kühn, D., Kirchner, F.: Development of the six-legged walking and climbing robot Space Climber. *J. Field Robotics* 29, 506–532 (2012)
- [21] Schilling, M., Paskarbeit, J., Schmitz, J., Schneider, A., Cruse, H.: Grounding an internal body model of a hexapod walker control of curve walking in a biologically inspired robot. In: IEEE IROS, pp. 2762–2768 (2012)
- [22] Gorner, M., Wimbock, T., Baumann, A., Fuchs, M., Bahls, T., Grebenstein, M., Borst, C., Butterfass, J., Hirzinger, G.: The DLR-Crawler: A testbed for actively compliant hexapod walking based on the fingers of DLR-Hand II. In: Proc. IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS 2008), Nice, France, pp. 1525–1531 (2008)
- [23] de Santos, P.G., Cobanoa, J., Garcia, E., Estremera, J., Armada, M.: A six-legged robot-based system for humanitarian demining missions. *Mechatronics* 17, 417–430 (2007)
- [24] Kenzo, N., Qing-Jiu, H.: Humanitarian mine detecting six-legged walking robot and hybrid neuro walking control with position/force control. *Mechatronics* 13, 773–790 (2003)
- [25] Belter, D., Labecki, P., Skrzypczynski, P.: Estimating Terrain Elevation Maps from Sparse and Uncertain Multi-Sensor Data. In: IEEE 2012 International Conference on Robotics and Biomimetics, pp. 715–722. IEEE (2012)
- [26] Georg, K., David, M.: Parallel Tracking and Mapping for Small AR Workspaces. In: Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2007), Nara, Japan (2007)

Towards Practical Implementation of an Artificial Force Method for Control of the Mobile Platform Rex

Mateusz Cholewiński, Krzysztof Arent, and Alicja Mazur

Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology
ul. Janiszewskiego 11/17
50-372 Wrocław

{mateusz.cholewinski,krzysztof.arent,alicja.mazur}@pwr.wroc.pl
<http://www.iiar.pwr.wroc.pl>

Abstract. The article describes skid steering mobile platform, which is a nonholonomic robot. For controlling such a robot artificial force method has been used. This method assumes that the number of control inputs is the same as the number of controlled variables. Simultaneously additional control signal is equal to zero equivalently. Considerable attention is paid to possibility of practical application of the described algorithm in the physical equipment.

Keywords: mobile manipulator, nonholonomic constraint, drive unit dynamics.

1 Introduction

In this paper a trajectory tracking problem for special type of mobile platform moving with slippage effect, namely for skid-steering mobile platform (SSMP), is considered. This work is a step towards development of substantially new low-level control strategy suitable for mobile manipulator designated to exploration and rescue. Such a robot is a subject of interest in the RobREx¹ project, which aims to increase the autonomy of this class of robots currently manufactured in Poland.

Objects utilizing slide phenomenon are difficult to describe in mathematical terms, nevertheless, they appeal interest of many researchers, see Caracciolo et al. in [1], Kozłowski and Pazderski [2], Mazur and Cholewiński [5] etc. Several control strategies has been proposed but they strongly rely on assumptions made with respect to slippage as well as to design of a mobile platform. In particular, longitudinal slippage is excluded, each wheel contact with the ground at one point, wheels are loaded evenly. The obtained results are of a theoretical nature. The control system behaviour was evaluated in simulations [1], [2], [5] or in laboratory environment [1].

¹ Funded by the Polish National Centre for Research and Development.

This article develops further the ideas from [5], where an *artificial force method* for control of a mobile platform was proposed. We are interested in these aspects that are important in the context of future implementation. But it has to be stressed that we do not aim to cover all the implementation issues here. This is a larger, multistage process with several phases consisting of design, implementation and evaluation. We rather try to figure out what are the consequences of implementing of a control strategy, based on the artificial force method, regarding drive units in the mobile platform Rex, which is now under design and is intended to serve as a tested at a Wrocław University of Technology. More precise, we would like to present a new version of a trajectory tracking control algorithm based on artificial force method for a mobile platform and next evaluate it by simulations. The evaluation criteria are twofold. Through analysis of a tracking error we evaluate a control system performance. Simultaneously, to assess implementation aspects, we analyse two types of input signals: indirect (torques on wheels) and direct (currents in the windings of motors).

The article is organised as follows. In Section 2, the concept of artificial force method is outlined. Next, in Section 3 a mathematical model for the mobile platform Rex is presented. Further, in Section 4, the model of a mobile platform is complemented by a formal model of drive units. In Section 5 the control problem is formulated and a control strategy is developed. The results of simulation studies are presented in Section 7. Finally, the conclusions are provided in Section 8.

2 Artificial Force Method

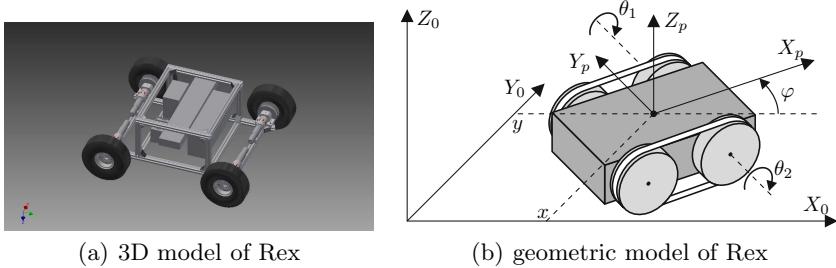
Skid steering mobile platforms (SSMP) are nonholonomic robots, where the lateral slippage is needed for changing the orientation of the platform. Thus, only constraints for the longitudinal slippage can be applied. It means that the number of controls is smaller than the number of state variables. Hence, the system is underactuated.

There are two solutions for dealing with this problem:

- adding another nonholonomic constraint, which would couple the uncontrolled variable ϕ with control inputs,
- adding another control input.

The first solution has been presented in several papers [1],[2], where authors proposed adding artificial constraint, frequently combining the platform's angular with linear speed using the constant value. However, adding artificial has not been proofed to be a good solution of this problem due to problems with describing such a constraint.

In this paper, the second solution is proposed. It assumes, that set of controls is increased with missing one, which identically equal to 0.

**Fig. 1.** Mobile platform REX

3 Mathematical Model of Platform Rex

3.1 Generalised Coordinates

Visualization of mobile platform Rex has been presented in the Figure 1a. This robot consists of one frame with four fixed, independently controlled, wheels, internal devices and the other elements e.g. batteries, motor controllers. Motors on the each side are coupled, what means that their torques can be summed.

To derive a formal model we use a geometric model of Rex, shown in Figure 1b. There are four parameters necessary to characterize this model: a , b , c , r . The values of these parameters can be derived from a 3D CAD model of a robot. The a and b dimensions are the distances from the mass centre to, accordingly, the front and back axle. Half of platform width is c and r is the radius of wheel. In the global coordinates, such a system can be described with the following state vector

$$q = (x \ y \ \phi \ \theta_1 \ \theta_2)^T \in R^5. \quad (1)$$

With such a state vector, kinematic constraints can be derived. For skid steering mobile platform, kinematic constraints consist only of two conditions for lack of a longitudinal slippage (each for one side) - lateral slippage is needed for changing the orientation. Thus, kinematics of such a mobile platform can be described in the Pfaffian form:

$$A(q)\dot{q} = \begin{bmatrix} \cos \phi & \sin \phi & -c & -r & 0 \\ \cos \phi & \sin \phi & c & 0 & -r \end{bmatrix} \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{pmatrix} = 0. \quad (2)$$

Dynamics can be derived form the Lagrange equations for coordinates (1). Rex platform is travelling on the equipotential plane, thus the potential energy is equal to zero ($E_p = 0$) and the Lagrangian is equal to the kinetic energy of the whole robot, which is composed of kinetic energy of the platform and its wheels. For nonholonomic robots, dynamics take the form

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + F(q, \dot{q}) = B(q)u + A^T(q)\lambda, \quad (3)$$

where elements of matrix equation (3) are equal to

$$M(q) = \begin{bmatrix} m_p & 0 & 0 & 0 & 0 \\ 0 & m_p & 0 & 0 & 0 \\ 0 & 0 & J_p & 0 & 0 \\ 0 & 0 & 0 & J_k & 0 \\ 0 & 0 & 0 & 0 & J_k \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (4)$$

Symbol m_p notes the platform mass (with the wheels), J_p is the moment of inertia of the platform, counted accordingly to Z axis, coming through the platform's centre of mass. J_k is the moment of inertia of wheel, counted accordingly to local X axis of the wheel. Inertia matrix is constant, thus the Coriolis matrix equals to zero

$$C(q, \dot{q}) = 0. \quad (5)$$

Friction and reactions forces from the ground can be presented in the form of vector F

$$F(q, \dot{q}) = [F_x \cos \phi - F_y \sin \phi \quad F_x \sin \phi + F_y \cos \phi \quad M_r \quad 0 \quad 0]^T, \quad (6)$$

where, F_x is the longitudinal resistive force, F_y is the lateral force acting on wheels and M_r is resistive moment of platform, described in the same force as in the [1].

3.2 Model in Auxiliary Coordinates

In equation (3), $A^T \lambda$ represents forces connected with nonholonomic constraints. Matrix A was presented before, in equation (2), λ represents Lagrange multipliers, which are very hard to compute. Therefore, dynamics of mobile platform Rex can be expressed in auxiliary coordinates, based on fact that the kernel of A matrix is spanned by the columns of G matrix

$$A(q) G(q) = 0.$$

Matrix G is an element of new kinematics, described in the form of driftless control system

$$\dot{q} = G(q)\eta,$$

Method of artificial force consist of extending original G matrix with additional, linearly independent column.

$$\dot{q} = G_e(q)\eta = \left[\begin{array}{c|c} \cos \phi & \cos \phi \\ \sin \phi & \sin \phi \\ \hline \frac{1}{c} & -\frac{1}{c} \\ 0 & \frac{2}{r} \\ \hline \frac{2}{r} & 0 \end{array} \right] \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_{3v} \end{pmatrix}. \quad (7)$$

Variables $\eta_{1,2,3v}$ are new coordinates of the model so-called auxiliary velocities. Model of mobile platform in auxiliary velocities can be computed in the following way [4]

$$\underbrace{G_e^T M G_e}_{M^*} \dot{\eta} + \underbrace{G_e^T (M \dot{G}_e + C G_e)}_{C^*} \dot{\eta} + \underbrace{G_e^T F}_{F^*} = \underbrace{G_e^T B}_{B^*} u. \quad (8)$$

Relevant elements of dynamics (8), have the following form

$$M^*(q) = \begin{bmatrix} m_p + \frac{J_p}{c^2} + 4\frac{J_k}{r^2} & m_p - \frac{J_p}{c^2} & 0 \\ m_p - \frac{J_p}{c^2} & m_p + \frac{J_p}{c^2} + 4\frac{J_k}{r^2} & 0 \\ 0 & 0 & m_p \end{bmatrix}, \quad (9)$$

$$C^*(q) = \begin{bmatrix} 0 & 0 & -m_p \dot{\phi} \\ 0 & 0 & -m_p \dot{\phi} \\ m_p \dot{\phi} & m_p \dot{\phi} & 0 \end{bmatrix}, \quad (10)$$

$$F^*(q, \dot{q}) = \begin{bmatrix} -Fx - \frac{M_r}{c} & -Fx + \frac{M_r}{c} & -F_y \end{bmatrix}^T. \quad (11)$$

Input matrix B^* and its inverse have the following forms

$$B^* = \begin{bmatrix} \frac{2}{r} & 0 & 0 \\ 0 & \frac{2}{r} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (B^*)^{-1} = \begin{bmatrix} \frac{r}{2} & 0 & 0 \\ 0 & \frac{r}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (12)$$

Variable η_{3v} is an artificial signal occurring as a result of adding third element to the control vector. Variable η_{3v} is used for generation of a reference signal, which would be presented in the section 6.2.

4 Drive Unit Model

Previous discussion was not addressed to a particular mobile platform. In fact, the equations (1) - (11) are valid for a wide class of skid-steering mobile robots. However, the input variable u in (3), that has a physical sense of torque, does not coincide with the real input variable. Usually, it is a current I (in a rotor winding) that is a subject of direct manipulation by a physical robot controller. The relationship between I in individual engines and u on individual wheels strongly depends on a robot design, in particular on the structure of a drive system.

The design requirements, with respect to mobile platform Rex, include support for low level model-based control. Because of them the design is very simple, in particular the axis of a wheel and the axis of an engine, driving this wheel, are overlapping.

As a result of a careful 3D CAD model analysis, carried out in [6], we have that

$$\begin{aligned} u = & \eta_G i k_M I - (\eta_G i^2 \frac{60}{2\pi} J_R + J_C) \frac{d\omega}{dt} \\ & - \eta_G i^2 \frac{60}{2\pi} \frac{k_M I_0 - F_c}{n_0} \omega - f_0 (v_0 i \frac{60}{2\pi})^{2/3} d_m^3 (\omega)^{2/3} \\ & - (f_1 d_m F + \eta_G i F_c) \text{sgn}(\omega) - \eta_G i \delta_R - \delta_G - \delta_B \end{aligned} \quad (13)$$

where u is the control input vector (control torque) from the equation (8), I is the motor's current, which generates electromotive force. Other symbols and definitions are presented in Table 1.

Table 1. Parameters of drive unit

Symbol	Definition	Symbol	Definition
η_G	gear's efficiency	i	transmission ratio
k_M	torque constant	J_R	shaft's moment of inertia
J_C	coupling's moment of inertia	ω	angular velocity of given wheel
F	bearings load	F_c	breaking moment
I_0	no load current	n_0	no load speed
f_0	bearing's friction coefficient	f_1	bearing's friction coefficient
v_0	oil kinematic viscosity	d_m	pitch diameter
δ_R	unknown motor's dynamics	δ_G	unknown gear's dynamics
δ_B	unknown bearing's dynamics		
I_{nom}	nominal current (max. continuous)	n_{nom}	nominal load speed
k_{M_n}	speed constant	$k_{M_{n\tau}}$	speed torque gradient

5 Statement of Control Problems

Nonholonomic systems in auxiliary coordinates have the cascade structure. Hence, in this paper, control system is divided into two subsystems:

- kinematic controller $\eta_r(t)$ – represents an embedded control input, which ensures the realizability of the trajectory tracking for the nonholonomic constraints. Kinematic controller can be treated as a solution to the kinematics, if the dynamics were not present. Such the controller generates a 'velocity profile' which can be executed in practice.
- dynamic controller u – as a consequence of cascaded structure of the model, the system's auxiliary velocities η cannot be commanded directly, as it is assumed in the design of kinematic control, and instead they must be realized as the output of the dynamics driven by u . The dynamic input u intends to regulate the real velocities η toward the reference control η_r , and therefore, attempts to provide control input necessary to achieve the desired task.

In the given case, SSMP should satisfy all the following goals, simultaneously:

1. find a control law u such, that the SSMP platform follows the desired trajectory, without longitudinal slippage,
2. value of control signal u_{3v} should be equal to zero,
3. needed motor torques, currents and rotations should not exceed the maximal admissible values (data obtained from **Maxon** guide [3] of a real engine drive unit).

6 Control

6.1 Kinematic Controller

Kinematics of given SSMP can be approximated using the model of monicycle kinematics, thus, Samson controller [8] could be used. Between (v, ω) and (η_1, η_2) following relationships holds

$$v = \eta_1 + \eta_2, \quad w = \frac{1}{c} (\eta_2 - \eta_1). \quad (14)$$

Then, the reference linear and angular velocities are determined

$$\begin{pmatrix} v_r \\ \omega_r \end{pmatrix} = \begin{pmatrix} k_1 x_e + v_d \cos \phi_e \\ \omega_d + k_2 \phi_e + v_d y_e \frac{\sin \phi_e}{\phi_e} \end{pmatrix}, \quad k_1, k_2 > 0, \quad (15)$$

with appropriate errors, defined as follows

$$q_e = \begin{pmatrix} x_e \\ y_e \\ \phi_e \end{pmatrix} = Rot(z, -\phi) \begin{pmatrix} x_d - x \\ y_d - y \\ \phi_d - \phi \end{pmatrix}. \quad (16)$$

Equations (14) - (16) determine the reference velocities for wheels, which are used by the dynamic controller

$$\eta_{r_1} = \frac{v_r - c\omega_r}{2}, \quad \eta_{r_2} = \frac{v_r + c\omega_r}{2}.$$

6.2 Dynamic Controller

For the fully known dynamics of SSMP, following control law can be applied

$$\begin{pmatrix} u_1 \\ u_2 \\ u_{3v} \end{pmatrix} = (B)^{* - 1} \left\{ M^* \begin{pmatrix} \dot{\eta}_{r_1} \\ \dot{\eta}_{r_2} \\ \dot{\eta}_{r_{3v}} \end{pmatrix} + C^* \begin{pmatrix} \eta_{r_1} \\ \eta_{r_2} \\ \eta_{r_{3v}} \end{pmatrix} + F^* - K_d \begin{pmatrix} e_{\eta_1} \\ e_{\eta_2} \\ e_{\eta_{3v}} \end{pmatrix} \right\}. \quad (17)$$

Relevant elements were defined as follows

$$e_{\eta_i} = \eta_i - \eta_{r_i}, \quad K_d = K_d^T > 0.$$

After extending vector of control inputs with additional virtual component, model of dynamics must be extended with third, virtual velocity signal as well. Only two reference velocity signals, η_{r_1} and η_{r_2} , can be obtained from the kinematic controller, therefore the third one must be generated in a special way. Having

$$u_{3v} = 0,$$

the analytic form of $\eta_{r_{3v}}$ signal can be computed from

$$\begin{aligned} u_{3v} &= (B^{*-1})_{3row} \{M^* \dot{\eta}_r + C^* \eta_r + F^* - K_d e_\eta\} = \\ &= M_{11}^* \dot{\eta}_{r_1} + M_{23}^* \dot{\eta}_{r_2} + M_{33}^* \dot{\eta}_{r_3} \\ &\quad + C_{31}^* \eta_{r_1} + C_{32}^* \eta_{r_2} + C_{33}^* \eta_{r_3} + F_3^* - K_d (\eta_3 - \eta_{r_3}) \equiv 0, \end{aligned} \quad (18)$$

After some transformations and taking into account fact that relevant elements have the following values

$$M_{13}^* = M_{23}^* = 0, \quad M_{33}^* = m_p, \quad C_{31}^* = C_{32}^* = m_p \dot{\phi}, \quad C_{33}^* = 0,$$

signal $\dot{\eta}_{r_3}$ can be obtained in the form

$$\dot{\eta}_{r_3} = -\frac{(m_p \dot{\phi}(\eta_{r_1} + \eta_{r_2}) - F_y - K_d(\eta_3 - \eta_{r_3}))}{m_p}.$$

This signal has to be integrated to obtain the numeric form of η_{r_3} signal, which is then used in the control algorithm.

7 Simulation Studies

This section presents results of a simulation analysis for a control system discussed in the sections 3, 4, 6. The main goal is twofold. Firstly, we study system performance and system stability. Secondly, we derive a number of parameter values that are essential from the perspective of drive unit selection. In this part we have in mind basic principles of mechatronic design concisely outlined in [7].

Desired trajectory of SSMP platform is a square, with a side length of 11 meters, which should be realized in 40 seconds (every 10 seconds per every side, linear desired velocity equals $1.1 \frac{m}{s}$). Parameters of object are presented in the Table 2, while control parameter are presented in Table 3. Parameters of drive unit model were taken from Maxon Motors data sheets [3] for a motor no. 370355 and a gear no. 223090.

REX platform is designed to be used with manipulator. For the design purposes it has been assumed that the robot already carries the manipulator - its mass is added to the mass of whole robot and moment of inertia of whole platform is increased too.

Parameters m_p and m_k are respectively mass of whole platform and mass of one wheel. Similarly, J_p and J_k are moment of inertia of platform and wheel. Parameter r is the radius of wheel, coefficients μ and f_r are, respectively, lateral friction coefficient and coefficient of rolling resistance.

Table 2. Values of robot and drive unit parameters

Parameter	Value	Parameter	Value
a	0.365 m	b	0.365 m
c	0.387 m		
m_p	42 kg	m_k	2.38 kg
J_p	5 kgm^2	J_k	0.15 kgm^2
r	0.127 m		
μ	0.9	f_r	0.1
η_G	75 %	i	53
k_M	$60.410^{-3} \frac{\text{Nm}}{\text{A}}$	J_R	5.610^{-5}kgm^2
J_C	0.1510^{-3}kgm^2	ω	actual value taken from object
F	30 N	F_c	6.410^{-4}Nm
I_0	0.147 A	n_0	5680 rpm
f_0	0.1	f_1	0.1
v_0	$22.810^{-6} \frac{\text{m}^2}{\text{s}}$	d_m	0.1 mm
δ_R	0.1	δ_G	0.1
δ_B	0.1		
I_{nom}	7.07 A	n_{nom}	5420 rpm
k_{M_n}	$158 \frac{\text{rpm}}{\text{V}}$	$k_{M_{n\tau}}$	$0.638 \frac{\text{rpm}}{\text{mNm}}$

Parameters v_d and ω_d are desired linear and angular velocities, k_1 , k_2 are regulation parameters of kinematic controller, used in equation (15) and K_d is also used in equation (17).

Table 3. Regulation parameters

Parameter	Value	Parameter	Value
v_d	$1.1 \frac{\text{m}}{\text{s}}$	ω_d	$0.1 \frac{\text{rad}}{\text{s}}$
k_1	0.2	k_2	1
K_d	500		

In the Figure 2(a) tracking of the desired trajectory has been presented. In the Figure 2(b) tracking errors of cartesian coordinates have been presented. It can be seen, that the biggest problem for mobile platform is to realise the corner of the square. It is associated with the characteristics of kinematic controller and the occurrence of the lateral slippage. However, the errors are minimizing and convergence finally to zero.

In the Figures 2(c) and 2(d) motor torques and current values has been presented. There are limit values marked with dotted lines. Both, the motor torques and the current values do not exceed those limits.

Figures 2(e) and 2(f) present the motor shaft rotations per minute and the linear velocity of the platform. It has been showed that in one moment, the rotations of motor exceed a little limit values and the linear velocity changes when there is an orientation change. However, physical limitation of motor's rotations can be implemented, which would affect the position errors but will not cause the failure of the control algorithm operation.

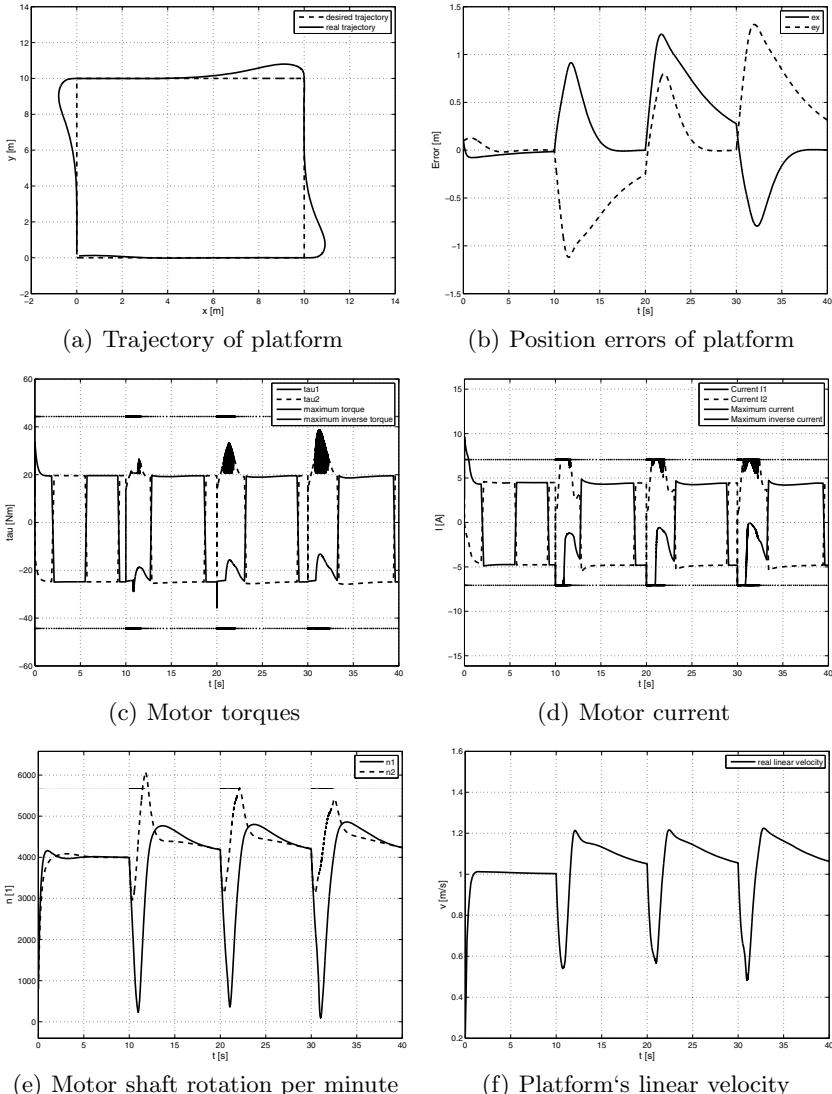


Fig. 2. Simulation figures

On the basis of obtained results we can conclude that the assembly of a motor no. 370355 and a gear no. 223090 is acceptable for the platform REX. In fact, it is the optimal choice from the set of DC motors offered by The Maxon Motor.

8 Conclusions

The article presents simulation results of a control system consisting of a mobile platform, that is close to Rex with respect to geometric and dynamic parameters, and a controller with implemented a low-level trajectory tracking control strategy, based on an artificial force method. The simulation results show, that if we accept a small reduction of the quality of a control performance then in the case of two drive units (composed of a motor and a gear) intended for Rex, practical implementation of the control strategy is feasible. Although the reference trajectory is demanding, the tracking position error is acceptable, motor torques, currents and shaft rotation per minute remain within required limits, with a small safety margin.

References

1. Caracciolo, L., De Luca, A., Iannitti, S.: Trajectory tracking control of a four-wheel differentially driven mobile robot. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, pp. 2632–2638 (1999)
2. Kozłowski, K., Pazderski, D.: Modeling and control of a 4-wheel skid-steering mobile robot. International Journal of Applied Mathematics and Computer Science, 477–496 (2004)
3. Maxon Motor: Program 2013/1014. High Precision Drives and Systems (2013), <http://www.maxonmotor.com>
4. Mazur, A.: Model-based control for nonholonomic mobile manipulators. Oficyna Wydawnicza Politechniki Wrocławskiej (2009) (in Polish)
5. Mazur, A., Cholewiński, M.: Robust control of differentially driven mobile platforms. In: Kozłowski, K. (ed.) Robot Motion and Control 2011. LNCS, vol. 422, pp. 53–64. Springer, London (2011)
6. Mazur, A., Cholewski, M., Arent, K., Malewicz, J., Szrek, J.: Selected Topics on Modelling, Control and Design of a Physical Model of a Robot for Rescue and Exploration RobREx, Technical report SPR No 8/2013 (2013) (in Polish)
7. Mianowski, K.: Selection of Drive Systems for a Light Manipulator Designed to be Mounted on Mobile Platforms. In: Problems of Robotics, vol. 1, pp. 347–356. Oficyna Wydawnicza Politechniki Warszawskiej (2008) (in Polish)
8. Samson, C., Ait-Abderrahim, K.: Feedback control of a nonholonomic wheeled cart in cartesian space. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, pp. 1136–1141 (1991)

Static Modeling of Multisection Soft Continuum Manipulator for Stiff-Flop Project

Jan Fraś, Jan Czarnowski, Mateusz Maciąś, and Jakub Główka

Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland
`{jfras, jczarnowski, mmacias, jglowka}@piap.pl`

Abstract. This paper describes the design and implementation of a static model used for position estimation of a flexible modular medical manipulator equipped with optic-fiber based sensors. Flexible manipulators are emerging technology in medical applications especially in minimally invasive surgery as it allows to perform the operation with tight space constraints without damaging other organs. Such option is often impossible with use of rigid surgical instrument. However one of the technical challenges in implementation of the flexible manipulator is to be able to determine the position of the manipulator during operation. A theoretical model of use of different information derived from optic-fiber based sensors to allow measurement of the position and deformation of the manipulator has been proposed. In comparison to typical constant curvature bending approach, proposed model allow to estimate deformation caused by external force applied to the structure. Simulation test has been carried out to present the advantages and possibilities of use of that model in data fusion algorithms to obtain precise positioning of the manipulator during the operation.

Keywords: Stiff-Flop, surgical manipulator, constant-curvature, continuum manipulator, static modeling, minimal invasive surgery.

1 Introduction

In modern medicine, minimal invasive surgery (MIS) is a surgical operation that is performed by the surgeon through small incisions. It is an established alternative to conventional open surgery [1]. Advantages are such as reduced post-operative pain, blood loss, tissue trauma and recovery time. Additional benefit is less chance of post-operative infection [2]. However, there are also several problems associated with this surgical technique. The surgeon has limited feedback – including visual and haptic. There is also a reduced number of degrees of freedom during the operation available to the surgeon. Use of current available technology result in relatively high chance of instruments damaging other tissues during transit of the MIS instruments. Current research conducted in project like Stiff Flop (EU FP7 founded project) is focused on introducing new flexible robotic manipulators into the MIS. This kind of structures are able to bend in a snake-like way and thanks to that the number of degrees of freedom is increased during MIS [7]. In results majority of the mentioned problems can be reduced or eliminated and surgeon is capable to reach targets which are

inaccessible with conventional rigid surgical instrument. Despite the soft manipulator construction issues the challenge lies also within the accuracy that is required to perform the surgical operation. However there is no one good solution to estimate the localization of the manipulator. To provide accurate and reliable positioning and control of the flexible structure multiple position estimation sources are required [6]. For that purpose several sensors were designed and developed [4]. This paper proposes a model that enable to estimate the shape and sizes of a flexible robotic manipulator module developed within Stiff-Flop project. The design of the STIFF-FLOP soft robot module is presented in Fig. 1. Whole robot is constructed from three similar modules. The end module is equipped with a tool for surgical operations. Each module is built with a braided silicone rubber tube, and current prototype has 2.5 cm in diameter. Inside the tube, there are three chambers [7]. The air or fluid pressure within each of these three chambers determine the bending, elongation and direction in 3D space. This design is inspired from a three-chamber micro actuator first proposed by Suzumori [8]. Different modeling methods have been proposed for continuum robots. A recent review of these methods can be found at [9]. Most of typically applied approaches are based on constant curvature assumption or are using numerical methods that has no direct physical relevance. This has not been thoroughly verified in experiments [10]. Additionally the Stiff-Flop robot is developed in order to interact with the patient. This implies the requirement to model not only robot movement and shape but also the deformation caused by external forces. These effects make its modeling more challenging. The proposed approach takes into account the pressure applied by the control system and forces measured by distributed tactile force sensors. It will be used for data fusion the provide accurate and reliable position (shape and size) of the manipulator for control system and as a main source of feedback information for surgeon.

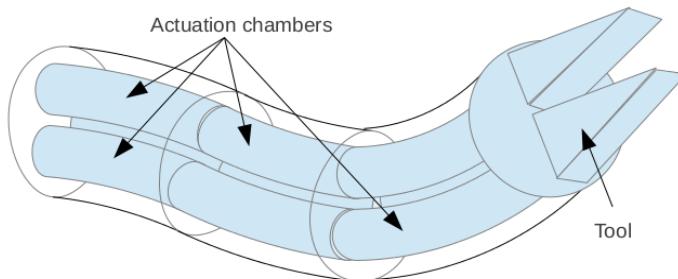


Fig. 1. Visualization of the Stiff-Flop manipulator

2 Construction of the Manipulator

The manipulator consists of three identical modules. Each module is made of silicon material with three pressure actuation chambers. To prevent deformation of the chambers, braiding is applied around the whole module. The module cross section is presented on Fig. 2. Real prototype of two segment manipulator is presented on Fig. 3.

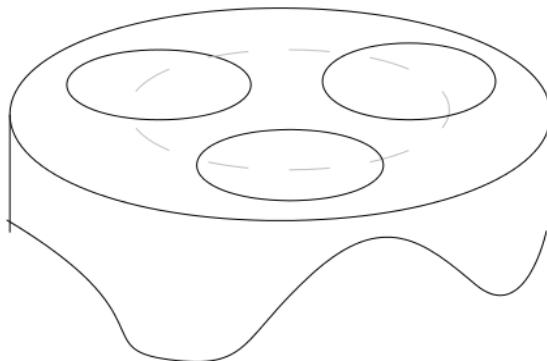


Fig. 2. Cross section of the Stiff-Flop module



Fig. 3. Photograph of a two segment arm manufactured by partner from the Stiff-Flop project

3 Modeling the Arm

In order to determine the shape of the manipulator, a proper model has to be designed and implemented. It has to be precise and its implementation cannot be computationally complex, because the results are to be used in real-time controlling of the arm. Therefore, techniques based on numerical simulation like Finite Element Method (FEM) are not directly applicable, because of their computation overhead. In order to achieve above mentioned goals, a tradeoff between simplifying the model and its precision has to be considered. On one side the precision requirement demands taking any factors that have significant influence on the manipulator shape into account. On the other side, modeling too many physical effects increases the amount of required calculations. Since the Stiff-Flop arm is designed to operate in tight spaces, there is high probability of contact with other bodies. This fact makes calculating the influence of external forces acting on the arm a must. This significant factor is not considered by the popular Constant Curvature model of continuous robots [9], what renders it unusable in this case. For the model described in this paper, following assumptions had been made: the pressures in each chamber and the values of external forces acting

on the module are measured at any point of the module; the segment is made of homogenous material of known stiffness, with three pressure chambers hollowed out; the dimensions of cross-section of the pressure chambers are constant (provided by the braiding); the pressure in chambers is constant at any point. The other influence of the braiding and optical fibers running through the module has been neglected. Nevertheless the influence of the omitted factors on the model is still subject of the research. Since the arm consists of three separately actuated segments (modules), equations for a single segment are presented.

4 Single Segment Model

The segment is controlled by pumping the work fluid in and out of the chambers. This causes the pressure to change its value accordingly. The pressure in each chamber is constant at any point of its volume. A cross section of the module showing the pressure changes is presented on drawing 2. Treating it as internal stress, the force acting at any cross section of the chamber can be described with the formula 1:

$$p = \frac{F}{A} \quad (1)$$

Forces in each cross section are parallel with its Z-axis (perpendicular to the cross section). Therefore, the Z-axis moment is zero. The resulting moment causes pure bending of the module. Because the chamber diameter and pressure is constant throughout the module length, the resulting moment is also constant. Using the force values and cross section geometry, the resulting bending moments can calculated:

$$M_{kp} = \vec{C}_{k1} \times \vec{F}_{k1} + \vec{C}_{k2} \times \vec{F}_{k2} + \vec{C}_{k3} \times \vec{F}_{k3} \quad (2)$$

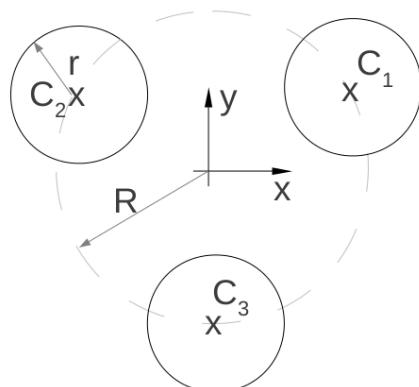


Fig. 4. Cross section of the module at point k, perpendicular to the chamber axis

The Euler-Bernoulli formula relates the bending moment with resulting curvature as follows [12]:

$$\kappa = \frac{1}{\rho} = \frac{M}{EI} \quad (3)$$

where E is the Young's modulus constant of the silicone, and I is the second moment of inertia of the module at cross section k. Assuming no presence of external forces, the bending moment has the same value at every cross section. Therefore, we calculate the total curvature:

$$\forall_k \kappa_k = \kappa_0, \quad \kappa_0 = \frac{M_0}{EI} \quad (4)$$

In this situation, the module shape can be described as a fragment of circle with radius rho. This is presented on Fig. 5.

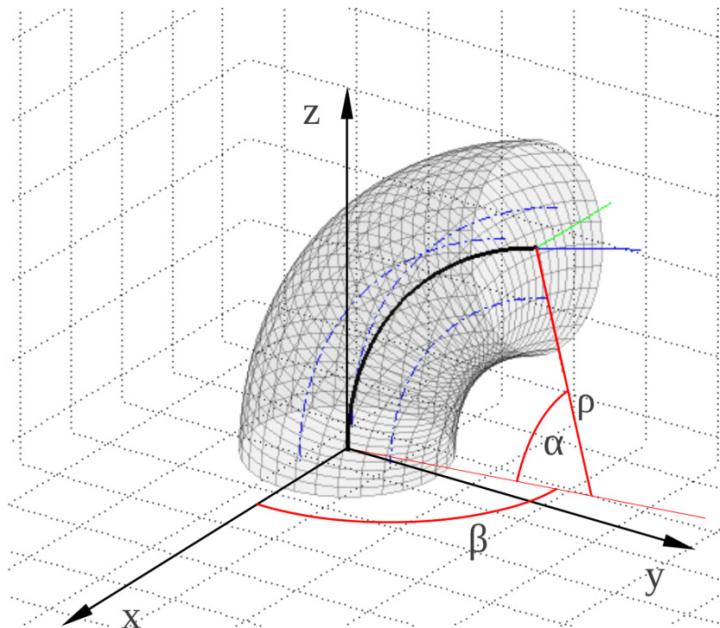


Fig. 5. Module bending with no external forces applied

This is the Constant Curvature case. The position and shape of the module can be described with only three parameters: the bending angle alpha, direction beta and curvature rho. Forces resulting from pressures also influence the module's length. Elongation at any point along the module's axis can be described using the Hooke's Law [12]:

$$\Delta dl = \frac{F_p}{AE} dl \quad (5)$$

where F_p is the overall force resulting from chamber pressures, at single cross section. It is calculated by adding forces resulting from each chamber pressure. A is the area of the silicone part of the cross section (cross section area minus the area of three chambers). Fig. 6 presents a fragment of the module elongated by the chamber pressures.

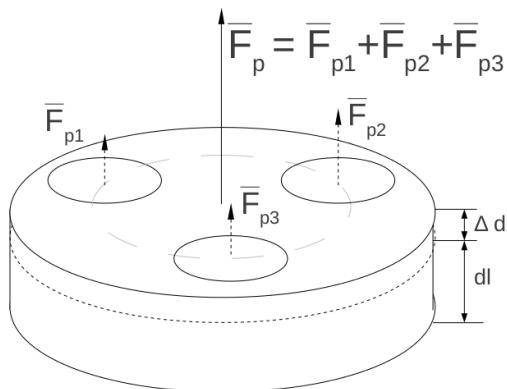


Fig. 6. Elongation of the module at point I

The overall change of length of the module can be calculated by integrating the previous formula from 0 to l_0 :

$$\Delta l = \int_0^{l_0} \Delta dl = \int_0^{l_0} \frac{F_p}{AE} dl \quad (6)$$

The value of the bending angle (α – see Fig. 7) can be evaluated by integrating the curvature from 0 to $l+\Delta l$ w:

$$\alpha = \int_0^{l_0+\Delta l} d\alpha = \int_0^{l_0+\Delta l} \frac{1}{\rho} dl \quad (7)$$



Fig. 7. Bend of the module at point I

In the case of external forces having any non-zero values constant curvature model does not apply, because of variable bending moment.

5 Modeling the Module with External Forces Applied

When taking the influence of external forces acting on the module into account, their values need to be included in previous equations. That determines the value of the bending moment to be variable throughout modules length. It can be generally expressed with the following equation:

$$M_k = M_{kp} + M_{kext} \quad (8)$$

The external forces influence has to also be represented in bending, torsion and elongation calculation. At the current state of the Stiff-Flop project, the force sensor is located on the tip of each module. Therefore, the equations for one force acting on the tip are presented next.

The measurement of tip force is expressed in terms of the tip frame. External force causes additional moment to appear in all module's cross sections. Its value in the tip frame can be calculated with the following formula:

$$\overrightarrow{M}_{kext} = \overrightarrow{F}_{ext} \times \vec{r}_k \quad (9)$$

where F_{ext} is the vector of the external force applied to the tip and r_k is the vector from center point of the k-th cross section to the tip center. Both vectors are expressed in coordinates of the tip frame. This situation is presented on Fig. 8 and Fig. 9.

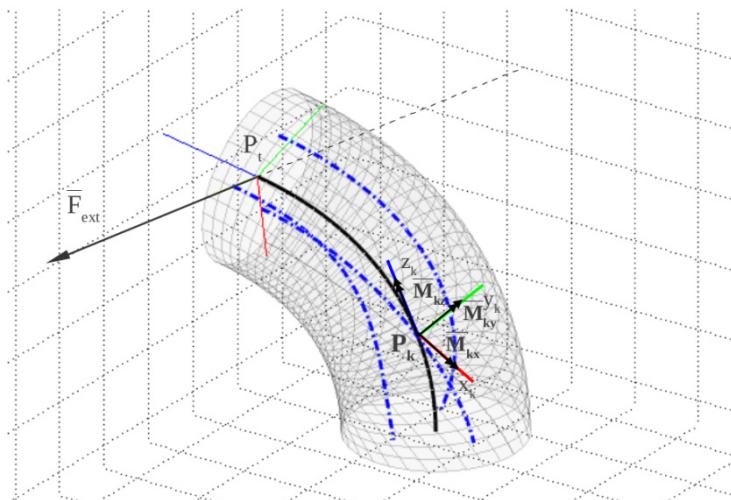


Fig. 8. Moments in cross section k

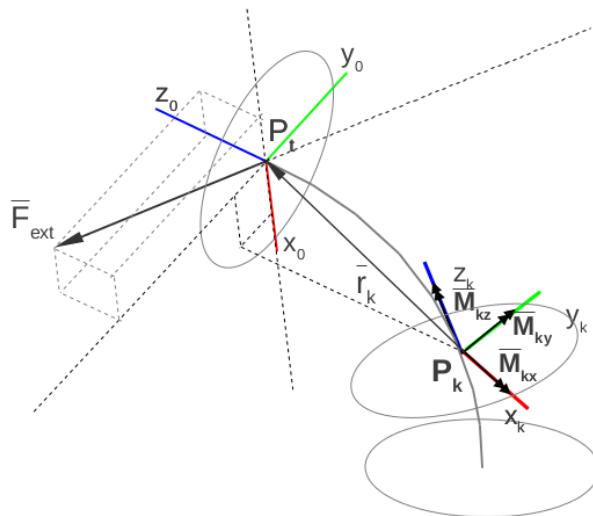


Fig. 9. Coordinate frames and vectors

As one can observe, the force and moment at cross section k depend on its position. Knowing the relative orientation of tip frame in the frame of cross section k and combining it with the force and moment resulting from pressures, one can calculate F_k and M_k – the overall force and moment acting on that cross section. Those values can be then used to determine the elongation and curvature of the module at point P_k (equations 10, 11)

$$\Delta dl_k = \frac{F_k}{AE} dl \quad (10)$$

where

$$F_k = F_p + F_{k ext, z}$$

and

$$\kappa_k = \frac{1}{\rho} = \frac{M_k}{EI} \quad (11)$$

where

$$M_k = M_{kp} + \vec{F}_{ext} \times \vec{r}_k$$

Twist can be calculated with [12]:

$$d \frac{\theta}{dl} = \frac{T}{GI_0} \quad (12)$$

where T is the torque value (the z part of force moment in k-cross-section) and GI_0 represents the torsional rigidity of the module

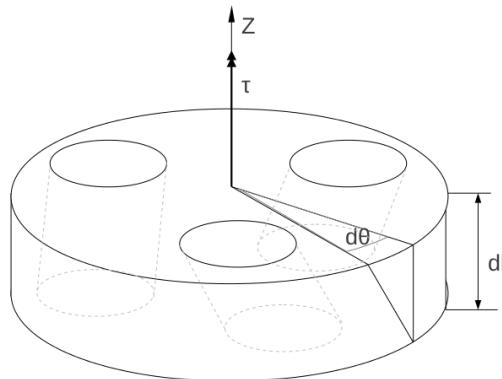


Fig. 10. Twist of the module fragment dl

Shape of the module can be determined by integrating the elongation, twist and bend of every point along the module axis.

6 Implementation

The model described in this paper has been implemented in order to use in the Stiff-Flop project system. For more convenient development, MATLAB had been used. For determining the overall shape of the arm, equations presented earlier are applied to all modules of the manipulator, starting from the last module. Elongation, bending and torsion are calculated in a fixed number of cross sections of the module in the direction from the tip to the base. At each iteration, coordinates (position and orientation) of the current cross-section in terms of the module's tip frame are used to calculate moments and forces. The coordinates of next iteration cross-section frame are calculated using the bending, elongation and twist of the current fragment. When next module is reached, the process is repeated, but the forces and moments from previous module are taken into account. The method of calculation used in the model causes the results to be expressed in terms of the module's tip frame. During the calculations, the resulting values have to be transformed between frames several times. For transformation of vectors between the frames Euler angles are used (Fig. 11).

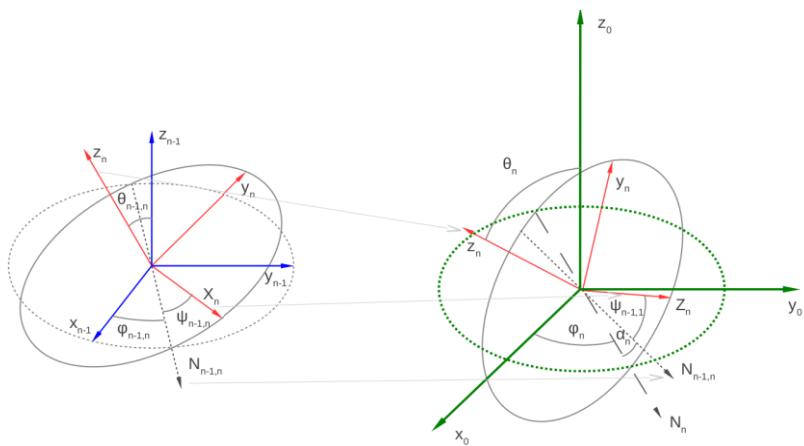


Fig. 11. Euler angles between n and $n-1$ cross section frame; n -th cross section frame shown with the global frame

Results are presented on Fig. 12.

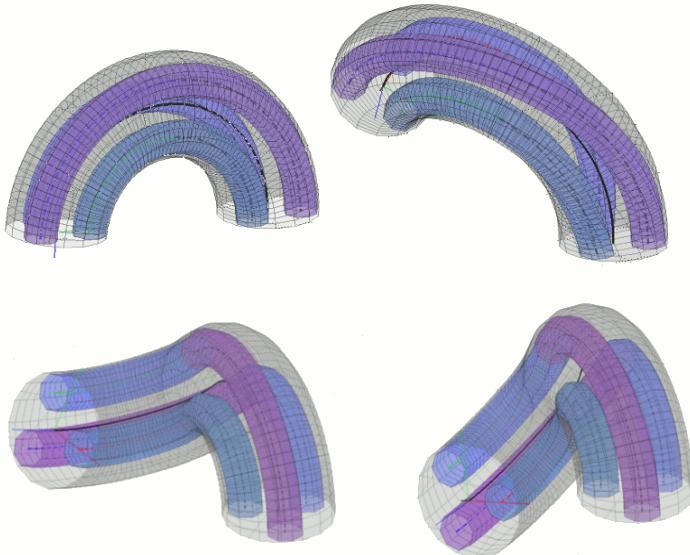


Fig. 12. Visualization of the simulated arm

7 Conclusion

The model described in this paper allows estimation of the continuum robot's shape. It has been successfully used as a simulation platform to work on data fusion in the Stiff-Flop project. Adjusting the number of calculation iterations allows choosing

between high precision and low computation time. More research about the optimal number of iterations for the task of estimating the Stiff-Flop manipulator shape is still to be carried out. The possibility of adjusting the model to describe the dynamic behavior of the arm is being explored.

References

1. Kuo, C.-H., Dai, J.S.: Robotics for Minimally Invasive Surgery: A Historical Review from the Perspective of Kinematics. In: International Symposium on History of Machines and Mechanisms, pp. 337–354. Springer Science+ Business Media (2009)
2. Zbyszewski, D., Challacombe, B., Li, J., Seneviratne, L., Althoefer, K., Dasgupta, P., Murphy, D.: A comparative study between an improved novel air-cushion sensor and a wheeled probe for minimally invasive surgery. *Journal of Endourological Society* 7(24), 1155–1159 (2010)
3. Degani, A., Choset, H., Zubiate, B., Ota, T., Zenati, M.: Highly Articulated Robotic Probe for Minimally Invasive Surgery. In: 30th Annual International IEEE EMBS Conference, Vancouver, British Columbia, Canada (August 2008)
4. Searle, T.C., Althoefer, K., Seneviratne, L., Liu, H.: An Optical Curvature Sensor for Flexible Manipulators. In: 2013 IEEE International Conference on Robotics and Automation, Karlsruhe (2013)
5. Jiang, A., Althoefer, K., Dasgupta, P., Nanayakkara, T.: The Core Snake, the variable stiffness laparoscopic camera. In: The Hamlyn Symposium on Medical Robotics, London, United Kingdom (2013)
6. Rozo, L., Calinon, S., Caldwell, D.G., Jimenez, P., Torras, C.: Learning collaborative impedance based robot behaviors. In: AAAI Conference on Artificial Intelligence, Bellevue, Washington, USA (2013)
7. Cianchetti, M., Ranzani, T., Gerboni, G., De Falco, I., Laschi, C., Menciassi, A.: STIFF-FLOP Surgical Manipulator: mechanical design and experimental characterization of the single module. In: IROS 2013 (2013)
8. Suzumori, K., Iikura, S., Tanaka, H.: Applying a flexible microactuator to robotic mechanisms. *IEEE Control Systems Magazine* 12, 21–27 (1992)
9. Webster III, R.J., Jones, B.A.: Design and kinematics modelling of constant curvature continuum robots: a review. *The International Journal of Robotics Research* 29(13), 1661–1683 (2010)
10. Suzumori, K., Iikura, S., Tanaka, H.: Development of flexible microactuator and its applications to robotic mechanisms. In: IEEE International Conference on Robotics and Automation (1991)
11. Jakubowicz, A., Orłoś, Z.: Wytrzymałość materiałów (Strength of Materials). WNT, Warsaw (1970)
12. Timoshenko, S.: Strength of Materials, Part I, Elementary Theory and Problems, 2nd edn. D. Van Nostrand Company (1940)
13. Webster III, R.J., Jones, B.A.: Design and Kinematic Modeling of Constant Curvature Continuum Robots: A Review. *The International Journal of Robotics Research* 10 (2010)

Estimation of Altitude and Vertical Velocity for Multirotor Aerial Vehicle Using Kalman Filter

Przemysław Gąsior, Stanisław Gardecki, Jarosław Gościński, and Wojciech Giernacki

Poznan University of Technology, Institute of Control and Information Engineering,
pl. M. Skłodowskiej-Curie 5 p.108, 60-965 Poznan, Poland
przemyslaw.m.gasior@student.put.poznan.pl,
{stanislaw.gardecki,wojciech.giernacki}@put.poznan.pl,
jaroslaw.a.goslinski@doctorate.put.poznan.pl

Abstract. Knowledge about precise robot localization is a key ingredient in controlling it, but the task is not trivial without any visual or GPS feedback. In this paper, authors concentrate on estimation of information about the robot's altitude. One of the ways to acquire it, is a barometer. This type of sensor returns atmospheric pressure from which the height above the sea level can be computed. These readings have some disadvantages e.i.: vulnerability to pressure jumps and temperature drift as well as delay on the output. These problems can be solved by using Kalman filter algorithm for estimating altitude and vertical velocity, based not only on barometer readings, but also on accelerometer data. In the paper, derivation of the Kalman equations for the process to estimated are shown. Also improvements of the algorithm are described. The results of tests of this algorithm on real flying robot proved that estimates calculated with this method are precise and noise resistant.

Keywords: Barometer, Accelerometer, IMU, Unmanned Aerial Vehicle UAV, Drone, Multirotor.

1 Introduction

In past few years there has been growing interest in the field of unmanned aerial vehicles. One of the problems is to get precise information about attitude and altitude. There are many ways to acquire attitude, using quaternions [1][2] or Kalman filter [3]. For attitude estimation, accelerometers, gyroscopes and magnetometers are used mostly, but for altitude estimation, only accelerometer and barometer readings are valid. Along with attitude and altitude data availability it is possible to create controllers stabilizing the drone. This part is crucial for correct stabilization, otherwise flying robots cannot be used in applications that they were designed for e.g.: commercial flying cameras, reconnaissance units in rescue missions or even moving inside buildings with implemented SLAM (Simultaneous Localization and Mapping).

Calculating altitude only from the barometer measurements is not accurate enough, therefore authors proposed Kalman filter based on basic physics equations. This filter estimates altitude and vertical velocity, taking two inputs: relative altitude and acceleration in Z axis.

Described algorithm will work only if robot is located parallel to the ground (Z axis is vertical to the ground). Therefore method for tilting of the robot's frame compensation was proposed. Thanks to that, estimation of the altitude was correct in every orientation. At the end of paper an experiment is presented. Results prove correct implementation of the algorithm.

Paper is divided into three main parts. In the first part, the research platform and on-board sensors are presented. Next one, contains description of all main steps of the algorithm with all needed equations. Finally, in the last part, the results of test during flights with the conclusions regarding the algorithm are shown.

2 Research Platform

All measurements and experiments in this paper are done on the research platform built in the Institute of Control and Information Engineering at Poznan University of Technology. The platform is shown in Fig. 1. This is a high quality robot constructed mostly with carbon fiber.

From perspective of this paper, the most important part is the AHRS (Attitude and Heading Reference System) module. It is a part of the on-board avionics, fully designed by researchers working on this construction. It consists of IMU sensor and microcontroller realizing computational algorithms and communicating with main controller board.



Fig. 1. Flaying robot Falcon

2.1 IMU

The main on-board sensor is IMU (Inertial Measurement Unit) from Analog Devices (type: ADIS16488). It is a high quality tactical grade sensor that integrates 3-axis accelerometer, gyroscope, magnetometer and barometer with filtering capabilities, Fig. 2. Based on its measurements, the attitude is computed and sent to main unit where stabilization of the robot is implemented.

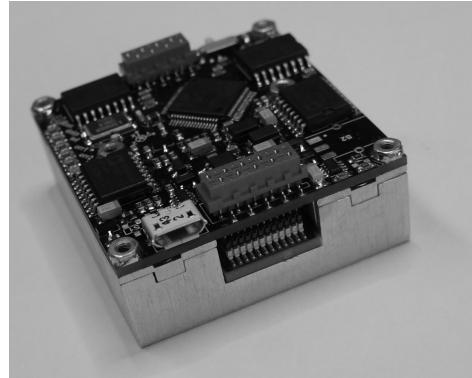


Fig. 2. AHRS module with tactical grade IMU sensor - ADIS16488

3 Description of the Algorithm

In the first part of this section, simple way of altitude computation is presented. This method is based only on barometer readings and returns height above the sea level. In the next part, problems with that type of measurement and a way to calculate the relative altitude from starting position of the robot are presented. In the third part, the simple introduction to the Kalman filter and measurement model is described. Finally, improvements to the algorithm and complete scheme of actions are shown.

3.1 Absolute Height above the Sea Level

Taking [5] into consideration, there is possibility to compute atmospheric pressure at a selected height, Eqn. 1.

$$p = p_0 \left(1 - \frac{L \cdot h}{T_0}\right)^{\frac{gM}{RL}} \quad (1)$$

From that equation the absolute height above the sea level can be derived (2).

$$h = -\left(\frac{p}{p_0}\right)^{\frac{RL}{gM}} - 1 \cdot \frac{T_0}{L} \quad (2)$$

where the coefficients are as follows: p - atmospheric pressure at defined level, p_0 - sea level standard atmospheric pressure, L - temperature lapse rate, T_0 - sea level standard temperature, g - Earth-surface gravitational acceleration, M - molar mass of dry air, R - universal gas constant.

Thanks to that equation, the height of the robot can be computed, though the result is not precise enough and depends mostly on barometer resolution. What is more, the effect of changing of the surrounding temperature is also not taken into consideration.

ADIS16488 has self-heating capability, which stabilizes temperature inside the case at constant level, independently of the outside conditions. Unfortunately, heating process takes very long time to end (approx. 1 hour) and has very significant influence on the barometer readings. Example data with temperature drift during constant altitude is shown in Fig. 3. The error grows up to 7 meters. Therefore authors decided to compensate this effect. Unfortunately height drift

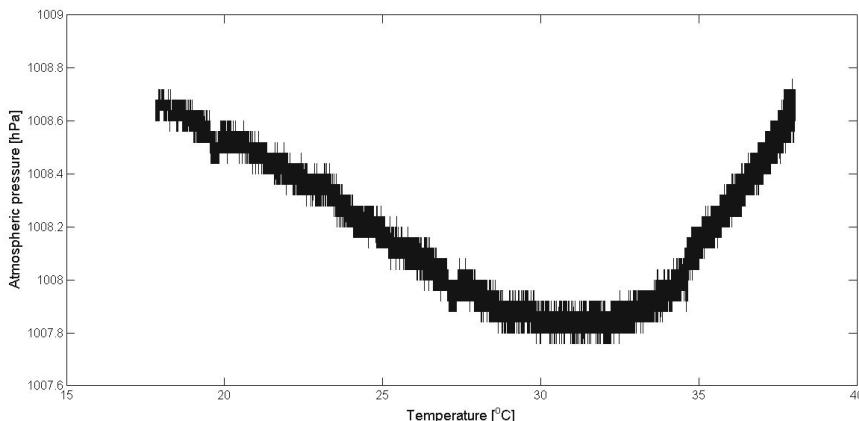


Fig. 3. Barometer temperature drift

depends not only on temperature but also on the weather conditions. In this case compensation was correct but the aim was not achieved.

Based on measurements when robot is still on the ground, the relative altitude during the flight phase can be calculated. This is done by taking one of the first height measurements and subtract it from next ones (3).

$$h_{rel} = h_{abs} - h_0 \quad (3)$$

where h_0 is a starting height and h_{abs} is the result from the (2)

3.2 Kalman Filter

Kalman filter is divided into two main parts: prediction and correction. In first part, the state vector values are computed based on the estimations from last iteration. Also covariance matrix is updated.

$$\hat{\mathbf{x}}(k|k-1) = \mathbf{A}\hat{\mathbf{x}}(k-1|k-1) + \mathbf{B}\hat{\mathbf{u}}(k-1) \quad (4)$$

$$\mathbf{P}(k|k-1) = \mathbf{A}\mathbf{P}(k-1|k-1)\mathbf{A}^T + \mathbf{Q} \quad (5)$$

The next step is an update based on predictions and measurements.

$$\mathbf{K}(k) = \mathbf{P}(k|k-1)\mathbf{C}^T(\mathbf{CP}(k|k-1)\mathbf{C}^T + \mathbf{R})^{-1} \quad (6)$$

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)(\mathbf{z}(k) - \mathbf{C}\hat{\mathbf{x}}(k|k-1)) \quad (7)$$

$$\mathbf{P}(k|k) = (\mathbf{I} - \mathbf{K}(k|k)\mathbf{C})\mathbf{P}(k|k-1) \quad (8)$$

where: $\hat{\mathbf{x}}$ - state vector estimate, \mathbf{z} - measurement vector, \mathbf{A} - process matrix, \mathbf{B} - control matrix, \mathbf{C} - input matrix, \mathbf{P} - covariance matrix, \mathbf{K} - Kalman gain matrix, \mathbf{Q} - process noise covariance matrix, \mathbf{R} - measurement noise covariance matrix [4].

3.3 Measurement Model

Robot equations in axis Z has been created using basic physics knowledge, based on available measurements - relative altitude and acceleration.

$$v_k = v_{k-1} + a_{k-1}T \quad (9)$$

$$h_k = h_{k-1} + v_{k-1}T + \frac{a_{k-1}T^2}{2} \quad (10)$$

Where T is a sample period. Taking altitude, vertical acceleration and vertical velocity as state variables, equations 9 and 10 can be written in a matrix representation:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_k \quad (11)$$

$$\begin{bmatrix} h_k \\ a_k \\ v_k \end{bmatrix} = \begin{bmatrix} 1 & \frac{T^2}{2} & T \\ 0 & 1 & 0 \\ 0 & T & 1 \end{bmatrix} \begin{bmatrix} h_{k-1} \\ a_{k-1} \\ v_{k-1} \end{bmatrix} \quad (12)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}_{k-1} + \mathbf{D}\mathbf{u}_k \quad (13)$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} h_{k-1} \\ a_{k-1} \\ v_{k-1} \end{bmatrix} \quad (14)$$

Since there is no control matrix, model is non-controllable and for $T \neq 0$ it is observable.

3.4 Robot Tilt Acceleration Compensation

When robot is tilting in Roll or Pitch axis, vertical accelerometer readings are changing and Kalman filter will do not work properly. Errors will appear mostly in velocity estimation. Therefore authors developed compensation algorithm to reduce errors caused by tilt it both axes. The algorithm adds correction to accelerometer measurement in the Kalman filter, calculated in following equations:

$$x = \sin(\alpha) \quad (15)$$

$$y = \sin(\beta) \quad (16)$$

$$z = \sqrt{x^2 + y^2} \quad (17)$$

$$a_{cor} = \sqrt{1 - z^2} \quad (18)$$

Where α is the Roll angle and β i the Pitch angle. In Fig. 4 example data, gathered in special testbed, allowing only rotary movement is shown. As it can be seen, the compensation method works as required.

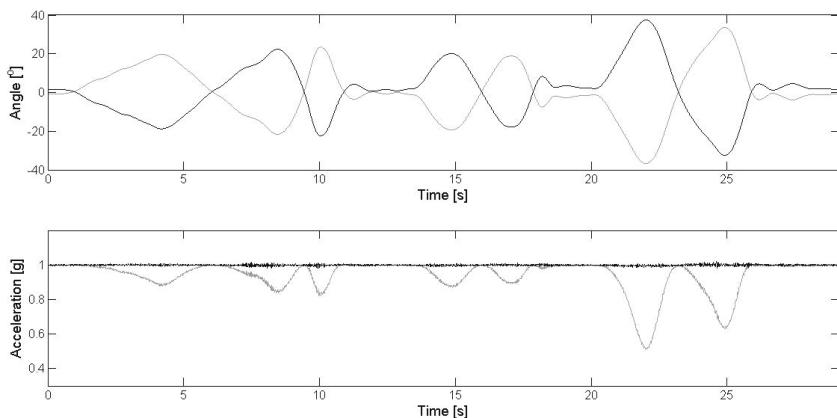


Fig. 4. Tilt compensation example. Angles α (gray) and β (black) in top graph. In bottom, accelerometer in z axis measurement (gray) and compensated accelerometer (black).

3.5 The Estimation Algorithm

The proposed algorithm consists of all above methods and it can be presented in points:

1. Calculation of the robot attitude
2. Measurement of barometric pressure, vertical acceleration and temperature
3. Temperature drift compensation of the barometer measurement
4. Calculation of relative altitude
5. Robot tilt compensation
6. Estimation of the altitude and the vertical velocity with the Kalman filter

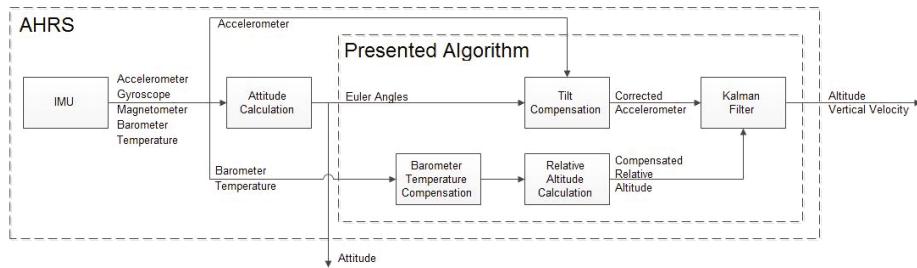


Fig. 5. Flowchart of the algorithm

Thanks to the algorithm, the altitude and the vertical velocity were estimated. Results during the test flight are presented in next the section.

4 Experiments

One experiment was performed in order to check accuracy and dynamics of estimation during the flight. At the beginning robot was on the ground with motors rotating. During the next phase the robot took off and had a few changes in altitude, finally robot landed. Data was sampled at frequency of approximately 200Hz. Results are presented in Fig. 6.

As it can been seen, temperature drift has substantial influence on altitude estimation. Difference between start and landing altitude was approximately 1 meter. After compensation, results are very satisfactory and error is lower than 0.1 meters. It is important to mention that temperature drift of barometric readings has minimal influence on velocity estimation. Disturbances in barometer's altitude during take off and landing were caused by air movement from propellers. Estimated altitude was very smooth and that is important while using it in altitude controller. Though, the accelerometer readings were very noisy 7, the algorithm gives desired results.

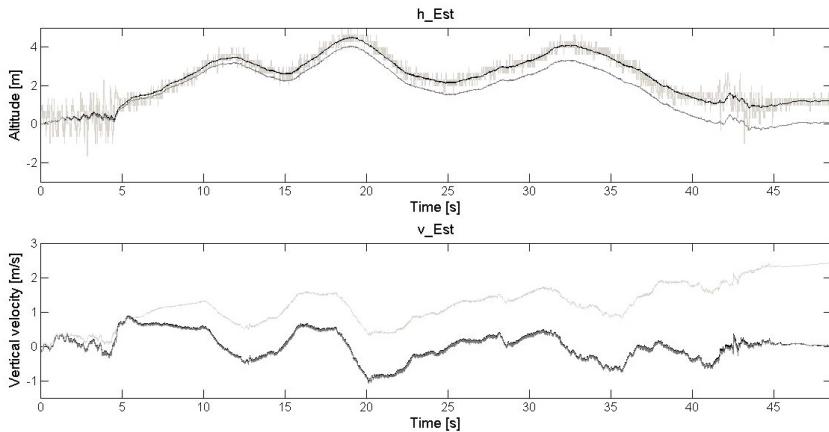


Fig. 6. Experiment results. Relative altitude (light gray), estimated altitude (black) and estimated altitude with temperature compensation (dark gray) in top graph. In bottom, vertical velocity from integrating accelerometer in Z axis (light gray), estimated vertical velocity (black) and estimated vertical velocity with temperature compensation (dark gray).

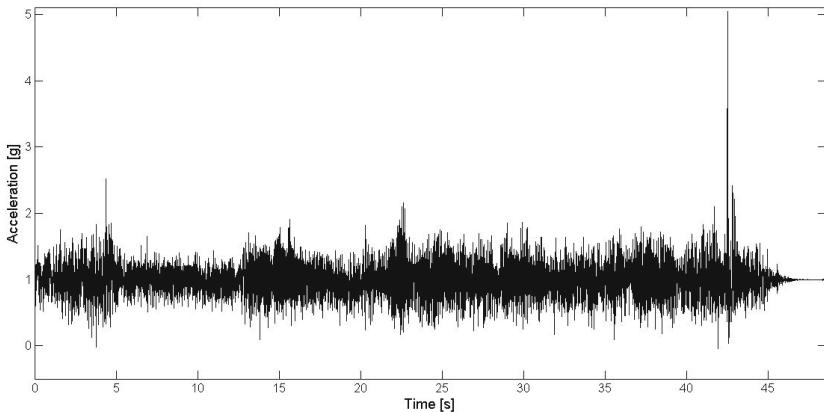


Fig. 7. Readings from accelerometer

5 Conclusions

As the experiment shown, presented algorithm estimates properly and can be reliable. Thanks to that, a regulator to hold desired altitude and the other one to hold velocity can be designed and implemented. These controllers will be capable to control robot more easily. Based on controllers for altitude and velocity, it will be possible to decrease or increase the vertical velocity with high precision.

The algorithm can be implemented in various machines (flying, driving or stationary) equipped with IMU, containing barometer and accelerometer.

To increase precision and reliability it is possible to fuse this algorithm with GPS receiver or visual system. Future works will be focused on controllers for the UAV, which has velocity and altitude estimators on the feedback signals.

References

1. Mahony, R., Hamel, T., Pflimlin, J.-M.: Nonlinear Complementary Filters on the Special Orthogonal Group. *IEEE Transactions on Automatic Control* 53(5) (June 2008)
2. Madgwick, S.O.H., Harrison, A.J.L., Vaidyanathan, R.: Estimation of IMU and MARG orientation using a gradient descent algorithm. In: Rehabilitation Robotics (ICORR) (2011)
3. Wang, S., Yang, Y.: Quadrotor aircraft attitude estimation and control based on Kalman filter. In: Control Conference (CCC), pp. 1072–1077 (2010)
4. Welch, G., Bishop, G.: An Introduction to the Kalman Filter. University of North Carolina at Chapel Hill (July 24, 2006)
5. Portland State Aerospace Society, A Quick Derivation relating altitude to air pressure (2004)

The Influence of the End Effector Gyroscopic Torques on a Base of the Manipulator

Jarosław Gośliński, Stanisław Gardecki, and Wojciech Giernacki

Poznań University of Technology,
Institute of Control and Information Engineering,
ul. Piotrowo 3A, 60-965 Poznań, Poland
jaroslaw.a.goslinski@doctorate.put.poznan.pl
{stanislaw.gardecki,wojciech.giernacki}@put.poznan.pl
<http://www.cie.put.poznan.pl/>

Abstract. In this paper, the model of the robotic arm with payload is presented. The payload is configured as a rotating mass, which imitate the tool on a robot's wrist. While manipulating with working tool the gyroscopic effect can occur. This leads to extra moments in each joint of the manipulator. In order to evaluate the scale of the process and the consequences in the robot's trajectory, mathematical model of the robot including rotating mass was derived. The system was simulated for different parameters. It was proven that the gyroscopic effect cannot be neglected, especially when the robot's movement are rapid and conducted simultaneously in more then one joint at once. For the purpose of the work, authors decided to use manipulator with five degrees of freedom, equipped with rotational joints only.

Keywords: gyroscopic effect, manipulator, mathematical modeling.

1 Introduction

Robotic arms are used in many applications especially in industry for assembling, welding, riveting etc. Some of other applications includes robotic arms in space (space manipulators). Space manipulators and special manipulators on assembly lines may have long arms or many joints. These features increase maneuverability as well as agility, though may lead to problems in control. The problem is particularly visible while working with rotating payloads. Torques and forces in each joint depend on arm's angular velocity, joint variables and external forces. One of external torques stem from tool mounted on a wrist, which is a spinning mass. In space manipulators, the problem is even more sophisticated - the gyroscopic effect form spinning rotors of the drive motors affect on the chassis of the spacecraft/base station [1]. On the other hand, the same principle of spinning mass can be used as a drives for actuators as a Control-Moment Gyroscopes (CMGs) [2]. CMGs can reduce the base reactions or even eliminate them, while using the same amount of power as a robotic system driven by conventional joint motors. Regardless of type of effects (desirable or side) the influence of spinning

mass in robot mechanical system cannot be neglected without its prior evaluation. In this paper authors derive the mathematical model of a manipulator with five degrees of freedom with payload attached on the last joint. In order to assess the level of influence of gyroscopic effect on the manipulator, simulations for payloads with different weights and angular velocities were shown. In the article, the first paragraph describes the robotic arm. In the next section the manipulator model is derived and described. In the third and fourth paragraph the model is enhanced with the gyroscopic effect. The fifth paragraph shows the simulations results. Finally in the last paragraph conclusions are presented.

2 The Manipulator Model

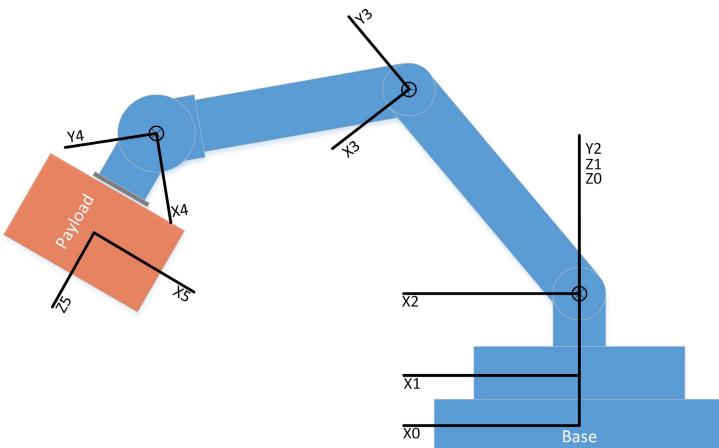


Fig. 1. The manipulator's kinematic chain

The kinematic model of the robotic arm with five degrees of freedom (DOF) can be expressed with five transitions matrices e.i. transitions from the base to the frame of the end effector. In solution depicted in figure 1 all joints are rotational. Base on this assumption, using Denavit-Hartenberg algorithm we can obtain:

$$T_1^0 = \begin{bmatrix} \cos(q_1) & -\sin(q_1) & 0 & 0 \\ \sin(q_1) & \cos(q_1) & 0 & 0 \\ 0 & 0 & 1 & l_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$T_2^1 = \begin{bmatrix} \cos(q_2) & 0 & -\sin(q_2) & 0 \\ \sin(q_2) & 0 & \cos(q_2) & 0 \\ 0 & -1 & 0 & l_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$T_3^2 = \begin{bmatrix} \cos(q_3) & -\sin(q_3) & 0 & l_3 \cos(q_3) \\ \sin(q_3) & \cos(q_3) & 0 & l_3 \sin(q_3) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$T_4^3 = \begin{bmatrix} \cos(q_4) & -\sin(q_4) & 0 & l_4 \cos(q_4) \\ \sin(q_4) & \cos(q_4) & 0 & l_4 \sin(q_4) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$T_5^4 = \begin{bmatrix} \cos(q_5) & 0 & \sin(q_5) & l_5 \cos(q_5) \\ \sin(q_5) & 0 & -\cos(q_5) & l_5 \sin(q_5) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Each transition matrix T_i^{i-1} describes coordinate i expressed in coordinate $i-1$. In order to derive the mathematical model of the manipulator, all of the manipulator's matrices must be defined. Matrices describe the robot dynamics and they are defined, base on previously shown transition matrices and inertia tensors. The manipulator equation can be written in one of many forms, here the standard form [3] is presented:

$$D(\underline{q})\ddot{\underline{q}} + C(\underline{q}, \dot{\underline{q}}) + G(\underline{q}) = \underline{\tau} \quad (6)$$

$D(\underline{q})$ denotes inertia matrix, $C(\underline{q}, \dot{\underline{q}})$ denotes centrifugal and Coriolis forces, vector $G(\underline{q})$ indicates gravitational forces. In purpose of simulations, the standard manipulator equation is extended to its real form with friction matrix B :

$$D(\underline{q})\ddot{\underline{q}} + C(\underline{q}, \dot{\underline{q}}) + G(\underline{q}) + B\underline{\dot{q}} = \underline{\tau} \quad (7)$$

The simplest form of friction matrix is a diagonal matrix with viscous friction elements for each joint:

$$B = \begin{bmatrix} b_1 & 0 & 0 & 0 & 0 \\ 0 & b_2 & 0 & 0 & 0 \\ 0 & 0 & b_3 & 0 & 0 \\ 0 & 0 & 0 & b_4 & 0 \\ 0 & 0 & 0 & 0 & b_5 \end{bmatrix} \quad (8)$$

Matrix D is symmetric and it is composed with d_{ij} (where $i, j \in 1, 2, \dots, 5$):

$$D(\underline{q}) = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} \end{bmatrix}, \quad (9)$$

where the elements on main diagonal are as follows:

$$\begin{aligned} d_{11} = & (Ix_3 + Ix_4 + Ix_5 + Iy_3 + Iy_4 + Iz_5)/2 + Iz_1 + Iy_2 - (Ix_4 \cos(2q_3 + 2q_4))/2 \\ & + (Iy_4 \cos(2q_3 + 2q_4))/2 + (l_3^2 m_3)/8 + (l_3^2 m_4)/2 + (l_3^2 m_5)/2 + (l_4^2 m_4)/8 + (l_4^2 m_5)/2 \\ & + (l_5^2 m_5)/8 - (Ix_3 \cos(2q_3))/2 + (Iy_3 \cos(2q_3))/2 - (Ix_5 \cos(2q_3 + 2q_4 + 2q_5))/2 \\ & + (Iz_5 \cos(2q_3 + 2q_4 + 2q_5))/2 + (l_3^2 m_3 \cos(2q_3))/8 + (l_3^2 (m_4 + m_5) \cos(2q_3))/2 \\ & + (l_5^2 m_5 \cos(2q_3 + 2q_4 + 2q_5))/8 + (l_4^2 m_4 \cos(2q_3 + 2q_4))/8 + (l_4^2 m_5 \cos(2q_3 + 2q_4))/2 \\ & + (l_3 l_5 m_5 \cos(2q_3 + q_4 + q_5))/2 + (l_3 l_5 m_5 \cos(q_4 + q_5))/2 + l_3 l_4 (m_5 + m_4/2) \cos(q_4) \\ & + (l_4 l_5 m_5 \cos(q_5))/2 + (l_4 l_5 m_5 \cos(2q_3 + 2q_4 + q_5))/2 + (l_3 l_4 m_4 \cos(2q_3 + q_4))/2 \\ & + l_3 l_4 m_5 \cos(2q_3 + q_4) \quad (10) \end{aligned}$$

$$\begin{aligned} d_{22} = & (Ix_3 + Ix_4 + Ix_5 + Iy_3 + Iy_4 + Iz_5)/2 + Iy_2 - (Ix_4 \cos(2q_3 + 2q_4))/2 \\ & + (Iy_4 \cos(2q_3 + 2q_4))/2 + (l_3^2 m_3)/8 + (l_3^2 m_4)/2 + (l_3^2 m_5)/2 + (l_4^2 m_4)/8 + (l_4^2 m_5)/2 \\ & + (l_5^2 m_5)/8 - (Ix_3 \cos(2q_3))/2 + ((Iy_3 \cos(2q_3)) - ((Ix_5 + Iz_5) \cos(2q_3 + 2q_4 + 2q_5)))/2 \\ & + l_3^2 (m_3/8 + m_4/2 + m_5/2) \cos(2q_3) + (l_5^2 m_5 \cos(2q_3 + 2q_4 + 2q_5))/8 \\ & + (l_4^2 m_4 \cos(2q_3 + 2q_4))/8 + (l_4^2 m_5 \cos(2q_3 + 2q_4))/2 + (l_3 l_5 m_5 \cos(2q_3 + q_4 + q_5))/2 \\ & + (l_3 l_5 m_5 \cos(q_4 + q_5))/2 + (l_3 l_4 m_4 \cos(q_4))/2 + l_3 l_4 m_5 \cos(q_4) + (l_4 l_5 m_5 \cos(q_5))/2 \\ & + (l_4 l_5 m_5 \cos(2q_3 + 2q_4 + q_5))/2 + (l_3 l_4 m_4 \cos(2q_3 + q_4))/2 + l_3 l_4 m_5 \cos(2q_3 + q_4) \quad (11) \end{aligned}$$

$$\begin{aligned} d_{33} = & Iy_5 + Iz_3 + Iz_4 + (l_3^2 m_3)/4 + l_3^2 m_4 + l_3^2 m_5 + (l_4^2 m_4)/4 + l_4^2 m_5 + (l_5^2 m_5)/4 \\ & + l_3 l_5 m_5 \cos(q_4 + q_5) + l_3 l_4 m_4 \cos(q_4) + 2l_3 l_4 m_5 \cos(q_4) + l_4 l_5 m_5 \cos(q_5) \quad (12) \end{aligned}$$

$$d_{44} = Iy_5 + Iz_4 + (l_4^2 m_4)/4 + l_4^2 m_5 + (l_5^2 m_5)/4 + l_4 l_5 m_5 \cos(q_5) \quad (13)$$

$$d_{55} = (m_5 l_5^2)/4 + Iy_5 \quad (14)$$

Elements that are non-diagonal can be expressed as follows:

$$\begin{aligned} d_{12} = d_{21} = & (Ix_3 + Ix_4 + Ix_5 + Iy_3 + Iy_4 + Iz_5) + Iy_2 - (Ix_4 \cos(2q_3 + 2q_4))/2 \\ & + (Iy_4 \cos(2q_3 + 2q_4))/2 + (l_3^2 m_3)/8 + (l_3^2 m_4)/2 + (l_3^2 m_5)/2 + (l_4^2 m_4)/8 + (l_4^2 m_5)/2 \\ & + (l_5^2 m_5)/8 + ((Iy_3 - Ix_3) \cos(2q_3))/2 + ((Iz_5 - Ix_5) \cos(2q_3 + 2q_4 + 2q_5))/2 \\ & + (l_3^2 (m_3/8 + m_4/2 + m_5/2) \cos(2q_3))/8 + (l_5^2 m_5 \cos(2q_3 + 2q_4 + 2q_5))/8 \\ & + (l_4^2 m_4 \cos(2q_3 + 2q_4))/8 + (l_4^2 m_5 \cos(2q_3 + 2q_4))/2 + (l_3 l_5 m_5 \cos(2q_3 + q_4 + q_5))/2 \\ & + (l_3 l_5 m_5 \cos(q_4 + q_5))/2 + (l_3 l_4 m_4 \cos(q_4))/2 + l_3 l_4 m_5 \cos(q_4) + (l_4 l_5 m_5 \cos(q_5))/2 \end{aligned}$$

$$+ (l_4 l_5 m_5 \cos(2q_3 + 2q_4 + q_5)) / 2 + (l_3 l_4 m_4 \cos(2q_3 + q_4)) / 2 + l_3 l_4 m_5 \cos(2q_3 + q_4) \quad (15)$$

$$d_{13} = d_{14} = d_{15} = d_{23} = d_{24} = d_{25} = 0 \quad (16)$$

$$d_{34} = d_{43} = Iy_5 + Iz_4 + (l_4^2 m_4) / 4 + l_4^2 m_5 + (l_5^2 m_5) / 4 + (l_3 l_5 m_5 \cos(q_4 + q_5)) / 2 \\ + (l_3 l_4 m_4 \cos(q_4)) / 2 + l_3 l_4 m_5 \cos(q_4) + l_4 l_5 m_5 \cos(q_5) \quad (17)$$

$$d_{35} = d_{53} = Iy_5 + (l_5^2 m_5) / 4 + (l_3 l_5 m_5 \cos(q_4 + q_5)) / 2 + (l_4 l_5 m_5 \cos(q_5)) / 2 \quad (18)$$

$$d_{45} = d_{54} = (m_5 l_5^2) / 4 + (l_4 m_5 \cos(q_5) l_5) / 2 + Iy_5 \quad (19)$$

The C matrix can be presented as a full matrix or as a vector. Vector of centrifugal and Coriolis forces is derived as product of multiplication: $C(\underline{q}, \dot{\underline{q}})\dot{\underline{q}}$ and can be written in a form:

$$C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix}, \quad (20)$$

where:

$$\begin{aligned} c_1 = c_2 = & -((dq_1 + dq_2)(4Iy_3 dq_3 \sin(2q_3) - 4Ix_5(dq_4 + dq_3) \sin(2q_3 + 2q_4 + 2q_5) \\ & + Ix_3 dq_3 \sin(2q_3)) - 4Ix_5 dq_5 \sin(2q_3 + 2q_4 + 2q_5) + 4Iz_5 dq_3 \sin(2q_3 + 2q_4 + 2q_5) \\ & + 4Iz_5 dq_4 \sin(2q_3 + 2q_4 + 2q_5) + 4Iz_5 dq_5 \sin(2q_3 + 2q_4 + 2q_5) - 4Ix_4 dq_3 \sin(2q_3 + 2q_4) \\ & - 4(Ix_4 dq_4 \sin(2q_3 + 2q_4) + Iy_4(dq_3 + dq_4) \sin(2q_3 + 2q_4)) + dq_3 l_3^2(m_3 + 4m_4) \sin(2q_3) \\ & + 4dq_3 l_3^2 m_5 \sin(2q_3) + (dq_3 + dq_4) l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) + dq_3 l_4^2 m_4 \sin(2q_3 + 2q_4) \\ & + dq_5 l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) + 4dq_4 l_4^2 m_5 \sin(2q_3 + 2q_4) + dq_4 l_4^2 m_4 \sin(2q_3 + 2q_4) \\ & + 4dq_3 l_4^2 m_5 \sin(2q_3 + 2q_4) + 4dq_3 l_3 l_4 m_4 \sin(2q_3 + q_4) + 8dq_3 l_3 l_4 m_5 \sin(2q_3 + q_4) \\ & + dq_4 l_3 l_4 (2m_4 + 4m_5) \sin(2q_3 + q_4) + (4dq_3 + 2dq_4) l_3 l_5 m_5 \sin(2q_3 + q_4 + q_5) \\ & + 2dq_5 l_3 l_5 m_5 \sin(2q_3 + q_4 + q_5) + 2dq_4 l_3 l_5 m_5 \sin(q_4 + q_5) + 2dq_5 l_3 l_5 m_5 \sin(q_4 + q_5) \\ & + dq_4 l_3 l_4 (2m_4 + 4m_5) \sin(q_4) + 2dq_5 l_4 l_5 m_5 \sin(q_5) + 4dq_3 l_4 l_5 m_5 \sin(2q_3 + 2q_4 + q_5) \\ & + 4dq_4 l_4 l_5 m_5 \sin(2q_3 + 2q_4 + q_5) + 2dq_5 l_4 l_5 m_5 \sin(2q_3 + 2q_4 + q_5))) / 4 \end{aligned} \quad (21)$$

$$\begin{aligned} c_3 = & (dq_1(dq_1 + dq_2)(4Iy_4 \sin(2q_3 + 2q_4) - 4Ix_4 \sin(2q_3 + 2q_4) - 4Ix_3 \sin(2q_3) \\ & + 4Iy_3 \sin(2q_3) + 4(Iz_5 - Ix_5) \sin(2q_3 + 2q_4 + 2q_5) + l_4^2(4m_5 + m_4) \sin(2q_3 + 2q_4) \\ & + l_3^2(m_3 + 4m_4 + 4m_5) \sin(2q_3) + l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) + 4l_3 l_5 m_5 \sin(2q_3 + q_4 + q_5) \\ & + 4l_4 l_5 m_5 \sin(2q_3 + 2q_4 + q_5) + 4l_3 l_4 m_4 \sin(2q_3 + q_4) + 8l_3 l_4 m_5 \sin(2q_3 + q_4))) / 8 \\ & + ((dq_4 + dq_5 - dq_4 dq_3) l_3 l_5 m_5 \sin(q_4 + q_5)) / 2 + dq_3 l_3 l_4 (m_4 / 2 + m_5) \sin(q_4) \\ & + dq_4 l_3 l_4 (m_4 / 2 + m_5) \sin(q_4) + (dq_5 l_4 l_5 m_5 \sin(q_5)) - dq_3 ((dq_4 l_3 l_5 m_5 \sin(q_4 + q_5)) / 2 \\ & + (dq_5 l_3 l_5 m_5 \sin(q_4 + q_5)) / 2 + (dq_4 l_3 l_4 m_4 \sin(q_4)) / 2 + dq_4 l_3 l_4 m_5 \sin(q_4)) \end{aligned}$$

$$\begin{aligned}
& + (dq_5 l_4 l_5 m_5 \sin(q_5)) / 2 + (dq_2(dq_1 + dq_2)(4Iy_4 \sin(2q_3 + 2q_4) - 4Ix_4 \sin(2q_3 + 2q_4) \\
& - 4Ix_3 \sin(2q_3) + 4Iy_3 \sin(2q_3) - 4Ix_5 \sin(2q_3 + 2q_4 + 2q_5) + 4Iz_5 \sin(2q_3 + 2q_4 + 2q_5) \\
& + l_4^2 m_4 \sin(2q_3 + 2q_4) + 4l_4^2 m_5 \sin(2q_3 + 2q_4) + l_3^2 m_3 \sin(2q_3) + 4l_3^2 m_4 \sin(2q_3) \\
& + 4l_3^2 m_5 \sin(2q_3) + l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) + 4(l_3 + l_4)l_5 m_5 \sin(2q_3 + q_4 + q_5) \\
& + 4l_3 l_4 m_4 \sin(2q_3 + q_4) + 8l_3 l_4 m_5 \sin(2q_3 + q_4))) / 8 - (dq_5 l_5 m_5 (l_3 \sin(q_4 + q_5) \\
& + l_4 \sin(q_5)) (dq_3 + dq_4 + dq_5)) / 2
\end{aligned} \quad (22)$$

$$\begin{aligned}
c_4 = & dq_3((dq_3 l_3 l_5 m_5 \sin(q_4 + q_5)) / 2 + (dq_3 l_3 l_4 m_4 \sin(q_4)) / 2 + dq_3 l_3 l_4 m_5 \sin(q_4) \\
& - (dq_5 l_4 l_5 m_5 \sin(q_5)) / 2 + (dq_1(dq_1 + dq_2)(4Iy_4 \sin(2q_3 + 2q_4) - 4Ix_4 \sin(2q_3 + 2q_4) \\
& - 4Ix_5 \sin(2q_3 + 2q_4 + 2q_5) + 4Iz_5 \sin(2q_3 + 2q_4 + 2q_5) + l_4^2 m_4 \sin(2q_3 + 2q_4) \\
& + 4l_4^2 m_5 \sin(2q_3 + 2q_4) + l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) + 2l_3 l_5 m_5 \sin(2q_3 + q_4 + q_5) \\
& + 2l_3 l_5 m_5 \sin(q_4 + q_5) + 2l_3 l_4 m_4 \sin(q_4) + 4l_3 l_4 m_5 \sin(2q_3 + 2q_4 + q_5) \\
& + 2l_3 l_4 m_4 \sin(2q_3 + q_4) + 4l_3 l_4 m_5 \sin(2q_3 + q_4))) / 8 \\
& + (dq_2(dq_1 + dq_2)(4Iy_4 \sin(2q_3 + 2q_4) - 4Ix_4 \sin(2q_3 + 2q_4) - 4Ix_5 \sin(2q_3 + 2q_4 + 2q_5) \\
& + 4Iz_5 \sin(2q_3 + 2q_4 + 2q_5) + l_4^2(m_4 + 4m_5) \sin(2q_3 + 2q_4) + l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) \\
& + 2l_3 l_5 m_5 \sin(2q_3 + q_4 + q_5) + 2l_3 l_5 m_5 \sin(q_4 + q_5) + 2l_3 l_4 m_4 \sin(q_4) + 4l_3 l_4 m_5 \sin(q_4) \\
& + 4l_4 l_5 m_5 \sin(2q_3 + 2q_4 + q_5) + 2l_3 l_4 m_4 \sin(2q_3 + q_4) + 4l_3 l_4 m_5 \sin(2q_3 + q_4))) / 8 \\
& - (dq_4 d_4 l_4 l_5 m_5 \sin(q_5)) / 2 - (dq_5 l_4 l_5 m_5 \sin(q_5) (dq_3 + dq_4 + dq_5)) / 2
\end{aligned} \quad (23)$$

$$\begin{aligned}
c_5 = & (dq_1(dq_1 + dq_2)(4(Iz_5 - Ix_5) \sin(2q_3 + 2q_4 + 2q_5) + l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) \\
& + 2l_3 l_5 m_5 (\sin(2q_3 + q_4 + q_5) + \sin(q_4 + q_5)) + 2l_4 l_5 m_5 (\sin(q_5) + \sin(2q_3 + 2q_4 + q_5))) / 8 \\
& + (dq_2(dq_1 + dq_2)(4(Iz_5 - Ix_5) \sin(2q_3 + 2q_4 + 2q_5) + l_5^2 m_5 \sin(2q_3 + 2q_4 + 2q_5) \\
& + 2(l_3 + l_4)l_5 m_5 \sin(2q_3 + q_4 + q_5) + 2l_3 l_5 m_5 \sin(q_4 + q_5) + 2l_4 l_5 m_5 \sin(q_5))) / 8 \\
& + (dq_3 l_5 m_5 ((dq_3 + dq_4)l_4 \sin(q_5) + dq_3 l_3 \sin(q_4 + q_5)) + dq_4 l_4 l_5 m_5 \sin(q_5) (dq_3 + dq_4)) / 2
\end{aligned} \quad (24)$$

Vector C has an explicit form, which is not valid in case of matrix $C(\underline{q}, \dot{\underline{q}})$. That matrix can be written in many different and correct configurations. The last part of the model, given by equation 7 is a vector of gravitational forces $G(\underline{q})$. The next elements of the G can be written as:

$$g_1 = g_2 = 0 \quad (25)$$

$$\begin{aligned}
g_3 = & -m_5 g (l_4 \cos(q_3 + q_4) + l_3 \cos(q_3) + (l_5 \cos(q_3 + q_4 + q_5)) / 2) \\
& - m_4 g ((l_4 \cos(q_3 + q_4)) / 2 + l_3 \cos(q_3)) - (l_3 m_3 g \cos(q_3)) / 2
\end{aligned} \quad (26)$$

$$g_4 = -m_5 g (l_4 \cos(q_3 + q_4) + (l_5 \cos(q_3 + q_4 + q_5)) / 2) - (l_4 m_4 g \cos(q_3 + q_4)) / 2 \quad (27)$$

$$g_5 = -(l_5 m_5 g \cos(q_3 + q_4 + q_5)) / 2 \quad (28)$$

In model's matrices and vectors, l_i denotes i-th arm's length, q_i , dq_i and ddq_i denote respectively: i-th joint angular position, i-th angular speed and i-th

angular acceleration. Inertia tensors were calculated with assumption, that the center of the mass for each arm is located in the middle of arm ($l_i/2$).

3 Gyroscopic Torques

The gyroscopic effect can be described as a cross product of precession's angular velocity ($\underline{\omega}$) and an angular momentum of the spinning mass (\underline{L}_m). When the gyroscope is mounted on the manipulator's end effector, then the precession is an angular velocity of the Tool Center Point (TCP) expressed in base coordinate system. The gyroscope general formula is given by:

$$\underline{\tau}_g = \underline{\omega} \times \underline{L}_m \quad (29)$$

In the case of the manipulator, an angular velocity is a combination of all joint's angular velocities in a kinematic chain [3]:

$$\underline{\omega} = \underline{\omega}_n^0 = \underline{\omega}_1^0 + R_1^0 \underline{\omega}_2^1 + R_2^0 \underline{\omega}_3^2 + R_3^0 \underline{\omega}_4^3 + \dots + R_{n-1}^0 \underline{\omega}_n^{n-1} \quad (30)$$

Each $\underline{\omega}_n^{n-1}$ vector has only one element corresponding to the Z-axis, in accordance with D-H notation e.i.:

$$\underline{\omega}_n^{n-1} = \begin{bmatrix} 0 \\ 0 \\ dq_n \end{bmatrix} \quad (31)$$

The last rotational joint has got spinning mass attached to it, and the last coordinate system in which the calculations of the gyroscope torque must be made is the static ($q_5 = 0, dq_5 = 0$) fifth frame. The total angular velocity, expressed in the base coordinate system is given as follows:

$$\underline{\omega}_4^0 = \begin{bmatrix} -\sin(q_1 + q_2)(dq_2 + dq_3 + dq_4) \\ \cos(q_1 + q_2)(dq_2 + dq_3 + dq_4) \\ dq_1 \end{bmatrix} \quad (32)$$

whereas the angular momentum of the spinning mass in the fifth joint is equal to:

$$\underline{L}_m = \begin{bmatrix} 0 \\ 0 \\ I_m dq_5 \end{bmatrix} \quad (33)$$

Finally, the cross product of angular velocity and the angular momentum is calculated:

$$\underline{\tau}_g = \begin{bmatrix} I_m dq_5 \cos(q_1 + q_2)(dq_2 + dq_3 + dq_4) \\ I_m dq_5 \sin(q_1 + q_2)(dq_2 + dq_3 + dq_4) \\ 0 \end{bmatrix} \quad (34)$$

Torque $\underline{\tau}_g$ is expressed in the last static coordinate system e.i. in coordinate of the end effector.

4 The Influence of Gyroscopic Effect on Robot Joints

In previous section, the torque of gyroscopic effect was derived. Here the influence of this torque on mechanical structure of the manipulator is evaluated. Firstly,

the overall formula for the torques projection from the joint $n+1$ to the joint n must be derived. Base on [3], it can be written that (without forces):

$$\underline{\tau}_n = R_n^{n+1} \underline{\tau}_{n+1} \quad (35)$$

where the rotation matrix R_n^{n+1} is a part of transition matrix T_n^{n+1} :

$$T_n^{n+1} = \begin{bmatrix} R_n^{n+1} & P_n^{n+1} \\ 0 & 0 & 1 \end{bmatrix}. \quad (36)$$

The equation 35 can be extended to all joints, making it possible to project the torque from the fourth to the base coordinate system (note that: $\underline{\tau}_5 = \underline{\tau}_g$):

$$\underline{\tau}_0 = R_0^1 R_1^2 R_2^3 R_3^4 R_4^5 \underline{\tau}_5 \quad (37)$$

Since the rotation matrix is orthogonal, its inverse is equal to transposition:

$$\underline{\tau}_0 = R_1^{0T} R_2^{1T} R_3^{2T} R_4^{3T} R_5^{4T} \underline{\tau}_5 \quad (38)$$

Having the formula given above, the total torque generated in end effector, projected on the base coordinate system, can be stated as:

$$\underline{\tau}_0 = \begin{bmatrix} \tau_x^0 \\ \tau_y^0 \\ \tau_z^0 \end{bmatrix} \quad (39)$$

where:

$$\begin{aligned} \tau_x^0 &= I_m dq_5 \sin(q_1 + q_2) \sin(q_1) (dq_2 + dq_3 + dq_4) + I_m dq_5 \cos(q_1 + q_2) \cdot \\ &\quad (\cos(q_4) (\cos(q_1) \cos(q_2) \cos(q_3) - \cos(q_1) \sin(q_2) \sin(q_3)) - \sin(q_4) (\cos(q_1) \cdot \\ &\quad \cos(q_2) \sin(q_3) + \cos(q_1) \cos(q_3) \sin(q_2))) (dq_2 + dq_3 + dq_4) \end{aligned} \quad (40)$$

$$\begin{aligned} \tau_y^0 &= I_m dq_5 \sin(q_1 + q_2) \cos(q_1) (dq_2 + dq_3 + dq_4) + I_m dq_5 \cos(q_1 + q_2) \cdot \\ &\quad (\cos(q_4) (\sin(q_1) \sin(q_2) \sin(q_3) - \cos(q_2) \cos(q_3) \sin(q_1)) + \sin(q_4) (\cos(q_2) \cdot \\ &\quad \sin(q_1) \sin(q_3) + \cos(q_3) \sin(q_1) \sin(q_2))) (dq_2 + dq_3 + dq_4) \end{aligned} \quad (41)$$

$$\tau_z^0 = -I_m dq_5 \sin(q_2 + q_3 + q_4) \cos(q_1 + q_2) (dq_2 + dq_3 + dq_4) \quad (42)$$

5 Simulations Results

Since the influence of gyroscopic torques on robot's joints is already determined by equations, the overall torque can be evaluated. Here, the experiment is narrowed only to the zero coordinate system (robot's base). It is due to the simplification of number of cases, where for one joint this task is straight. As it was written in equations (39)-(42) the base coordinate system is under the influence of gyroscopic torques in all three axis. Since the rotational joint of robot's base is only in Z-axis, in the rest two axis, acting torques force only the mechanical structure of the robot. Nevertheless, authors decided to show all figures, stressing the range of torque values. The experiment was conducted on virtual robot manipulator with following parameters:

$$l_1 = 0.2m, l_2 = 0.4m, l_3 = l_4 = 1.0m, l_5 = 0.4m, m_1 = 20, m_2 = 40, m_3 = 40, m_4 = 25, m_5 = 10; Ix1 = 0.4, Ix2 = 0.4, Ix3 = 0.3, Ix4 = 0.15, Ix5 = 0.20, Iy1 = 0.22, Iy2 = 0.4, Iy3 = 0.4, Iy4 = 0.18, Iy5 = 0.2, Iz1 = 0.3, Iz2 = 0.5, Iz3 = 0.5, Iz4 = 0.2, Iz5 = 0.2; b_1 = 0.9, b_2 = 0.8, b_3 = 0.7, b_4 = 0.8, b_5 = 0.7$$

The mathematical model given by equations 7, was discretized by using Forward Euler method with sample time $T_s = 0.0001s$. During the experiment, the manipulator was set in starting positions with joints velocities equal to zero. In the first part of research the robotic arm was subjected only to the force of the gravity and consequential gravity moments. In the second part, the experiment was performed again with additional gyroscope torques (close-loop model). The results were shown in figure 2, 3 and 4. In the first and second situation, the

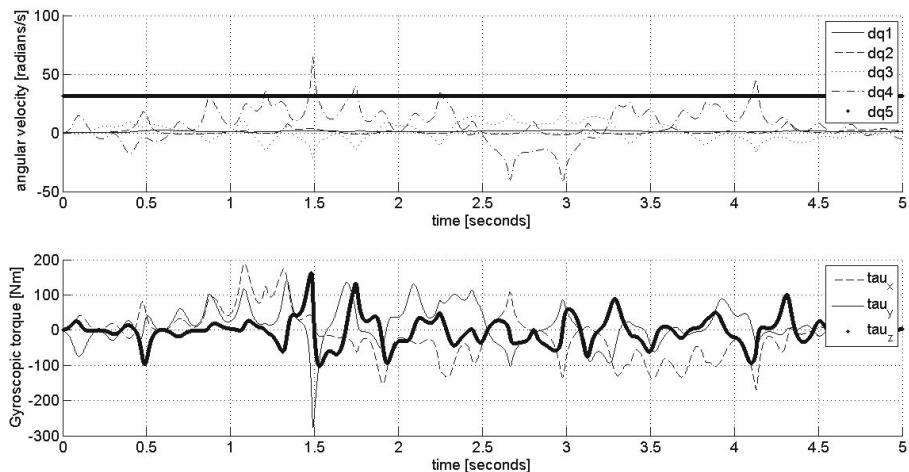


Fig. 2. Joints velocities and gyroscopic torques in open-loop model

total gyroscopic moments were calculated for the inertia $I_m = 0.2$. It is approximately equal to inertia of a car's wheel. In the third experiment the inertia was less than $0.8 [kgm^2]$ (e.g. a saw blade). In case of figure 2 the gyroscopic torques were calculated, but had no influence on the base coordinate system - the model was open-loop. The situation changed in figure 3, where for the same parameters, model was under the influence of the gyroscopic effect in the Z-axis of the first joint. As it can be seen, velocities in the manipulator's kinematic chain were different, especially in joint one, two and four. The same applies to gyroscopic torques, which goes up to $200 [Nm]$ but has different shape in all axes. In the figure 4, a magnitude of the gyroscopic effect was reinforced. For high speed (1400 RPM) of the mass attached on the last joint, gyroscopic torques goes up to $500 [Nm]$ which can be destructive for the whole mechanical construction. What is more, the torques have impulsive nature, which has huge impact on the velocities in all joints.

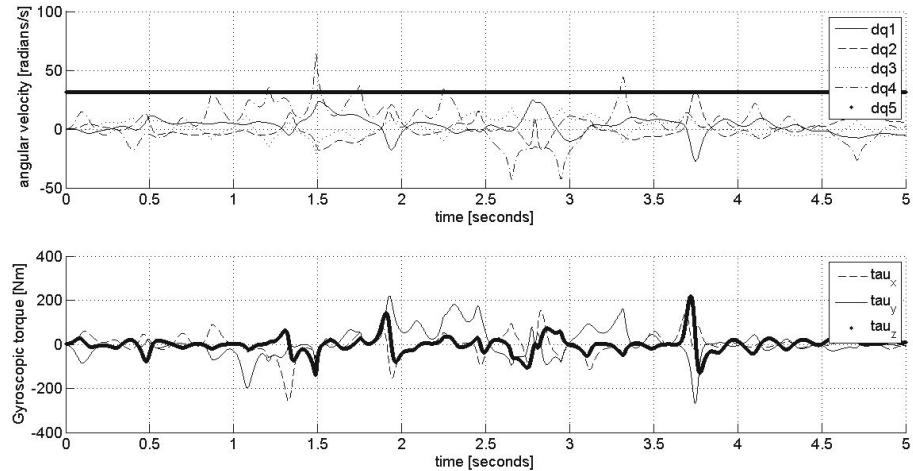


Fig. 3. Joints velocities and gyroscopic torques in close-loop model

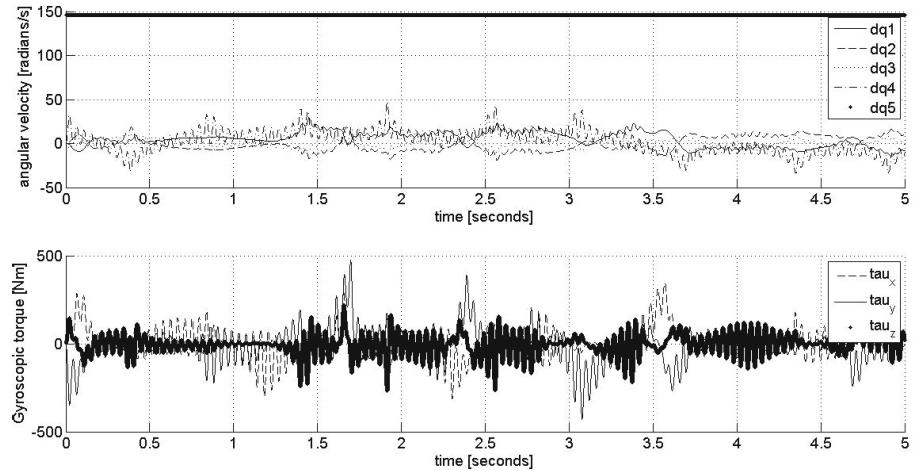


Fig. 4. Joints velocities and gyroscopic torques in close-loop model - high speed in fifth joint

6 Summary

In the article, the mathematical model of 5 DOF manipulator was derived. The model was enhanced with additional equations, describing the gyroscopic effect. The gyroscopic effect was presented as a set of analytical equations (with torques on the output). In accordance with assumptions, torques were projected from the end effector to the base coordinate system. Thanks to that, the simulations of the manipulator with additional spinning mass, attached to the end effector were conducted. Three main dependencies have been studied: the influence of

inertia and velocity of spinning mass on the gyroscopic torques, the influence of manipulator's joints velocities on overall torques and the model dynamics with and without the gyroscopic effect. Torques in the base coordinate system can go up to hundreds of Nm, while increasing the inertia and velocity of spinning mass. The dynamics of the manipulator is different with the modeled gyroscopic effect. It is important to keep in mind that the spinning objects in the end effector can have extreme impact on a control system of manipulators. The gyroscope effect can be neglected, but only if velocities in joint as well as velocity of tool or object in the end effector are small. In any other cases the gyroscopic effect should be modeled and evaluated prior to application.

References

1. Murphy, S.H., Wen, J.T.: Analysis of Active Manipulator Elements in Space Manipulation. *IEEE Transactions on Robotics and Automation* 9(5), 544–552 (1993)
2. Carpenter, M.D., Peck, M.A.: Reducing Base Reactions With Gyroscopic Actuation of Space-Robotic Systems. *IEEE Transactions on Robotics and Automation* 25(6), 1262–1270 (2009)
3. Spong, M.W., Vidyasagar, M.: Robot Dynamics and Control, 1st edn. John Wiley and Sons, Canada (1989)

A Virtual Receptor in a Robot Control Framework

Włodzimierz Kasprzak¹, Tomasz Kornuta², and Cezary Zieliński²

¹ Industrial Research Institute for Automation and Measurements,
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
wkasprzak@piap.pl

² Warsaw University of Technology, Institute of Control and Computation
Engineering, Nowowiejska 15/19, 00-665 Warsaw, Poland
{T.Kornuta,C.Zielinski}@elka.pw.edu.pl

Abstract. We propose a general-purpose virtual receptor for 3D robot vision based on RGB-D sensor data. The application independent robot vision framework performs two basic tasks: it creates a 3D metric map of the environment and it recognizes basic 3D solids and 2D textures and shapes. The design methodology follows the principle of knowledge-based systems, as the virtual receptor is structured into a knowledge base (including the model, data and inference rules) and a control strategy. Procedural semantic networks are chosen as the knowledge representation language. Their main features - an object-oriented modeling of the environment and non-monotonic logic of inferences - makes them specially suitable for 3D object recognition in RGB-D images. The interfaces to other modules of a autonomous robot control structure are discussed also - these are: the main control and ontology modules.

Keywords: robot control system, visual perception, virtual receptor.

1 Introduction

Service and field robots, that are the subject of current intensive research, need to be autonomous - automatic perception and decision making is crucial for them. Perception and ontology are two robot control modules that are relevant to implement autonomous behavior (Fig. 1).

There exist various software packages being implemented under the ROS system for the purpose of image analysis and visual object recognition. In contrast to them our approach fulfills new objectives.

1. First, an integrated segmentation procedure for color image and depth map has been developed [1]. The detection of regions, edges and textures in the color image is integrated with the estimation of surface patches in the depth map, in order to reliably generate textured surface patches of different types. Current perception packages in ROS only seldom cover both types of input data (video camera and laser scanner) at the same time. For example, the package COP (cognitive perception) can process both types of data, but

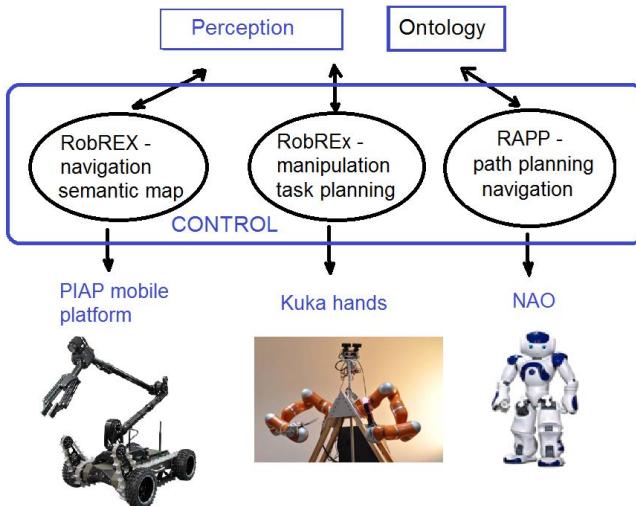


Fig. 1. A general-purpose control structure for different autonomous robots: a mobile, a manipulating and a social robot

not integrated, e.g. planar surfaces are based on geometric features without texture. The project RoboEarth provides two distinct modules for object recognition: "Vision" and "KinectObjectDetector" [2]. Other packages, such as "The Blocks World Robotic Vision Toolbox (BLORT)" [3], ODUFinder [4] and "Interactive Perception" (IAP) [5], process video images only.

2. Second, in video image segmentation not only characteristic points and their features will be detected, as in most of the available ROS packages, but also other typical discrete segments such as edges, regions and textures. Perception packages in ROS, e.g. BLORT, ODUFinder, IAP and RoboEarth, rely mostly on point-like features and their descriptors (e.g. SIFT or SURF, and Lucas-Kanade or ICP point correspondences). In standard image processing, besides points or corners, there are also other well-known segment types detected: edges, regions and textures. They are needed to design a general-purpose object recognition system.
3. Third, our aim is to create general-purpose model-based object recognition using generic object models (representing object types and not only instances, as in current ROS perception) [6]. A limited generic-model approach (only to planar surfaces) is available in the COP package and some variability of segments used for object recognition (lines LINEMOD or textures TOD) is available in the "WG Object Recognition" framework [7], but again for instance-based and not generic object models. The other packages use instance-based models only, e.g. point-cloud models.
4. Fourth, we use a multi-representation approach to object modeling (a 3D model for a foreground object with depth information, or a 2D shape for a distant object. This is not yet available in ROS perception. Deformable

objects can also be represented in our model as different modalities of a concept - they can be interpreted as different views of solids or as different templates of deformable shapes [8].

5. Fifth, complex object types are application-dependent and they can be freely defined in terms of basic solids (like cuboid or cylinder) and textures, recognized by the perception module. The perception module has an open architecture - it is a knowledge-based framework that can be filled with application-dependent object models. We define an interface between a general-purpose perception module and an application-dependent ontology, by which various object types and texture types defined by the ontology can be passed to and be recognized by the perception module.
6. Finally, we can have many objects of different types in the scene. Hence, object recognition is not limited, different than in majority of the ROS packages, to a verification that a given object instance exists in the scene. A reliable detection of multi-instance scenes is made possible by a bi-directional inference scheme [6].

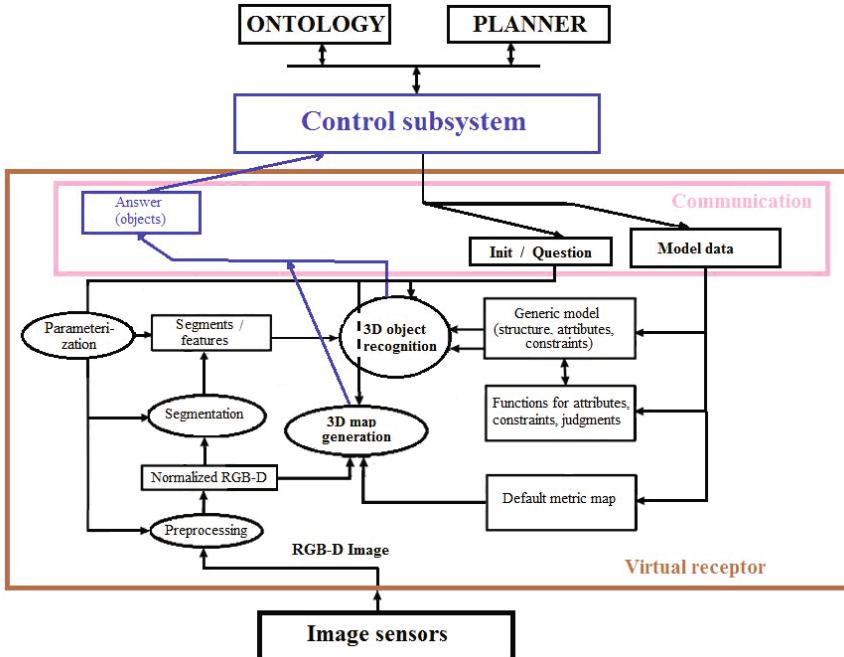


Fig. 2. Virtual receptor is a module of a robot control system

2 The Structure of Virtual Receptor

The following tasks will be solved by the virtual receptor (Fig. 2):

- 3D object recognition of foreground objects (by integrated RGB-D image analysis),

- Foreground and background distance measurement - environment map creation (in the depth map),
- 2D object recognition (in RGB image, by texture and shape features).

The virtual receptor is serving the ontology module, whose main goal is to provide a high-level (abstract) semantic map of the environment. From this map various action plans can be built by the use of a Planning module, i.e. autonomous robot navigation, grasping and manipulation, providing dialogue capabilities for a robot-to-human communication, etc. Among others, the virtual receptor directly supports the creation of various environment maps performed by the ontology and planning modules [9]:

- semantic map creation (scene recognition) (the semantics of concepts: locations, objects, situations, operations);
- topology map creation and path planning (a classification of places due to geometrical and texture features; e.g. door, corridor, room, furniture);
- navigation map (various methods for connecting places; e.g. by Voronoi diagram, visibility graph, use of markers);
- metric 2D map (the geometry of rooms can be obtained by direct projection of a metric 3D map).

3 Knowledge Representation. Object Model

The key advantage of a knowledge-based approach to computer vision is the existence of a declarative or partially declarative knowledge representation language. Different modeling languages have been proposed for 3D object recognition, among them frames and semantic networks [10], and attributed graph- and relation structure-grammars [11].

3.1 Ontology and Perception Levels

Although the virtual receptor and ontology modules communicate via the control sub-module, their models can be made consistent by the use of the same KRL (knowledge representation language), e.g. the semantic network (Fig. 3).

The ontology levels are: the scene (it represents object groups, a background context, abstract actions and properties), a complex object (a set of simple objects assembled into one physical system), a simple object (that consists of basic solids with textured faces). The object representation levels distinguished by the receptor are: solids (also called shells, volumes), faces, surface patches/loops/textures and edges/regions/points/depth points. The lowest level corresponds to segments obtained by the segmentation of a RGB-D image.

The solids, out of which all the basic 3D objects have to be defined by using the aggregation operation, are the following: cuboid (with textured faces), generalized cone (with textured faces).

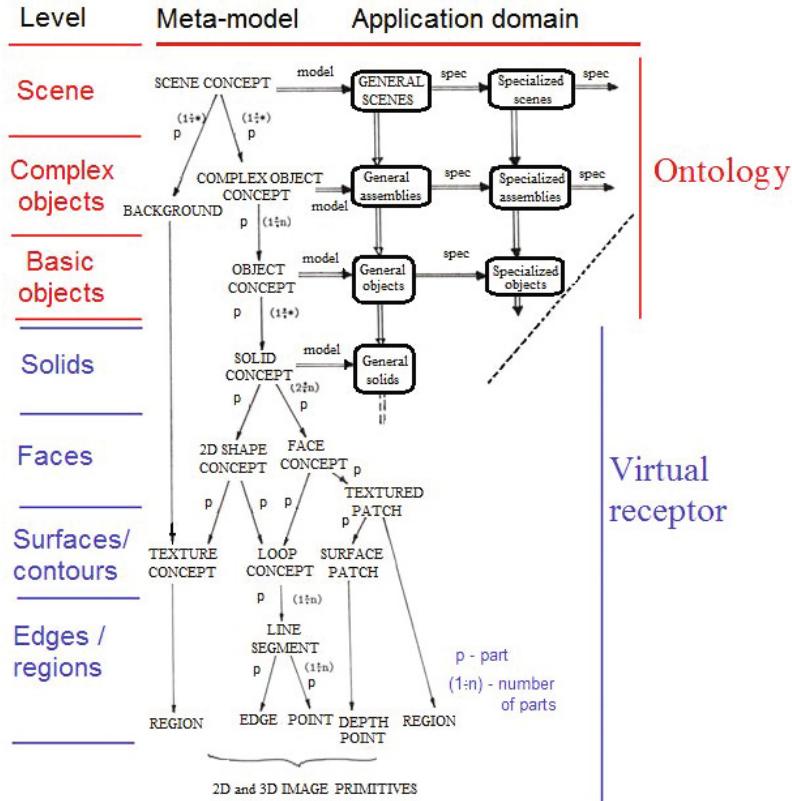


Fig. 3. A hierarchy of concepts in a semantic network considered for world modeling by the ontology and virtual receptor

3.2 Basic Solids

The virtual receptor exports basic solid types to other modules and accepts for recognition any 3D object type that is defined by means of a semantic network concept in terms of parts being these basic solids. These 3D primitives are: a cuboid type (with textured faces) and a generalized cone type (also with textured faces).

Basic objects and texture classes have to be passed to the virtual receptor during the initialization phase, after the learning stage is completed. Examples of objects that can be defined in this way are:

- small objects that can be grasped by hand, e.g. bottle, box, apple, door handle;
- corridor, desk, door, chair, furniture;
- a human body.

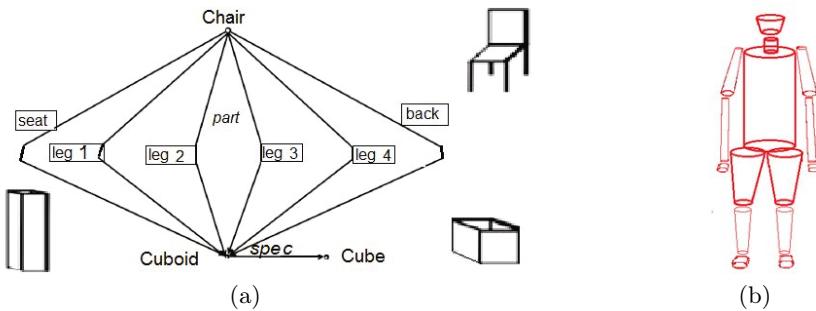


Fig. 4. An example of simple object definition in terms of basic solids: (a) a chair is defined by cuboids, (b) a human body is modeled as an aggregation of generalized cones

For example a "chair" object consists of cuboid parts that play the role of "seat, leg, back" while a human body can be modeled as an aggregation of generalized cones (Fig. 4).

3.3 Meta-type "Concept"

The object model expressed in terms of semantic networks will be integrated with another two general-purpose tools: a modified search based on constraint satisfaction problems [12] defined for the purpose of partial model-to-data matching, and the Bayesian approach to statistical inference [13] will allow evaluation of matched results. All three abstract tool parts are of dominating declarative nature and there exist well-known machine learning approaches for them, e.g. ML- or MAP-estimation of Bayesian probability distributions [14].

A single concept is defined by the following elements:

1. attributes - a vector of values being unique to every instance of a given concept,
2. parts - which must be concepts of types recognized by the virtual receptor, i.e. basic solids and known textures,
3. properties - predicates that must be satisfied by parts of a given concept,
4. functions for the computation of attribute values and the evaluation of properties.

The base class of all terms is called **Concept**. The main fields of this class are:

- **I_atrNum** – number of attributes,
- **T_atr** - an array of attributes (elements of type **Attribute**),
- **I_partsNum** – number of parts,
- **T_parts** – an array of part references and roles (elements of type **Part**, referring to other concepts that are parts of the current concept and play some defined role in the structure of current concept),
- **I_propNum** – number of properties (constraints),
- **T_properties** – an array of properties (elements of type **Property**),

- I_modNum – number of modalities of current concept,
- T_mods – an array of modalities (elements of type **Modality**; each modality declares some subset of parts as obligatory, while the remaining parts are optional),
- O_score – the score of instance of current concept (elements of type **Score**).

3.4 Properties - Constraints

Example. The graph of constraints for a concept "cuboid", that consists of 6 parts of type "rectangular face", can use three types of constraints, denoted as (A,B,C), to allow an aggregation of "face" instances into a "cuboid" instance (Fig. 5).

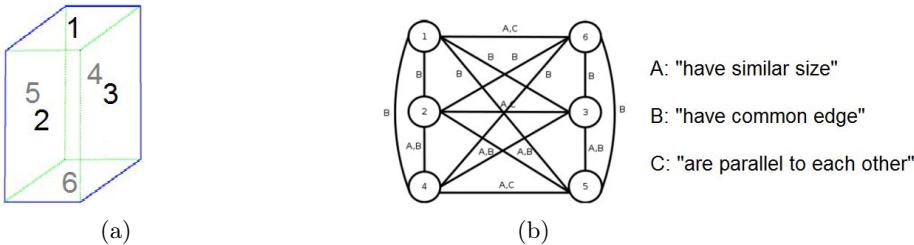


Fig. 5. An example of the graph of constraints for a cuboid: (a) there are 6 parts of type "face", (b) three types of binary constraints between faces are defined

3.5 Scoring by Stochastic Inference

Example. A causal Bayesian network for the "cuboid" concept is defined. Stochastic discrete variables (binary or many-valued variables) correspond to the concept, its modalities and parts and to the constraints between parts (Fig. 6). The evidence variables are given by instances associated with the obligatory parts (e.g. face edges, region textures) and the evaluations of constraints (between edges). The intermediate variables are: concepts for parts and modalities of the "cuboid" concept. The hidden cause is represented by the "cuboid" concept itself. Discrete conditional pdfs are associated with links of the Bayesian network. These pdfs are created automatically from the description of modalities, parts and properties of a concept. Every such entry is characterized by a priority value and these values are converted into discrete pdfs.

4 Inference Rules and Recognition Strategy

Several inference rules are proposed for the creation of instances and modified concepts of the model concepts. In order to activate an inference rule a model-to-image data matching process needs to be performed. We view this matching as

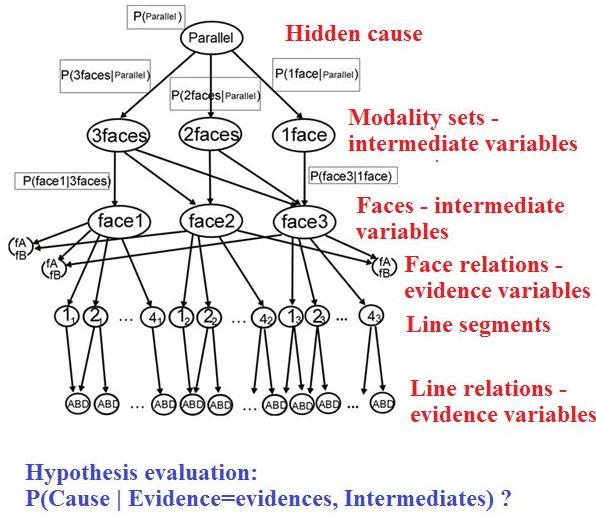


Fig. 6. The structure of a Bayesian network created for a simplified cuboid concept

a solution to constraint satisfaction problem (CSP), supported by Bayesian net-based evaluation of partial variable assignments. A modified incremental search for CSP solution is designed. It allows partial solutions and it calls for stochastic inference in order to provide judgments of partial states [6]. Hence the detection of partial occlusion of objects is handled consistently with Bayesian inference over evidence and hidden variables.

4.1 Inference Rules

Processing by the virtual receptor can be in one of three stages - in the initialization stage, in the question-answering stage (which means an active recognition process) or in the waiting stage.

During the initialization stage the client (control or ontology) passes object models and eventually a 2D default metric map of the environment.

A question can take one of two forms: either "verify an existing object (hypothesis) of given type and extend the 3D map" or „perform full object recognition and get an initial 3D map". The image recognition strategy uses several "IF-THEN" inference rules:

- Rule 1 (bottom-up inference): create a partial instance of a concept (w.r.t. to some modality set) on the basis of an established assignment (due to a CSP search) between obligatory parts of a given modality and data instances;
- Rule 2 (top-down inference): given a partial instance create modified concepts of its optional parts w.r.t. a given modality;
- Rule 3 (bottom-up inference): create a full instance from a partial one after an assignment for its optional parts has been found (by CSP search);

- Rule 4 (top-down inference): propagate constraints from a modified concept to its parts.

4.2 Recognition Strategy

1. Get the goal instance or concept from the question statement
2. Extend the goal in a top-down manner:
 - Apply Rule 4 – recursively create modified concepts of parts of existing modified concepts related to the goal.
3. Creation of hypotheses (partial instances)
 - Apply Rule 1 - partial instances of a concept $I_{\text{partial}}^{(i)}(C)$ are created when its obligatory parts have been assigned to important data instances (e.g. large and well detected image segments); for every $I_{\text{partial}}^{(i)}$ estimate its attributes $a'(i) = [S_k, R_k, t_k, \text{shape}]$ (scale, rotation, translation, shape).
4. Verification of hypotheses (complete instances):

FOR every new hypothesis $I_{\text{partial}}^{(i)}(C)$ DO

 - Apply Rule 2 - propagate its $a'(i)$ to optional parts of a given concept C (make modified concepts of such parts).
 - CSP search - matching of modified concepts of optional parts to data instances.
 - Apply Rule 3 - make a complete instance from a partial one - estimate its attributes $a(i) = (S_k, R_k, t_k, \text{shape})$.
5. Set an object lattice - set a disjunctive relation between every two competing object hypotheses.

Thus, the answer returned by the virtual receptor consists of two data structures: 1) the current updated 3D metric map of the environment, and 2) a lattice of object hypotheses generated for the sensor data - the types of instances are consistent with the question.

5 Conclusions

The aim of this work has been to design a general-purpose perception module for robot systems. It meets new objectives that, to our knowledge, have not yet been available in the Robot Operating System. Some elements of this concept have already been implemented, some others are in the development stage. First, an integrated segmentation procedure for color image and depth map has been developed [1]. The image segmentation process is a robust one as not only characteristic points and their features are detected, as in most of the available ROS packages, but also other typical discrete segments such as edges, regions and textures. The general-purpose object modeling framework uses generic object models (representing object types and not only instances, as in the current ROS perception) [6]. We implement a multi-representation of objects (a 3D model for a foreground object with depth information, or a 2D shape for a distant object and alternative templates for deformable shapes [8]). Complex object types are

application-dependent and they need to be defined in terms of basic solids and textures, recognized by the perception module. The recognition of multi-object scenes is possible - competitive hypotheses are returned in form of a hypothesis lattice (the final selection of best hypotheses is expected to be done by the caller module).

Acknowledgement. The financial support of National Center for Research and Development, Warsaw, Poland, grant no. PBS1/A3/8/2012, is gratefully acknowledged.

References

1. Stefańczyk, M., Kasprzak, W.: Multimodal segmentation of dense depth maps and associated color information. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 626–632. Springer, Heidelberg (2012)
2. RoboEarth, <http://www.ros.org/wiki/roboearth>
3. The Blocks World Robotic Vision Toolbox (BLORT),
http://ros.org/wiki/perception_blort
4. ODUFinder, http://www.ros.org/wiki/objects_of_daily_use_finder
5. Interactive Perception, <http://www.ros.org/wiki/iap>
6. Kasprzak, W.: Integration of different computational models in a computer vision framework. In: CISIM 2010, CFP1040C-CDR, @2010 IEEE, pp. 13–18 (2010)
7. WG Object Recognition,
http://wg-perception.github.io/object_recognition_core/
8. Wilkowski, A., Kasprzak, W.: Hand gesture modeling using Dynamic Bayesian Networks and Deformable Templates. In: SITIS 2011, pp. 390–397. IEEE Computer Society (2011)
9. Mozos, O.M., Triebel, R., Jensfelt, P., Rottmann, A., Burgard, W.: Supervised semantic labeling of places using information extracted from sensor data. *Robotics and Autonomous Systems* 55(5), 391–402 (2007)
10. Niemann, H., Sagerer, G., Schroder, S., Kummert, F.: ERNEST: A semantic network system for pattern understanding. *IEEE Trans. PAMI* 12, 883–905 (1990)
11. Kasprzak, W.: A Linguistic Approach to 3-D Object Recognition. *Computers & Graphics* 11(4), 427–443 (1987); Pergamon Journals, London, UK
12. Russel, S., Norvig, P.: *Artificial Intelligence. A modern approach*, 2nd edn. Prentice Hall (2002)
13. Murphy, K.P.: *Dynamic Bayesian Networks: Representation, Inference and Learning*, Ph.D. thesis, University of California, Berkeley (2002)
14. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification and Scene Analysis*, 2nd edn. J. Wiley, New York (2001)

Kinematic Structures of Functional Assemblies of the Table for Patients Verticalization with Lower Limbs Rehabilitation Functions

Wojciech J. Klimasara and Dariusz Grabowski

Industrial Institute for Automation and Measurements PIAP, Warsaw, Poland
`{wklimasara, dgrabowski}@piap.pl`

Abstract. The paper presents review of kinematic structures of mechanisms and drives of the selected tilt-tables for patients verticalization used in the process of their diagnostics, treatment and rehabilitation, considering the tilt-tables equipped with mechanisms for exercising lower limbs.

Keywords: Tilt-table, medicine, improvement exercises performance, cardiologic rehabilitation and brain-trauma rehabilitation.

1 Introduction

Tables for patients verticalization are medical devices used in the process of diagnostics, treatment and rehabilitation of patients. They are used to change position of bedridden patient, lying on the table, who is properly secured with belts against movement and falling down. Medical practice shows that even periodical change of a bedridden patient from horizontal to vertical position (or close to it) can positively affect the circulatory system, nervous system and other internal organs. Tilt-tables applied in the medicine are of various designs. They differ with equipment, kinematic structure of applied mechanisms, range of angular movements of the lying part as well as controls and drives. The simplest tilt-tables allow to change the angular position of the lying part (and of the patient lying on it) using manually driven mechanism, e.g. screw and crank mechanism. Tilt-table mechanism drives are commonly executed using button-controlled (start-stop) electrical motors. There are also servomotors. More complex designs allow the lying part to be “bent” e.g. to two or more segments, this way the angular position of individual segments can be individually set by the therapist. To do so, each segment is equipped with separate drive. The most advanced designs of the verticalization tables employ devices, which allow the lying patient to exercise the limbs. Hocoma’s Ergo tilt-table is equipped with a mechanism that allows lying patient to exercise lower limbs in the form of stepping. Such exercises are supported by an exercising device. The device is equipped with controlled and programmable drives and force sensors. The paper reviews and analyses the kinematic structures of tilt-tables as well as drive and control systems.

2 Application of Tilt-Tables in Therapy

2.1 Purpose of Tilt-Tables

In medical terms, tilt-tables belong to mechanical medical devices group called static verticalizers. Verticalizers allow a handicapped person to stay in a standing position for a long time, which is more natural posture. Long-term failure to take the standing position by a patient may lead to the following afflictions:

- decubital ulcers,
- halosterosis, especially of lower limbs,
- impairment of peristalsis,
- reduction of lungs vital capacity,
- internal organs degeneration,
- reduction of cardiac output,
- reduction of muscle heating function.

Verticalization is performed in order to:

- minimize the aforementioned afflictions,
- improve psychical condition of verticalized patient,
- improve life comfort by making any operations in standing position,
- allow moving in vertical position.

Verticalizers are applied in the following afflictions:

- lower limbs paralysis (paraplegia),
- four-limb paresis (tetraparesis),
- four-limb paralysis (tetraplegia),
- SM – sclerosis multiplex,
- demyelinating neurological disorder,
- cerebral palsy,
- myelomeningocele,
- spinal trauma condition.

Tilt-tables are devices applied in bedridden patients therapy. They do not require any cooperation of a patient. Angle of inclination of the lying part of the table can be changed manually or using electrical cylinder button-controlled by a therapist. In many cases electrical drive systems are used, operation of which is supported by gas springs. During verticalization, patient is properly secured against sliding or falling down. To do so, he or she is strapped to the table by his or her chest, hips and knees. In some cases, it may be necessary to support a patient by armpits. Patient's feet should be supported on footrests and secured with straps against sliding out. The table can be equipped with functional tabletop or arm supports e.g. to perform operations related to eating meals. During verticalization, it might be useful to check patient's pulse and blood pressure and observe his / her condition (including observation of eyes, face).

In some medical examinations related to diagnostics and efficiency of treating syncope in disordered autonomous control of cardio-vascular system, called

vasovagal syncope [7], verticalization tests [4] are performed that require to apply so called mobile inclination tables, which allow gradual, controlled attaining the vertical position of the table at specified rate, which can be changed within a vast range (from low to high, e.g. 45°/s). Such tables are not part of the paper.

2.2 Tilt-Table Construction

Utility features of tilt-tables are best described by the following functional parameters:

- dimensions of the lying part: width and length in cm,
- method of bending the lying part,
- height of the lying part: min, max in cm,
- headrest lifting angle: min, max in degrees,
- inclination angle: max in degrees,
- Trendelenburg position: expressed in degrees.

Figure 1 presents an example of simple tilt-table of Techmed.



Fig. 1. Tilt- table SP-1/E of TECHMED

The tilt-table consists of the following assemblies:

- frame of the base equipped with a stand as well as swivel wheels and brakes,
- lying part frame installed using articulated joints to the stand,
- linear movement electrical cylinder installed with one end to the lying part frame and with the other to the stand,
- footrest for patient's feet.

The tilt-table is also equipped with top and elements for patient's head stabilization. Table construction does not allow changing the lying part position height.

3 Kinematic Schemes of the Selected Tilt-Tables

Kinematic system structures of the tilt-tables and their drives can be divided to the following groups:

- a) lever and joint systems,
- b) scissors systems,
- c) other structures, e.g. telescopic.

3.1 Lever and Joint Structure

Figure 3 presents a scheme of tilt-table of lever and joint structure.

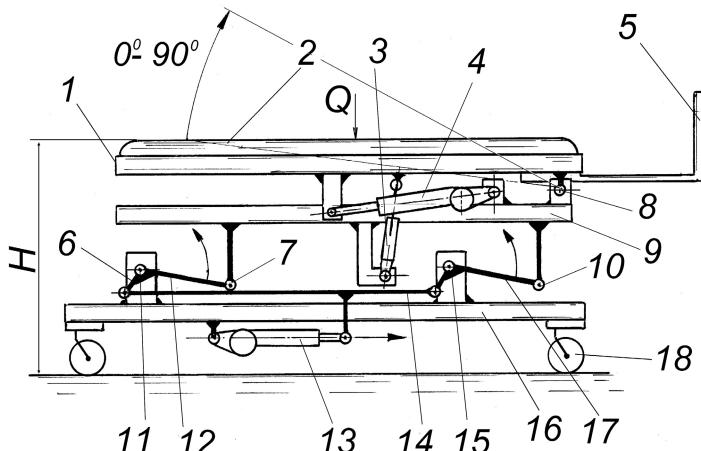


Fig. 2. Kinematic scheme of tilt-table of lever and joint structure: 1 – lying part frame, 2 – lying part, 3 – compression gas spring, 4 – electric actuator for lifting the lying part, 5 – footrest, 6 – angular lever arm of the lying part lifting mechanism, 7 – lever joint, 8 – joint fixing the lying part frame – pos. 1 to the frame – pos. 9, 9 – table frame, 10 – joint of lever – pos. 17, 11 – angular lever joint, 12 – angular lever arm, 13 – electrical actuator of the frame lifting mechanism drive – pos. 9, 14 – connector, 15 – angular lever joint, 16 – frame of the base, 17 – angular lever, 18 – wheel.

Two following functional assemblies are applied in the table kinematic system:

- lifting / lowering the lying part assembly, and
- lying part frame angular position change assembly – pos. 1.

Lying part frame – pos. 1 is installed using an articulated joint to the frame of the table – pos. 9 in the lying part lifting/lowering assembly. The assembly allows controlling the lying part height in relation to the floor.

These assemblies are driven by self-braking, electrical linear-movement actuators, which are button-controlled (start – stop) by an operator.

The lying part lifting/lowering assembly houses two two-arm angular levers pos. 12 and 17, which at locations pos. 11 and 15 are fixed using articulated joint to the table base frame – pos. 16. One pair of the angular levers arms is fixed using articulated joint with each other using a connector pos. 14. The connector is connected with the spindle of the electric actuator – pos. 13. Linear movement of the actuator spindle causes simultaneous rotation of both angular levers by the same angle leading to identical shifting of the joints pos. 7 and pos. 10 which are connected to the frame of the table – pos. 9 using connectors. These elements create spatial rhomboid, where the side connected to the lying part frame remains parallel to its base. This allows maintaining fixed angular position of the lying part when changing its height.

Kinematic assembly for lifting/lowering the lying part frame is driven by self-braking, linear-movement electrical actuator – pos. 13. This assembly consists of two identical angular levers – pos. 12 and 17, which are fixed using articulated joint in points 6 and 15. These levers are connected using a connector pos. 14 which ensures that angular motions of both levers are synchronized. This allows maintaining fixed horizontal position of the frame 10 during its lifting and lowering.

The kinematic assembly for changing angular position of lying part 2 consists of a frame – pos. 1, self-braking, linear-movement electrical cylinder – pos. 4, which is fixed with one end to the frame – pos. 9 and with the other. Gas spring – pos. 3 is to aid the electrical cylinder – pos. 4 while verticalization of the lying part – pos. 1. Operation of the gas spring significantly supports the actuator, especially during the initial phase of the verticalization motion, when the actuator spindle operating arm in relation to the table turn axis is small.

Construction of many tilt tables of lever and joint structure is similar.

Fig. 3 presents kinematic scheme of GYMNA tilt-table of lever and joint structure. Three drives are installed in this table which are provided by self-braking, linear-movement electrical actuators. Lying part is separated. It consists of two segments. There is a headrest with own drive. Kinematic structure of the lifting/lowering mechanisms of the lying part and its angular positions (verticalization) is similar to the one presented in the Fig. 2.

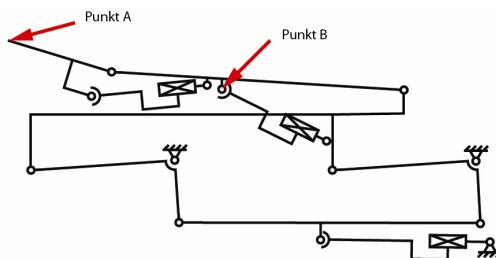


Fig. 3. Kinematic scheme of GYMNA tilt-table

Fig. 4 presents spatial scheme of GYMNA tilt-table construction

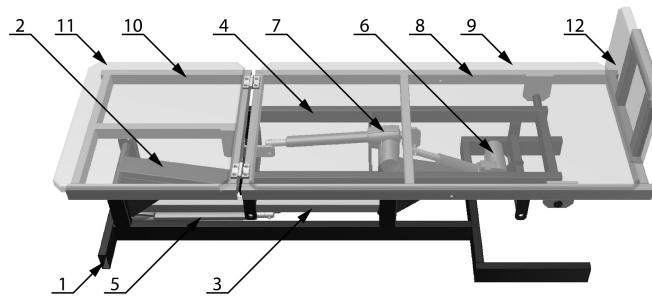


Fig. 4. Construction scheme of GYMNA TILTING ADVANCED table: 1 – frame of the base, 2 – level of the table lifting/lowering mechanism, 3 – connector, 4 – frame, 5 – linear-movement electrical actuator for lifting/lowering the table, 6 – linear-movement electrical actuator of verticalization mechanism of the lying part, 7 – headrest control cylinder, 8 – lying part frame, 9 – lying part cover, 10 – headrest frame, 11 – headrest cover, 12 – footrest

3.2 Scissors Structure

Lying part lifting mechanism can also consist in other principle, e.g. scissors mechanism. Kinematic scheme of such mechanism is presented in the Fig. 5.

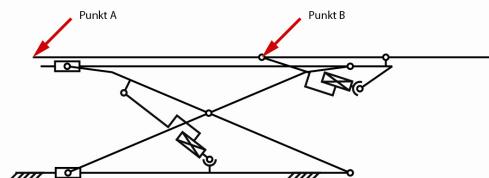


Fig. 5. Kinematic scheme of tilt-table with lying part lifting/lowering scissors mechanism

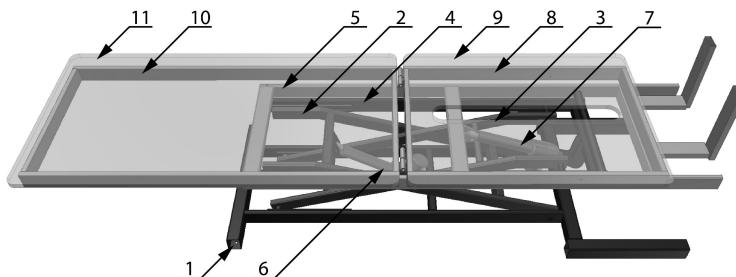


Fig. 6. Construction scheme of tilt-table with lying part lifting / lowering scissors mechanism: 1 – frame of the base, 2, 3, 4 – elements of scissors mechanism frame, 5 – frame, 6 – electrical cylinder for lifting / lowering the lying part, 7 – electrical actuator for lying part verticalization, 8 – lying part frame, 9 – lying part cover, 10 – backrest segment frame, 11 – backrest segment cover

Figure 6 presents spatial construction scheme of table with lying part lifting/lowering scissors mechanism. Two-segment lying part is applied in case of this table. For better comprehension, the figure does not present backrest segment drive.

3.3 Other Structures

Lever and joint as well as scissors structures belong to the most common kinematic structures of tilt-tables. There are also other structures, e.g. telescopic structures. Tables built this way allow tilting the lying part by a specified angle in relation to longitudinal axis of the table. These tables are complex and expensive.

4 Mechanisms of Exercising Lower Limbs

4.1 Purpose

As mentioned in the introduction, advanced designs of modern tilt-tables are equipped with kinematic mechanisms for passive and active performance of rehabilitation exercises of lower limbs. These tables allow performing passive exercises of patient's lower limbs, who is secured to the lying part using proper methods. The patient is protected against sliding or falling down providing simultaneous freedom of lower limb movements when exercising. Patient's position during the exercises (lying, semi-lying or standing) as well as intensity of exercises is determined by a therapist using proper positions of the lying part and settings of the exercising mechanism. During exercises, lower limbs movement is forced by mechanisms of proper kinematics. Twin mechanisms are employed, one per each limb.

These mechanisms are driven with electrical servomotors. Control system of these mechanisms allows setting the rate and range of movement and synchronization of movement necessary to perform the exercises. The performed exercises are called stepping. These are not gaits, similar to the ones performed by a healthy person. Patient, on the tilt-table, performs stepping motions. Therapeutic value of such exercises is considerable. These exercises are performed by patients with severe neurological [9] or cardiologic [3] affections, within passive, safe mode, and when necessary with real-time control of the parameters of cardio-vascular system (pulse, blood pressure, etc.). Purpose of these exercises is to activate lower limbs muscles and joints as well as overall activation of cardio-vascular system. During the exercises, patient's condition (heart beat, blood pressure) can be monitored. Proper dosing of exercises activate lower limb muscles, affect blood vessels, cause better blood flow through veins in the patient's limbs leading to better perfusion.

4.2 Mechanism of Performing Exercises on ERIGO Tilt-Table

Tilt-table ERIGO of HOCOMA is an example of tilt-table equipped with lower limbs rehabilitation exercises mechanism. The table is presented in the Fig. 7. The lying part lifting and lowering mechanism of this tilt-table is of lever and joint structure, alike the one presented in section 3.1. Lying part consists of two segments. Angular position of the headrest is controlled with separate electrical drive.



Fig. 7. Tilt-table ERIGO of Hocoma

This table employs two, symmetrically located, in relation to longitudinal table axis, mechanisms for performing passive exercises of lower limbs. Both mechanisms are equipped with electrical servomotors. Patient's limbs are secured just above knees in clamps, which are positioned on the ends of levers of the mechanisms for performing movements of the left and right leg. The control system allows synchronization of both mechanisms. During exercises, these mechanisms force alternative bending of patient's knees leading to stepping movements. Patient's feet (in shoes) are fixed to the footrest located on single-sided levers composing the mechanisms for performing movements of feet in tarsus. Exercise performance mechanism is the integral part of the ERIGO tilt-table construction.

4.3 Exercise Performance Mechanism acc. to PIAP Concept

Within the scope of industrial examinations of the targeted project No: ROW-III-224-2012, called "Table for patients verticalization with lower limbs rehabilitation functions" exercises performance mechanism concept was developed at PIAP, destined to be applied as an option in lot produced tilt-tables without lower limb rehabilitation function. Figure 8 presents conceptual scheme of this mechanism.

The mechanism consists of the following assemblies:

- lever mechanism with drive provided by electrical actuators (pos. 2) applied to force the motion bending the legs in knees,
- footrest mechanisms, forcing movement of feet in tarsus.

When straightening the legs in knees, patient (without his/her will), presses the footrest with heels, overcoming the resistance of pressure springs (pos. 4), causing rotation of the lever on which the footrest is seated. While bending the legs in knees, alternative motion of heels upwards occurs, caused by the force of compressed springs. Mechanism acc. to Fig. 8, alike the one applied in ERIGO tilt-table, is used to perform stepping exercises, during which bending and straightening of lower limbs in the following joints takes place: hip, knee and ankle.

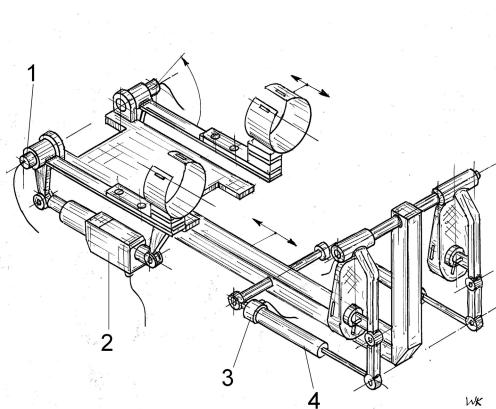


Fig. 8. Conceptual scheme of exercises performance mechanism. acc. to PIAP design: 1 – rotation angle sensor, 2 – electrical actuator, 3 – force sensor, 4 – enclosed pressure spring.

Figure 9 presents exploded view of the stepping mechanism acc. to solution presented in the Fig. 8.

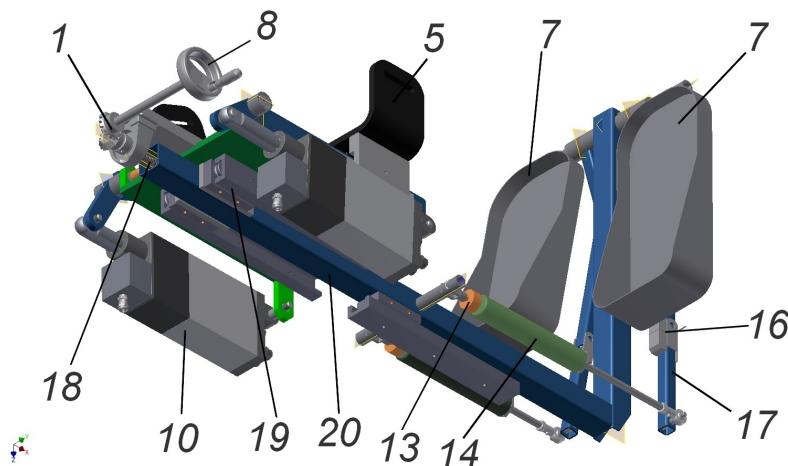


Fig. 9. Mechanism for performing the stepping exercises: 1 – toothed, angular gear for lifting/lowering the rests (pos. 7), 5 – patient's thigh clamp, 7 – footrests, 8 – footrest (pos. 7) lifting/lowering adjustment knob, 10 – servo-drive, 13 – patient's feet pressure force measurement sensor, 14 – tension spring (enclosed), 16 – lever clamp, 17 – lever, 18 – screw mechanism for footrest height, pos. 7, 19 – slide for adjustment of the position of knees lifting motion forcing mechanism, 20 – support frame

5 Tilt-Table VERTIMO

Within the scope of development works of the targeted project No: ROW-III-224-2012, called: "Table for patients verticalization with lower limbs rehabilitation

function”, Meden-Inmed company developed a construction of VERTIMO (Fig. 10) tilt-table, in which the stepping mechanism is integrated with the tilt-table. This tilt-table is of scissors structure.



Fig. 10. Tilt-table VERTIMO developed by Meden-Inmed [10]

Mechanisms for performing exercises are driven with electrical servomotors controlled by a control system developed by Meden-Inmed. This allows selection and adjustment of a series of exercising options, including: rate and scope of motions as well as mutual synchronization of motions of both legs.

6 Conclusion

The paper presents benefits resulting from application of tilt-tables during patients rehabilitation. It also shows advantages provided by lower limbs rehabilitation exercises while treating some affections. It explains issues related to mechanical design and drives of tilt-tables, including tables equipped with functions allowing execution of lower limbs exercises. Mechanism model construction for performing exercises is described, which is used as an option in the tilt-tables without the exercises function. It describes also the construction of tilt-tables: ERIGO of Hocoma and VERTIMO developed by Meden-Inmed. It is expected that due to benefits and advantages resulting from application of tilt-tables with lower limbs rehabilitation functions in the treatment of post-trauma patients and patients with cardio-vascular affections, their application will systematically increase. This will also be affected by relative reduction of tilt-table prices caused by relative reduction of construction components prices as well as competition on the market resulting from developing new construction solutions.

References

1. Bichsel, L., Sommer, M., Hunt, K.J.: Weiterentwicklung eines automatisierten Kipptisches mit Beinantrieb für die Regulierung der Patientenbelastung. Papers od Workshop AUTOMED 2010, Eidgenössische Technische Hochschule Zurich, Swiss Federal Institute of Technology, October 29-30 (2010)
2. Colombo, G., Schreier, R., Mayr, A., Plewa, H., Rupp, R.: Novel Tilt Table with integrated robotic stepping mechanism: Design Principles and Clinical Application. In: Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics, Chicago, IL, USA, pp. 227–230 (June 28, 2005)
3. Craven, C.T., Gollee, H., Coupaud, S., Purcell, M.A., Alln, D.: Investigation of robotic-assisted tilt-table therapy for early-stage spiral cord injury rehabilitation. *J. Rehabilit. Res. Dev.* 50(3), 367–378 (2013)
4. Drużbiński, M., Kwolek, A., Depa, A.: Pionizacja w procesie rehabilitacji chorych z objawami ogniskowego uszkodzenia ośrodkowego układu nerwowego – nowe możliwości aparaturowe. *Przegląd Medyczny Uniwersytetu Rzeszowskiego*, Rzeszów 1, 14–20 (2008) (in Polish)
5. Grubb, B.P.: Neurocardiogenic syncope. W: *Syncope*. In: Grubb, B.P., Olshansky, B. (eds.), pp. 73–106. Futura Publishing Comp., New York (1998)
6. Hesse, S., Schmidt, H., Werner, C.: Machines to support motor rehabilitation after stroke: 10 years of experience in Berlin. *Journal of Rehabilitation Research & Development* 43(5) (August/September 2006)
7. Kuznetsov, A.N., Rybalko, N.V., Daminov, V.D., Luft, A.R.: Early Poststroke Rehabilitation Using a Robotic Tilt-Table Stepper and Functional Electrical Stimulation. *J. Stroke Research and Treatment* 2013, 1–9 (2013)
8. Luther, M., Kreuer, C., Muller, F., Koenig, E.: Passive Beinbewegung bei der Mobilisation auf dem Kipptisch verhindert die Orthostasereaktion bei appallischen und minimal responsiven Patienten. *Neurochirurgie und Neurorehabilitation im Dialog*
9. Łukowicz, M., Kuczma, W., Hoffman, J.: Aktywna pionizacja pacjentów we wczesnej rehabilitacji neurologicznej. *Acta Bio-Optica et Informatica Medica. żynieria Biomedyczna* 14(3), s. 213–s. 216 (2008)
10. MEDEN-INMED – Materiały informacyjne
11. Müller, F.: New Technologic Approach to Minimizing Immobilization Effect of Patients with Brain Injury. International Brain Injury Association
12. Wieser, M., Gisler, S., Sarabadani, A., Ruest, R.M., Buetler, L., Vallery, H., Klamroth-Marganska, V., Hund-Georgiadis, M., Felder, M., Schoenberger, J.L., Gutknecht, C., Riener, R.:
13. Cardiovascular control and stabilization via inclination and mobilization during bed rest. *Med. Bio. Eng. Comput.* (2013)
14. Yoshida, T., Masani, K., Sayenko, D., Miyatani, M., Fisher, J., Popovic, M.: Cardiovascular Response of Individuals With Spiral Cord Injury to Dynamic Functional Electrical Stimulation under Orthostatic Stress. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21(1), 37–46
15. Zygmunt, A., Stańczyk, J.: Współczesne metody diagnostyki i leczenia omdleń wzorcowalnych u dzieci. *Przegląd Pediatriczny* 33(4), 255–260 (2003)

Basic 3D Solid Recognition in RGB-D Images

Tomasz Kornuta, Maciej Stefańczyk, and Włodzimierz Kasprzak

Warsaw University of Technology, Institute of Control and Computation Engineering,
Nowowiejska 15/19, 00-665 Warsaw, Poland
{T.Kornuta,M.Stefanczyk,W.Kasprzak}@elka.pw.edu.pl

Abstract. The paper deals with the problem of recognition of 3D objects for the purpose of their subsequent grasping and manipulation by a two-handed robot. We describe the idea of a general framework for object recognition rooted in the compositional model of the world. This approach treats complex objects as entities constructed of simpler, elementary ones, termed solids. In particular, we focus on recognition of two types of such solids: cuboids and generalized cones. We present details of their operation, starting from the low-level processing of RGB-D images and ending with the generation of hypotheses regarding the presence and parameters of those types of solids.

Keywords: RGB-D images, object recognition, recognition-by-parts, object primitives, solids, cuboids, generalized cones.

1 Introduction

The description of the objects present in the scene has been the major objective of computer vision since the advent of this branch of science. The combination of color and depth information on the one hand allows to overcome a number of classical computer vision problems, whereas on the other creates new problems and challenges. It also enables the revision of some of the old, abandoned ideas, which now, due to the computational power of modern computers combined with RGB-D sensors such as Microsoft Kinect or Asus Xtion, are able to work efficiently and in real-time.

In particular, this paper presents our recent progress in the development of a framework for 3D object recognition in RGB-D images (i.e. images containing three RGB channels enhanced with a fourth channel containing depth), which originates with one of our ideas from the mid-80s of the twentieth century [1]. In principle, the framework is used to produce perception subsystems of a robotic system realizing two-handed manipulation tasks. Hence we focus our attention on objects consisting of several simpler ones and potentially containing internal degrees of freedom (i.e. articulated objects), such as cupboards with doors, cabinets with drawers, jars with lids etc.

1.1 The General Concept of the Framework

The recognition strategy pursued by our framework is derived from the OBREC system (Object Recognition System), proposed in the work [1]. The original

strategy was defined as a process of generation of hypotheses, their evaluation, explanation of occlusions and verification of hypotheses. As a result, it merged both the bottom-up (data-driven) and top-down (model-driven) strategies.

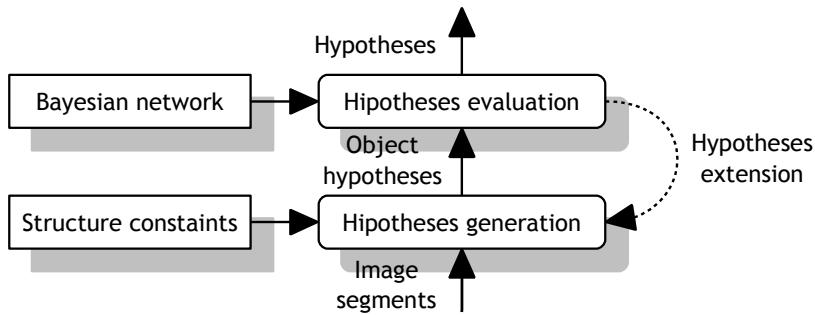


Fig. 1. General strategy of iterative 3D object recognition

The general idea of our recent recognition strategy is presented in fig. 1. There are two major operations distinguished, i.e. generation of hypotheses regarding present objects and evaluation of those hypotheses, ordered in a data-driven manner. The hypotheses generation bases on a declarative knowledge. Formulation of a single rule containing structure constraints covering the diversity of objects, their poses and possible occlusions would result in its incredibly high complexity. Hence, after the evaluation step we added a top-down transition, which goal is the extension of hypotheses. As a result the process forms a loop that enables building complex objects out of simpler ones, with the use of a number of simple rules. For the purpose of object modelling we use Semantic Network, i.e. a representation of knowledge in the form of a graph, with nodes (called concepts) corresponding to the types of objects and arcs describing the relationships between them. In our Semantic Network we distinguish two types of relations between concepts: Spec, representing the relationship between general and specialized objects (inheritance), and Part, constituting the relationship linking parts and complex objects. The structure of a Semantic Network determines the relationships between objects and their parts. The recognition process utilizes a discrete model of Constraint Satisfaction Problem (CSP) for the search of a (sub)optimal image segment to object model matching, in a way similar to the one presented in the work [2], coupled with the probabilistic inference in Bayesian Network used for the evaluation of hypotheses [3].

1.2 Contribution of This Paper

The aforescribed framework for 3D object recognition assumes that all objects are composed of a number of elementary three-dimensional objects, termed solids. Those objects are detected in RGB-D images. Fig. 2a and fig. 2b present the idea of decomposition of two exemplary solids into elements facilitating the

process of their recognition. Because we cannot rely on the texture of solids, the proposed solid recognition procedures consider only object shape, hence depth information. However, we store the textures of all faces of solids for the purpose of recognition of specific objects in the future.

It is worth noting that every image segment can belong to several hypotheses of solids at once – the subsequent inference performed in the Bayesian Network will select only the ones that fit the complex object hypotheses the most. In this paper we focus on two types of solids: cuboids and generalized cones.

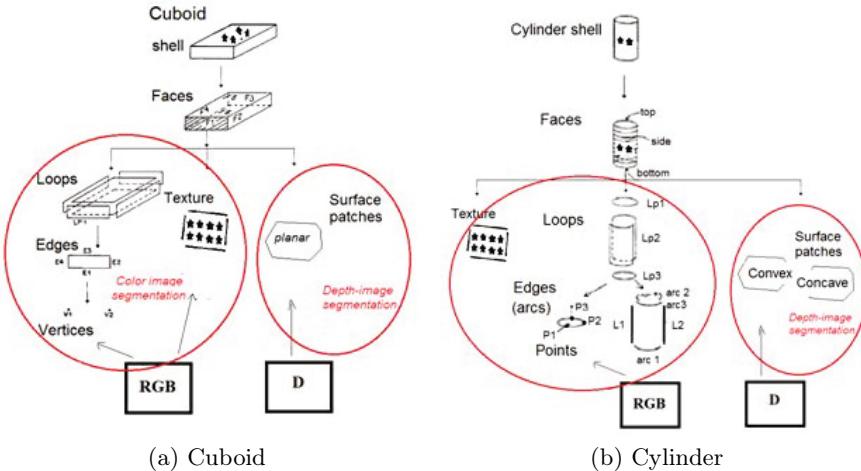


Fig. 2. Recognition of the exemplary solids - presentation of a general idea

The paper is organized as follows. In Section 2 we briefly present the origins and current research in the area of recognition of objects containing parts. In Section 3 we briefly describe the operations leading to the segmentation of RGB-D image. Sections 4 and 5 present rules enabling the generation of the extended hypotheses regarding cuboids and generalized cones, respectively. We end with the summary and short presentation of directions of our future research.

2 Related Work

The pattern recognition theory utilizes the compositional view of the world [4], by creating complex structures on the base and by the combination of simpler ones, usually called primitives. One of the first solutions using such an approach for analysis of a scene was described in the work [5], where the objects of interest (faces, human poses etc.) were decomposed into a set of basic elements specific to a particular class of objects, such as hair, eyes, nose and mouth for facial recognition or head, arms, torso and legs in the case of recognition of people. Aside of the decomposition into a set of primitives it was also important to

determine the spatial relationships that these primitives must fulfil. This work has in fact led the way to further research in primitive-based pattern recognition.

On the basis of representation of objects by the combination of primitive ones a theory called Recognition-By-Components (RBC) [6] emerged. In this approach object recognition was executed by a sequence of operations organized in layers. As a result of the operations performed in the first layer the image edges were extracted. Those in turn were integrated into segments, nodes, etc. in the second layer. The purpose of the third layer was the recognition of three-dimensional primitives (such as spheres, cones and cuboids) called geons (from geometrical ions), combined in the fourth layer into complex objects.

A number of works was devoted to general-usage representations of primitives. For example, the work [7] proposed to use two-dimensional shapes called silhouettes. Among the most interesting examples of three-dimensional primitives one can list deformable primitives called superquadrics [8], pictorial structures [9] and generalized cones [10]. Superquadrics enable the definition of a family of various geometric shapes (such as cubes or cylinders) with a single mathematical formula. Pictorial structures were used to identify human postures using collection of primitives arranged in a changing (i.e. deformable) configurations. Generalized cones were used e.g. in [11] for creation of a hierarchical structures, enabling representation of objects at different levels of detail.

Several of the recent works utilize the RBC-based approach to recognition of articulated objects. For example, in [12] the authors focused on the problem of human pose estimation and proposed Articulated Part Model (APM) with object represented by a mixture of non-oriented pictorial structures. The same problem was addressed in the work [13], where authors introduced representation of relations between object components with string-like deformable relations.

Another field of development of this type of object recognition methods is related to the utilization of depth information (RGB-D images). For example, in [14] the authors proposed the so called Hypotheses object Foreground Masks (HFM_s), which represent proposals of object parts detected in the RGB image by means of bounding boxes. Those hypotheses are next validated against the associated depth map. Such a combination of appearance and depth values within the HFM_s allows the determination of object part locations in 3D.

3 RGB-D Image Segmentation

Data flow diagram presenting the preprocessing operations leading to the segmentation of RGB-D images is presented in fig. 3. In the first step for each depth map point we estimate a normal vector, based on its neighbours. The depth map with the associated normals and colour image are next passed to the multi-modal segmentation procedure, described in details in [15]. The resulting image segments, aside of the pixels constituting given segment, contain several useful features, including center of the mass, mean normal vector, standard derivation

from the mean normal vector and rough classification of the segment shape (flat, convex, concave, indistinct), computed on the base of the latter. Those segments are passed to the procedure, which extracts the RGB features of the textures associated with them.

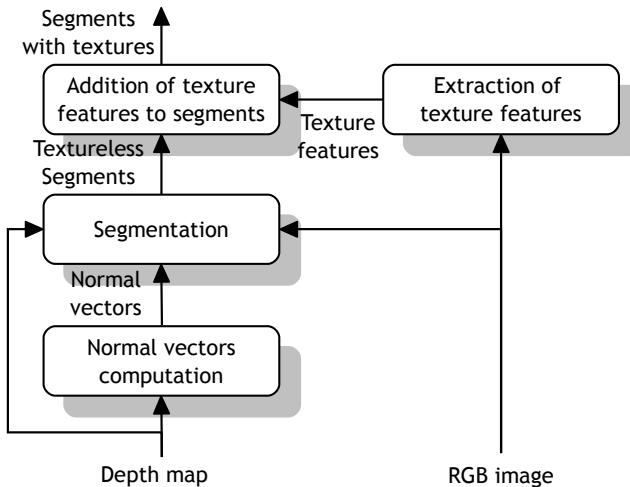


Fig. 3. Data flow diagram of RGB-D image segmentation

It is assumed that the resulting segments are the most elementary structures, constituting arguments for rules responsible for recognition of all types of objects, including the two aforementioned types of solids. Details of the rules responsible for generation of the hypotheses regarding those solids are presented in the following sections.

4 Cuboids Recognition

4.1 Model of the Solid

The cuboid is represented by three parameters (width w , height h and depth d) and a homogeneous matrix T describing the pose of the object in the camera reference frame. The model of a cuboid is presented in fig. 4a. We assume that the cuboid reference frame is associated with the center of its mass. In addition, each of the cuboid facets (image segments) stores a vector of features describing its texture.

In the case of cuboids three major cases concerning facet visibility (object modalities) are of interest to us, i.e. when one, two or three cuboid facets are visible. We assume that a flat, quadrilateral segment is treated as a basic cuboid hypothesis. First hypothesis extension merges two of neighbouring image segments (two-facet cuboid), whereas the second extension combines pairs of segments into triplets, i.e. three-faces hypotheses.

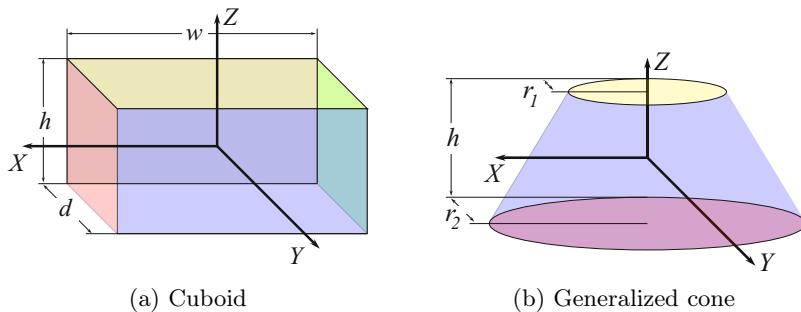


Fig. 4. Parameters of the models of solids

4.2 Basic Hypothesis

Generation of a basic hypotheses starts from finding a flat image segment – rough determination of the segment shape is realized during the preceding multi-modal segmentation. For every flat segment we extract its external edges and transform them into polar coordinates, with center of the segment mass used as the center of the transformation. Subsequently, we analyse the polar edges representation in order to find peaks (local maximas) regarding the distance from center – if exactly four maximas are found, we treat this segment as a cuboid facet. Because only one facet does not let us to determine the depth of the cuboid, it is assumed that the depth is equal to the mean of the length of facet diagonals.

4.3 Extended Hypothesis I – Two Facets

The main concept underlying the first extension of basic hypothesis concerns merging of neighbouring segments. We consider four major extension directions, indicated by vectors starting from the center of the mass and led through the centres of the sides of the quadrilateral. For every one of the found image segments we check its flatness and compare its normal vector with the normal vector of our basic hypothesis. If those vectors are orthogonal, we try to find two corners of the considered segment contour that are more or less adjacent to the corners of the segment constituting the basic hypothesis. If this condition is fulfilled, we generate and extended hypothesis. For every basic hypothesis four extended hypotheses might be generated.

4.4 Extended Hypothesis II – Three Facets

The goal of the second hypothesis extension is the generation of three-facet cuboid hypotheses. This rule is quite simple – every found triplet of extended hypotheses concerning the same three segments generates a single three-facet hypothesis.

5 Generalized Cones Recognition

5.1 Model of the Solid

A cone is defined as a three-dimensional geometric shape that tapers smoothly from a flat base to a point called apex. In our approach we assume that the base is circular and, aside of that, that the cone can be truncated. In such a case the apical part is removed and a second plane, parallel to the base appears. Besides that, we assume that the axis of the cone is perpendicular to its base. The model of a generalized cone is presented in fig. 4b. The cone is described by three parameters: height h and radii r_1 and r_2 of the upper and lower circular planes respectively. In particular, the upper circle radius r_1 can be equal to zero. Besides that, the model also stores the homogeneous matrix T , describing the pose of the cone in the camera reference frame. The frame is located half way from the lower to the upper plane (or to the vertex in the case of a cone with apex), on the axis of the cone.

In the case of a generalized cone there are five modalities that enable us to determine all of the required parameters of the model: two for the cone with the apex (we observe only the conical segment or the conical segment with a base) and three for the truncated cone (i.e. conical segment, conical segment with the upper plane or conical segment with the lower plane). Hence our most basic hypotheses (the ones composed of a single segment) are generated for all segments that passed the conical surface test.

5.2 Basic Hypothesis

The most important part of the generalized cone is the segment constituting the conical surface. For the purpose of its detection we utilize one of the most standard approaches, i.e. RanSaC (Random Sample Consensus) [16]. In the sampling phase we select three points. On the basis of those points, along with the associated normals, the parameters of equations of three planes are computed. Their intersection determines the pose of the apex of the cone (which is of interest to us even if it is only virtual, i.e. the considered cone is truncated).

On the other hand, those three random points lie on a plane, which intersects with the cone creating (typically) a conic section of an ellipsoid shape. Computation of the ellipse parameters enables us to find its center. On the basis of this center point and the cone apex we can compute the parameters of the line constituting the cone axis. Finally, the axis with one out of those three points is used for the computation of parametric description of the conic surface.

This function is used in the RanSaC consensus step for selection of the points fitting the hypothesis. If a number of inliers is sufficient, we classify the shape of the segment as conical and generate a basic hypothesis. If not, we repeat the process for several times.

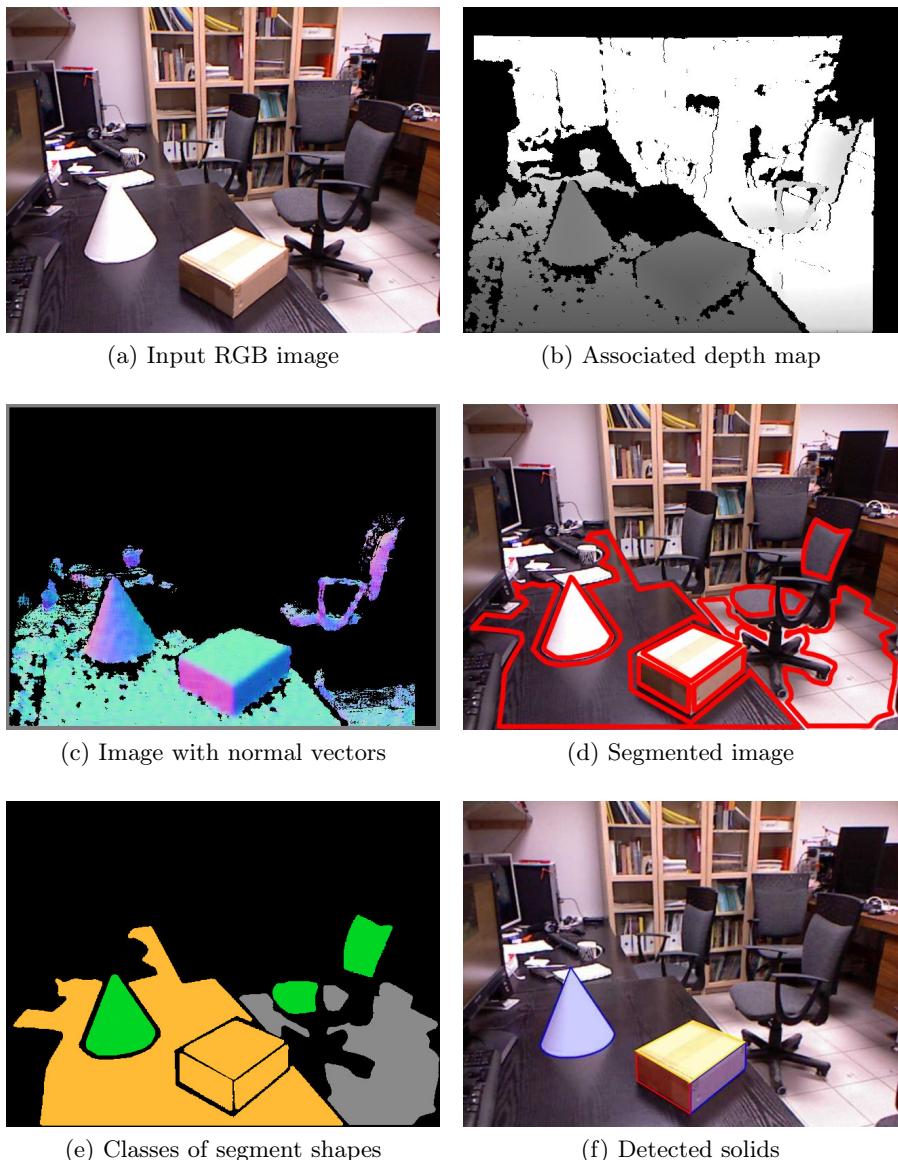


Fig. 5. Results of the operation of the presented recognition of solids

5.3 Extended Hypothesis

The knowledge of the cone axis enables us to search for neighbouring segments that are intersected by this axis and possess normals parallel to it. If those two

conditions are met for a given segment, we use RanSaC once again to fit it to an ellipse. In the case of success, we extend the hypothesis and treat the given segment as a lower or upper plane, taking into consideration the cone tapering direction.

6 Summary

In this paper we report our recent work, being a part of a bigger image – a general framework for recognition of 3D objects in RGB-D images. We shortly presented our motivation, described the overall idea of the framework, and finally we focused our attention on the recognition of two types of solids: cuboids and generalized cones. The described system was implemented in DisCODe (Distributed Component Oriented Data Processing) [17], a framework facilitating development of robot perception subsystems. In our implementation we used components encapsulating several open-source projects, including OpenKinect [18] (for acquisition of RGB-D images from the Kinect sensors), OpenCV [19] (data types, 2D data processing, extraction of textures) and PCL [20] (3D data processing, RanSaC). Fig. 5 presents exemplary results of the operation of our system.

Our future plans include the development of probabilistic inference in Bayesian Network enabling the evaluation of the hypotheses regarding the presence of 3D objects in the scene. Aside of that, because the recognition framework will have to work with the grasp planning system that currently is also under development, we also consider the addition of two new solids: spheres and cylinders. This is related to the fact that the planner distinguishes four types of primitives of the grasped objects, i.e. cuboids, cones, spheres and cylinders. However, it is worth pointing out that these objects are much simpler than the generalized cone, hence development of the rules enabling the recognition of the two new solids will be quite straightforward.

Acknowledgement. This project was financially supported by the National Centre for Research and Development grant no. PBS1/A3/8/2012.

References

1. Kasprzak, W.: A linguistic approach to 3-D object recognition. *Computers & Graphics* 11(4), 427–443 (1987)
2. Kasprzak, W., Stefańczyk, M.: 3D semantic map computation based on depth map and video image. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 441–448. Springer, Heidelberg (2012)
3. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River (1995)
4. Grenander, U., Miller, M.I.: Pattern Theory: From Representation to Inference. Oxford University Press (2007)
5. Fischler, M., Elschlager, R.: The representation and matching of pictorial structures. *IEEE Transactions on Computers* C-22, 67–92 (1973)

6. Biederman, I.: Recognition-by-components: a theory of human image understanding. *Psychological Review* 94(2), 115–147 (1987)
7. Blum, H.: Biological shape and visual science. *Journal of Theoretical Biology* 38, 205–287 (1973)
8. Krivic, J., Solina, F.: Part-level object recognition using superquadrics. *Comput. Vis. Image Underst.* 95(1), 105–126 (2004)
9. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(9), 1627–1645 (2010)
10. Nevatia, R., Binford, T.: Description and recognition of curved objects. *Journal of Artificial Intelligence* 8 (1), 77–98 (1977)
11. Marr, D., Nishihara, H.K.: Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 200(1140), 269–294 (1978)
12. Sun, M., Savarese, S.: Articulated part-based model for joint object detection and pose estimation. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 723–730. IEEE (2011)
13. Yang, Y., Ramanan, D.: Articulated pose estimation with flexible mixtures-of-parts. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1385–1392. IEEE (2011)
14. Kim, B.S., Xu, S., Savarese, S.: Accurate localization of 3d objects from rgbd data using segmentation hypotheses. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3182–3189. IEEE (2013)
15. Stefańczyk, M., Kasprzak, W.: Multimodal Segmentation of Dense Depth Maps and Associated Color Information. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 626–632. Springer, Heidelberg (2012)
16. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), 381–395 (1981)
17. Kornuta, T., Stefańczyk, M.: DisCODE: a component framework for sensory data processing. *Pomiary Automatyka Robotyka* 16(7–8), 76–85 (2012) (in Polish)
18. OpenKinect: libfreenect – drivers and libraries for the xbox kinect device on windows, linux, and os x (2010)
19. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library, 1st edn. O'Reilly (September 2008)
20. Rusu, R.B., Cousins, S.: 3D is here: Point Cloud Library (PCL). In: International Conference on Robotics and Automation, Shanghai, China (2011)

Selected Issues of Collecting Forensic Evidence with a Mobile Robot

Grzegorz Kowalski, Mateusz Maciąś, and Adam Wołoszczuk

Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland
`{gkowalski,mmacias,awoloszczuk}@piap.pl`

Abstract. The aim of this paper is to present the approach to the problem of collecting the forensic evidence within the CBRN contaminated crime scene with the use of teleoperated mobile robot. The idea of the whole system is presented first. While presenting a holistic approach to the problem is not feasible within a single paper, the article focuses on two selected solutions elaborated in R&D project entitled “Design of mobile platform for supporting forensic surveys in places where CBRN threats may occur”. The first solution is intended to gather chemical samples with the use of the SPME (Solid Phase Microextraction) adsorber. The second solution is designed to reveal the forensic evidence with the use of luminol solution deployed by the mobile robot. Conclusions, major issues addressed by described system and development trends are presented at the end of the paper.

Keywords: CBRN, forensics, mobile robots, robotic payloads.

1 Introduction

Mobile robots are commonly used in response to crisis events, with CBRN events being no exception [1]. Some cases, however, demand very specific approach, such as the CBRN events of criminal origin. In this kind of event, forensic technicians often face the need of collecting the samples in environments that pose a threat to human life or health. The contaminated crime scene or hazardous forensic samples cause a serious risk of technician exposure to CBRN agents. The sample itself has a critical value as an evidence and thus it requires careful handling and safe transport. In such case, the task of collecting a forensic sample is a complex issue combining safety during inspection of the contaminated crime scene, effectiveness of sample picking, security of the sample itself and safety of the technicians.

The R&D project entitled “Design of mobile platform for supporting forensic surveys in places where CBRN threats may occur” addresses issues stated above. In its main goals, the project aims to deliver a robotic system that will enable the scene inspection, sample collection and transport by remote operator. The technician operates the mobile robot from safe distance. The robot is equipped with a set of dedicated payloads for collecting forensic samples in various forms and physical states. Main tasks to be performed by the mobile robot include visual inspection with spherical

camera, sensor readings for CBRN substances presence in the crime scene, collection of the forensic samples and their transportation with respect to safety (in terms of not affecting the evidence integrity). As any object or substance within the crime scene may be a forensic evidence, the samples to be collected come in many forms and different physical states. In its basics the robotic system will handle samples of liquids, powders as well as physical objects. As some evidence (such as blood stains, cracks or bullet holes) may also be integral to unmovable entities such as ground or walls, the system will have the ability to reveal the evidence on such entity. The dual nature of the sample, being both an evidence as well as hazardous material, implies specific approach in regard of robotic platform, payload design and operators perspective.

2 General Principle of Operation

There are many design constraints that influence payloads construction and principles of operation. One of them is the preservation of the robot mobility and its manipulator movement capabilities. This is the reason for designing the system to be modular, so only payloads that are required are installed. Modularity requirements also come from scenario requirements – there could be a need to put multiple payloads of the same kind on the robot. For example, SPME adsorber can take only one sample at the time, so if more samples are required multiple devices have to be carried by the robot. There are various types of the payloads:

- samplers – payloads that collect objects and samples of suspicious liquids, powders, solids, volatile particles, etc. They are collected and transported back to the forensic technician and transferred for further laboratory analysis,
- sensors – payloads that provide information from environment directly to the operator. This information can be used for further mission planning or can be archived for post-mission analysis of the crime incident,
- support tools – tools supporting first two kinds of the payloads. Work of these devices is necessary for e.g. the acquisition of the sample.

Depending on the type of samples to be taken, the robot can be equipped with few different tools, which combined can carry out specific process. For example, taking blood sample from carpet often includes cutting part of the fabric, then picking it up and packing. For taking samples and picking up evidence robot has to be equipped with manipulator. Due to the fact that the manipulator can hold only one item at a time, payloads are carried on robot's base. They are picked up by gripper when needed and after usage are put back on robot's base. Components for holding payloads on robot base are called armory. After reaching the decision that the robot will be used in a mission (when there is a risk for humans), robot takes multiple trips from safe location to mission area and back. After every trip robot's payload set can be changed.

Due to the fact that robots take multiple payloads, and have to manipulate them on scene, the decision has been made to use radio communication between robot and

payloads. This has additional advantage that payloads can be sealed for easier decontamination. Robot and its payloads create 6LoWPAN (RFC 6282) network. Data from and to this network is relayed using main robot communication system to the robot control console (fig. 1).

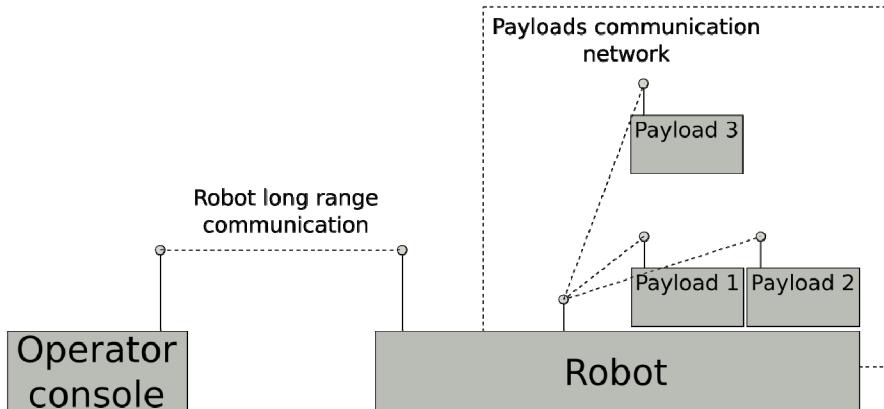


Fig. 1. System communication network layout

3 Collection of the Chemical Particles

During the forensic procedures many kind of samples could be collected. Chemical samples could be collected in various ways, dependent on the physical state of the sample. On the crime scene one can face fluid or solid state chemicals in the containers or spilled on the surfaces. Also, volatile and semi-volatile particles could be faced. One of the method of collecting the latter is adsorption, based on the polymer coated fused silica fiber. The process is called Solid-Phase Microextraction (SPME) and can be used to acquire particles from air and liquids [2]. The general rule of collecting particles using the fiber is reaching the point of chemical equilibrium between the fiber's coating and the medium that contains chemical particles. This process can take up to few minutes and runs itself without the intervention of a forensic technician. Afterwards, the fiber with collected particles can be put directly into the device, such as gas chromatograph, for particle analysis without the use of solvents or mechanical preparation of the fiber. In this way, the use of the SPME method for evidence analysis is faster and does not need any special treatment of the sample [3–5]. SPME method is used in forensics to acquire samples that could prove the use of chemical substances for criminal purpose. Samples, which can be collected for further analysis are: drugs, poisons, explosives and aggressive substances [6–8]. As the method is widely exploited by forensic technicians, the proposed system is equipped with the SPME fiber.

The commercial SPME fiber is hidden in the tool, which protects fine fiber from mechanical destruction (fig. 2). The fiber is exposed by a forensic technician, who

ejects it from the tool in the same way as a refill is ejected from a pen. The mobile robot is incapable of ejecting the fiber the same manner the person does. The decision was taken to use only the fiber in the needle (fig. 3, top), which is a part of the whole tool (fig. 3, middle).

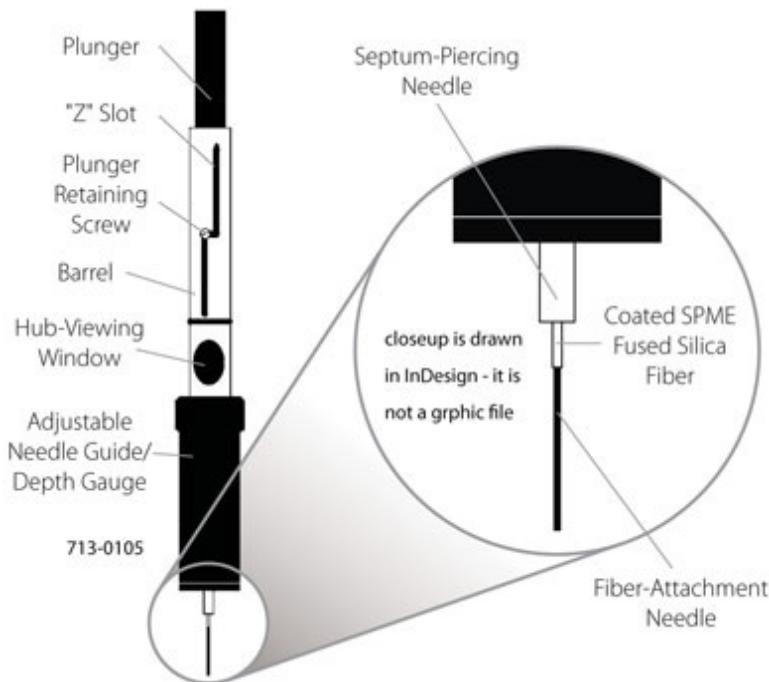


Fig. 2. SPME fiber in the housing (Source: www.sigmaaldrich.com)

The main part of the device is the needle with its propulsion mechanism – the electromagnet with the pushing shaft. When powered, the electromagnet ejects the fiber from the housing by the shaft. Whole needle is protected from damaging by the cap, which allows its fast exchange with the tightness of the device taken into account. On the top of the device there is the user control unit, which consists of capacitive buttons and status lights. The unit allows the user to power-on the device and eject the fiber with the push of the button. This functionality makes the device a dual-use one, as it can be used remotely on the robot and by the forensic technician. For simplification of the remote usage, the device was equipped in its own power source and radio communication. This solution enables the device to be used wherever the operator wants to within the range of the robot manipulator without exposing the power and communication wires on the risk of destruction. The battery used in the device is able to keep it in stand-by mode (ready to power-on) for almost half a year. For simplification of the charging process the device was equipped with wireless charging unit.

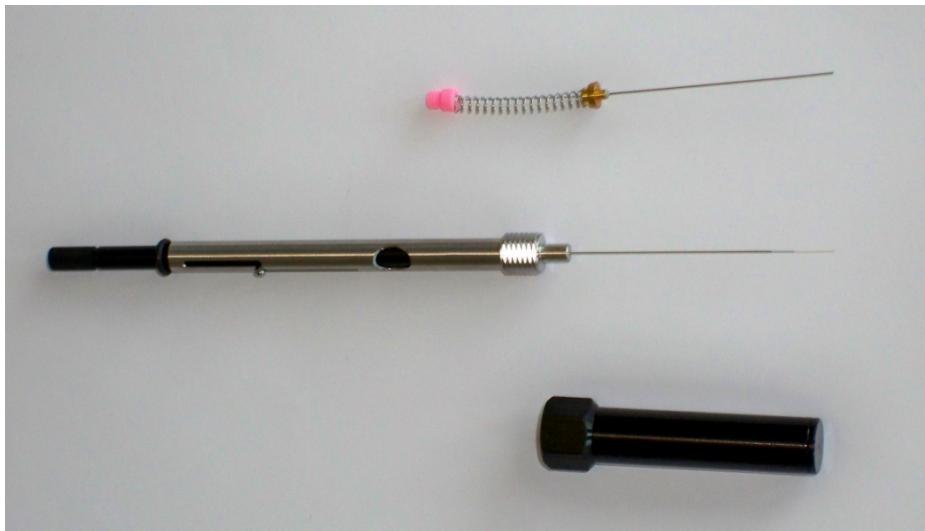


Fig. 3. The elements of the SPME tool: the fiber in the needle (top), needle in the housing (middle), protective cap (bottom) (Source: people.whitman.edu)

When presence of suspicious liquids or chemical pollution of air is expected or recognized, SPME adsorber (fig. 4) will be used.

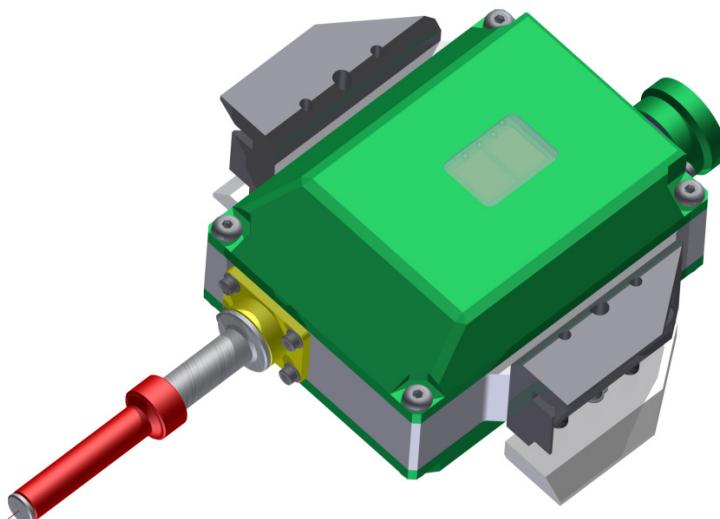


Fig. 4. CAD render of the SPME adsorber

The device will be mounted in one of the armory's slots in the back of the robot. The operator will grab the device using robot's manipulator and put it close to the object of interest or partially dip it in the suspicious liquid. Then, from the control

console the ejection of the fiber will be started. After few minutes the operator will hide the fiber, ordering it from the console, and put the device into to armory's slot. After completing the mission, which could involve collection of the other samples, and coming to the operator the device will be taken from the robot and disassembled in order to acquire the fiber from the device's interior. The fiber itself could be then analyses in the gas chromatograph. The body of the device and the fiber's housing could be flushed with a decontaminant to avoid contamination of the personnel and device's storage box.

4 Searching for Blood

During the investigation of some types of the crime scenes, a need to search for some biological evidence such as blood is likely to occur [9]. When found, blood traces are documented by forensic technicians with a camera. The blood can also be taken for later laboratory tests, like PCR or ABO group test. These tests help forensic technicians deliver another evidence for the investigation purposes. Sometimes, a criminal tries to cover the tracks of the crime and clean the blood stains with a liquid, like bleach. To reveal such tracks and hidden blood stains (e.g. dried blood on clothes, furniture, soil, etc.) technicians use liquid based on luminol, which causes that blood "lights" [10] (fig. 5).



Fig. 5. Blood stains sprayed with Bluestar® Forensic – luminol based liquid (Source: www.bluestar-forensic.com)

This is an effect of the luminescence, which is the product of the alkaline oxidation of the luminol solution with the iron, kept in the haemoglobin of the red blood cells, acting as a catalyst. The process produces blue light visible for humans for 30–60 seconds. After this period the intensity of the illumination is too weak to notice. Due to the fact that the luminol solution reacts only with a part of the blood cell components, the discovered blood can be afterwards used for ABO blood group tests [9, 11–13]. Negative aspects of the use of the luminol solution in the process of searching for the blood stains are related to the nature of the oxidation process and the interferences of some chemical compounds. The first problem can cause that the animal blood stains can be taken for human blood stains, because the luminol solution reacts with mammalian blood. The latter problem is the interference with commonly used chemicals (e.g. bleach) as well as with vegetables (e.g. carrot, horseradish) and fruits (e.g. watermelon) [11, 14].

Nevertheless, the use of the luminol is common. It is recommended to prepare the solution earliest few hours before usage, due to the short time of the chemical properties of the solution. It is dispersed on the suspected plains by spraying. The reaction – blue glow occurrence – is immediate. During this time it is recommended to take a picture of the evidence, due to the fading of the effect and the fact that every another spraying causes decrease of the intensity of the glow [15, 16]. Mobile robot equipped with dedicated tools is able to perform such activities. There are two tools, which can operate independently, depending on the type of the mission. One of the device is the sprayer for dispersion of the luminol solution, the other one is the camera for documenting the scene.

Main part of the sprayer is mounted to the robot's manipulator and contains the container with 0,5 l of the luminol solution. The idea of mounting the sprayer to the manipulator instead of creating it as one of the armory's tools was to preserve the robot's gripping functionality. This will be described later. The luminol solution is transported to the exterior of the device by pump through elastic tube and the nozzle, mounted to the manipulator's gripper. The device, just as robotic SPME adsorber, has got: its own power supply, wirelessly charged, radio communication and the user control unit. In the case of this device, the control unit enables manual spraying as well as filling the internal fluid container. This process will involve disconnection of the spraying tube from the main part of the device and connecting short tube with its second end dipped in the new luminol solution. Triggering the control unit button will start the filling process. The same functionalities will be available for the robot operator through the control console. Another tool – the camera – is the COTS (commercial-of-the-shelf), equipped with the functionality of real-time photo feed and cable control and added radio communication with the rest of the system. These functionalities enable to control the process of documenting evidence. While the operator sees the scene he is documenting he will be able to take pictures with accurate crop, zoom and focus.

The sprayer itself can be used to discover blood by the operator of the robot. Glowing stains will be recognized through the robot's on-board video camera and e.g. marked with the forensic number card. Additional camera is a tool, which can be taken from the robot's armory's slot and used to document evidence that are on the

crime scene. Combined, these two tools can perform the task of documenting hidden blood stains. This is the reason for mounting the sprayer to the manipulator (fig. 6). While having the gripper free, it can carry the camera for documenting short-lived glowing effect of the chemiluminescence. During the search for the evidence, first, the operator will spray the area with luminol solution and after the occurrence of the glow he will take a picture of the effect. Then, the area can be marked with forensic number card for further swabbing of the blood.

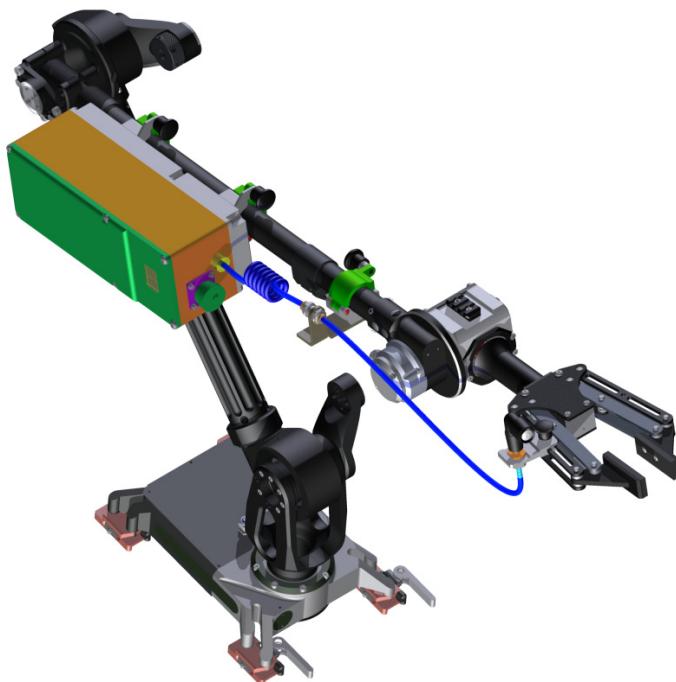


Fig. 6. CAD render of the luminol sprayer mounted to the robot's manipulator

5 Summary

The system is designed with strong end-users support, who are the members of the consortium. The requirements they state and feedback they provide make the system purposeful, while the scientific and engineering partners make it technically feasible in scope of the given requirements. Described solutions are the examples from a range of the forensic CBRN payloads for mobile robot covered by the project. With a system still being in development, it can already be noted that the use of the mobile robots equipped with dedicated payloads features a potential of significantly increasing technician's safety without hindering the critical aspects of operating within CBRN contaminated crime scene. As the time is not an essential issue during the crime scene inspection the operator may allow himself careful and precise control of robot

movement. However, during the project, major areas of technical risks affecting systems effectiveness were spotted. They are related to the fact that operator is controlling the system from a distance, which is both a system's flaw (in the context of operator's perception) and a virtue (in the context of operator's safety). The key issue of described solution is a trade-off between the forensic technician's safety and difficulty of the operating the robotic system in conditions of the contaminated crime scene. Remote operator's capabilities and perception are limited to those of the robotic system. The key issues that, in the context of current state-of-art, make the operation difficult are: operators perspective limited to what he can see on the console and evidence handling by a machine. The system addresses the operator's perspective issue by providing him ergonomic and legible user interface. The evidence handling issue is mitigated by dedicated payloads, designed with respect to requirements provided by the end user. Still, there is a vast space for improvement in terms of systems effectiveness, both on the robot's side and operator's side. The operator's perception and the facility of the operation can be enhanced by developing the intuition and multi-modal human-machine interface improving operator's awareness by visual (stereoscopic vision) and non-visual (tactile interface, force feedback, audio feedback) feedback. There is also a strong potential of implementing the supported teleoperation functionality to the system, that would bring operator's robot control to a higher level of abstraction allowing him to operate the robot by simple high-level commands (such as: grip indicated object, store the evidence, etc.), leaving the low-level task execution to systems autonomy.

Acknowledgements. The research described in this paper was funded from the polish National Centre for Research and Development under grant agreement no. 0015/R/ID 1/2011/01.

References

1. Humphrey, C.M., Adams, J.A.: Robotic Tasks for CBRNE Incident Response. *Advanced Robotics* 23(9), 1–14 (2009)
2. Ahmad, U., Selvaraju, G.: Forensic Detection of Fire Accelerants Using a New Solid Phase Microextraction (SPME) Fiber. *The Internet Journal of Forensic Science* 2(1) (2006)
3. fIELD Forensics (November 1, 2013),
<http://www.fieldforensics.com/forensics-sampling.html>
4. SPME/GC for Forensic Applications: Explosives, Fire Debris, and Drugs of Abuse. *Supelco Bulletin* 922 (1998), <http://www.sigmaaldrich.com/Graphics/Supelco/objects/4600/4546.pdf> (November 17, 2013)
5. Helmenstine, A.M.: Equilibrium Constant. *About.com/Chemistry* (2013),
<http://chemistry.about.com/od/equilibrium/a/Chemical-Equilibrium.htm> (November 18, 2013)
6. Pietryszyk, M., Matyjasek, Ł., Wachowicz, M.: Badania identyfikacyjne wybranych substancji niszczących powłoki malarskie. *Problemy Kryminalistyki* 260, 5–23 (2008)

7. Furton, K.G., Wang, J., Hsu, Y., Walton, J., Almirall, J.: The Use of Solid-Phase Microextraction-Gas Chromatography in Forensic Analysis. *Journal of Chromatographic Science* 38, 297–306 (2000)
8. Solid Phase Microextraction Application Guide. Supelco Bulletin 925D (2001), http://www.sigmaldrich.com/etc/medialib/docs/Supeco/General_Information/1/t199925.Par.0001.File.tmp/t199925.pdf (November 17, 2013)
9. Quickenden, T.I., Ennis, C.P., Creamer, J.I.: The forensic use of luminol chemiluminescence to detect traces of blood inside motor vehicles. *Luminescence* 19(5), 271–277 (2004)
10. Stene, I., Shimamoto, S., Gabel, R., Tewes, R., Adair, T.: Using Luminol to Detect Blood in Soil Eight Years after Deposition. *Journal of the Association for the Crime Scene Reconstruction* 19(1), 1–4 (2013)
11. Barni, F., Lewis, S.W., Berti, A., Miskelly, G.M., Lago, G.: Forensic application of the luminol reaction as a presumptive test for latent blood detection. *The International Journal of Pure and Applied Analytical Chemistry* 72(3), 896–913 (2007)
12. Lytle, L.T., Hedgecock, D.G.: Chemiluminescence in the Visualization of Forensic Bloodstains. *Journal of Forensic Sciences* 23(3), 550–562 (1978)
13. Rogiski da Silva, R., Agustini, B.C., Lopes da Silva, A.L., Frigeri, H.R.: Luminol in the forensic science. *Journal of Biotechnology and Biodiversity* 3(4), 172–177 (2012)
14. Quickenden, T.I., Creamer, J.I.: A study of common interferences with the forensic luminol test for blood. *Luminescence* 16(4), 295–298 (2001)
15. Dilbeck, L.: Use of Bluestar Forensic in Lieu of Luminol at Crime Scenes. *Journal of Forensic Identification* 56(5), 706–720 (2006)
16. Rohrig, B.: The Forensics of Blood. *Chem. Matters* 26, 4–7 (2008)

Lesson Learned from Eurathlon 2013 Land Robot Competition

Karol Majek, Paweł Musiałek, Piotr Kaczmarek, and Janusz Będkowski

Institute of Mathematical Machines, ul. Krzywickiego 34, 02-078 Warsaw, Poland
{karolmajek,pjmusialik,piotrakaczmarek1991,januszbedkowski}@gmail.com

Abstract. This paper shows evaluation result of the mobile robotic system for Urban Search and Rescue performed during Eurathlon 2013 robotic competition by IAIR-IMM team. Our team was competing in two scenarios: a) Reconnaissance and surveillance in urban structures (USAR), b) Search and rescue in a smoke-filled underground structure. The main task for this system from our team point of view was to build 3D metric map of the environment and to find OPIs (Objects of Potential Interest). Therefore in this paper we described the vision system for objects recognition and 3D map building. The system is composed of mobile robot equipped with camera, 3D laser measurement system and base station composed of computer equipped with NVIDIA GPU for parallel processing of derived clouds of points. The main focus of the work was to improve the performance of the operator controlling the robot in harsh environment. We achieved satisfactory results that could be still improved in many aspects. In experimental part we demonstrated validation of vision recognition system and 3D maps built during preparation trials and during final competition. The best quantitative result of this work was 3rd place in USAR scenario. Unfortunately, we could not build the map in a smoke-filled underground structure, but the result is also very interesting for future developments.

Keywords: Eurathlon, mobile robot.

1 Introduction

One of the applications of mobile robots is reducing devastation caused by natural or man-made disasters. In some situations, further damage could be prevented if there was a way of intervening directly at the disaster site. However it is often too dangerous or impossible to send human due to nuclear or chemical contamination, intense pressure or structural instability. Robots that are currently available are useful in such situations, but they are not always able to navigate in all environments and to perform required tasks. For these reason many institutions are developing mobile robots that can work in hush environment. For example US government Defense Advanced Research Projects Agency (DARPA) seeks solution to that problem and organizes the Disaster Robotics Challenge. The goal of the DARPA DRC is to develop a robot, which is capable

of operating in human engineered environment freely, using tools and vehicles designed for human. While there is no specific requirement that the designed robot should be humanoid, it seems that walking robot is the way to achieve the goal. The first stage of the competition, the Virtual Robotics Challenge, will take place entirely in simulator, therefore many of researchers can join the competition because of the decreased budget needed to obtain satisfactory results. In Europe we have ELROB (European Land Robot Trial) that is divided into M-ELROB (military purpose) and C-ELROB (civilian purpose). From 2013 new EU FP-7 project EURATHLON is running as an extension of ELROB to test multi robotic system (land , water, air). First EURATHLON 2013 has already finished and some satisfactory results and observations were taken during realistic trials. Our team (IAIR-IMM) was competing in two scenarios: a) Reconnaissance and surveillance in urban structures (USAR), b) Search and rescue in a smoke-filled underground structure. The result in scenario a) was successful because we could deliver 3D metric map of the environment, unfortunately the smoke in scenario b) successfully eliminated the use of 3D laser scanner (rotated LMS SICK 100) that was not parameterized for working in fog conditions. We observed that it is very difficult to use autonomous algorithms for almost all robotic tasks. But our methods efficiently helps operator in providing correct control decision. Therefore in this paper we are focused on support tools for robot operator. In section **Eurathlon competition** the Eurathlon 2013 competition is described. Section **Mobile Platform** shows our robot. Section **3D map building** briefly describes our mapping system. Section **OPI (Object of Potential Interest) detection** describes the vision system for object recognition. Final trials are demonstrated in section **Results**. Section **Conclusion** finalizes the paper.

2 Eurathlon Competition

The Eurathlon is a new robotics competition which provides realistic emergency-response scenarios. Eurathlon is divided into three competitions. First is land robot competition in 2013. Second is underwater robot competition in 2014. Last is "Grand challenge" in 2015 where team of land, underwater and flying robots will work together to survey the scene and collect environmental data. Eurathlon 2013 competition has taken place between 23rd and 27th September in Berchtesgaden in Germany. There were following scenarios:

- Reconnaissance and surveillance in urban structures (USAR)
- Search and rescue in a smoke-filled underground structure
- Mobile manipulation for handling hazardous material
- Autonomous navigation using GPS, GLONASS and GALILEO
- Reconnaissance and disposal of bombs and explosive devices

This paper shows IMM-IAIR team approach in the first two scenarios.

Reconnaissance and Surveillance in Urban Structures (USAR)

USAR scenario is a key emergency service task and can be divided into two tasks. First task is the search of the area of interest for OPI as described in [1]. Second task is the rescue of the designated target. Scenario take place in a dilapidated building which have two levels (ground and basement). If robot can not pass through the stairs to the basement it can use the elevator. Main part of the task is situated in the basement.

Objective. Approach the area of interest which is located in up to 10m distance from starting position. Enter structure, search for OPI, acquire imagery and reference the position of OPI in the geometric representation. Report gathered data to the control station. Whole search task should be done with highest autonomy possible.

Search and Rescue in a Smoke-Filled Underground Structure

The second search and rescue scenario have taken place in tunnel. The tunnel was filled with very dense smoke. The objective was the same as before - enter structure, search for OPI, acquire imagery and reference OPI position in map.



Fig. 1. Mobile robot during the USAR trial

3 Mobile Platform

We evaluated our Small Mobile Mapping System (SMMS) in realistic conditions. We are motivated by the fact that small mapping system can be easier deployed

on the side in USAR scenario rather than large machine. There may happen major difficulties to deploy large UGV when the transportation infrastructure is demolished. SMMS is supposed to be deployed in the field with the effort of one operator per robot. Our mapping system is composed of low cost 3D laser measurement unit (range 20m), inertial measurement unit (XSENS MTI-G), mobile platform (DrRobot robot) and computer equipped with NVIDIA GPGPU GF540M for parallel computing (fig. 1). Used hardware:

- Mobile Platform Dr Robot Jaguar 4x4 Wheel
- Laptop with NVIDIA GT540M with 96 CUDA cores
- 3D scanning system with SICK LMS100
- Camera
- Wi-Fi (IEEE 802.11ac) router

We presented major improvements into 6DSLAM algorithm based on real-time data registration and loop closing using semantic matching. Thus, our approach efficiently improves the applicability of 6DSLAM in realistic scenarios. Unfortunately we could not use all prepared functionality during Eurathlon, only scan matching was used as support tool for robot operator.

4 3D Map Building

ICP (Iterative Closest Point) algorithm is originally introduced by Besl and McKay [5]. ICP algorithm minimizes error function to align two 3D data sets. GPGPU ICP (General-Purpose computing on Graphics Processor Units ICP) implementation presented in [3] and detailed discussed in [4] was used for 3D data matching and map building purposes. It improves SoA algorithm shown in [8] what is also related with approach shown in [9]. The core component of 6DSLAM algorithm is ICP where KNN problem has to be solved to find pairs of points. Several implementations of KNN was shown in [7]. Our implementation provides almost the same convergence as state of the art algorithm, but GPGPU ICP is several times faster as described in [3]. Localization and mapping was our main research effort. We developed low cost shock resistant 3D laser range finder based on LMS SICK 100 mounted onto rotated head. To map environments without occlusions, multiple 3D scans have to be registered. After the registration phase performed in real time a semantic loop closing technique finds matchable scans. If their registration result falls below a threshold, these scans are considered as overlapped. Local maps and robot poses are organized into a graph. Once loop is detected and loop error is distributed over all scans, a 6-DoF graph optimization algorithm for global relaxation is used. The result is globally consistent 3D metric map. To deploy our SMMS into the field we decided to extend SoA 6DSLAM by

additional functionalities to cope with problems occurring in real task scenarios. The additional functionalities are:

- data decomposition (we implemented regular grid decomposition to improve Nearest Neighbor Search procedure by decreasing the computational time),
- data filtering and sub sampling (instead of filtering existing in SoA approach, our methods efficiently eliminates noise and dynamic obstacles such walking people),
- 6DoF GPUICP (we implemented new ICP methods using parallel computing to obtain real-time execution),
- new loop detector - Complex Shape Histogram (it is efficient loop detector using semantic information),
- loop error distribution over all scans (it helps LUM procedure to create consistent map).

5 OPI (Object of Potential Interest) Detection

One of the tasks in each scenario of the Eurathlon Competition was detection of OPI. To represent such objects ERICard number plates were used (fig. 2). According to the rules, for the detection to be accepted, a picture of OPI had to be made along with position and timestamp (accuracy of position should be 5m). For this task 3 different approaches were considered:

- Using SURF [2] descriptor of example OPI
- Using HOG(Histogram of Oriented Gradient) [6] method with an SVM trained on a set of positive and negative samples
- Creating a decision tree based on simple characteristics of the OPI

First two approaches were abandoned after initial tests. SURF works well only for OPI's very similar to the example one. On the other hand HOG method requires a large dataset to work correctly. The effort needed for constructing such a set was considered too high. Finally the third method was chosen. To



Fig. 2. Object of Potential Interest marker

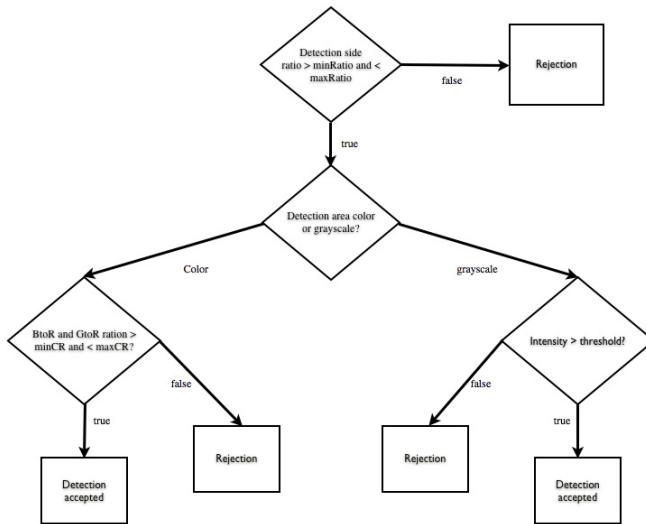


Fig. 3. Algorithm for Object of Potential Interest marker

create the decision tree the characteristic features of the OPI had to be taken into account. The choice was made to concentrate on the following:

- rectangle shape
- specific color scheme (black numbers and borders on orange background)
- high reflexivity of the surface

The base of the method is detection of rectangles in a camera image. Implementation from OpenCV library is used for this task. After finding potential detection areas each of them is tested according to the decision tree (fig. 3). At first step the ratio of the rectangle sides is calculated. The result has to be inside $\langle A, B \rangle$ boundaries. The second step depends on the type of image in which the OPI is detected. If the image is in color the ratios of averaged blue and green color to the red in the detection area are calculated. If they are inside given boundaries $\langle \text{minCR}, \text{maxCR} \rangle$ the detection is considered as correct. If the image in which the detection is made is gray scale the average intensity of the potential detection area must be larger than threshold to be considered correct. The parameters of the method were chosen to give best results for the camera of the robot. Chosen values are: $A=1.75$, $B=3.5$, $\text{minCR}=0$, $\text{maxCR}=0.75$, $\text{threshold}=180$. The last parameter had to be assigned even though the camera is a color one. This was because of camera behavior in low light environments in which it would start giving gray scale images. Detection method was tested on three different film samples gathered with the camera of the robot. The first one was made during the day in team's tent. Second one was from the cave that could be used for testing the system. the cave was completely dark, which

allowed to test the method only with robot's on-board light. The third film was taken during the night with minimal outside light. The results are shown in table 1. Cave and Night sets were tested for two values of the maxCR parameter. The change had no affect on the Day set. During the test number of positive and missed detections was counted. For the frame to be counted the OPI had to be recognizable for a human operator and at least half of it had to be visible in the image. It is important to note that most of missed frames had the OPI in large distance from the camera ($>6m$) where it was still recognizable for the operator but the detailed features were considerable blurred. Maximum distance of detection increased when the algorithm was tested on better hardware. Figure 5 shows the results of detection for partially covered and rotated OPI.



Fig. 4. Left: Detection in day sample; Right: Detection in cave sample; Bottom: Detection in night sample

Table 1. Results of detection algorithm testing. The results are show for 3 data sets with 2 different values of maxCR parameter

Set	Positive Frames	Negative Frames	Detection Rate	False Detection
Day Set - 0.6	414	173	0.705	22
Cave Set - 0.6	31	40	0.436	1
Night Set - 0.6	55	51	0.518	1
Cave Set - 0.75	41	30	0.577	2
Night Set - 0.75	77	27	0.74	3



Fig. 5. Left: Partial overlap detection; Right: Rotation 90°; Bottom: Rotation 180°

6 Results

Test Environments

During Eurathlon 2013 two test was done before the trials. First was done in the Team Tent that was set by organizers. Several scans was matched to check if mobile robot's and operator's side software works well. Model of team tent is shown on figure 6. Second test was done in cave near the Eurathlon site. This test was performed under real task conditions. Operator base station was deployed about 20m from the cave. There was no problems with Wi-Fi communication between robot and operator. Data were collected from several locations inside the cave and aligned to make one accurate model (fig. 6). IMM-IAIR as the only team provided 3D data.

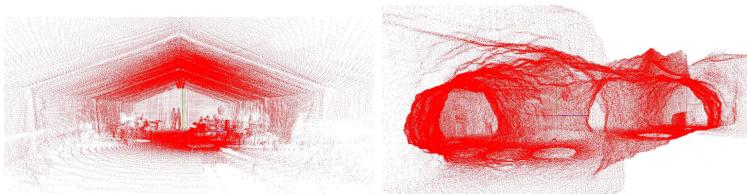


Fig. 6. Left: Laser data from teams tent at Eurathlon 2013; Right: Cave mapped during test before trial

Reconnaissance and Surveillance in Urban Structures

USAR scenario started about 5m from the entrance to the building. Robot entered the building and acquired 3D data. After first 3D scan robot moved toward stairs and wait for technician intervention. He moved robot half way down and operator executed second scan. Afterwards operator was aligning data sets to create consistent model (fig. 8). There where no odometry data from robot excluding raw encoder measures. Data was initially transformed by operator (initial translation and rotation) and finally corrected by GPGPU ICP. There were four more scans aligned into one model (fig. 7). When operator lost communication tried two times to manually move robot back to restore the connection. Unfortunately these tries failed and the task ended.

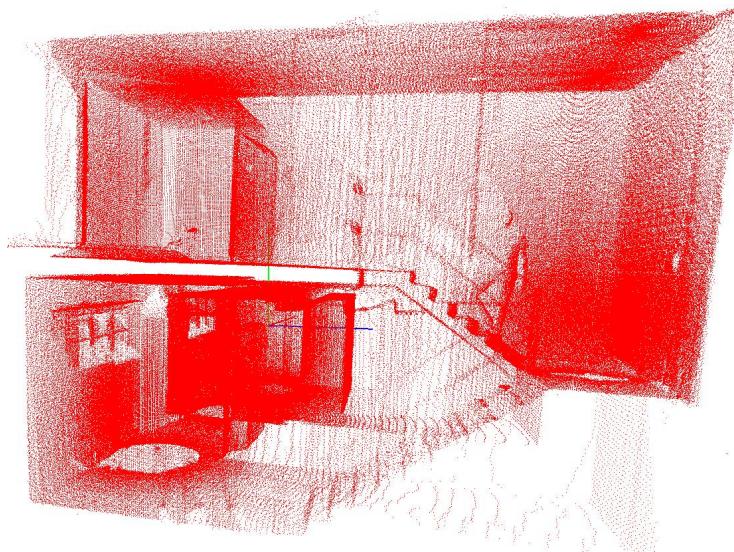


Fig. 7. Six point clouds aligned by the operator during USAR trial using GPGPU ICP

Search and Rescue in a Smoke-Filled Underground Structure

This scenario was very difficult because of the dense smoke inside the structure. Robot entered the tunnel and operator had tried to get 3D data from laser system. Unfortunately smoke was too dense and the distances measured by scanner was incorrect (fig. 8). 3D data was useless therefore operator started to look for OPI. Finally two potentially hazardous objects was found (fig. 9). Images of these objects was reported. In very last minute of the task the operator has found OPI marker, but there was not enough time to report it.

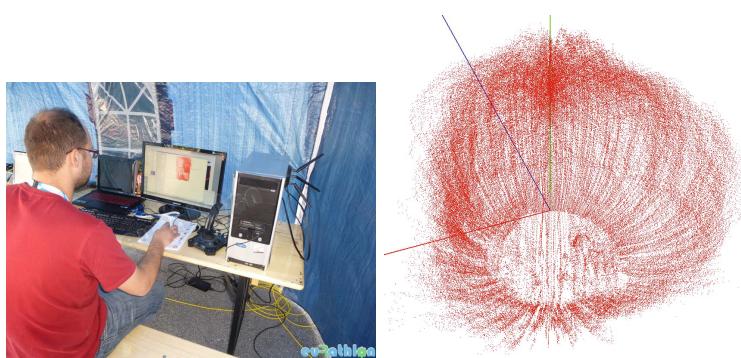


Fig. 8. Left: Mobile robot operator during the competition; Right: Laser data from smoke-filled tunnel



Fig. 9. Images of dangerous materials found in the tunnel

7 Conclusion

This paper shows our contribution into Euroathlon 2013 - robotic competition that helps evaluation of the robotic systems quantitatively and qualitatively. We can not claim that we are satisfied of our result, but lesson learnt from this work is that it is very difficult to use autonomy in realistic conditions. Support tool (3D map builder) efficiently helps operator in providing navigation decision. The parallel implementation of ICP algorithm allows an operator to acquire 3D maps on-line during the USAR task. To use this mapping strategy we should improve laser technology to be able to cope with smoke and fog. Another problem was with vision system that was very sensitive for realistic conditions such as illumination, fog, smoke. During the preliminary trials we observed satisfactory results, but realistic competition shows the influence of stress and some unpredictable factors that drastically decreases the performance of the system. We observed that simplified methods work better in realistic conditions. Complicated algorithms with sophisticated calculations are very sensitive for complex tasks.

Acknowledgments. This work has been supported by the European Union in the framework of European Social Fund through the Warsaw University of Technology Development Programme. Work supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 285417 Integrated Components for Assisted Rescue and Unmanned Search operations.

References

1. Eurathlon 2013 competition, <http://www.eurathlon2013.eu>
2. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
3. Bedkowski, J., Majek, K., Nüchter, A.: General purpose computing on graphics processing units for robotic applications. *Journal of Software Engineering for Robotics* 4(1), 23–33 (2013)
4. Bedkowski, J., Maslowski, A., de Cubber, G.: Real time 3D localization and mapping for USAR robotic application. *Industrial Robot* 39(5), 464–474 (2012)
5. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2) (1992)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, pp. 886–893 (2005)
7. Elseberg, J., Magnenat, S., Siegwart, R., Nüchter, A.: Comparison of nearest-neighbor-search strategies and implementations for efficient shape registration. *Journal of Software Engineering for Robotics (JOSER)* 3(1), 2–12 (2012)
8. Nüchter, A., Lingemann, K., Hertzberg, J., Surmann, H.: 6D SLAM-3D mapping outdoor environments. *Journal of Field Robotics* 24(8-9), 699–722 (2007)
9. Qiu, D., May, S., Nüchter, A.: GPU-Accelerated Nearest Neighbor Search for 3D Registration. In: Fritz, M., Schiele, B., Piater, J.H. (eds.) ICVS 2009. LNCS, vol. 5815, pp. 194–203. Springer, Heidelberg (2009)

Analysis of Thrust of Underwater Vehicle with Undulating Propulsion

Marcin Malec and Marcin Morawski

Cracow University of Technology,
Mechanical Engineering, Cracow, Poland
{malec,morawski}@mech.pk.edu.pl

Abstract. This article presents a preliminary study of determining the effect of the control parameters, shape and stiffness of the tail fin and the lateral fins on the thrust of the fifth version of polish biomimetic unmanned vehicle (BUV) called CyberFish. In the second paragraph the BUV's construction and its rule of motion is shortly mentioned. The next paragraph presents results of thrust measurements and conclusions from conducted tests. Conclusion summarizes influence of control parameters, tail fin and lateral fins on the average thrust of the CyberFish.

Keywords: Biomimetic Unamnned Vehicle, BUV, Underwater Mobile Robot, Thrust Force Measurements, Undulating Propulsion, Robotic Fish, CyberFish.

1 Introduction

In the recent years a rapid growth in biomimetic robotic constructions can be noticed. Particular attention is drawn to mobile robots that mimics the behavior of organisms inhabiting water environment. Some biomimetic underwater robots construction is presented in [1]. The fastest growing group of biomimetic underwater vehicles with subcarangiform or carangiform type undulating propulsion which are reflecting salmon or perch family fish. A characteristic feature of these vehicles is to perform the undulating motion using at least half the length of the body. Publication, [2 3] presents details of similar prototypes of BUVs. Some experiments on biomimetic underwater vehicles maximal speed was done by Wai Leung Chan et. all. [4]. The only difference is the type of the robot was ostraciiform which means that the tail consisted of only one segment. Similar experiments but with three-segment subcarangiform type fish was done by authors in [5]. In the available literature there is the lack of thrust analysis for different BUV's and control parameters influence on that thrust. Some research on influence of frequency and amplitude of osculation of robot's drive and type of the tail fin on the robot's thrust was done by C. W. Chong et. all. in [6]. In this paper authors presents results of experimental studies on average thrust of carangiform type biomimetic unmanned vehicle called CyberFish in extend

range of frequencies and amplitudes of robot's drive. Furthermore the influence of the tail fin and lateral fins on the thrust was examined.

2 The Experiment

2.1 Fifth Version of CyberFish

CyberFish is the biomimetic unmanned vehicle (BUV) which mimics Carp fish both in the external shape and its kinematics [7]. The hull of the robot was designed based on the 3D scan of a real fish in CAD software. All the necessary mechanical parameters (density of material, masses and arrangement of the internal devices etc.) to obtain close to neutral final buoyancy of the vehicle was estimated during the design process. DFM (Design For Manufacturing) and DFA (Design For Assembly) methodology was also taken into account. After the 3D model of the robot was done all the elements of the hull where manufactured from polyethylene (PE500) by means of CNC milling. Some internal parts for equipment integration was made of laser cut and welded stainless steel. Some fin elements was made of 3 mm rubber with the ability to detach them from the hull and replace them with different material. The robot's mechanical structure contains the payload segment (the head) in which the control board, communication modules and sensors are installed. Additionally in this section the variable buoyancy system (artificial swim bladder), the lateral fins servomotor and the first segment of the tail servomotor are placed. Tail section of the vehicle contains three segments connected with each other with class V rotary kinematic pairs. The first and the second segment of the tail also contains servomotors for their drive as well as additional batteries. The third segment of the tail is solid block of polyethylene with slot for mounting tail fin. Each servomotor of the vehicle (4 in total) is connected to RS-485 bus and enables for precise position and speed control. Drives over-load and overheating can be detected. Variable buoyancy system is driven by separate BLDC motor. The CyberFish is powered with three packets of 11.1 V Lithium-Polymer batteries with total capacity of 2400 mAh which enables the robot to operate for three hours (dependent on the control parameters). Batteries can be recharged without removing them from the hull, using terminals located on top of the first segment of the tail. During robots operation terminals are disconnected from the batteries by means of magnetic switches. The CyberFish is also equipped with MEMS based inertial measurement unit (3-axis accelerometer, 3-axis gyro, 3-axis magnetometer) which was used for research on robot's control parameters influence on its tilt, heel and sway. The results of these work are the subject of different paper [8]. The robot is manually controlled but future work is concerned with development of robots partial autonomy Figure 1 presents the view of the CyberFish during free swimming in the pool.

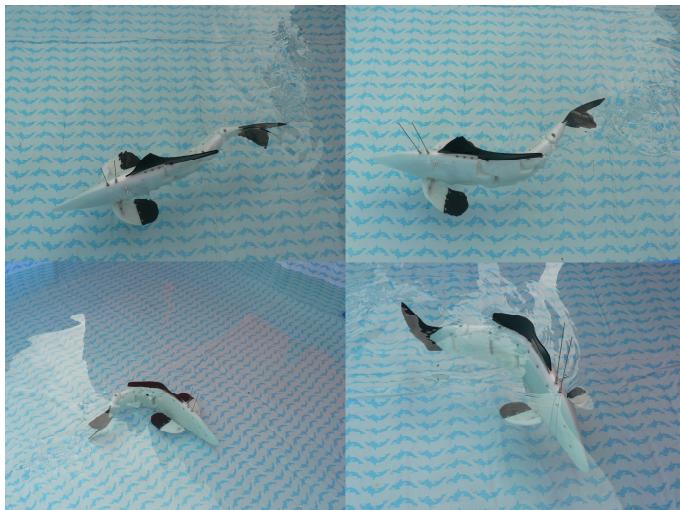


Fig. 1. BUV during free swimming in the pool (own elaboration)

2.2 The Test Stand

In order to measure BUV's thrust the test stand was build (Fig. 2). The test stand contains: the CyberFish (1), the non-malleable links (2), the stainless steel frame (3), the tensometric small force transducer AST KAP-S (4) of maximum load 10 N and resolution of 0.001 N, computer for data acquisition (5) and a water tank (4 m length, 2 m width, 1 m depth). Links where connected to the CyberFish through a pulley in order to eliminate adverse force components. The robot, the links and the lower part of the frame where immersed in water while the upper part of the frame with mounted force transducer extends above water surface.

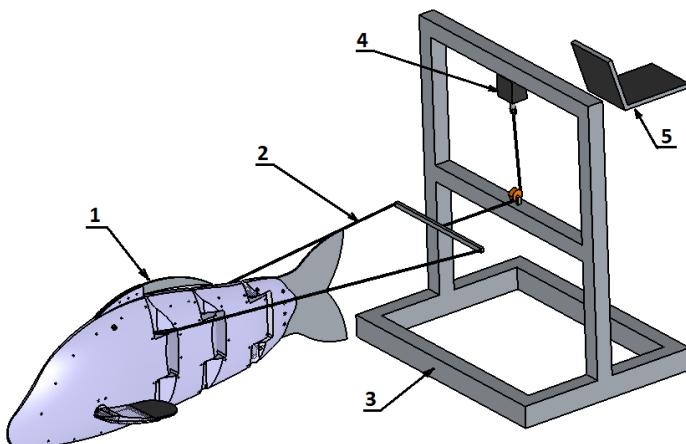


Fig. 2. The scheme of a test stand for vehicle's thrust measurements (own elaboration)

Robot's motion along straight line can be achieved in three ways:

- using variable buoyancy system and lateral fins angle of attack change only;
- using lateral fins oscillations only or with variable buoyancy system for depth control;
- using tail undulations only or with variable buoyancy system and lateral fins angle of attack change for depth control;

There is the possibility to changing the volume of water in variable buoyancy system connected with adjusting angle of attack of lateral fins thus the vehicle will move along using lift force generated on lateral fins by variable buoyancy force. In this case the motion of the fish seems to be the most energetically efficient and desirable in situations when the hull oscillations should be significantly reduced.

CyberFish is able to move by means of oscillations of lateral fins. In this case the robot's thrust depends on the shape and stiffness of lateral fins and control parameters (neutral point, amplitude and frequency of oscillations). Tests were carried out with symmetrical lateral fins which fore part was made of stiff polyethylene. Additionally 3 mm thick elastic rubber endings was added to the lateral fins and tests were repeated. The area of the flexible part longitudinal cross-section of one fin is 4010 mm^2 , while the area of rigid part longitudinal cross-section is approximately 3960 mm^2 . Both lateral fins are driven from the same servomotor. Taking into account limitations of lateral fins servomotor, the control parameters was also limited. When the lateral fins amplitude of oscillation is from 2.5 deg to 20 deg then the frequency of oscilation does not exceed 6 Hz. When the amplitude is less than 10 deg then the frequency could be increased to 9.6 Hz.

The third mode of CyberFish's locomotion is by means of undulation of its tail while the lateral fins does not oscillate. In this case, the tail undulates accordingly to traveling wave which is typical for carangiform type fish. More detailed description of realization of tail movement with short mathematical description can be found in [5]. The only change is that in this work authors introduced a coefficient which associate six parameters of tails motion (amplitudes and phase shifts of deflection of each segment). This coefficient (namely R coefficient) is the ratio of the length of a Cyber-Fish's tail to the wavelength of traveling wave to which the tail is fitted. Figure 3 presents schematically fitting the simplified tail into traveling wave for different R values.

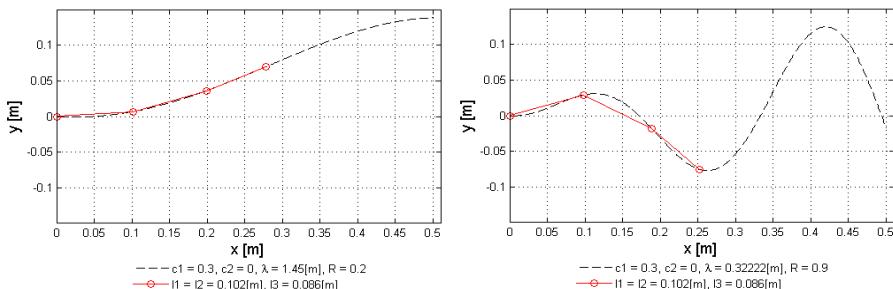


Fig. 3. Fitting CyberFish's tail into traveling wave with different R coefficient

Thrust measurements in this case was done for tail frequency changing from 0.2 Hz to 1.6 Hz (0.2 Hz step) and for R coefficient from 0.1 to 0.9 keeping in mind tail servomotors limitations.

2.3 Results and Conclusion

Vehicles thrust measurements while CyberFish was in motion using lateral fins movement only was done in two ranges of change of control parameters. In the first range the frequency of oscillations was from 1.2 Hz to 6 Hz while amplitude of oscillation was form 2.5 deg to 20 deg. In the second range, the frequency was from 1.2 Hz to 9.6 Hz but the amplitude was limited and was from 2.5 deg to 10 deg. Frequency change in both cases was with the step of 0.2 Hz and amplitude change was with the step of 2.5 Hz. The tests was done for lateral fins with and without rubber endings. Thrust was measured during vehicle's 18 s long steady state motion then the measurements were averaged. In the figure 4 there are surface plots of measured forces.

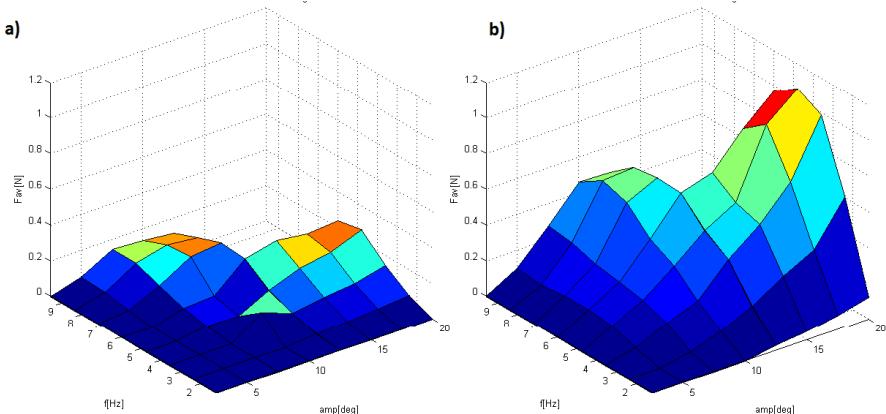


Fig. 4. Surface plots of vehicles thrust, a) without lateral fins rubber endings, b) with lateral fins rubber endings

The maximum force $F_{av} = 1.063$ N achieved for $f = 4.8$ Hz and $A = 20$ deg. Both the frequency and amplitude increase gives the robot's thrust increase but when lateral fins are equipped with rubber endings the thrust increase is more rapid. The average thrust increase for lateral fins with rubber endings is by 62% while the maximum force is three times higher than with lateral fins without rubber endings. It is worth notice that achieving given vehicle's thrust using low amplitude and high frequency of lateral fins is more desirable than with low frequency and high amplitude. It is because in the first case the heave of the robot is lower. Lo heave of underwater unmanned vehicle is important to maintain good quality of data registered by onboard devices (IMU, sonar, video camera etc.). It also can be seen in the figure 4 b) that there is the peak on the surface (for $A = 20$ deg, $f = 4.8$ Hz). After that point further

increase of the frequency does not give the increase on the thrust. It is expected that the thrust in high amplitude and high frequency of lateral fins will even decrease due to increasing water drag.

The next part of the experiment involved measurements of the vehicle's thrust while the robot was in motion using its tail only. In this case the tail undulation frequency was changing from 0.2 Hz to 1.6 Hz and R coefficient was changing from 0.1 to 0.9. Furthermore the test was repeated three times: without mounted tail fin, with stiff 1mm thick acrylic tail fin and with elastic 3 mm thick rubber tail fin of the same side area. The tail fin area equals 18670 mm^2 and its shape is homocercal, typical for carangiform type fish. Thrust was measured during vehicle's 12 s long steady state motion then the measurements were averaged. Measured thrust for selected motion parameters and elastic tail fin was shown in the figure 5. It can be seen that depending on control parameters the character of thrust is different. Observed force is characterized by different standard deviation with different control parameters. Standard deviations where the lowest with no tail fin and does not exceed 0.3 N while the biggest standard deviations was noticed for elastic tail fin driven by tail undulation with low frequency and high R coefficient.

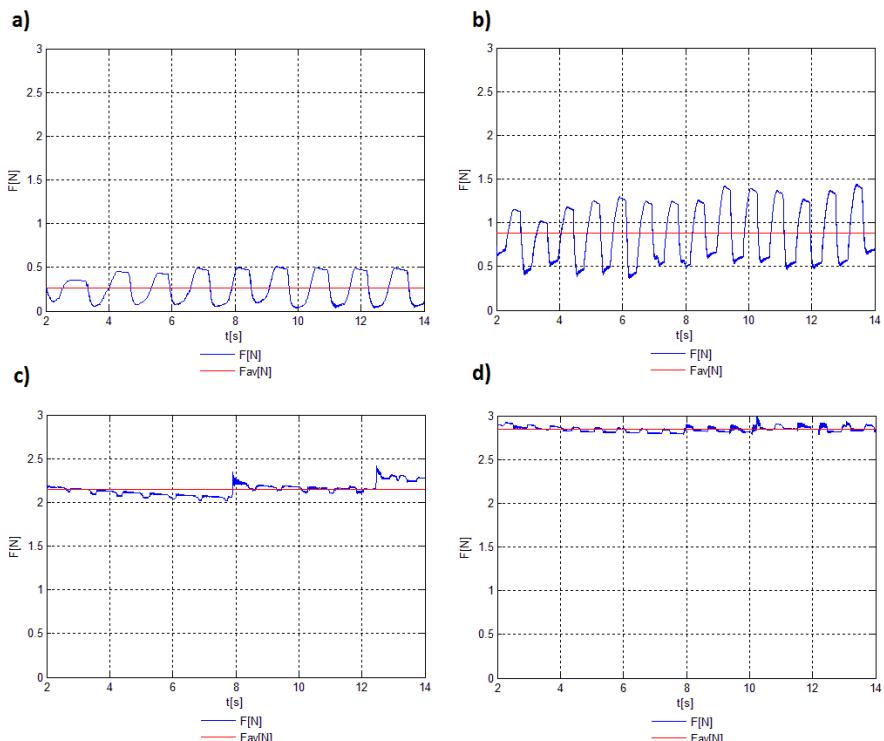


Fig. 5. Selected results of thrust measurements for tail with elastic fin and different control parameters, a) $f=0.8 \text{ Hz}$, $R=0.3$, b) $f=1.2 \text{ Hz}$, $R=0.4$, c) $f=1.2 \text{ Hz}$, $R=0.9$, d) $f=1.4 \text{ Hz}$, $R=0.9$

Figures 6 presents surface plots of the CyberFish's thrust with different tails. The average thrust increase with the increase of frequency of undulations and R coefficient as it was noticed with lateral fins. When the tail fin is removed the total thrust of the vehicle is rather low and does not exceed 0.9 N (Fig. 6 a). Maximal force in this case is $F_{av} = 0.8724$ N (for $f = 1.6$ Hz, $R = 0.7$). When the tail fin is attached to the last segment of the vehicle's tail then the thrust of the robot significantly increases with increasing frequency and R coefficient. It can be noticed that elastic fin gives greater thrust boost at the same control parameters than the stiff one. Maximal registered average thrust was for rubber tail fin driven at $f = 1.4$ Hz and $R = 0.9$ and was equal to 2.848 N while the maximum thrust for stiff acrylic fin is only 1.937 N (for $f = 1.2$ Hz, $R = 0.6$).

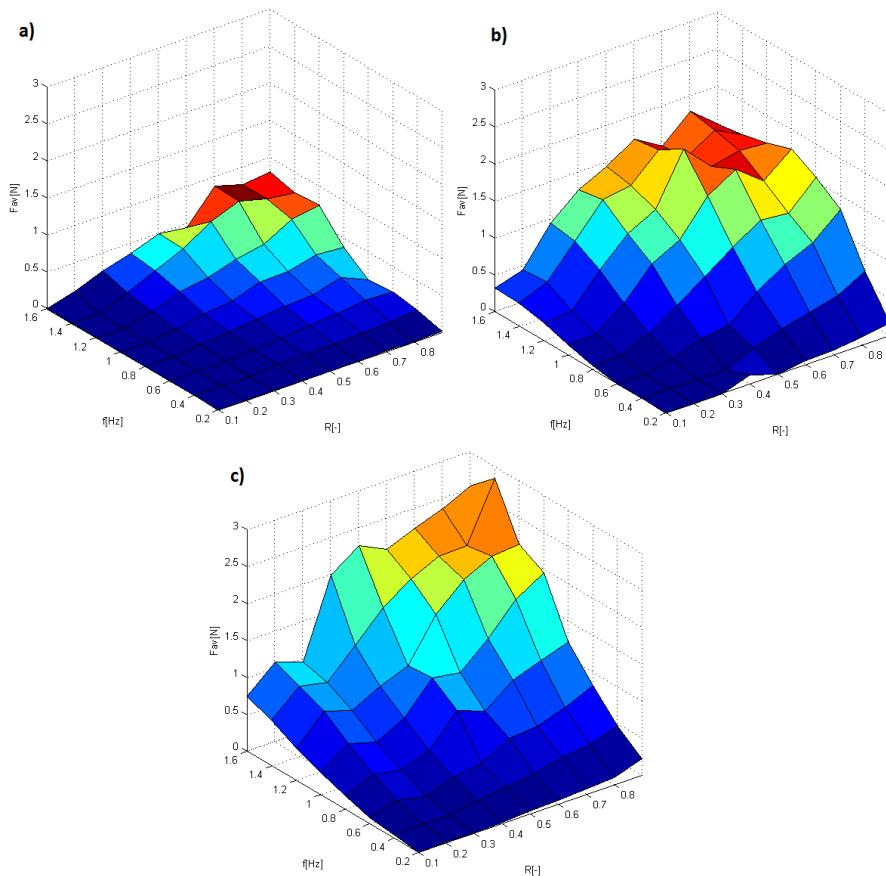


Fig. 6. Surface plots of vehicles thrust, a) tail without a fin, b) tail with stiff fin, c) tail with elastic fin

3 Summary and Future Work

Based on the conducted experiments it can be concluded that the maximum thrust is generated by undulating tail equipped with elastic tail fin rather than stiff one. The highest registered average thrust was in this case almost 3 N. The frequency and introduced R coefficient increase also increases thrust but in some ranges causes greater standard deviations of the vehicles thrust. There is the possibility to use lateral fins oscillations for generating vehicle's thrust and using the tail as a big rudder. In this case the average thrust is almost three times lower than the one generated by the tail. Also frequencies and amplitudes have to be much higher than the tails but this type of motion significantly reduces undesirable sway of the vehicle. Greater sway while the robot is driven by the tail is caused by shifting center of gravity of the construction in respect to center of buoyancy. Lateral fins oscillations has another advantage. The thrust can be generated in any direction of vertical plane while the tail generates thrust in limited directions of horizontal plane. This could be useful to increase vehicle's depth control dynamic or even generate reverse thrust when needed.

In the nearest future authors plan to carefully examine dependencies between tail fin and maximum thrust of BUV. Furthermore the optimal parameters of the tail fin are going to be studied and its influence on the robot's energetic efficiency. The appropriate experiment plan will be chosen and object function will be determined. The next step will be to optimize the function using either genetic algorithms or gradient methods [9].

Acknowledgements. Some of the work was not able to be done without financial aid from Małopolskie Centrum Przedsiębiorczości which is the organizer and coordinator of scholarship "Doctus – Małopolski Fundusz Stypendialny dla Doktorantów". Some of the work related to CyberFish BUV was also done in cooperation with Polish Naval Academy.

References

1. Low, H.K.: Modelling and parametric study of modular undulating fin rays for fish robots. *Mechanism and Machine Theory* 44, 615–632 (2008)
2. Hu, H.: Biologically Inspired Design of Autonomous Robotic Fish at Essex. In: Proceedings of the IEEE SMC UK-RI Chapter Conference 2006 on Advances in Cybernetic Systems, pp. 1–8 (2006)
3. Szymak, P., Malec, M., Morawski, M.: Conception of Research on Bionic Underwater Vehicle with Undulating Propulsion. *Solid State Phenomena* 180, 160–167 (2012)
4. Chan, W.L., Kang, T., Lee, Y.J.: Experiments and Identification of an Ostraciiform Fish Robot. In: Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics, pp. 530–534 (2007)
5. Malec, M., Morawski, M., Szymak, P.: Analysis of Parameters of Traveling Wave Impact on the Speed of Biomimetic Underwater Vehicle. *Solid State Phenomena* 210, 273–279 (2014)

6. Chong, C.W., et al.: Can the Swimming Thrust of BCF Biomimetics Fish be Enhanced? In: Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics, Bangkok, pp. 437–442 (2009)
7. Polish biomimetic unmanned vehicle CyberFish, <http://www.cyberryba.pl>
8. Malec, M., Morawski, M., Szymak, P.: Pomiar parametrów ruchu pojazdu podwodnego z napędem falowym z wykorzystaniem inercyjnego czujnika położenia. Zeszyty Naukowe Akademii Marynarki Wojennej (185A), 275–284 (2011)
9. Osyczka, A., Krenich, S.: Evolutionary Algorithms for Global Optimization. Nonconvex Optimization and Its Applications 85, 267–300 (2006)

Direct Local Communication for Distributed Coordination in a Multi-robot Team

Marta Rostkowska, Michał Topolski, and Piotr Skrzypczyński

Institute of Control and Information Engineering, Poznań University of Technology,
ul. Piotrowo 3A, 60-965 Poznań, Poland
 {marta.a.rostkowska,michal.m.topolski}@doctorate.put.poznan.pl,
piotr.skrzypczynski@put.poznan.pl

Abstract. This paper addresses problems of communication and coordination in the system of simple mobile robots that have limited sensing capabilities and computing resources. We propose a framework based on the direct, local communication and the finite-state machine encoding the coordination rules. We briefly describe our robots and then define the hardware and software architecture of the communication system. The proposed solution is validated experimentally on the task of passing through a gap in a wall.

Keywords: multi-robot team, communication, finite-state machine.

1 Introduction

Multi-robot systems are an active research field in robotics for many years. The introduction of multiple physical agents increases robustness of the whole system through redundancy, and enables to complete some tasks faster because of the spatial distribution of the individual agents [10]. Moreover, several resource-limited robots can be cheaper and easier to build than a single machine with many sensors, actuators, etc.

Experiments in multi-robot systems are challenging because of the level of complication and cost of a large group of robots. One possibility to alleviate these problems for experiments with many mobile robots is to keep the robots simple and to reduce the cost per unit. Recently, we have presented the SanBot (*Sandwich roBot*), which is an inexpensive, desktop-size, modular mobile robot intended especially for research and education in multi-robot systems. The motivation to build this robot and details of its design have been presented in [16]. This paper deals with the integration and use of the SanBots in a multi-robot system, focusing on the inter-robot communication mechanism and the formalization of the local control laws, which ensures deadlock-free operation of the system in absence of the global communication and coordination.

2 Related Work

A team of mobile robots can be organized as an intentionally cooperative system, or a collective swarm system [5]. In the former type of multi-robot systems

cooperation is achieved through explicit, intentional communication between agents. However, in some tasks (e.g. cleaning of large areas) implicit, indirect coordination between agents through traces left in the environment (called stigmergy) is possible. In contrast, the swarm-style systems involve agents, which are often very simple and rely mostly on stigmergy to achieve collective behaviour. Swarming behaviours based on stigmergy are exhibited by many organisms, particularly insects [11].

The control architecture of cooperative multi-robot systems is often hierarchical and centralized to some degree [18]. This means that there is a central unit/station or a leader agent that coordinates tasks of the entire team. The agents can be treated just as distributed physical effectors with a central controller [3], or they can only use the central unit for cooperation [19]. The central controller that has full knowledge about the system can generate optimal solutions to such problems like path planning or formation control [14], but it requires real-time and reliable communication between the agents and the central unit. Therefore, the centralized approach can fail easily under some environment conditions that limit range of the communication, render it unreliable, or cause significant delays in message passing [7]. An alternative is the fully distributed control paradigm. If this case agents use their own local controllers and do not depend much on the global communication system. Because in such a system there is no global information the plans and control decisions are inherently suboptimal. However, examples from Nature provide evidence that coherent behaviours of the whole group can be achieved through the interaction of agents lacking global information. Such coherent behaviours are observed e.g. in ants [11] and bees [9]. Unfortunately, in a physical implementation of a multi-robot team efficient stigmergy is often hard to achieve, due to technical problems with leaving recognizable but unobtrusive “traces” in the environment. Several researchers have studied the problems of communication in teams of behaviour-based robots. Balch and Arkin [4] investigated the importance of explicit communication in robotic societies. An early example of a physical multi-robot team that relied on the concept of stigmergy is ALLIANCE [15], which, however, simulated the implicit communication by means of radio transmissions.

For our SanBot II robots we have chosen fully distributed model of coordination due to its robustness to failures of particular agents and global communication. We believe that coordination and cooperation algorithms developed on such robots can be of some interest in real multi-robot applications, particularly for search and rescue, where reliability is a crucial factor [10]. Thus, we decided to implement a direct local communication mechanism, where the communication is restricted to a situation when two robots are in close proximity and have an unobstructed line of sight between them. This can be viewed as a functional analogy of the situation when insects transfer information by direct contact, like antennation [13] or mandibular contact in ants, or when they communicate by performing specific sequences of movements (like waggle “dance” of honey bees [9]). Such communication takes place in a distributed and concurrent manner, avoiding the bottleneck of a central unit managing the information sharing. The

short-range, line-of-sight communication can be implemented by various physical means, and it is possible to choose the media which is robust in the given environment, e.g. ultrasonic in the water or infrared in the air.

In the SanBot II robots the local communication has been implemented by means of infrared transmitters/receivers. While this concept of a communication device for a mobile robot is not entirely new, in this paper we show how the direct local communication model influenced our coordination strategy, which at the logical level is represented by a finite-state machine. The finite-state machine approach to coordination of mobile robots was also considered in [12] for multi-robot patrol – one of the most important applications of multi-robot systems. Similarly to our work, authors of [12] experiment with real robots, but they do not use explicit communication. Instead, implicit communication through observation of other robots is implemented. In this paper we show how the explicit, local communication between the agents can be employed by the finite-state automaton describing the behaviour of the system. This approach is inspired by our earlier work on communication in a team of robots and monitoring sensors [17], where we also used the finite-state automaton to describe and analyse the communication, which was however global and Ethernet-based.

3 Mobile Robot Design

SanBot II is the second, improved version of the SanBot robot, described in more details in [16]. Both versions are shown in Fig. 1.

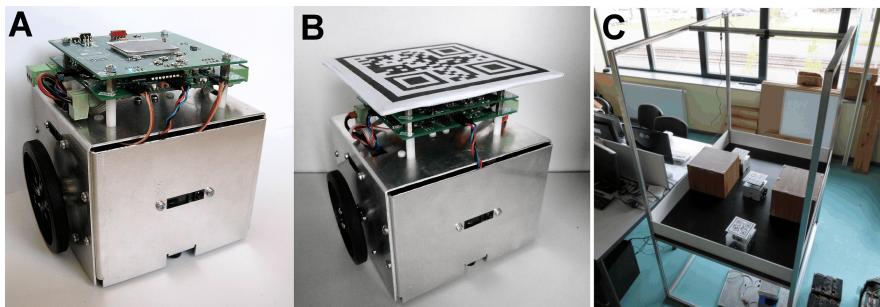


Fig. 1. Desktop-size mobile robots: SanBot I (A) and SanBot II (B, and the SanBot II multi-robot test bed with the overhead vision system (C)

SanBot is a small mobile robot. It has a differential drive mobility system with two DC motors. Perception of the environment is accomplished mainly using four Sharp infrared distance sensors, which are located on each side of the robot. The main application area intended for SanBot are multi-robot systems. Keeping this application in mind we have designed some extension modules for SanBot. The robot is built in a modular fashion, so adding specific modules does not cause a problem [16].

The main difference between the SanBot I and II is the change the dead reckoning sub-system. The SanBot I used cheap optical sensors from computer mice. Unfortunately, this form of odometry turned out to be inaccurate and unreliable in practice. Therefore, in SanBot II, optical sensors are replaced by classic encoders mounted on the motor shafts. The encoder-based odometry is more precise, but sporadically also provides erroneous readouts, which is attributed to the low-cost motors and shafts used. Due to the insufficient reliability of the odometry-based self-localization, an external positioning system consisting of an overhead camera and passive markers on robots was used. The SanBot II has a communication module installed, that allows communication between the robot and the external vision system, as well as the short-range communication between robots.

4 Coordination in a Multi-robot System

4.1 Positioning of the Agents Using the Overhead Vision System

The SanBot II robots are used in an experimental setup of a multi-robot system, which is installed on a table with a single-camera overhead vision system (cf. Fig. 1C). An overhead camera sensor is standard in many desktop multi-robot applications, like the robot-soccer, but can be also effectively used with full-scale mobile robots, as demonstrated in our previous work [17].

The camera is located quite high over the table, in such a way, that its field of view covers the entire working area for robots. The acquired image of the scene is filtrated, in order to improve the quality of the picture. Then, in the process of segmentation and outlier elimination the specified objects are extracted from this image, and their position and orientation is computed (Fig. 2B). The position of each robot agent is equivalent to the position of its marker, which is ensured by proper robot-camera calibration. All this processing takes place on an external PC computer. The computed poses of agents are sent to the them via Bluetooth, and used to correct the odometry-based pose estimate. From the experimental evidence we assume that the vision-based pose estimate is much more certain (accurate and reliable) than the odometry-based one. Thus, we do not attempt to fuse these two estimates, but the vision data override the odometry-based pose each time they are available.

The passive markers are based on the QR matrix codes [8], employed in mobile devices to acquire information via their embedded cameras. An example marker is shown in Fig. 2A. It has a shape of a square with a side length of 14 cm. Inside a 10×10 QR code is placed, with a 0.5 cm thick black frame around and three small rectangles (pointed by arrows in Fig. 2A) that define the orientation. Our choice is motivated by the need to embed more information in the marker, and to make these markers robust to partial damage or occlusion. The SanBot robots are modular, and the team can consist of heterogeneous robots, which differ as to the number and function of the modules. Because we do not assume any central unit to which the robots can report their configuration, we required the markers to contain more information than just the unique ID. The QR

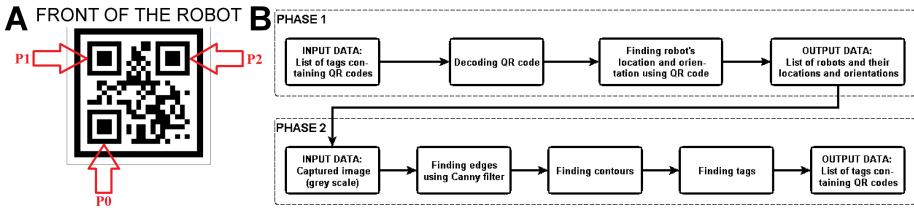


Fig. 2. Example QR code marker (A) and the image processing block scheme (B)

codes can accommodate an amount of data large enough for information about the configuration and selected parameters of the robot, e.g. actual dimensions (height in particular), maximum velocity. Owing to this feature it is possible to add or remove heterogeneous agents to/from the system on-the-fly, without reconfiguration or reset of the vision software. The implemented overhead vision system enables real-time localization of a large number of robots with satisfying precision. In the experiments we have found that the position error is bounded to 1 cm, and orientation error to 5°. As it can be noticed from Fig. 3 QR recognition is robust to blur (e.g. from motion), and the code can be correctly recognized even if 30% of its surface is unreadable.



Fig. 3. Examples of robust QR code marker identification in the presence of blur

4.2 Coordination with the Finite-State Machine

The coordination scheme we propose for our multi-robot system favours an idea, that complex tasks can be solved by exchanging short, direct signals among agents employing simple, behaviour-based control algorithms. There are many multi-robot systems that follow a similar idea of coordination by passing messages for specific behaviour activation. However, architectures using hand-crafted rules for coordination are hard to analyse, and it is not easy, or even not possible to check and prove that such a network of discrete events leads to a correct behaviour of the team, e.g. if it is deterministic and non-blocking. Therefore we decided to use a finite-state machine to encode the logic of the coordination scheme. This enables to use a proven mathematical apparatus to check the correctness of the design [6].

The control system of the multi-robot team can be divided into the high-level control layer, and the low-level controllers of the particular agents. The low-level controllers are hand-crafted following the behavioral paradigm. From the point of view of the high-level control layer the behaviors (activated behavioral modules) in the agents can be treated as separate states of a finite-state machine. Each state is responsible for a particular motion type or another action of the agent – sensing or direct communication. Also, the whole team's state can be defined as superposition of the states of particular agents. These states and transition events between them constitute the finite-state machine, which plays the role of the high-level control layer. The transitions are made according to external signals depending on the sensed state of the environment or the state of other agents communicated by them.

A finite-state machine for coordination can be formulated for each particular task and then implemented in the agents. We show one particular example of such a formulation for the task of concurrent passing through a narrow passage. Robots are grouped in two areas on the opposite sides of a wall. These areas are connected by a narrow gap in the wall. The gap is big enough for one robot. Both groups initially roam on their sides of the wall, but eventually they find the gap, and try to pass to the other side. Thus, a problem how to cross without mutual blocking arises. This problem was formulated in [2], and solved by introducing a centralized area manager that controls the critical resource (the gap area), and grants the right to pass to the agents. The solution of [2], although efficient can easily fail if the centralized agent cannot do its job due to problems in either communication or perception. Therefore, we show that this problem can be solved in a completely distributed manner using the finite-state machine and direct local communication to coordinate the agents. To facilitate the design of the state machine four cases in the task under study have been specified:

1. **Case 1 - one robot.** We assume that there is only one robot in the scene. The solution is simple – when the gap is found the agent goes through.
2. **Case 2 - one robot at each side.** We assume that there are two robots, one at each side of the wall. They meet each other when trying to pass through the gap. To determine the right of way, the IDs of agents are compared. The one with the lower ID passes first, and the other one turns back to form one-unit queue along the wall, waiting for its turn.
3. **Case 3 - existing queue.** We assume that there are one or two queues on one or on both sides of wall. The incoming robot discovers the queue and communicates with the agent last in the queue, thus getting to know that it should join the queue and wait. Robots go through the passage one-by-one from each side. The robot leaving the gap communicates to the robot on the other side the number of agents left on the entry side.
4. **Case 4 - long passage** This case is similar to case 2. The main difference is, that it assumes that the passage is not only narrow, but also long. Therefore, robots meeting in the middle of it can have other robots moving behind them, which blocks their way back. The scenario is exact as in case 2, but if

it occurs, that retreating agent is blocked by others, and it has the highest priority to go, forcing other robots to move back and form a queue.

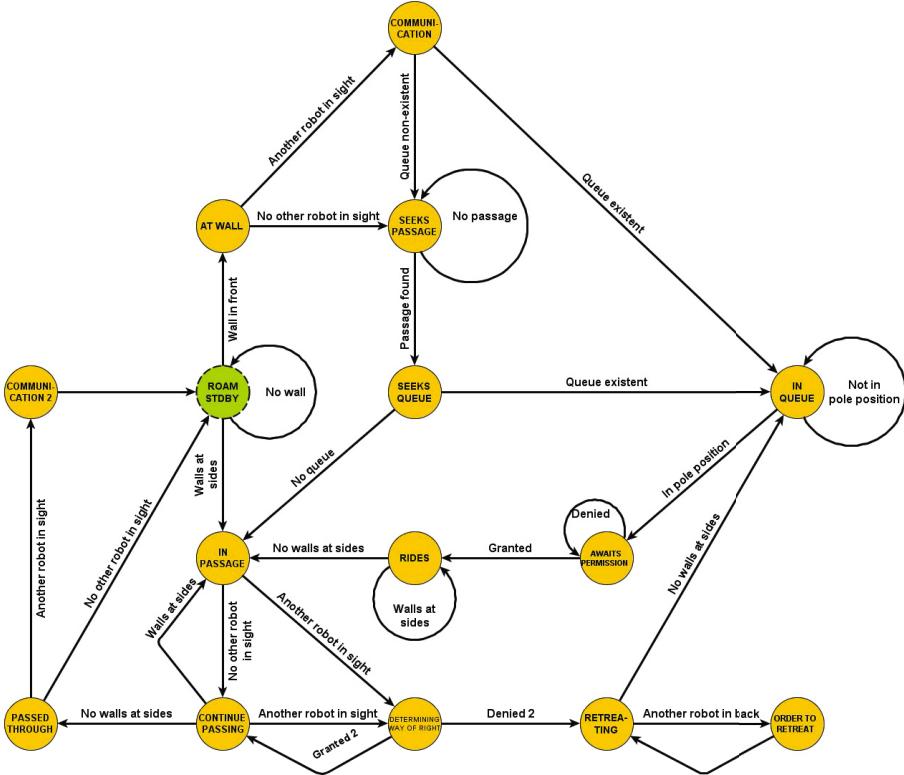


Fig. 4. Finite-state machine of the high-level coordination scheme in the narrow passage passing task

The above-mentioned local strategies have been consolidated to formulate a finite state machine shown in Fig. 4. The designed automaton consists of 15 states and 29 transitions and has been verified as non-locking and minimal using *Supremica* [1] software. Each agent in the team has implemented the same coordination algorithm based on this automaton.

5 Implementation of the Direct Local Communication

As already mentioned, the direct local communication between agents is intended to mimic the one known from Nature. Therefore, a short-range, line-of-sight communication using infrared waves has been designed. Each robot carries

a transmission module equipped with six infrared transmitting (Tx) diodes and the same number of receivers (Rx). The working wavelength is 940 nm. The parameters are tuned to achieve communication on distances up to 40 cm. Each Tx-Rx pair creates a separate transmission channel, arranged in a circular pattern, with a 60° gap between each of them, as seen in Fig. 5.

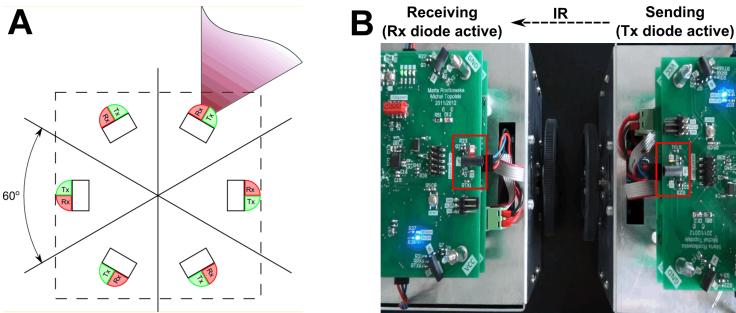


Fig. 5. Add-on communication module: concept (A) and implementation (B)

Data sent via infrared is carried by 36 kHz frequency wave, due to the usage of receivers which are immune to constant signal, but interpret presence of modulated signal as logical 0 and its absence as logical 1. Modulation is a mechanism used to protect against unwanted disturbances from the environment. The next shield in line is data encoding using Manchester code. The last level of protection is a custom protocol used to transmit data between the agents. It consists of three phases: (i) Initial synchronization of each frame, (ii) supervision over connection, (iii) verification of every incoming frame. The protocol ensures, that frames (their format is presented in Fig. 6) are re-sent until correct data is received.

Part	Header (4 bytes)							
Name	Preamble (2 bytes)				Frame size (1 byte)	Address (1 byte)		
Hex	0xAA		0xAA		-	Sender		
Bits	1	0	1	0	1	0	1	0
	total frame size					s7	s6	s5
						s4	s3	s2
						s1	s0	

Part	Data (1+n bytes)					Ending (2 bytes)	
Name	Command (1 byte)		Data (n bytes)			Ending (2 bytes)	
Hex	-	-	-	-	-	0x55	0x55
Bits	c7	c6	c5	c4	c3	c2	c1

data

Fig. 6. Frame structure in the direct local communication protocol

6 Experimental Validation

The coordination mechanism based on direct local communication and finite-state machine has been implemented for the above-described exemplary task

using a team of three SanBot II robots. The experimental set up consists of a laboratory table (130×100 cm) with bands and a single camera placed above. The camera is used for positioning of the robots, as described in section 5. Two box-like obstacles divide the table into two areas. A narrow passage is created between these obstacles. Figure 7 shows frames from the overhead camera depicting selected phases of the experiment¹, and at the same time, illustrating the operation of the finite-state machine. The subimages present:

1. Initialization – due to the quite restricted space, the agents immediately switch from the state "ROAM STDBY" into "AT WALL", in which they move in counter-clockwise direction, following the wall. "ROAM STDBY" state would make robots move in random direction.
2. Robots do not meet other agents of the system, so there is no "COMMUNICATION" state. They go into the state "SEEKS PASSAGE" by transition "No other robot in sight", and continue the counter-clockwise move.
3. Robot on the right side (yellow) has detected the gap by its sensor, and continues to move forward to confirm whether it is really a passage,
4. After finding the end of the gap, the yellow robot goes into state "SEEKS QUEUE". No other agent is in sight, so the robot goes directly into state "IN PASSAGE" by "No queue" transition. It moves to the middle of the gap, so it can turn around there and move into the passage.
5. When reaching center of the gap, the yellow robot turns to aim at the passage. Red robot is still looking for a gap.
6. While passing through the passage, the yellow robot is in the state "IN PASSAGE". Red robot has started recognizing the passage.
7. Both robots have met when in states "IN PASSAGE". The "Another robot in sight" transition has been activated and robots went into the state "DETERMINING WAY OF RIGHT". Yellow robot has lower ID number, thus having higher priority, so by transition "Granted 2" goes into the state "CONTINUE PASSING" and exits from the gap by turning around. Meanwhile, the red robot went into the state "RETREATING" using transition "Denied 2" and has retreated into an one-robot queue position. It instantly goes from the state "IN QUEUE" to "AWAITS PERMISSION", as it automatically is "IN POLE POSITION".
8. The yellow robot has finished the phase of passing through, and went into the state "PASSED THROUGH". As there is no other robot in sight, it instantly goes into "ROAM STDBY" state. Then, because there is already a wall at its side and front, the robot goes into state "AT WALL" using transition "Wall at front". Meanwhile, red robot continues measuring the gap by going into states "RIDES" and "IN PASSAGE" by transitions "Granted" and "No walls at sides". After the red robot passes through, the whole cycle repeats.

¹ See also a video available at <http://lrm.cie.put.poznan.pl/aut2014.wmv>

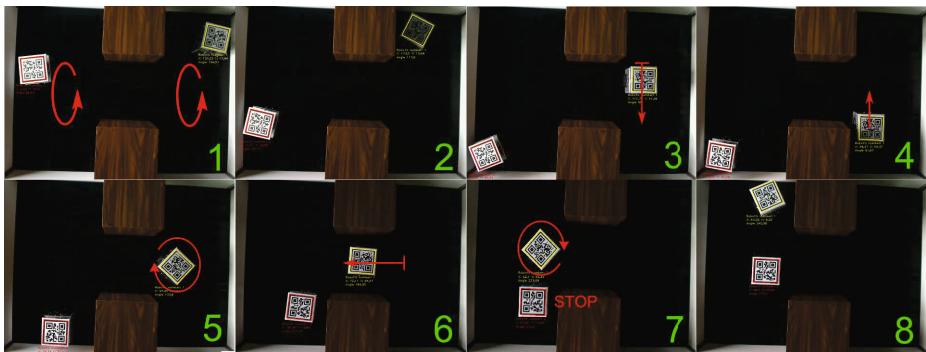


Fig. 7. Actual implementation of the passing through a narrow passage scenario

7 Conclusions

In this paper we have shown that short-range, direct local communication can be used to implement coordination schemes for tasks that previously have been solved by introducing centralized units prone to failures. The proposed coordination was described formally as a finite-state machine, which provided the possibility to automatically check such important features like determinism and absence of deadlocks by using off-the-shelf software. The proposed physical implementation of the robot team was tested experimentally, and turned out to work reliably.

References

1. Akesson, K., Fabian, M., Flordal, H., Malik, B.: Supremica – An Integrated Environment for Verification, Synthesis and Simulation of Discrete Event Systems. In: Proc. IEEE Int. Workshop on Discrete Event Syst., Ann Arbor, pp. 384–385 (2006)
2. Ambroszkiewicz, S., Borkowski, A., Cetnarowicz, K., Turek, W.: Adaotive Bilateral Control of Multiple Mobile Robots. In: Proc. Int. Symp. on Measurement and Control in Robotics, Warsaw, pp. 31–50 (2007)
3. Antonelli, G., Chiaverini, S.: Kinematic Control of Platoons of Autonomous Vehicles. *IEEE Transactions on Robotics* 22(6), 1285–1292 (2006)
4. Balch, T., Arkin, R.C.: Communication in Reactive Multiagent Robotic Systems. *Autonomous Robots* 1, 1–25 (1994)
5. Balch, T., Parker, L.E. (eds.): *Robot Teams: From Diversity to Polymorphism*. Natick, A.K. Peters (2002)
6. Christos, C.G., Lafourture, S.: *Introduction to Discrete Event Systems*. Springer, New York (2008)
7. Cui, X., Ragade, R.K., Elmaghriby, A.S.: A Collaborative Search and Engage Strategy for Multiple Mobile Robots with Local Communication in Large-Scale Hostile Area. In: Proc. Int. Symp. on Collaborative Technologies and Systems, San Diego (2004)

8. Denso Wave Inc., QR code information homepage (2013),
<http://www.qrcode.com/index-e.html>
9. von Frisch, K.: Bees: Their Vision, Chemical Senses, and Language. Cornell University Press, Ithaca (1976)
10. Heppner, G., Roennau, A., Dillman, R.: Enhancing Sensor Capabilities of Walking Robots Through Cooperative Exploration with Aerial Robots. *Journal of Automation, Mobile Robotics & Intelligent Systems* 7(2), 5–11 (2013)
11. Latty, T., Ramsch, K., Ito, K., Nakagaki, T., Sumpter, D., Middendorf, M., Beekman, M.: Structure and Formation of Ant Transportation Networks. *J. R. Soc. Interface* 8, 1298–1306 (2011)
12. Marino, A., Parker, L.E., Antonelli, G., Caccavale, F.: Behavioral Control for Multi-Robot Perimeter Patrol: A Finite State Automata Approach. In: Proc. IEEE Int. Conf. on Robotics and Automation, Kobe, pp. 831–836 (2009)
13. McCabe, S., Farina, W.M., Josen, R.B.: Antennation of Nectar-receivers Encodes Colony Needs and Food-Source Profitability in the Ant *Camponotus mus*. *Insectes Sociaux* 53(3), 356–361 (2006)
14. Milutinović, B., Lima, P.: Modeling and Optimal Centralized Control of a Large-Size Robotic Population. *IEEE Transactions on Robotics* 22(6), 1280–1285 (2006)
15. Parker, L.E.: Heterogeneous Multi-Robot Cooperation, PhD Thesis, MIT (1994)
16. Rostkowska, M., Topolski, M., Skrzypczyński, P.: A Modular Mobile Robot for Multi-Robot Applications. *Pomiary Automatyka Robotyka* 2, 288–293 (2013)
17. Skrzypczyński, P.: A Team of Mobile Robots and Monitoring Sensors – from Concept to Experiment. *Advanced Robotics* 18(6), 583–610 (2004)
18. Uhl, T., Buratowski, T.: The 2-Wheeled Mobile Robots Capable of Cooperating in Group of Robots. *Journal of Automation Mobile Robotics & Intelligent Systems* 1(1), 24–31 (2007)
19. Zlot, R., Stentz, A.: Market-Based Multirobot Coordination for Complex Tasks. *Int. Journal of Robotics Research* 25(1), 73–101 (2006)

TALOS – Mobile Surveillance System for Land Borders and Large Areas

Agnieszka Sprońska, Jakub Główka, Mateusz Maciąś, and Tomasz Rokosz

Industrial Research Institute for Automation and Measurements PIAP, Warszawa
`{aspronska, jglowka, mmacias, trokosz}@piap.pl`

Abstract. This paper presents the concept of an innovative, transportable and fast deployable land border and large area surveillance system, based on mobile unmanned vehicles (UVs). The system has been designed, developed and tested, within the course of realisation of a research project entitled Transportable Autonomous Patrol for Land Border Surveillance System – TALOS, executed under the EU 7th Framework Programme, in Security priority. In result of the TALOS project a technology demonstrator was developed, which consists of two unmanned ground vehicles (UGVs) and a command and control centre. The system robots feature a high degree of autonomy and are equipped with the multi-sensor data acquisition systems. The data are processed and integrated, including their fusion with navigation device information and accurate terrain data-bases. The Command, Control and Communication Centre, where large amounts of system information is displayed and managed, utilises advanced Man Machine Interface (MMI). This paper describes the developed system demonstrator and its features, as well as the TALOS system integration process.

Keywords: unmanned vehicles, UGV, robots, autonomy, sensors, data processing, man-machine interface, border security.

1 Introduction

In consequence of the European Union (EU) extension in the past few years, the character of the European borders has changed significantly. The new external EU border appears different from the one before the accession. It extends between Finland in the North and Bulgaria in the South of Europe, and it's very diversified with respect to topographic characteristics, climatic conditions, and probability and intensity of illegal activities. The borders of the new member states, shared with the former Soviet Union countries, are particularly exposed to illicit trafficking. It is known to be used as a backdoor to EU by illegal immigrants, possibly involved in international terrorism, and as an area of illicit activities, such as trafficking in people. TALOS system was designed to address the problems of surveillance of large land border areas. The aim of the system is to help in detection, tracking and intercepting persons trying to cross the land border illegally between border crossing points. To meet the requirements resulting from the diversified nature of the border, such a

system needs to be easy transportable and adaptable, to be quickly moved and deployed in the particular sections of the border, which require strengthen supervision, due to the intensified illegal activity. The needed characteristics are ensured by application of the semi-autonomous unmanned vehicles, which allow efficient surveillance of large areas without engaging large human resources.

All the elements of the TALOS system are installed in standard containers and can be easily transported via roads, tracks or water for fast deployment in selected border zones, within the single country and between countries, all over Europe.

2 TALOS Project

TALOS system has been developed within an international research project co-funded by the European Commission (EC) under the 7th Framework Programme in Security priority, by fourteen partners from eight EU countries (Belgium, Estonia, Finland, France, Greece, Poland, Romania, Spain) as well as from Turkey and Israel. The main objectives and challenges of the project were:

1. designing the overall architecture of the system, as well as main system components, i.e. Command& Control, UGVs, communication system;
2. designing the interfaces between elements of the TALOS system as well as between the whole system and external systems;
3. designing, developing and integrating the main subcomponents, i.e. UGV, Command&Control Centre, Communication;
4. designing and setting-up the TALOS system demonstrator.

The project realisation was planned to be a two-phase process, as some of the system elements (described in the following chapter) were not to be implemented in the system demonstrator. The project's second phase shall include the development and integration of those elements with the system. That strategy allowed to concentrate on the challenge of development of the unmanned and autonomous ground robotic system, whereas the results could be ready for extension in the second phase. Such extension should cover integration with other available systems, in order to provide more operational capabilities and greater interoperability.

3 TALOS High Level System Architecture

TALOS high level system architecture was developed basing on the functional and technical requirements gathered – both from the end-users (Border Guards) and the TALOS Project Consortium's own knowledge and experience. It has been designed according to the Joint Architecture for Unmanned Systems (JAUS) standards. The following system nodes have been identified in the high level architecture description:

- Headquarters
- Theatre Command Centre
- Theatre Surveillance Unit

- Unmanned Unit Command Centre (UUCC)
 - Commander Operator Control Unit (Commander OCU)
 - UGV Commander Operator Control Unit (UGV OCU)
- UGV - Observer
- UGV - Patroller
- UGV - Interceptor
- UAV System
 - UAV Operator Control Unit (UAV OCU)
 - UAV
- Sensor Tower

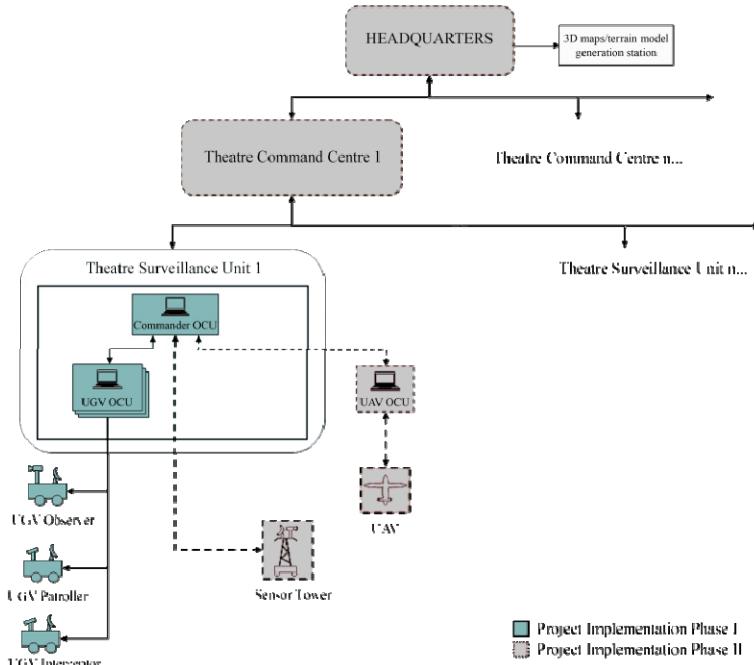


Fig. 1. TALOS system architecture

In the first phase of TALOS both the Headquarters and Theatre Command Centre were not implemented, but their functionalities were assumed to be fulfilled so that the Unmanned Unit Command Centre behaved accordingly. Also, the functionalities of Patroller UGV and Interceptor UGV were joined in a single UGV (Interceptor UGV). The UAV and Sensor Tower were simulated, in order to show the capability of the system to have those elements integrated at any time, like any other unmanned units, such as USVs. TALOS system architecture is presented on figure 1.

4 TALOS System Demonstrator

TALOS Project Phase-I was aimed at the development of the TALOS system demonstrator, by using the already existing solutions and development of the parts which are not existing yet. It has been composed of the Unmanned Units Command Centre (UUCC) and two Unmanned Ground Vehicles (UGVs). For the purpose of the system demonstrator, communication with Unmanned Air Vehicle (UAV) and transportable Sensor Tower was simulated on the dedicated computers, placed in the UUCC. The architecture of the TALOS demonstrator is presented in figure 2.

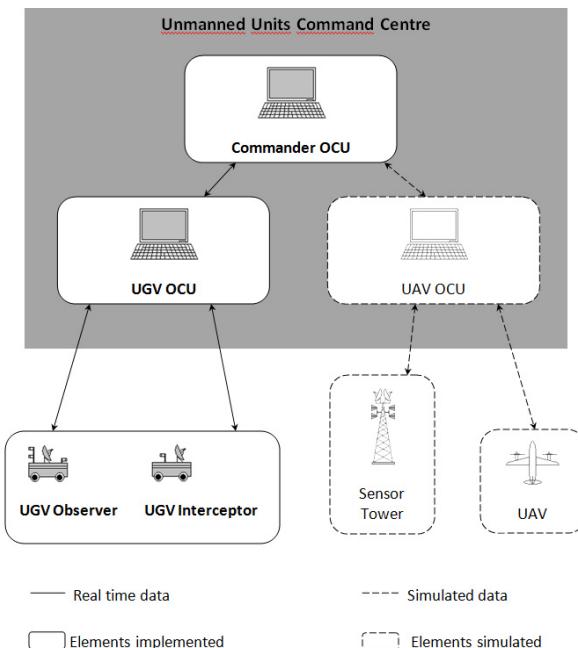


Fig. 2. TALOS system demonstrator architecture

The Unmanned Units Command Centre (UUCC) controls and monitors operation of all the unmanned units. It contains of the Operator Control Units (OCUs) for the Commander and the Operator of the UGVs. The Commander is presented with a visualisation of the mission area, position of the system elements, patrol routes and data from all the sensors on different map layers. UGV Operator has the possibility of viewing the information from sensors carried by the vehicles, planning vehicle routes and operating the vehicles manually.

4.1 Usage Description

The TALOS system demonstrator contains of two UGVs, able to operate simultaneously, based on the same mobile platform, and equipped with the sets of

payloads, which correspond with the vehicle's function, as described in the next section. First UGV (Observer) is designed for the performance of the surveillance and detection missions (preset patrolling route and observation tasks), the second vehicle (Interceptor) is intended for interception of the suspicious objects (individual, vehicle etc.) and following it until the manned Border Guard patrol will arrive to intervene, so the Observer UGV can continue its surveillance mission. Even though the unmanned units are equipped with autonomy, the constant supervision of the human operators is essential. In the system vehicle routes can be calculated automatically, but the operator always needs to accept the proposed route. Similarly, the identification of the trespasser (after he is detected by the sensors) is always performed by the personnel, with use of the optic devices carried by the vehicle. No weapons or any coercive measures are being used by the robots towards the detected and followed objects. In case of a need the communication with the tracked intruder is possible via the interrogation system.

5 TALOS Subsystems

The three core components of the TALOS system demonstrator are: the UGV (Unmanned Ground Vehicle) subsystem, UUCC (Unmanned Units Command Centre) subsystem and the Communication subsystem. All the subsystems were developed independently, but derived from the TALOS system architecture. The biggest challenge to overcome in the development process was the integration of the various subsystems and their components into a one working system, which could be subjected to a field test and the real-time demonstration.

5.1 UGV Subsystem

The UGV subsystem is composed of the tracked platform of a high mobility, integrated with a set of payloads, allowing it to perform autonomously the tasks assigned by the operator via the UUCC operator panels. The TALOS system contains of two UGVs, as presented in Figure 3.

Both UGVs are able to operate simultaneously, based on the same mobile platform, and equipped equally with the following set of payloads:

1. High and low level computers, responsible for steering the platform and controlling the sensors;
2. Navigation devices, primarily based on precise Global Positioning System (DGPA, RTK) providing horizontal accuracy (RMS) of 0.01 m, that is aided by Inertial Navigation System (INS) providing in real time the attitude, azimuth of the vehicle, its velocities, as well as acceleration and angular rates in three axes; as well as the 3D mapping laser scanner, with a 360 degrees horizontal field-of-view and a 26.8 degrees vertical field-of-view, scanning rate of 15 Hz and the measurement accuracy better than 0.04 meters, giving more than 1.333 million points per second;
3. Safety sensor, ie. laser scanner located on the front-side of the vehicle to prevent the UGV from hitting obstacle if the autonomous obstacle avoidance function fails;



Fig. 3. TALOS UGVs

4. Driving sensor, ie. day/night camera installed on a two axis pan/tilt platform, additionally allowing to detect and support recognition of the objects from up to 3 km line of sight in the daytime and up to 0.14 km during the night.

First UGV (Observer), as designed for the performance of the patrolling and detection missions, is equipped with the specialized surveillance sensors, ie.: the Doppler radar, capable to detect movements of individuals within the range of 4 km and vehicles within the range of 7 km, under both day and night conditions; the surveillance/observation camera, additionally equipped with a FLIR capability for night operation and laser range finder for determining the exact position of observed object, as well as the automatic video tracker (AVT), enabling auto tracking of the objects during day and night operation.

The second vehicle's (UGV Interceptor) function is to intercept the objects (individuals, vehicles etc.) detected by the UGV Observer and identified by the Operators as suspicious, and to follow it until the arrival and intervention of the manned Border Guard patrol. When appropriate, the communication with the tracked person or group is possible via the interrogation system, composed of a microphone and a loud speaker, capable both to replay the pre-recorded tracks of voice and to establish a live two-way communication with the personnel of the UUCC.

As mentioned above, both TALOS vehicles have been build on the identical platform, so the design of the robotic system enables interchangeability of the equipment. Therefore, the platform itself does not have a dedicated role in the TALOS system. It is the installed equipment and the software configuration that decides the role of the vehicle.

Both UGVs can be easily stored and transported in standard sea containers (20 ft and / or 40 ft), either separately or in one.

5.2 UUCC Subsystem

Unmanned Units Command Centre (UUCC) has been specially designed to allow an easy transport and fast deployment of the unit at the desired area. It has been placed within the standard container of the following dimensions: length: 3,7 m, height: 2,2 m, width: 2,1 m, and the capacity of 28,27 m³. The exterior and interior of the UUCC is shown on Figure 4.



Fig. 4. TALOS UUCC

Within the UUCC two Operator Control Units (OCU) are located:

1. UGV Operator OCU - designed to execute the missions and assign the tasks for the UGVs, and in case of a need to take a direct control (via the tele-operation mode) over the vehicles, by the means of a steering wheel and pedals;
2. Commander OCU – dedicated to mission planning and execution, assigning tasks and missions for the Operator OCU as well as for unmanned units, having the capability to directly control the vehicle payloads according to the needs, via the joystick.

The software installed to the OCU's computers integrated with the UGV's computers allows the performance of both pre-defined and ad-hoc assigned tasks, as well as the recording of the acquired data for further analysis.

In addition to that the UUCC includes the following subsystem nodes:

1. World Model Knowledge Store (WMKS)
2. Video/Audio Recorder (NVR)
3. RTK Base Station
4. Emergency Stop
5. Communication Units
6. Simulated Subsystems (UAV, Static Sensor Tower)
7. 3D Maps/Terrain Model Generator

The simulated subsystems were intended to be implemented as the real subsystems in the scope of the future system development (Phase-II). TALOS demonstrator

includes the simulated elements, in order to show the system is equipped with the appropriate interfaces and is capable to have the subsystems integrated at any time.

5.3 Communication Subsystem

In the TALOS system the communication between system elements is a critical issue, as the UGVs are to operate within the long distances from the UUCC and they need to constantly and simultaneously transmit huge amounts of data from different sources (video, voice, data, etc) in the uplink and downlink among TALOS nodes. Therefore, the TALOS communication network needs to ensure the security of the communications, so its architecture must provide the redundancy by using heterogeneous communications networks, integrated through mesh capabilities and protocols. It also has to provide the sufficient quality of service for feasible operation of the TALOS system, with a special focus on data rate, packet loss, latency, availability, priorities depending on traffic type, etc. and it should also be capable of diagnosing the communication problems.

TALOS network architecture is based on two types of network nodes: the TALOS Base Stations (TLS) and TALOS Remote Stations (TLR).

For the purpose of communications security based on redundancy, the TALOS network has three communication subsystems, based on three selected radio access technologies, which interconnect the different system nodes:

1. Primary communications, main subsystem for connecting base station nodes and remote station nodes, based on Mobile WIMAX technology, which transmits:
 - a. Radar data from the UGV to the UUCC
 - b. Commands and Mission information data between the UGV and UUCC
 - c. Surveillance video from the UGV to the UUCC
 - d. Driving video from the UGV to the UUCC
 - e. Audio data between the UGV and UUCC
 - f. Emergency stop command from the UUCC to the UGV;
2. Backup communications in order to provide redundant communication among nodes when primary communications fail, based on CDMA cellular technology at 450MHz, which transmits:
 - a. Commands and Mission information data between the UGV and UUCC
 - b. Driving video from the UGV to the UUCC
 - c. Emergency stop command from the UUCC to the UGV;
3. Safety Communications for emergency purposes, connecting the remote station nodes (that is, the unmanned ground vehicles) with the escort vehicles, based on two different technologies, Wi-Fi and UHF-FM, which in the current state of system development transmits the driving video and emergency stop command from the UGV to the Escort Vehicle (which was used for the purposes of the integration, tests and demonstration of the system).

The Communication equipment mounted on the mast attached to the UUCC is presented in figure 5.



Fig. 5. TALOS Communication equipment

The communication subsystems on the vehicles are integrated within the main system case. Each communication link has a separate omnidirectional antenna, that enables the TLR to connect to TLS in any position.

The communication is made in direct links between UGVs and UUCC. There is no option of data transmission from one UGV over another UGV, since the state of the art mesh technologies do not provide a sufficient throughput for required data and reliability to work in the security applications.

6 TALOS System Integration

As mentioned before, one of the biggest challenges, both in the course of the project realisation and in the development of the TALOS system was the phase of system integration. It was caused by the multiplicity of the elements and subcomponents needed to be integrated, but also by the fact that all the three subsystems were developed independently, by different project partners in various parts of Europe and beyond, ie.: Israel, Spain, Greece and Turkey. Moreover, some of the subsystems components were developed or delivered by other Project Partners, coming from Belgium, Estonia, Finland, France, Poland and Romania. The UGV subsystem was developed under the leadership of Israeli partner, UUCC subsystem development was leaded by the Turkish partners in close collaboration with Greece one, while Communication subsystem design and development was supervised by partner from Spain. Polish Industrial Institute for Automation and Measurements PIAP was the Project Coordinator.

The integration process was divided into two main stages – the integration of the subsystems and the whole system integration. Both stages contained several integration sessions, heading towards the full system integration, which took place in Poland and was finalised with the official demonstration in field of the Project results.

6.1 Integration of the Subsystems

At first all the subsystems needed to be integrated, as composed of numerous H/W and S/W parts. Some of them were developed by the Partners and some were COTS solutions, chosen from the relevant markets for the application in the system.

Israel, Spain and Turkey were the main integration centres, where the three system core elements were assembled. As they were derived from one system architecture, they needed to be developed in a consistent way, with a special focus on the compatibility of the interfaces, which was ensured by following the JAUS standards. For the purposes of the UGV and UUCC subsystems internal integrations the relevant subsystem simulators were developed – UGV simulator for the UUCC development and integration, as well as the temporary Command and Control Centre (located in a vehicle) to support the integration and testing of the UGVs.

Every stage of particular subsystem integration consisted a part, that was designated to check, test and possibly improve the interface between that subsystem and the other ones, specified in the architecture. That was accomplished by using a VPN connection between partners involved in subsystem development.

The integration of subsystems was divided into blocks of functionalities, with every partner and every element being accompanied by the acceptance test report, that checked item's readiness for integration.

All the bugs and problems encountered were recorded and checked regarding the influence on the subsystem development, system integration and project in general. The information was stored in the common tracking system and was available to all the partners in the TALOS consortium.

6.2 System Integration

The final integration of the whole system (the demonstrator) took place in Poland and covered several phases, including workshop and in field sessions, and was performed in 2-5 weeks cycles within the period of eleven months. It was executed partly as an iterative process to enable identification and fixing of bugs, and to gradually prepare the system for the final demonstration.

First phase was the integration between one UGV (Observer), UUCC and Communication subsystems, then – only after the successful integration of those – the second UGV (Interceptor) was also integrated within.

Plan of the integration sessions was updated after every integration phase, so that the constraints of the project were addressed.

The integration periods consisted of large amount of test cases. The system was tested in near real live scenarios and environment over long time. It enabled the system constructors to answer to the questions raised by the users and testers, and provide well improved user interfaces, TALOS use cases and instructions.

After five weeks of the last integration session a live-in-field final demonstration of the TALOS system was performed. A live operational scenario was used to present the functionalities of the TALOS technology demonstrator.

7 Summary

The concept of an innovative, transportable and fast deployable land border and large area surveillance system, based on mobile unmanned vehicles (UVs) has been implemented and tested as a demonstrator in real environment. Technology demonstrator, serving as the proof-of-concept, consisting of two unmanned ground vehicles (UGVs) and a command and control centre, with the capability to detect and locate the movements of single individuals, groups of people, and vehicles was presented and evaluated by interested end users in April 2012 on the military training ground in Poland.

The technology that followed the project concept and objectives was based on commercially available Components-Off-The-Shelf (COTS) and partners' own developments, both were integrated with respect to project scope and user requirements. That approach enabled an intensive research over the current state of the art technologies and narrowed the gap between robotic systems developed by industry and research institutions, as well as demands and needs of end users in security applications.

Concept of the system based on high degree of autonomy serving as support for protection of European land borders was examined. One of the main conclusions was that, the currently available sensors that allow much better performance in detecting and tracking, require a base system, that supports operation of border guard officers. The application of autonomous robots enables the use of technologies, that are too complicated for direct human operation. Moreover, integrating them into one system gives an opportunity to use them in a more efficient and approachable way.

The TALOS concept of providing a complete system supporting border guards activities on land borders was appreciated by officials involved in the project and present at its final demonstration.

The paper above describes the developed TALOS system demonstrator and its features, as well as the integration phase of the project.

Acknowledgements . The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement N° 218081.

References

1. Andrzejczak, M., Koziński, M., Sprońska, A.: The concept of TALOS – Transportable Autonomous Patrol for Land Border Surveillance System. In: Proceedings of Trust and Confidence in Autonomous Systems Workshop, Orlando, Florida, USA (2008)
2. Libson, Y.: TALOS – Transportable Autonomous patrol for Land Border Surveillance System. In: Proceedings of The 3rd Israeli Conference on Robotics, Herzlia, Israel (2010)
3. Virtanen, A., Parmes, E., Rainio, K., Peussa, P.: Simulation models in the concurrent development of a land border surveillance robot system. In: Proceedings of Automaatio XIX 2011, Helsinki, Finland (2011)

4. Barbe, S., Krapez, J.C., Louvet, Y.: Performance modeling and assessment of infrared-sensors applicable for TALOS project UGV as a function of target/background and environmental conditions. In: Proceedings of SPIE Defense, Security, and Sensing, Baltimore, USA (2012)
5. Tanaś, M., Hołubowicz, W., Adamczyk, A., Taberski, G.: Projekt TALOS. Rola systemów informatycznych w zapewnieniu bezpieczeństwa granic lądowych, Przedsiębiorczość i Zarządzanie. Tom X, Zeszyt 6, 101–118 (2009)
6. Tanaś, M., Hołubowicz, W., Adamczyk, A., Taberski, G.: The TALOS Project. EU wide robotic border guard. In: Proceedings of 16th International Conference on Methods and Models in Automation and Robotics, Międzyzdroje, Poland (2011)
7. Gierszal, H., Hołubowicz, W., Kiedrowski, Ł., Romanowski, K.: A Testing Platform Supporting Application Development For Ad Hoc And Challenged Networks. In: Proceedings of 14th International Multiconference Information Society, Ljubljana, Slovenia (2011)
8. Sprońska, A., Główka, J., Maciąś, M., Tanaś, M., Hołubowicz, W., Taberski, G., Samp, K.: The TALOS project – autonomous robotic patrol vehicles. In: Proceedings of the Workshop on Innovation in Border Control 2012, Odense, Denmark (2012)
9. Tanaś, M., Hołubowicz, W., Taberski, G., Adamczyk, A.: The TALOS project - Robotic Guards of the UE borders. In: Proceedings of the Safety of the Modern World - Challenges and Threats, Poznań, Poland (2011)
10. European Security Research Advisory Board. Meeting the challenge: the European Security Research Agenda. A report from the European Security Research Advisory Board. European Commission (September 2006), http://ec.europa.eu/enterprise/security/doc/esrab_report_en.pdf (web. July 28, 2008)
11. European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Examining the creation of a European Border Surveillance System (EUROSUR). Brussels (February 13, 2008) (Web. July 28, 2008)
12. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0068:FIN:EN:PDF>
13. Joint Architecture for Unmanned System, Reference Architecture (RA),
<http://www.openjaus.com/products/openjaus33/79-jaus-documents>

Localization of Essential Door Features for Mobile Manipulation

Maciej Stefańczyk and Michał Wałęcki

Warsaw University of Technology, Institute of Control and Computation Engineering,
Nowowiejska 15/19, 00-665 Warsaw, Poland
{M.Stefanczyk,M.Walecki}@elka.pw.edu.pl

Abstract. In this paper, we present a control system of a robot actively detecting and labeling door features, such as handle, lock or door plate. System is described using agent-based approach, with clear division into control, effector and receptor subsystems, with most important parts deeply explained. Presented approach supplements gap between door detection and opening algorithms, giving a robot the ability to take more actions, like recognizing room number or opening door lock.

Keywords: active vision, door identification, key hole, door handle, fuzzy inference.

1 Introduction

1.1 Motivation

Indoor spaces are natural environments for service robotics. Although there are applications in which a robot operates only in a single room, vast majority of them requires traversing through at least a part of building, where the solution of problem of detecting doors, operating them and passing through is crucial. In this paper we present a method of recognizing essential door features, such as accurate dimensions, opening direction as well as candidate regions for detecting door parts (such as a handle, key hole, text plates) with additional hints.

1.2 Related Work

Door Localization. As a crucial part in many applications, many approaches to door localization exist, used either in mobile robotics or blind people assistant applications. Solutions based on ultrasonic sensors [1] or a pointcloud analysis [2, 3] are of no use in our application due to limited usable range of sonars and 3D sensor used (Kinect). Thus the only reliable data source is color camera.

Some researchers focus on a limited subset of door localization problem, where a camera is mounted rather low, on mobile base [4–6]. This limitation renders their algorithms unusable in the case where lower door part is occluded by some furniture elements. Working only with color information is another limiting factor [7]. Algorithms based on color distribution histograms are prone to lighting

condition changes and hard to adapt in new environments, needing additional learning step for every new kind of object.

On the other hand, using only edge information [8] may lead to many false positives, caused by shelves or other rectangular objects in robot's field of view. Complicated algorithms give better results using advanced line segments analysis [9], but it takes additional processing time. Those observations lead to a conclusion, that using many weak classifiers gives better results and takes less time, due to avoiding complicated processing. That's why approaches using for example AdaBoost [5, 6] emerged, integrating simple features like parallel door frame lines, untextured kickplate, gap under door, proper proportions and size, color different from surrounding wall etc.

Handle Recognition. All algorithms presented above return estimated door position, that can be used as a goal for a robot to approach near them and take some action, with opening being the most popular. Trajectory generation for door opening is a problem itself [7, 10], and is out of scope of this article. Localization of door handle, however, is a crucial part in door opening. In this task, as robot is close to the door, both RGB and 3D information can be used. Some researchers focus on lidar scans only [3], looking for specifically for handle in dense pointcloud, rejecting all points except fixed height range. Others [7] use color images only, with some strong assumptions about handle appearance (must contain linear segment perpendicular to door frame and close to it). Eventually, mixed approach is also used [11], in which, after initial detection in RGB image, depth data is used to filter out false positives.

Problem Formulation. Previously presented algorithms solve the problems of locating doors and opening them. However – there is still a gap between those two steps. How to check, whether the door the robot is facing is this the robot wants to open and cross? How to open locked doors? The first problem can be solved by using additional information from map or localization and recognition of door plates with text on them. There are approaches to this problem [8] working successfully. The problem of locked doors is much harder to deal with – there is a wide spectrum of different door locks available with different kind of keys to handle. On doors and around them there are many features that can help robot choose a proper action (whether to open this door or find another, how to open it etc.). Door plates, handles, locks, additional devices – they all carry some useful information.

1.3 Paper Structure

This article is organized as follows. Following section describes robotic system Velma used in our experiments, focusing mainly on the robot's head and its sensory system, which is formally described in section 3. The same section outlines the algorithm used in our works. Experiments results, preceded by experimental system setup, are presented in section 4. The paper ends with summary and an outline of future works.

2 Two-Arm Robotic System Velma

Velma (fig. 1a) is a two-arm robotic system, developed in the Institute of Control and Computation Engineering, Warsaw University of Technology in 2013. It consists of two Kuka LWR-4+ manipulators with BarrettHand grippers, mounted on a 2-dof torso and a head equipped with vision sensors. It is a research platform for the field of service robotics, intended to operate in environment adapted to human needs. Velma is 180cm high and has human-like proportions which allow it to perform typical activities of human, such as opening of doors and kitchen cabinets, unscrewing jars and other manipulation tasks. Kuka LWR-4+ manipulators, combined with custom made torso, form a powerful system capable of manipulation with impedance control [10], useful in complicated, unstable environment and essential when performing tasks with possible contact with human.

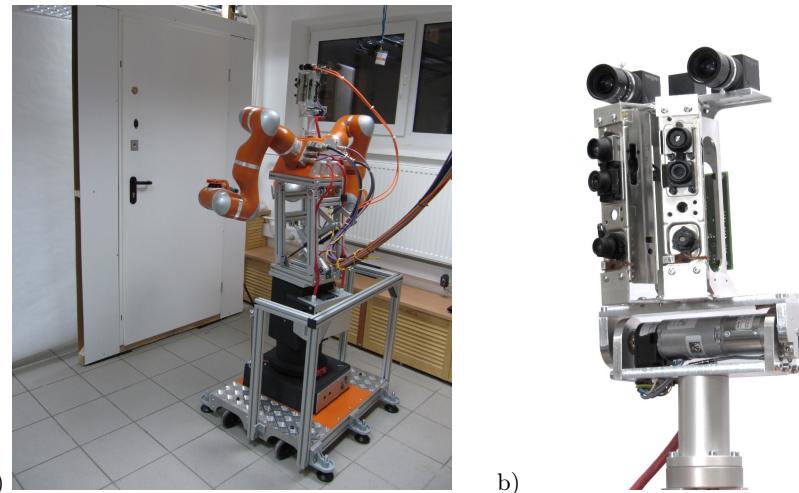


Fig. 1. Velma robot and a close-up of its head

2.1 Velma's Robotic Head

The head of Velma robot (fig. 1b) itself is a complex robotic device. It is attached to Velma's torso with a fast pan-tilt unit [12]. It gives the head the range of motion -90° to $+90^\circ$ in both horizontal and vertical direction respective to "look ahead" position. Applied drives enable the head to move with speed over $180^\circ/s$, obtaining position measurements with accuracy better than 0.1° . Main goal of employing a moving head is to extend the sensors' range. Standard cameras mounted in robot's mechatronic head can cover the whole manipulation space and actively explore entire visual field for the information that is essential to the ongoing task. In this particular case head motion is used to scan the whole door while standing next to it.

2.2 Visual Sensory System

Visual information from the environment is gathered using a set of 4 sensors calibrated with each other. On the top of the head there are two PointGrey BlackFly cameras (color, $1280px \times 1024px$), mounted with $9cm$ baseline. Each camera has $8mm$ lens giving $50^\circ \times 38^\circ$ field of view. Beneath them there are two Microsoft Kinect sensors mounted vertically, with 43° angle between them. This setup gives combined field of view of $86^\circ \times 57^\circ$ for both color and depth images from Kinect. Another advantage of depth sensors mounted this way is that their infrared projectors don't interfere with each other giving clean measurements from both units.

3 Robot Control System Specification

In the control system design and implementation an *agent* approach was utilized [13]. According to it, the whole robot with its control system is considered an embodied agent, which is defined as a device or program that has the ability to perceive its surroundings for subsequently influencing the environment state, can communicate with other agents and has an internal imperative to achieve its goal. The agent's goal in this case is to identify and localize features of the door that are essential from the point of view of further task of door opening. Fig. 2 shows internal structure of a generic agent, distinguishing five subsystems. Control subsystem c represents a set of high level algorithms that pilot the task. It communicates with virtual effectors e_i and virtual receptors r_i that form an abstraction layer for real hardware. Virtual effectors and virtual receptors provide convenient software interface to corresponding real effectors E_i and real receptors R_i , which represent agent's corporeal body. Control subsystem, virtual effectors and virtual receptors form a control system that runs on a PC. Each of subsystem owns a set of behaviors, which consist of transfer functions, start condition and terminal condition.

There is a variety of effectors which represent Velma's manipulators and drives of the torso, as well as a variety of receptors, including all the sensors mounted on Velma's head. However, most of them do not play any role in this work and will not be considered.

3.1 Real Effector and Virtual Effector

Effector E_1 represents the pan-tilt unit, moving the robot's head. Its joints are driven with electrical motors powered by custom motor controllers, specially designed for robotic research applications [14], that communicate with PC.

During the task of door recognition only one of behaviours of pan-tilt unit virtual effector e_1 is utilized. The operation of transfer functions corresponding to it proceeds as follows.

Virtual effector e_1 internal memory ${}^e e_{dr,h}$ contains a vector of last measured motors positions m_c and a scheduled reference trajectory, written as a vector of

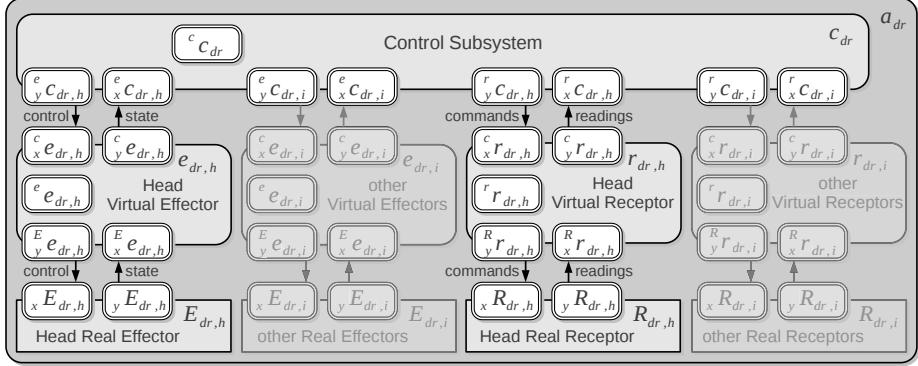


Fig. 2. General agent structure

triplets of joint positions q_d , speeds \dot{q}_d and accelerations \ddot{q}_d to be executed in every subsequent control cycle. Whenever a new target joint position q_t is set by control subsystem through virtual effector's communication buffer $c_x e_{dr,h}$, a new reference trajectory is computed, beginning from the first scheduled triplet to be executed $q_{d,t+1}, \dot{q}_{d,t+1}, \ddot{q}_{d,t+1}$ and leading to q_d joint position with zero speed and acceleration. When virtual effector receives new readings of motor positions m_c from motor controllers through $x e_{dr,h}$, their values are stored in e_1 's internal memory.

Generation of control data for motor controllers is triggered by external clk signal. Its period of $10ms$ determines the control cycle duration. In each control cycle a subsequent position $q_{d,t+1}$ of joint reference trajectory and current motors position are read from memory and on this basis desired motors positions and speed limits for the next cycle are computed. Together with desired mode of operation they are sent to motor controllers via $y e_{dr,h}$ communication buffer. During the task of door recognition both motor controllers operate in the mode of position regulation with configurable speed limit.

Triggered with the same signal, every control cycle the latest values of motor positions are converted to joint position values and sent to control subsystem through $c_y e_{dr,h}$ communication buffer.

3.2 Receptor

Gathering Photos. Similarly to effector, receptors $R_{dr,i}$ represent image sensors mounted on robot's head and virtual receptor $r_{dr,r}$ provides a software interface to them, aggregating data and processing it. Although in our system we have more than one real receptor (presented in sec. 2.2), in presented task only one camera is used. To create detailed door image, the robot must be very close to it, which results in acquiring the whole door width at once, but a only limited fraction of its height. Virtual receptor $r_{dr,r}$ has to create a single image from few gathered, so it has different behaviors. One is responsible for gathering

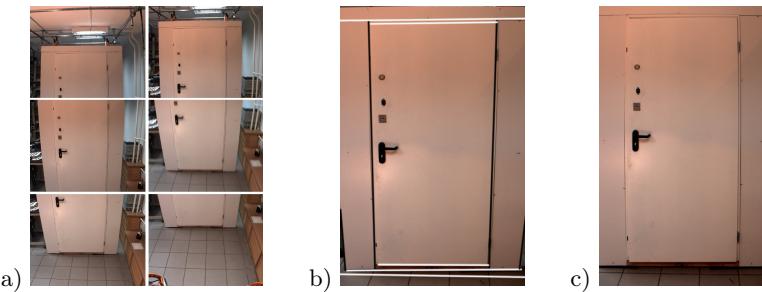


Fig. 3. Input photos (a), stitched, distorted image (b) and corrected image (c)

a series of photos for different head positions – after each trigger from control subsystem (generated after setting a new head position, in 10° steps down from straight-ahead position) a new photo is gathered and stored in internal buffer $r_{dr,h}$ (fig. 3a).

After obtaining all 6 photos, the next behavior is selected, which is responsible for stitching them to create full door view. In general, this task is rather hard to do, but as our scene is mostly flat (doors are only slightly recessed into the wall) it can be done with good results using built-in OpenCV functionality. To retain straight lines in resulting image, remapping must be done using planar warper. Resulting image is almost always perspective distorted (fig. 3b), so, in next step, this distortion is estimated and corrected. Hough transform is used to detect lines in image, which are then analyzed and grouped into horizontal (less than 30°) vertical (greater than 60°) and other (presented on fig. 3b as white, black and gray, respectively). Horizontal and vertical lines are used to estimate vanishing points in image, according to following steps:

1. pick two random lines
2. if angle between them is less than threshold - go to (1)
3. calculate and remember their intersection
4. if there is less than 20 intersection points calculated - go to (1)
5. calculate mean of all points
6. if all points lay in ϵ neighborhood of calculated mean - select it as vanishing point and finish
7. otherwise remove furthest point and go to (5)

Vanishing points are then used to generate remapping from input, distorted image to straight one (fig. 3c).

Regions of Interest Segmentation. Having a straight, complete door image, the next behavior of virtual receptor is responsible for segmenting the door itself and all regions of interest on them and around its frame. Both steps use binary image created by adaptive segmentation and some morphological operations (to remove noise). It makes the algorithm insensitive to lighting conditions, door

color and texture (partly), and only slight shadow or deviation in intensity is needed to properly detect edges of door frame or other features (fig. 4b). Line segments, extracted from binary image, are again grouped in horizontal and vertical lines, this time allowed deviation is only 3° , as image is rectified and door frame should be parallel to x and y axis. Door frame segmentation proceeds as follows:

1. pick the longest vertical line
2. find horizontal lines ending close to the vertical line's upper end and pick the longest one
3. if there is no such line - remove the first line and start over
4. find vertical lines close to the other end of a lintel and pick the longest one
5. if there is no such line - remove lintel line and go to (2)
6. find horizontal line connecting both vertical lines bottom ends

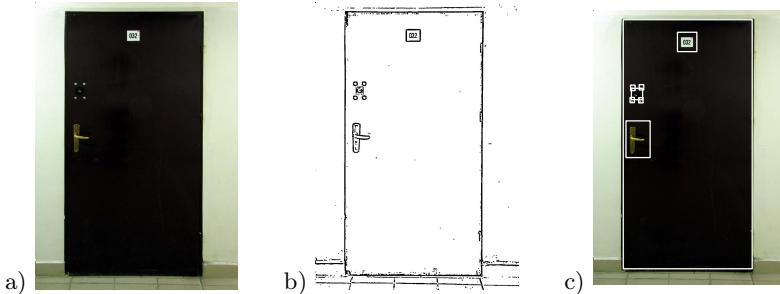


Fig. 4. Input image (a), thresholding (b) and final segmentation (c)

The obtained closed loop is considered as a door frame and acts as a reference border for the next algorithm stages. From all other edges, those with sufficient size and line width are selected and act as regions of interest passed to labeling step. Selected ROIs are presented on fig. 4c.

Regions of Interest Labeling. The regions detected in the previous step can be of many different kinds – door locks, handles, light switches, door plates etc. To make use of it and pick the best algorithm to detect its type, some initial guess is needed about its purpose. In our system, we defined classes for said objects, supplemented with unknown objects. A thing, that is important in our approach is that we don't limit any region to only one class, giving them confidence score for any existing possibility. As different objects can be found in different locations and there are no simple rules to distinguish between them, fuzzy rules were prepared to describe every class, based on few, basic properties – location in a door plane (horizontal and vertical), size and shape regularity. Corresponding fuzzy sets are presented on fig. 5.

For object's horizontal position we have two possibilities – **Edge** and **Center**, which are based on a normalized distance from the door frame. Similarly, vertical

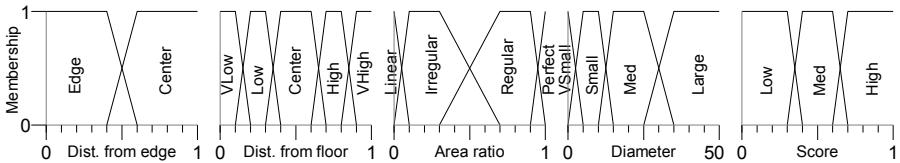


Fig. 5. Fuzzy sets for linguistic variables in presented system

position is based on normalized distance from the floor and possible values are **VeryLow**, **Low**, **Center**, **High** and **VeryHigh**. Shape regularity is computed using ellipse and rectangle circumscribed on segment according to given formula:

$$Ratio = \max \left\{ \frac{A_S}{A_R}, \frac{A_S}{A_C} \right\}$$

where A_S , A_R and A_C is area of segment, circumscribed rectangle and ellipse accordingly. A segment can have **Perfect** shape if it's elliptical or rectangular, **Regular**, **Irregular** or **Linear** shape (when it's composed from linear parts only). Size could be **VerySmall**, **Small**, **Medium** and **Large**, depending of diameter of circumscribed circle. All output variables (each class scores) have similar distribution presented as Score on fig. 5, and could be **Low**, **Medium** or **High**.

Another part of our system is a rule database describing relationships between input and output variables. Few rules are presented below:

```
if Size is Med and Vert is High then Plate is High
if Size is Small and Shape is Regular and Vert is High then Plate is High
if Size is VSmall and Horiz is Edge and Vert is High then Lock is High
```

Our system is composed of more than twenty rules at the moment. Operators used were basic **min** and **max** for AND/OR, **min** for implication and **max** for accumulation. Bisection was used as a defuzzifier.

Final decision about object class was made based on a defuzzified result given by system described earlier. If final score for object for all classes was below given value S_{low} , then this object was labeled as unknown. Also, from every segment placed on the wall Lock and Handle classes were removed and from segments placed on the door Switch class was removed.

4 Experiments

To confirm the validity of the approach, tests on a set of different doors were carried. Initially, the robot is placed near the door we're interested in and looks roughly forward at them (fig. 1a). Then series of 6 photos is taken, moving the head (in 10° increments down from straight forward position), which are then merged and analyzed according to the algorithm presented earlier.

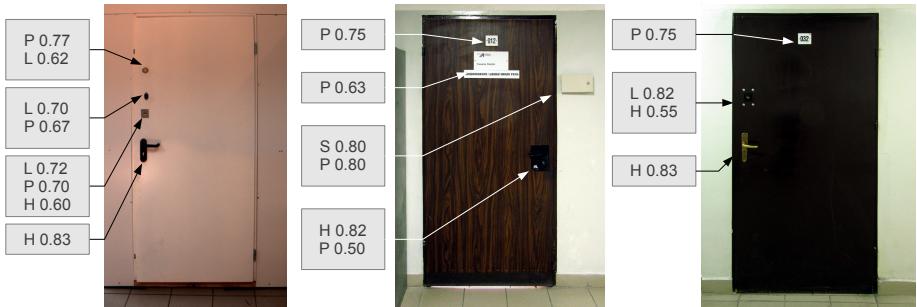


Fig. 6. A sample of test results. Labels shown for classes with membership bigger than 0.4. H - handle, L - lock, P - plate, S - switch

Algorithm was tested on photos of doors varying in color and texture, different handles, door plates and other features. Three of them, with resulting labeling, are presented on fig. 6. For every segment its membership to classes is presented (for score bigger than 0.4).

5 Summary

In this paper we presented a new approach to detection of regions of interest in autonomous door opening task. The analysis of current solutions yielded uncovered area between door detection, handle identification and door opening, which was then analyzed and solution was proposed. Resultant robotic system was presented using agent approach, with clear division into control, effector and receptor subsystems.

Developed vision system is capable of detecting door frame and different types of features (handles, locks, plates, switches) providing clues for succeeding recognition modules. The usage of fuzzy inference gave our system robustness, which resulted in good labeling even for some not typical cases. This system is planned to be extended with some other variables, like texture diversity as input or other object classes as output. As future work we are going also to create a larger database of different doors, with bigger diversity of features. Another thing is determination of exact door parameters (i.e. size) and their opening direction, which involves the usage of three dimensional sensing (using point clouds gathered by a stereo-pair and Kinect sensors), which will also enable the usage of multimodal image segmentation prepared earlier [15].

Acknowledgment. This project was supported by the grant 2012/05/D/-ST6/03097 funded by National Science Centre.

References

- Carinena, P., Regueiro, C.V., Otero, A., Bugarin, A.J., Barro, S.: Landmark detection in mobile robotics using fuzzy temporal rules. *IEEE Transactions on Fuzzy Systems* 12(4), 423–435 (2004)
- Anguelov, D., Koller, D., Parker, E., Thrun, S.: Detecting and modeling doors with mobile robots. In: Proceedings of the 2004 IEEE International Conference on Robotics and Automation, ICRA 2004, vol. 4, pp. 3777–3784. IEEE (2004)
- Rusu, R.B., Meeussen, W., Chitta, S., Beetz, M.: Laser-based perception for door and handle identification. In: International Conference on Advanced Robotics, ICAR 2009, pp. 1–8. IEEE (2009)
- Aude, E., Lopes, E., Aguiar, C., Martins, M.: Door crossing and state identification using robotic vision. In: 8th International IFAC Symposium on Robot Control (Syroco 2006), Bologna, Italy (2006)
- Chen, Z., Birchfield, S.T.: Visual detection of lintel-occluded doors from a single image. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008, pp. 1–8. IEEE (2008)
- Hensler, J., Blaich, M., Bittel, O.: Real-time door detection based on adaBoost learning algorithm. In: Gottscheber, A., Obdržálek, D., Schmidt, C. (eds.) EU-ROBOT 2009. CCIS, vol. 82, pp. 61–73. Springer, Heidelberg (2010)
- Brooks, R., Aryananda, L., Edsinger, A., Fitzpatrick, P., Kemp, C.C., O'Reilly, U.M., Torres-Jara, E., Varshavskaya, P., Weber, J.: Sensing and manipulating built-for-human environments. *International Journal of Humanoid Robotics* 1(1), 1–28 (2004)
- Tian, Y., Yang, X., Yi, C., Arditi, A.: Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. *Machine Vision and Applications*, 1–15 (2013)
- Yang, X., Tian, Y.: Robust door detection in unfamiliar environments by combining edge and corner features. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 57–64. IEEE (2010)
- Winiarski, T., Banachowicz, K.: Opening a door with a redundant impedance controlled robot. In: 9th Workshop on Robot Motion & Control (RoMoCo), pp. 221–226 (2013)
- Klingbeil, E., Saxena, A., Ng, A.Y.: Learning to open new doors. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2751–2757. IEEE (2010)
- Walęcki, M., Stefańczyk, M., Kornuta, T.: Control system of the active head of a service robot exemplified on visual servoing. In: 9th Workshop on Robot Motion and Control (RoMoCo), pp. 48–53 (2013)
- Kornuta, T., Zieliski, C.: Robot control system design exemplified by multi-camera visual servoing. *Journal of Intelligent & Robotic Systems*, 1–25 (2013)
- Walęcki, M., Banachowicz, K., Winiarski, T.: Research oriented motor controllers for robotic applications. In: Kozłowski, K. (ed.) Robot Motion and Control 2011. LNCIS, vol. 422, pp. 193–203. Springer, Heidelberg (2012)
- Stefañczyk, M., Kasprzak, W.: Multimodal segmentation of dense depth maps and associated color information. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 626–632. Springer, Heidelberg (2012)

Motion Planning for the Mobile Platform Rex

Krzysztof Tchoń, Krzysztof Arent, Mariusz Janiak, and Łukasz Juszkiewicz

Institute of Computer Engineering, Control and Robotics,
Wrocław University of Technology,
Janiszewskiego Street, 11/17, 50-372 Wrocław, Poland
{krzysztof.tchon,krzysztof.arent,
mariusz.janiak,lukasz.juszkiewicz}@pwr.wroc.pl

Abstract. This paper deal with the motion planning of a wheeled, skid-steering platform, with four independently actuated wheels that slip laterally and longitudinally. Two methods of motion planning are proposed: the Endogenous Configuration Space Approach, and the Optimal Control Method. Their performance is illustrated by numeric computations.

Keywords: wheeled mobile platform, dynamics, motion planning.

1 Introduction

A major challenge of the rescue and exploration robotics is endowing these kind of robots with some amount of autonomy [1], in order to increase so called neglect time of the robot operator [2]. For practical reasons, a semi-autonomous control scheme is recommended, where robot executes low-level tasks under the operator's supervision. Increasing the degree of autonomy of the rescue and exploration robots has been the main objective of the project RobREx funded by the Polish National Centre for Research and Development. In order to achieve this objective a mobile platform Rex will be built at the Wrocław University of Technology. This platform will serve as a testbed for model-based motion planning and control algorithms of a wheeled mobile manipulator dedicated to rescue and exploration operations. Currently, the platform is in the phase of mechatronic design including computer implementation of its dynamics model and performance simulations of its motion planning and tracking control algorithms. The platform's general view is shown in Figure 1. The design methodology of the Rex platform is standard, patterned on the procedure of designing a robotic manipulator, described in [3]. First, the user's requirements should be defined, as presented in [4], then individual platform components are designed. To this aim, the methods of 3D CAD have been employed, that allow to simultaneously develop the mechatronic design, conduct on-line necessary analyses and computations concerned with the platform dynamics and strength parameters, and deploy in the most rational way the platform on-board equipment. Looking from the perspective of mathematical modeling of the Rex platform dynamics, a correctly structured 3D platform model is able to provide representative values of parameters for the equations of motion. Specifically, this refers to the masses

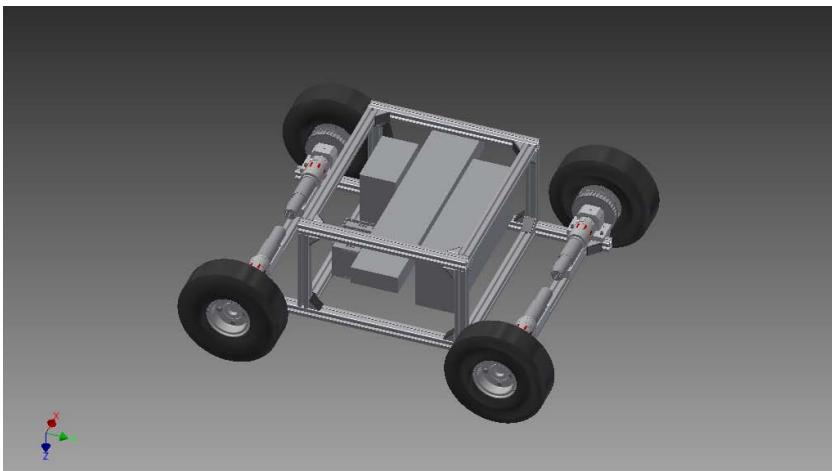


Fig. 1. Rex: general view

of the platform body and wheels, coordinates of centers of mass, and the corresponding inertia tensors. A non-standard element of the Rex design process is taking into account, already at the stage of defining the user's requirements, constraints imposed on the platform construction by typical assumptions made at the mathematical modeling. For this reason it has been decided that the Rex construction should be as simple as possible. In particular, in order to decrease the friction in the platform power units, it has been assumed that each wheel should be actuated driven by an individual motor, and that the motor and wheel axles are collinear. The intertwined 3D and mathematical modeling adopted in this paper allowed to establish adequacy of the geometric model 2 with the 3D model 1, and provided realistic values of dynamic parameters to the dynamics model of the platform.

A detailed derivation of dynamics models of the Rex platform, under various possible slip/non-slip conditions of the wheels has been presented in the technical study [5]. By recommendation of this study, special attention should be paid to the dynamics model allowing for the slip of all platform wheels. This paper is devoted to the motion planning problem of the platform Rex. The dynamics model of the platform is endowed with physical parameters resulting from the mechatronic design using the Inventor package. Two methods of solving the motion planning problem have been proposed: the Endogenous Configuration Space Approach (ECSA) [6], and the Optimal Control Method (OPCM) as implemented numerically in the ACADO toolkit [7,8]. Performances of both these methods have been illustrated and compared via computer simulations. This paper focuses solely on motion planning whose aim is to determine a nominal trajectory of the platform. The tracking of this trajectory will be achieved by a tracking control algorithm based on the predictive control scheme. It is hypothesized that frequently enough updates of the predictive control algorithm will

cope with the uncertainty of the traction conditions of the platform's wheels. Further results on this subject will be reported separately.

The paper is organized in the following way. Section 2 introduces a control theoretic model of the platform dynamics. In Section 3 the motion planning problem is defined, along with two methods of its solution. An application of these methods to an example motion planning problem is presented in Section 4. Section 5 concludes the paper.

2 Modeling

A geometric schematic of the Rex platform is shown in Figure 2. It is assumed

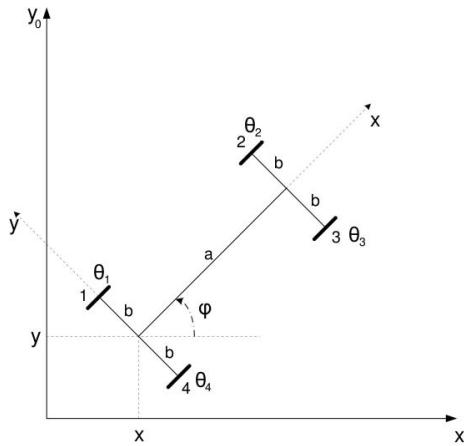


Fig. 2. Rex: geometric schematic

that the platform moves on the horizontal plane, its wheels are identical, actuated independently, and may slip laterally and longitudinally. The platform motion will be described with respect to the space reference frame (X_0, Y_0, Z_0) with Z -axis pointing upward. The body frame (X, Y, Z) has been attached in the middle of the platform rear axle, with Z -axis going upward and X axis set along the platform, directed forward. Natural generalized coordinates of the platform are $q = (x, y, \varphi, \theta_1, \theta_2, \theta_3, \theta_4)^T \in \mathbb{R}^7$, and have been explained in Figure 2. For computational reasons these coordinates can be made dimensionally consistent by setting $w = (x, y, a\varphi, R\theta_1, R\theta_2, R\theta_3, R\theta_4)^T$, R denoting the wheel radius, so each coordinate has the same dimension (of length). The platform dynamics can be described by the following control system

$$P(w)\ddot{w} + D(w, \dot{w}) = F(w, \dot{w}) + \frac{1}{R}Bu, \quad y = k(w), \quad (1)$$

where terms P and D refer to the inertial and centripetal/Coriolis components of the dynamics, F describes counteraction forces against the slip, $u = (u_1, u_2,$

$u_3, u_4) \in \mathbb{R}^4$ represents actuation torques, and the matrix $B = [0_{4 \times 3}, I_4]^T$ transmits torques to the platform coordinates. The output function selects coordinates subject to the motion planning. Specifically, we have the following data

$$P(w) = \begin{bmatrix} Q_{11} & 0 & \frac{Q_{13}}{a} & 0 & 0 & 0 & 0 \\ 0 & Q_{22} & \frac{Q_{23}}{a} & 0 & 0 & 0 & 0 \\ \frac{Q_{13}}{a} & \frac{Q_{23}}{a} & \frac{Q_{33}}{a^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{Q_{44}}{R^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{Q_{55}}{R^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{Q_{66}}{R^2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{Q_{77}}{R^2} \end{bmatrix}, D(w, \dot{w}) = \frac{\dot{w}_3^2}{a^2} (-Q_{23}, Q_{13}, 0, 0, 0, 0, 0)^T, \quad (2)$$

where $Q_{11} = Q_{22} = m_p + 4m_w$, $Q_{44} = Q_{55} = Q_{66} = Q_{77} = I_{w33}$, $Q_{13} = -m_p (a_{p1} \sin \frac{w_3}{a} + a_{p2} \cos \frac{w_3}{a}) - 2m_w a \sin \frac{w_3}{a}$, $Q_{23} = m_p (a_{p1} \cos \frac{w_3}{a} - a_{p2} \sin \frac{w_3}{a}) + 2m_w a \cos \frac{w_3}{a}$, $Q_{33} = I_{p33} + m_p (a_{p1}^2 + a_{p2}^2) + 4(I_{w11} + m_w b^2) + 2m_w a^2$. The symbols have the following meaning: m_p and m_w = mass of the platform and of the single wheel, a_{p1} and a_{p2} = position of the platform mass center in body frame, I_{w11}, I_{w33} = wheel principal moments of inertia with respect to X and Z axis of the body frame, I_{p33} = the platform principal moment of inertia with respect to Z axis of the body frame, a, b and R have already been defined. A computation supported by the Autodesk Inventor 3D CAD software has resulted in the following values of the Rex platform parameters: $m_p = 21.107\text{kg}$, $m_w = 2.380\text{kg}$, $a_{p1} = 0.377\text{m}$, $a_{p2} = 0.008\text{m}$, $a = 0.730\text{m}$, $b = 0.350\text{m}$, $R = 0.127\text{m}$, $I_{p33} = 1.991\text{k}gm^2$, $I_{w11} = 0.015\text{k}gm^2$, $I_{w33} = 0.009\text{k}gm^2$. In w coordinates the Pfaffian matrix characterizing the slip of the wheels takes the following form

$$H(w) = \begin{bmatrix} -\sin \frac{w_3}{a} \cos \frac{w_3}{a} & 0 & 0 & 0 & 0 & 0 \\ -\sin \frac{w_3}{a} \cos \frac{w_3}{a} & 1 & 0 & 0 & 0 & 0 \\ -\sin \frac{w_3}{a} \cos \frac{w_3}{a} - \frac{b}{a} & -1 & 0 & 0 & 0 & 0 \\ -\sin \frac{w_3}{a} \cos \frac{w_3}{a} - \frac{b}{a} & 0 & -1 & 0 & 0 & 0 \\ -\sin \frac{w_3}{a} \cos \frac{w_3}{a} & \frac{b}{a} & 0 & 0 & -1 & 0 \\ -\sin \frac{w_3}{a} \cos \frac{w_3}{a} & \frac{b}{a} & 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} H^1(w) \\ H^2(w) \\ H^3(w) \\ H^4(w) \\ H^5(w) \\ H^6(w) \end{bmatrix}, \quad (3)$$

$H^i(w)$ denoting the rows of $H(w)$. The slips of the wheels can be computed as follows: the lateral slip of the rear wheels $s_{14} = H^1(w)\dot{w}$, the lateral slip of the front wheels $s_{23} = H^2(w)\dot{w}$, and the longitudinal slip of the wheel i equals $s_i = H^{i+2}(w)\dot{w}$, $i = 1, 2, 3, 4$. The presence of slipping induces counteraction forces at the contacts of wheels with the ground. The linear model of these forces has been adopted, according to which these forces directed along the transposed rows of the matrix $H(w)$, while their values are proportional to the amount of slip [4,5]. In this setting, the force $F(w, \dot{w})$ on the right hand side of (1) takes the form

$$F(w, \dot{w}) = R_{14} \frac{H^{1T}(w)}{\|H^{1T}(w)\|} + R_{23} \frac{H^{2T}(w)}{\|H^{2T}(w)\|} + \sum_{i=1}^4 R_i \frac{H^{i+2T}(w)}{\|H^{i+2T}(w)\|}. \quad (4)$$

In the above expression $R_{14} = -(\varepsilon_1 N_1 + \varepsilon_2 N_2)s_{14}$, $R_{23} = -(\varepsilon_2 N_2 + \varepsilon_3 N_3)s_{23}$, $R_i = -\tau_i N_i s_i$, $i = 1, \dots, 4$, and $\|\cdot\|$ is the Euclidean norm. N_i denotes the normal

force exerted by the wheel i on the ground. The slip coefficients ϵ_i (lateral) and τ_i (longitudinal) characterize the dynamic friction of the wheels against the ground, and should be determined experimentally.

By combining the identities (2)-(4) and substituting them into (1) the dynamics equations of the Rex platform are obtained in the form of the control affine system with output function

$$\dot{x} = f(x) + g(x)u = f(x) + \sum_{i=1}^m g_i(x)u_i, \quad y = k(x), \quad (5)$$

where $x = (w, \dot{w}) = (x^1, x^2) \in \mathbb{R}^{14}$, $u \in \mathbb{R}^4$, and with a slight abuse of notations, $f(x) = (x^2, -P^{-1}(x^1)(D(x) + F(x)))$, and $g(x) = [0_{7 \times 4}, \frac{1}{R}P^{-1}(x^1)B]$. The output function selects the platform position and orientation, $k(x) = (w_1, w_2, \frac{w_3}{a})$. By definition, all vector fields and functions appearing in (5) are smooth.

3 Motion Planning

In order to facilitate the presentation of general concepts, we let the dynamics model of the Rex platform have the form (5) with $u \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^r$. Let $T > 0$ denote a control time horizon. It will be assumed that admissible control functions in (5) belong to the space $L_m^2[0, T]$ of Lebesgue square integrable functions. For an initial state x_0 and a time instant t , let $\varphi_{x_0, t}(u(\cdot))$ be the trajectory of (5) subject to a control $u(\cdot)$. Then, the end point map of system (5)

$$K_{x_0, T}(u(\cdot)) = k(x(T)) = k(\varphi_{x_0, T}(u(\cdot))). \quad (6)$$

computes the system output response at T to the control function $u(\cdot)$. Given the end point map, the motion planning problem in system (5) consists in determining a control function $u(t)$ that drives the system's output at T to a prescribed point y_d , so that $y(T) = K_{x_0, T}(u(\cdot)) = y_d$.

3.1 Endogenous Configuration Space Approach (ECSA)

A basic concept of ECSA is the endogenous configuration space $\mathcal{X} = L_m^2[0, T]$ identified with the Hilbert space $L_m^2[0, T]$ equipped with the inner product, $\langle u_1(\cdot), u_2(\cdot) \rangle = \int_0^T u_1^T(t)u_2(t)dt$. The derivative of (6) with respect to the control is computed by means of the linear approximation of system (5) along $(u(t), x(t))$, in the following way

$$DK_{x_0, T}(u(\cdot))v(\cdot) = J_{x_0, T}(u(\cdot))v(\cdot) = \rho(T) = C(T) \int_0^T \Phi(T, s)B(s)v(s)ds, \quad (7)$$

where

$$\dot{\xi}(t) = A(t)\xi(t) + B(t)v(t), \quad \rho(t) = C(t)\xi(t), \quad \xi(0) = 0, \quad (8)$$

whereas

$$\begin{aligned} A(t) &= \frac{\partial(f(x(t)) + g(x(t))u(t))}{\partial x}, & B(t) &= g(x(t)), \\ C(t) &= \frac{\partial k(x(t))}{\partial x}, & \frac{\partial \Phi(t, s)}{\partial t} &= A(t)\Phi(t, s), & \Phi(s, s) &= I_n. \end{aligned} \quad (9)$$

The map $J_{x_0, T}(u(\cdot)) : \mathcal{X} \rightarrow \mathbb{R}^r$ defined by (7) is referred to as the Jacobian of system (5). Within the ECSA the motion planning problem can be solved using Jacobian algorithms [6], originated in the homotopy (continuation) method [9]. If the Moore-Penrose inverse of the Jacobian is chosen, the Jacobian motion planning algorithm is determined by the trajectory of the dynamic system

$$\frac{du_\theta(t)}{d\theta} = -\gamma B^T(t)\Phi^T(T, t)C^T(T)\mathcal{G}_{x_0, T}^{-1}(u_\theta(\cdot))(K_{x_0, T}(u_\theta(\cdot)) - y_d) \quad (10)$$

evolving in the endogenous configuration space, $\gamma > 0$ being the error decay rate. If $u_\theta(t)$ is a trajectory of this system, a solution of the motion planning problem is defined as a limit $u_d(t) = \lim_{\theta \rightarrow +\infty} u_\theta(t)$. The matrix $\mathcal{G}_{x_0, T}(u(\cdot))$ appearing in (10) is the output Gram matrix of the linear approximation (8), referred to as the mobility matrix of system (5), [10]. This matrix can be computed by integrating the Lyapunov differential equation

$$\dot{M}(t) = B(t)B^T(t) + A(t)M(t) + M(t)A^T(t), \quad (11)$$

with $M(0) = 0$, and setting $\mathcal{G}_{x_0, T}(u(\cdot)) = C(T)M(T)C^T(T)$. The presented Jacobian motion planning algorithm for system (5) refers to the unconstrained problem. For a motion problem with constraints this algorithm may be replaced by either the imbalanced Jacobian [11] or the prioritized Jacobian algorithm [12].

For computational reasons the control functions of system (5) will be represented by the truncated Fourier series, in the form

$$u_\lambda(t) = P_s(t)\lambda, \quad (12)$$

where $P_s(t) = \text{diag}\{P(t), P(t), \dots, P(t)\}$ is a block diagonal matrix built of m copies of a row matrix $P(t) = [1, \sin \omega t, \cos \omega t, \dots, \cos p\omega t]$, $\omega = 2\pi/T$, containing $2p+1$ basic functions, $s = m(2p+1)$, and $\lambda \in R^s$ collects control parameters. After the parameterization of controls the Jacobian (7) converts to a matrix

$$J_{q_0, T}(\lambda) = C_\lambda(T) \int_0^T \Phi_\lambda(T, t)B_\lambda(t)P_s(t)dt,$$

where matrices $A_\lambda(t)$, $B_\lambda(t)$, $C_\lambda(t)$, and $\Phi_\lambda(t, w)$ are computed in accordance with (9) along $(u_\lambda(t), q_\lambda(t))$. It is easily seen that Jacobian $J_{q_0, T}(\lambda) = C_\lambda(T)J_\lambda(T)$, where $J_\lambda(t)$ solves the differential equation

$$\dot{J}_\lambda(t) = A_\lambda(t)J_\lambda(t) + B_\lambda(t)P_s(t), \quad (13)$$

initialized at $J_\lambda(0) = 0$. In the parametric setting the basic equation (10) is

$$\lambda_{\theta+1} = \lambda_\theta - \gamma J_{q_0, T}^T(\lambda_\theta)\mathcal{G}_{q_0, T}^{-1}(\lambda_\theta)(K_{x_0, T}(\lambda_\theta) - y_d), \quad \theta = 0, 1, \dots, \quad (14)$$

for the parametric mobility matrix $\mathcal{G}_{q_0,T}(\lambda) = J_{q_0,T}(\lambda)J_{q_0,T}^T(\lambda)$. An alternative, non-parametric solution of equation (10) has been proposed in [13].

3.2 Optimal Control Method (OPCM)

Alternatively, the motion planning problem can be stated and solved as an optimal control problem of minimizing the quadratic cost functional

$$\min_{u(\cdot)} \int_0^T ((y(t) - y_d)^T \mathcal{P}(y(t) - y_d) + u(t)^T \mathcal{R}u(t)) dt \quad (15)$$

with a symmetric, positive matrix \mathcal{P} , and a symmetric, positive definite matrix \mathcal{R} , subject to system equations (5), and possibly, control or state constraints in the form $s(x(t), u(t)) \leq 0$ (permanent constraints) or $r(x(0), x(T)) \leq 0$ (boundary constraints). The numeric solution of the motion planning problem will employ Sequential Quadratic Programming and Direct Multiple Shooting approach [7], as implemented in the ACADO (Automatic Control and Dynamic Optimization) toolkit [8].

4 Results

The methods ECSA and OPCM will be applied to the following motion planning problem of the mobile platform Rex represented by (5): Given an initial state $x_0 = (0, 0, a\frac{\pi}{2}, 0_{1 \times 11})^T$, a time horizon $T = 4$, and a desired output $y_d = (10, 0, \frac{\pi}{2})^T$, find a control $u(t) \in \mathbb{R}^4$ such that $y(T) = (w_1(T), w_2(T), \frac{w_3(T)}{a})^T = y_d$. The slip coefficients in (4) are set to $\epsilon_i = 1$, $\tau_i = 3$, $i = 1, \dots, 4$. Computations stop after the error norm $\|y(T) - y_d\|$ drops below 10^{-2} . The problem is a sort of the parking maneuver performed with a mean lateral speed of $2.5m/s$.

In the case of ECSA the controls have the form of truncated Fourier series containing the constant and up to the 3rd order harmonics. Vector of initial values of control parameters: $\lambda_{0i} = (0.5, 0.01, 0.01, 0.001, 0.001, 0.0001, 0.0001)^T$, $i = 1, \dots, 4$, the error decay rate $\gamma = 0.05$. A solution provided by ECSA is demonstrated in Figure 3. It was computed using MATLAB, version R2010a, in 137 steps of the algorithm. On CPU Intel(R) Core(TM) i7-2600K CPU @ 3.40GHz with RAM: 8GB DDR3 1333MHz, running Ubuntu 10.04 LTS 64bit the computations took approximately 114s.

OPCM uses piecewise constant controls on 100 time intervals, initialized with $u_i(t) = 5$, $i = 1, \dots, 4$. Computations have been run for the weight matrices $\mathcal{P} = \text{diag}\{1, 1, 2\}$ and $\mathcal{R} = I_4$. The condition of reaching the desired point y_d with prescribed accuracy 10^{-2} is encoded as a boundary constraint. Results obtained by OPCM are displayed in Figure 4. In the top part of both these figures the platform orientation has been marked. Control torques applied to the left (1 and 2) and to the right (3 and 4) wheels are identical, so the platform behaves as if its side wheels were coupled. The ACADO Toolkit v1.1.0 beta, gcc version 4.4.3 (Ubuntu 4.4.3-4ubuntu5.1) was implemented on the same CPU. The solution was found in 39 iterations, within approximately 228s.

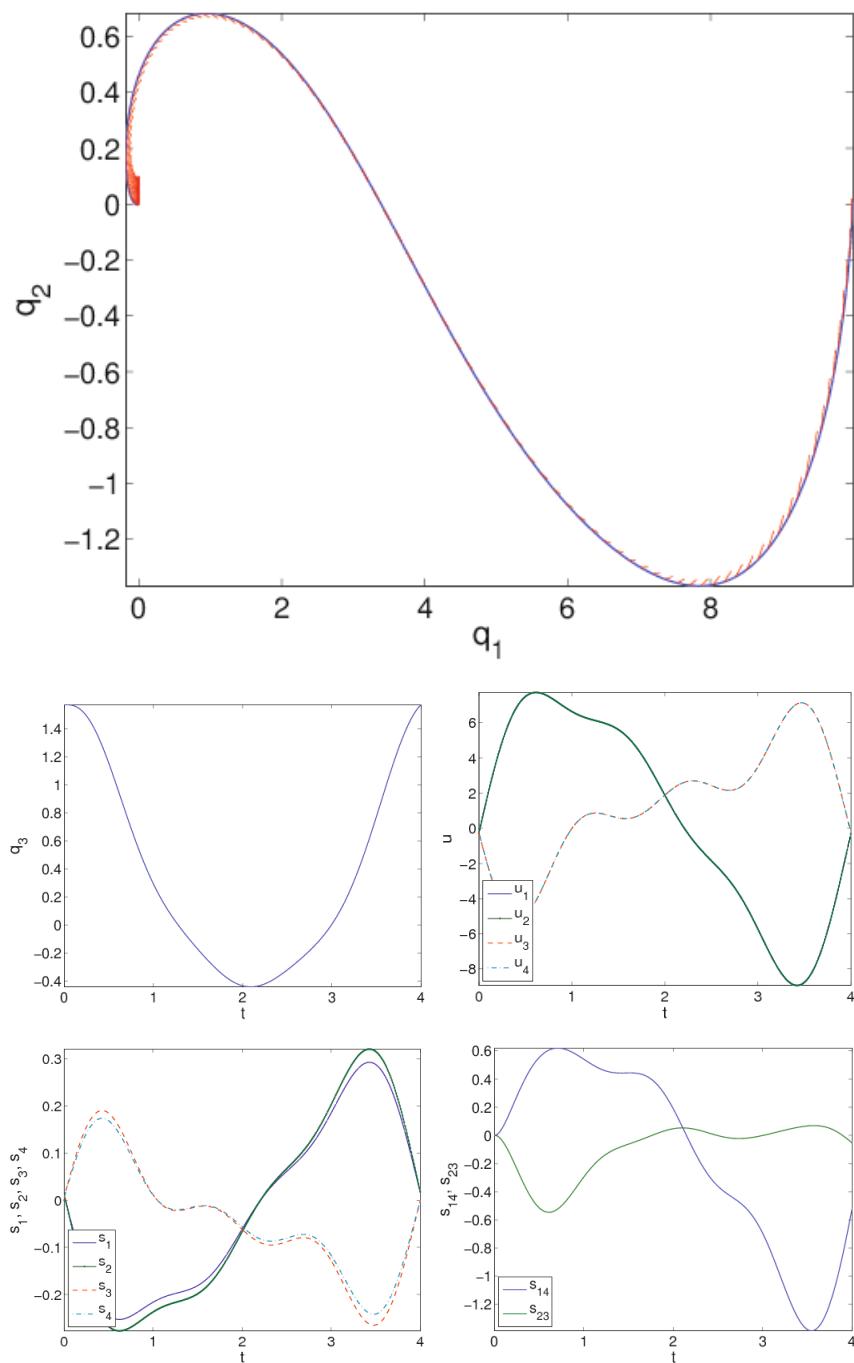


Fig. 3. ECSA: platform path, orientation, controls, longitudinal slips, lateral slips

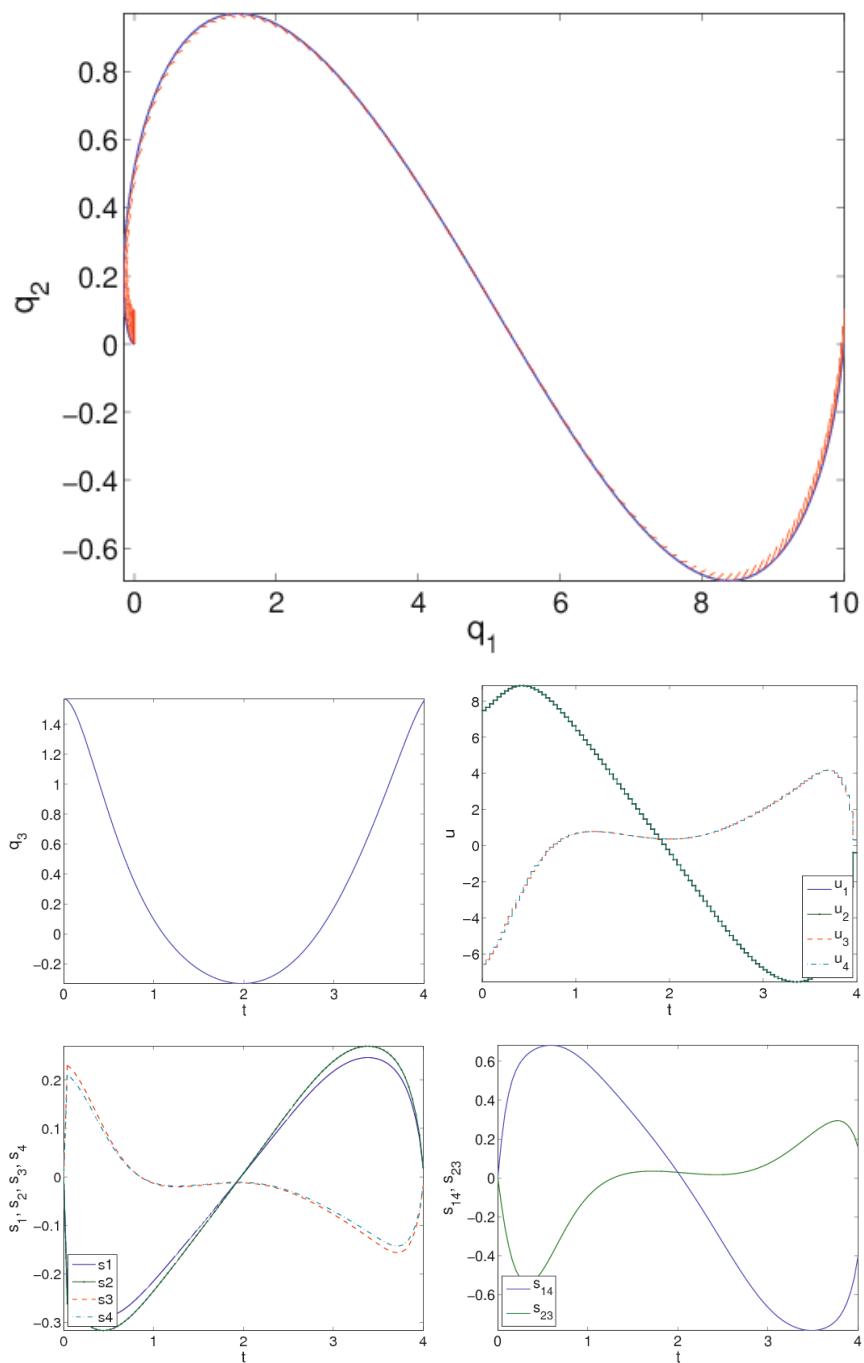


Fig. 4. OPCM: platform path, orientation, controls, longitudinal slips, lateral slips

5 Conclusion

This paper focuses solely on the motion planning whose aim is to determine a nominal trajectory of the platform. The tracking of this trajectory can be achieved by a dedicated control algorithm based on the predictive control scheme. It is hypothesized that frequent enough updating of the predictive control algorithm will be able to cope with the uncertainty of the traction conditions of the platform's wheels. Further results on this subject will be reported separately.

Acknowledgments. This work was supported by the Polish National Centre for Research and Development under the project RobREx.

References

1. Murphy, R., et al.: A search and rescue robotics. In: Siciliano, B., Khatib, O. (eds.) Springer Handbook of Robotics, pp. 1151–1173. Springer, Berlin (2008)
2. Liu, Y., Nejat, G.: Robotic Urban Search and Rescue: A Survey from the Control Perspective. *J. Intell. Robot. Syst.* 72, 147–165 (2013)
3. Mianowski, K.: Selection of servomotors for lightweight manipulator to assembly on mobile platforms. *Problems of Robotics*, pp. 347–356. Publishing House of the Warsaw University of Technology, Warsaw (2008) (in Polish)
4. Arent, K., K., et al.: General concept of the mathematical and physical model of the experimental mobile platform Rex. Institute of Computer Engineering, Control and Robotics, Wroclaw University of Technology, Technical report PRE 9 (2013) (in Polish)
5. Janiak, M., Tchoń, K., Zadarnowska, K.: Analysis of models and motion planning algorithms of the mobile platform Rex. Institute of Computer Engineering, Control and Robotics, Wroclaw University of Technology, Technical report SPR 4 (2013) (in Polish)
6. Tchoń, K., Jakubiak, J.: Endogenous configuration space approach to mobile manipulators: A derivation and performance assessment of Jacobian inverse kinematics algorithms. *Int. J. Control* 76, 1387–1419 (2003)
7. Diehl, M., et al.: Fast Direct Multiple Shooting Algorithms for Optimal Robot Control. In: Diehl, M., Mombaur, K. (eds.) *Fast Motions in Biomechanics and Robotics*, pp. 65–94. Springer, Berlin (2006)
8. Houska, B., Ferreau, H.J., Diehl, M.: ACADO Toolkit - An Open-Source Framework for Automatic Control and Dynamic Optimization. *Optimal Control Methods and Application* 32, 298–312 (2011)
9. Sussmann, H.J.: A continuation method for non-holonomic path finding problems. In: Proc. 32nd IEEE CDC, pp. 2718–2723. IEEE Press, New York (1993)
10. Zadarnowska, K., Tchoń, K.: A control theory framework for performance evaluation of mobile manipulators. *Robotica* 25, 703–715 (2007)
11. Janiak, M., Tchoń, K.: Constrained motion planning for nonholonomic systems. *Syst. Control Lett.* 60, 625–631 (2011)
12. Ratajczak, A., Tchoń, K.: Multiple-task motion planning of non-holonomic systems with dynamics. *Mechanical Sciences* 4, 153–166 (2013)
13. Ratajczak, A., Tchoń, K.: Parametric and non-parametric Jacobian motion planning for non-holonomic robotic systems. *J. Intell. Robot. Syst.* (2013), <http://dx.doi.org/10.1007/s10846-013-9880-04>

Trajectory Tracking Control of a Four-Wheeled Mobile Robot with Yaw Rate Linear Controller

Maciej Trojnacki¹, Przemysław Dąbek¹, Janusz Kacprzyk¹, and Zenon Hendzel²

¹ Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland

{mtrojnacki, pdabek, jkacprzyk}@piap.pl

² Rzeszów University of Technology, Rzeszów, Poland

zenhen@prz.edu.pl

Abstract. The paper concerns the problem of trajectory tracking control of a four-wheeled PIAP SCOUT mobile robot with non-steered wheels. For this kind of wheeled robots, it is impossible to find kinematic relationship between robot's body motion and motion of driven wheels, because of inherent sliding of wheels on the ground during turning. This is an important problem from the point of view of control of the robot. The approach followed in the present work relies on introducing a simple linear controller with feedback of actual yaw rate of robot's body. The yaw velocity is measured by inexpensive MEMS gyroscope. Experiments were conducted on two kinds of floor typical for office buildings: PVC flooring and carpet flooring. Measurements of motion parameters were possible with INS technique. It was found that the proposed yaw rate controller significantly reduces the angular error of path tracking for 90 degrees turn maneuver.

Keywords: wheeled mobile robot, tracking control, linear controller, motion investigations, wheels' slip, INS technique, MEMS gyroscope, empirical research.

1 Introduction

Tracking control, which is the kind of control where chosen point of a robot has to move on certain desired motion trajectory, belongs to important problems of control theory and robotics. In the present article this problem is considered on example of trajectory tracking control of a four-wheeled mobile robot with non-steered wheels.

Nowadays among mobile robots available on the market, the largest group consists of tracked and wheeled robots. Until recently, quite popular propulsion system intended for movement in various terrains was the tracked system [1]. However, observation of trends on the market of mobile robots renders conclusion that market share of wheeled robots is increasing [2–3].

This trend may be mainly the result of relative simplicity of design and control of mobile robots with this kind of propulsion system. Moreover, in many cases wheeled robots perform in open terrain at least as good as tracked robots, but at the same time they have less impact on the ground surface on which they drive.

From the point of view of synthesis of control systems, wheeled mobile robots can be divided in two groups: robots for which in typical operating conditions there is almost no wheel sliding on the ground during motion, and robots for which wheel sliding is an inherent feature of motion.

Robots with steered wheels belong to the first group. An example of such design is the mining inspection mobile robot described in [4]. A three-wheeled Pioneer 2-DX robot, shown in Fig. 1a, may be considered a special case, because although it does not have steered wheels, its kinematic structure allows movement without sliding of wheels in wide range of conditions.

The second group is comprised of robots with all wheels non-steered. Usually robots like that are intended for use in open terrain (outdoor robots). Examples of this kind of design are PIAP SCOUT and IBIS robots shown in Fig. 1b-c [5], both designed in Industrial Research Institute for Automation and Measurements PIAP.



Fig. 1. Examples of wheeled mobile robots: a – Pioneer 2-DX, b – PIAP SCOUT, c – IBIS

In case of wheeled robots from the second group, motion of robot's body is not defined by kinematic dependency on rotation of driven wheels, as it is in the case of robots from the first group. This difference is associated with occurrence of wheel sliding always during change of direction of motion. In this case forces generated between tire tread and ground depend in non-linear manner on various factors, e.g.: velocity of tire tread with respect to ground, material and shape of the tread and the ground, etc.

In typical robot's operating conditions, some of the factors connected with the ground may change in rather arbitrary fashion. For this reason, it is impossible to find analytical expression that would allow to associate desired motion of robot's body with motion parameters of driven wheels necessary to its realization, and would be valid in any conditions.

One approach to solve this problem is to introduce a controller, whose task is to calculate on-line the parameters of motion for driven wheels so as to realize the desired trajectory of motion of some characteristic point of a robot.

In work [6] for solution of a trajectory tracking problem, a fuzzy control, with rules based on typical behavior of a human driving a car, was implemented. Experimental research was completed using the Pioneer 3-AT robot. Necessary input quantities were measured by gyroscope and wheel encoders. The applied control algorithm may be convenient while solving the path planning problem, because it relies on a set of sparsely distributed points, close to which the robot must pass, imposed as input. Then continuous trajectory is generated by the control algorithm.

Thus, the article is concerned with higher level of a robot motion control system than the system analyzed in the present work.

A controller with adaptive motion trajectory tracking is presented in work [7]. The coefficient describing slope of a straight line that approximates dependency between tire-ground adhesion and longitudinal slip is the estimated unknown parameter. The control system does not require feedback. It relies on robot's dynamics model and certain adaptation rule to estimate the unknown parameter. Presented preliminary result of the experiment on a four-wheeled robot shows effectiveness of this approach. However, error of path tracking reaches nearly 0.5 m in case of circular motion with radius of 2 m. It is not clear if the lack of feedback will not lead to situation where in real conditions the tracking error grows with elapsed time.

In work [8] a four-wheeled robot was equipped with sliding-mode controller. Its algorithm relies on dynamics model of the robot. Linear velocity of the robot's body obtained from GPS receiver is used as feedback input to the controller. Experiments were carried out on soil and improvement in errors of trajectory tracking was reported as compared to experiment without the investigated controller. However, because of relatively low frequency of GPS data updates and relatively low accuracy of data in typical real operating conditions where differential corrections are unavailable, general effectiveness of this approach may be limited.

Aim of the present work is to investigate effectiveness of simple linear controller with feedback of angular velocity of a robot body about an axis normal to the ground (yaw rate) with objective of improvement of robot's trajectory tracking. The yaw rate is measured by inexpensive Microelectromechanical (MEMS) gyroscope. Advantages of this approach are low cost of implementation and well developed theory of linear PID controllers, which is important because of the prospect of industrial application of the solution.

Experimental research was conducted on two kinds of ground common in office buildings (PVC flooring and carpet flooring) for several desired linear velocities of robot motion. Motion of the robot in configurations with the proposed controller and without it was investigated. Results of experiments with the proposed controller were quantitatively compared to results without this controller using introduced quality indexes.

2 Four-Wheeled Mobile Robot and Experimental Setup

In the present work a four-wheeled PIAP SCOUT mobile robot designed and produced in Industrial Research Institute for Automation and Measurements PIAP, Poland [4] is analyzed. The robot has been adapted to realization of experimental research in such way that it was equipped with additional frame to allow installation of laptop and sensors (Fig. 2a).

The PIAP SCOUT robot has two rear wheels driven independently by DC servomotors with gear units and encoders. Drive is transmitted from rear to front wheels by means of toothed belts. Tires are made of viton rubber with foam filling (non-pneumatic tires). Block tread depth is about 5 mm.

Kinematic structure of the robot is shown in Fig. 2b. It is possible to distinguish the following main components of the robot: 0 – body with frame for installation of research equipment, 1-4 – wheels, 5-6 – toothed belts.

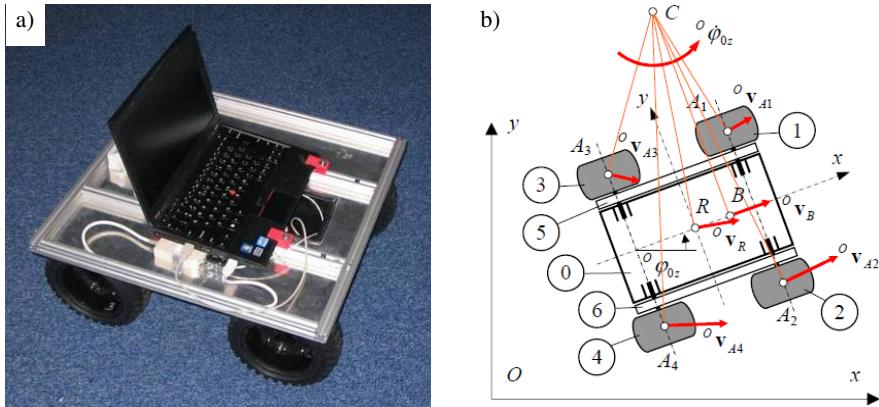


Fig. 2. PIAP SCOUT robot: a – robot adapted to experimental research, b – its kinematic structure ($A_1A_3 = A_2A_4 = L$, $A_1A_2 = A_3A_4 = W$)

The most important robot's kinematic parameters are ($i = \{1, \dots, 4\}$):

- dimensions (see Fig. 2b for explanation of symbols): $L = 0.35$ m, $W = 0.386$ m,
- masses of the components: $m_0 = 15.02$ kg, $m_i = 0.66$ kg, $m_5 = m_6 = 0.17$ kg,
- tire unloaded radius $r_i = 0.0965$ m.

On the frame was installed laptop computer connected by means of USB-CAN adapter to robot's CAN bus (Controller Area Network), which allowed to:

- impose desired velocities of spin of wheels $\dot{\theta}_d = [\dot{\theta}_{3d}, \dot{\theta}_{4d}]^T$,
- determine actual velocities of spin of wheels $\dot{\theta} = [\dot{\theta}_3, \dot{\theta}_4]^T$.

Sensors mounted to the frame include:

- iNEMO sensors' module connected to laptop via USB interface (Universal Serial Bus),
- Phidget acceleration sensor connected to laptop via USB.

Schematic diagram of the measurement and control system is shown in Fig. 3.

The iNEMO sensors' module enabled measurement of: linear acceleration ${}^R\mathbf{a}_R^O = [{}^R\mathbf{a}_{Rx}^O, {}^R\mathbf{a}_{Ry}^O, {}^R\mathbf{a}_{Rz}^O]^T$ of the point R (see Fig. 2b) and angular velocity ${}^R\dot{\phi}^O = [{}^R\dot{\phi}_x^O, {}^R\dot{\phi}_y^O, {}^R\dot{\phi}_z^O]^T$ of the robot's body 0, both in robot's coordinate system. The Phidget sensor mounted at the front part of the robot's frame allowed measurement of acceleration ${}^R\mathbf{a}_P^O = [{}^R\mathbf{a}_{Px}^O, {}^R\mathbf{a}_{Py}^O, {}^R\mathbf{a}_{Pz}^O]^T$. Its main role was to provide means of verification of signal from iNEMO's accelerometers.

Both iNEMO and Phidget devices are based on components manufactured in the Microelectromechanical Systems (MEMS) technology. Recent study [9] showed that MEMS inertial accelerometers already provide adequate accuracy of measurements of motion parameters, and that the inherent errors of accelerometers may not be the most significant as compared to other sources of errors in the inertial method of measurements of motion parameters. Some problems associated with this technique will be briefly mentioned in section 5, where discussion of empirical data is presented.

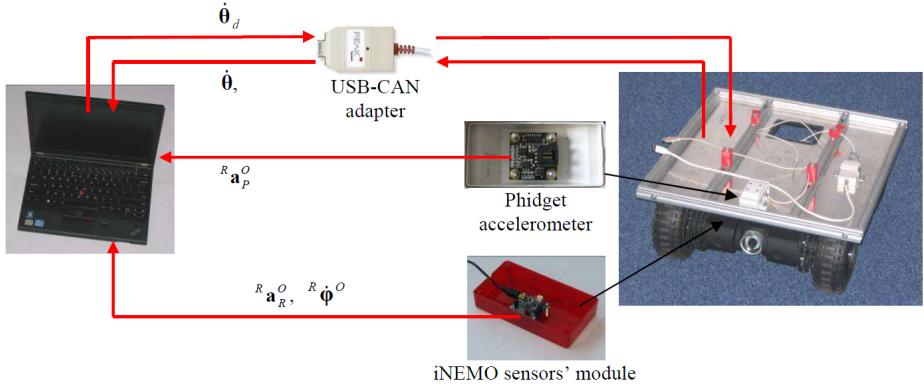


Fig. 3. Schematic diagram of measurement and control system used in experimental research

3 Motion Description and Desired Trajectory of the Robot

It is assumed that robot's motion is realized in Oxy plane of the fixed coordinate system $\{O\}$. The moving coordinate system, considered as rigidly connected to the robot, is denoted with symbol $\{R\}$. Position and orientation of mobile platform are described by the vector of generalized coordinates:

$${}^O\mathbf{q} = [{}^Ox_R, {}^Oy_R, {}^O\phi_{0z}]^T, \quad (1)$$

where: Ox_R , Oy_R are coordinates of point R belonging to mobile platform, and ${}^O\phi_{0z}$ denotes angle of spin of mobile platform about z axis with respect to fixed coordinate system $\{O\}$.

In turn, vectors of generalized velocities respectively in $\{O\}$ and $\{R\}$ coordinate systems can be written as:

$${}^O\dot{\mathbf{q}} = [{}^O\dot{x}_R, {}^O\dot{y}_R, {}^O\dot{\phi}_{0z}]^T, \quad {}^R\dot{\mathbf{q}} = [{}^R\dot{x}_R^O, {}^R\dot{y}_R^O, {}^R\dot{\phi}_{0z}^O]^T, \quad (2)$$

where: ${}^Ov_{Rx} = {}^O\dot{x}_R$, ${}^Ov_{Ry} = {}^O\dot{y}_R$, ${}^Rv_{Rx}^O = {}^R\dot{x}_R^O$, ${}^Rv_{Ry}^O = {}^R\dot{y}_R^O$.

If one makes assumption that ${}^Rv_{Ry}^O = 0$ then vector of generalized velocities ${}^O\dot{\mathbf{q}}$ can be defined on the basis of kinematic equations of motion in the form:

$${}^o\dot{\mathbf{q}} = \begin{bmatrix} {}^o\dot{x}_R \\ {}^o\dot{y}_R \\ {}^o\dot{\phi}_{0z} \end{bmatrix} = \begin{bmatrix} \cos({}^o\phi_{0z}) & 0 \\ \sin({}^o\phi_{0z}) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} {}^Rv_{Rx}^O \\ {}^R\dot{\phi}_{0z}^O \end{bmatrix}, \quad (3)$$

where vector ${}^R\mathbf{v} = [{}^Rv_{Rx}^O, {}^R\dot{\phi}_{0z}^O]^T$ contains respectively component of velocity of the point R of the robot on the x direction of $\{R\}$ coordinate system and yaw rate of mobile platform.

The above equation is valid in case when robot moves on horizontal ground.

It should be noted that in case of robot turning or spinning about vertical axis, because robot's wheels are non-steered, its motion largely depends on longitudinal and lateral forces acting within contact of wheels with ground, that is, on system dynamics. For this reason, robot spins about vertical axis with smaller angular velocity ${}^R\dot{\phi}_{0z}^O$ than it follows from angular velocities of spinning of driven wheels, which can be determined from kinematics model of the robot. In view of that, in order to solve inverse kinematics problem for the robot, one may use the following relationship as the first approximation:

$$\begin{bmatrix} \dot{\theta}_3 \\ \dot{\theta}_4 \end{bmatrix} = \frac{1}{r} \begin{bmatrix} 1 & -W/2 \\ 1 & W/2 \end{bmatrix} \begin{bmatrix} {}^Rv_{Rx}^O \\ k_{\omega 0z} {}^R\dot{\phi}_{0z}^O \end{bmatrix}, \quad (4)$$

in which coefficient $k_{\omega 0z} \geq 1$ depends in particular on: the layout of robot propulsion system (especially, number of driven wheels with respect to total number of wheels), properties of applied tires and ground on which the robot moves.

Because of initial lack of knowledge about interaction conditions of wheels with the ground, initially coefficient $k_{\omega 0z} = 1$ is assumed, which corresponds to the case when robot moves without sliding. As a result of conducted research it will become possible to select optimal value of this coefficient with respect to particular conditions.

In this work the idea of additional controller responsible for realization of robot motion with velocities ${}^Rv_{Rx}^O$ and ${}^R\dot{\phi}_{0z}^O$ will be presented. The maneuver of 90° turning to the right will be analyzed. In this case characteristic point R of the robot should move forwards along straight line for the distance of 1 m, then along circular arc of radius of 0.15 m and finally again along straight line for the distance of 1 m.

Investigations of robot motion will be conducted with the assumption that desired velocities of robot motion, defined as a generalized velocities vector ${}^R\mathbf{v}_d = [{}^Rv_{Rxd}^O, {}^R\dot{\phi}_{0zd}^O]^T$, change in a step manner, which will allow to determine maximum accelerations reachable by the robot in the given tire-ground interaction conditions.

Experimental investigations will be conducted in several variants characterized by maximum velocities of robot motion $v_{Ru} = \max({}^Rv_{Rxd}^O)$ and $\omega_{0zu} = \max({}^R\dot{\phi}_{0zd}^O)$ and types of ground, as listed in Table 1.

Table 1. Variants of desired motion parameters and types of ground for turning maneuver (used symbols: i – internal controller, e – external controller, c – carpet flooring, p – PVC flooring)

Variant symbol		v_{Ru} (m/s)	ω_{0zu} (rad/s)	Ground type
internal controller	external controller			
ic0.3	ec0.3	0.3	-2.00	carpet flooring
ic0.5	ec0.5	0.5	-3.33	carpet flooring
ic0.7	ec0.7	0.7	-4.67	carpet flooring
ic1.0	ec1.0	1.0	-6.67	carpet flooring
ip0.3	ep0.3	0.3	-2.00	PVC flooring
ip0.5	ep0.5	0.5	-3.33	PVC flooring
ip0.7	ep0.7	0.7	-4.67	PVC flooring
ip1.0	ep1.0	1.0	-6.67	PVC flooring

In the variant symbol included in the table there is coded information about the type of control system, kind of ground and value of desired linear velocity used in particular experiment. It should be noted that proposed ground materials, i.e., carpet flooring and PVC flooring, are seldom analyzed in investigations of mobile robot motion, but they should be of interest as they represent typical kinds of floor used in office buildings.

4 Proposed Yaw Rate Controller and Quality Indexes

In the present work investigations of two versions of robot control system will be conducted. The first version will contain only controller of wheels' velocity, which is the internal robot controller and will not be an object of investigations by itself, thus being treated as a “black box”. Input quantities for this controller will be desired angular velocities of driven wheels defined as a vector $\dot{\theta}_d = [\dot{\theta}_{3d}, \dot{\theta}_{4d}]^T$ and output quantities, the control signals for DC motors which drive the wheels described by a vector $\mathbf{u} = [u_3, u_4]$.

Schematic diagram of robot control system with only internal controller of wheels' velocity is shown in Fig. 4.

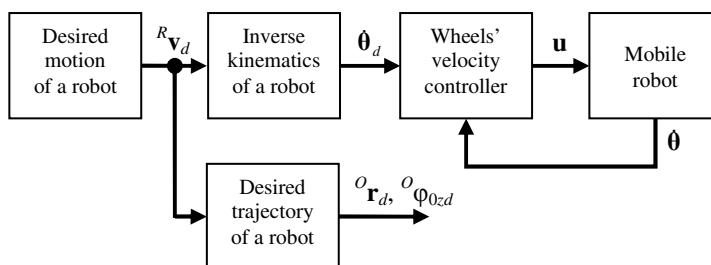


Fig. 4. Schematic diagram of robot motion control system with only internal controller of wheels' velocity

The following blocks are distinguished in the diagram:

- Desired motion of a robot – responsible for generation of a vector of desired velocities of robot's mobile platform ${}^R\mathbf{v}_d = [{}^Rv_{Rd}^O, {}^R\dot{\phi}_{0zd}^O]^T$, that is, desired time histories of velocity of characteristic point R of robot ${}^Rv_{Rd}^O = {}^Rv_{Rd}^O$ and yaw rate of its body ${}^R\dot{\phi}_{0zd}^O$ depending on imposed maximum velocities v_{Ru} and ω_{0zu} (according to Table 1).
- Inverse kinematics of a robot – determines a vector of desired angular velocities of spin for driven wheels $\dot{\theta}_d = [\dot{\theta}_{3d}, \dot{\theta}_{4d}]^T$ on the basis of a vector of desired velocities of robot's mobile platform ${}^R\mathbf{v}_d = [{}^Rv_{Rd}^O, {}^R\dot{\phi}_{0zd}^O]^T$, from equation (4) (with $k_{\omega_0z} = 1$), in which in place of actual values of velocity are substituted desired values.
- Desired trajectory of a robot – calculates the vector of desired generalized velocities ${}^o\dot{\mathbf{q}}_d = [{}^o\dot{x}_{Rd}, {}^o\dot{y}_{Rd}, {}^o\dot{\phi}_{0zd}]^T$ based on a vector of desired velocities of robot's mobile platform ${}^R\mathbf{v}_d = [{}^Rv_{Rd}^O, {}^R\dot{\phi}_{0zd}^O]^T$, from equation (3), and then determines by integration, the vector of desired generalized coordinates ${}^o\mathbf{q}_d = [{}^ox_{Rd}, {}^oy_{Rd}, {}^o\phi_{0zd}]^T$, that is desired path of motion of point R , i.e., ${}^o\mathbf{r}_d = [{}^ox_{Rd}, {}^oy_{Rd}]^T$ and desired yaw angle of the robot ${}^o\phi_{0zd}$.
- Wheels' velocity controller – on the basis of desired and actual velocities of wheels, that is, respectively vectors $\dot{\theta}_d = [\dot{\theta}_{3d}, \dot{\theta}_{4d}]^T$ and $\dot{\theta} = [\dot{\theta}_3, \dot{\theta}_4]^T$, it determines control signals for robot motors as a vector $\mathbf{u} = [u_3, u_4]$.
- Mobile robot – the investigated four-wheeled PIAP SCOUT mobile robot with measurement and control equipment described in section 2.

In the second version of the control system, the velocity controller of mobile platform will be used, which is a higher level controller with respect to wheels' velocity controller. Both controllers will cooperate in this version of the control system, which is shown in Fig. 5.

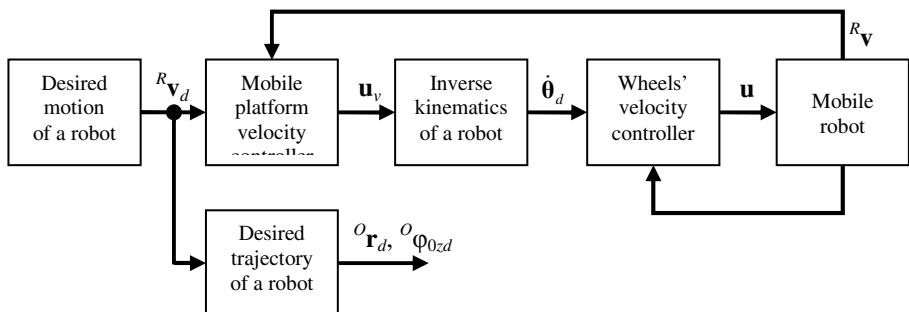


Fig. 5. Schematic diagram of robot motion control system with additional mobile platform velocity controller

On the basis of desired and actual robot velocities, that is, respectively ${}^R\mathbf{v}_d = [{}^Rv_{Rd}^O, {}^R\dot{\phi}_{0zd}^O]^T$ and ${}^R\mathbf{v} = [{}^Rv_{Rx}^O, {}^R\dot{\phi}_{0z}^O]^T$, error of generalized velocities is determined as:

$${}^R\mathbf{v}_e = \begin{bmatrix} {}^R\mathbf{e}_{vR} \\ {}^R\mathbf{e}_{\omega 0z} \end{bmatrix} = {}^R\mathbf{v}_d - {}^R\mathbf{v} = \begin{bmatrix} {}^Rv_{Rd}^O - {}^Rv_{Rx}^O \\ {}^R\dot{\phi}_{0zd}^O - {}^R\dot{\phi}_{0z}^O \end{bmatrix}. \quad (5)$$

In the work simplified version of this controller will be analyzed, in which it is assumed that because of the presence of non-steered wheels the error of angular velocity ${}^R\mathbf{e}_{\omega 0z}$ will be relatively greater than error of linear velocity ${}^R\mathbf{e}_{vR}$. For this reason, it is assumed that ${}^Rv_{Rd}^O \approx {}^Rv_{Rd}^O$ and ${}^R\mathbf{e}_{vR} \approx 0$, therefore the control law of controller of mobile platform velocity has the form:

$$\mathbf{u}_v = \begin{bmatrix} u_{vR} \\ u_{\omega z} \end{bmatrix} = \begin{bmatrix} {}^Rv_{Rd}^O \\ k_{\omega 0z} s_{\omega 0z} \end{bmatrix}, \quad (6)$$

where: $s_{\omega 0z} = {}^R\mathbf{e}_{\omega 0z} + \lambda_{\omega 0z} {}^R\mathbf{e}_{\phi 0z}$, ${}^R\mathbf{e}_{\omega 0z} = ({}^R\dot{\phi}_{0zd}^O - {}^R\dot{\phi}_{0z}^O)$, ${}^R\mathbf{e}_{\phi 0z} = \int {}^R\mathbf{e}_{\omega 0z} dt$.

The adopted simplification follows also from the fact that for calculation of linear velocity error one should determine actual linear velocity of the point R , that is ${}^Rv_R^O$, which in case of use of inertial method of measurement requires integration of accelerations. Velocity determined in this way has the worse quality the more uneven is the ground on which robot moves, which inhibits application of full version of the controller at present. However, works necessary to implement this fully-featured controller are planned in near future.

Optimal settings of the mobile platform velocity controller, i.e., factors $\lambda_{\omega 0z} = 13$ and $k_{\omega 0z} = 1$ were selected after realization of series of separate experiments where robot was spinning in place about its vertical axis z . During mentioned experiments various combinations of those factors were analyzed for values of factors $\lambda_{\omega 0z}$ and $k_{\omega 0z}$ that changed by fixed differences within certain assumed ranges.

Having control signal \mathbf{u}_v of mobile platform velocity controller, desired velocities of spin of driven wheels can be determined using the relationship which results from robot's kinematics:

$$\dot{\theta}_d = \begin{bmatrix} \dot{\theta}_{3d} \\ \dot{\theta}_{4d} \end{bmatrix} = \frac{1}{r} \begin{bmatrix} 1 & -W/2 \\ 1 & W/2 \end{bmatrix} \mathbf{u}_v, \quad (7)$$

which is analogous to relationship (4).

To assess performance of both versions of the robot control system the following errors of realization of motion by the robot were defined:

- errors of x and y coordinates of the characteristic point R of the robot (Fig. 2b) and error of robot's position, respectively:

$${}^O\mathbf{e}_{Rx} = {}^Ox_{Rd} - {}^Ox_R, \quad {}^O\mathbf{e}_{Ry} = {}^Oy_{Rd} - {}^Oy_R, \quad {}^O\mathbf{e}_R = \| {}^O\mathbf{e}_R \| = \sqrt{{}^Oe_{Rx}^2 + {}^Oe_{Ry}^2}, \quad {}^O\mathbf{e}_R = [{}^O\mathbf{e}_{Rx}, {}^O\mathbf{e}_{Ry}]^T, \quad (8)$$

- error of yaw angle (robot's heading):

$${}^o e_\psi = {}^o \Psi_d - {}^o \Psi, \quad {}^o \Psi_d = {}^R \varphi_{0zd}^o, \quad {}^o \Psi = {}^R \varphi_{0z}^o, \quad (9)$$

- errors of rotation angles for driven wheels:

$$e_{\theta i} = \theta_{id} - \theta_i, \quad i = \{3, 4\}. \quad (10)$$

In order to compare in a comprehensive way accuracy of realization of motion by the robot for different experiment results the following quality indices were introduced:

- maximum errors of: robot's position, heading and rotation angles for driven wheels:

$${}^o e_{Rmax} = \max_{t \in \langle 0, T \rangle} (\| {}^o \mathbf{e}_R \|), \quad {}^o e_{\psi max} = \max_{t \in \langle 0, T \rangle} (| {}^o e_\psi |), \quad e_{\theta imax} = \max_{t \in \langle 0, T \rangle} (| e_{\theta i} |), \quad (11)$$

- integrals of squared errors of: robot's position, heading and rotation angles for driven wheels:

$$E_R = \int_0^T \| {}^o \mathbf{e}_R \|^2 dt, \quad E_\psi = \int_0^T {}^o e_\psi^2 dt, \quad E_{\theta i} = \int_0^T e_{\theta i}^2 dt, \quad (12)$$

where: $\| {}^o \mathbf{e}_R \| = \sqrt{({}^o e_{Rx})^2 + ({}^o e_{Ry})^2}$, $i = \{3, 4\}$ and T is the analyzed time period of robot's motion.

5 Experimental Investigations of Controller Performance

At first, experimental investigations of robot motion for only the internal controller of wheels' velocity were conducted (Fig. 4). Quality indices obtained for results of experiment variants from Table 1 are shown in Table 2.

Table 2. Quality indices obtained for results of experimental investigations with use of internal controller of wheels' velocity for variants of turning maneuver

Variant symbol	${}^o e_{Rmax}$ (m)	${}^o e_{\psi max}$ (rad)	${}^o e_{\theta 3max}$ (rad)	${}^o e_{\theta 4max}$ (rad)	E_R ($m^2 \cdot s$)	E_ψ ($rad^2 \cdot s$)	$E_{\theta 3}$ ($rad^2 \cdot s$)	$E_{\theta 4}$ ($rad^2 \cdot s$)
ic0.3	1.3016	0.8443	1.2833	0.5576	3.6490	2.6787	5.2398	0.8700
ic0.5	1.1242	0.9106	1.6969	0.6840	1.7471	2.0494	4.4585	0.7387
ic0.7	1.1891	0.9426	2.1490	0.7331	1.8673	1.8368	5.4437	0.7364
ic1.0	1.1265	1.0787	3.2256	0.9357	1.3892	1.9306	12.0443	0.8742
ip0.3	0.8975	0.8513	1.1557	0.7660	1.4143	2.7932	4.7434	1.7486
ip0.5	0.9444	0.9022	1.4366	0.6606	1.0848	1.9890	3.5478	0.8867
ip0.7	0.9776	0.9887	2.0050	0.7542	1.2450	2.0645	4.1413	0.9142
ip1.0	1.0697	1.1193	2.8211	0.9362	1.2272	2.0532	7.9052	0.9334

Next, experiments with enabled external controller (of velocity of mobile platform, Fig. 5) were carried out. The obtained quality indices for experiment variants from Table 1 are summarized in Table 3.

Table 3. Quality indices obtained for results of experimental investigations with use of internal controller and additional external velocity controller of mobile platform for variants of turning maneuver

Variant symbol	$\sigma_{e_{Rmax}}$ (m)	$\sigma_{e_{\psi max}}$ (rad)	$\sigma_{e_{\theta 3max}}$ (rad)	$\sigma_{e_{\theta 4max}}$ (rad)	E_R ($m^2 \cdot s$)	E_ψ ($rad^2 \cdot s$)	$E_{\theta 3}$ ($rad^2 \cdot s$)	$E_{\theta 4}$ ($rad^2 \cdot s$)
ec0.3	0.3453	0.2820	1.6070	0.5591	0.2290	0.0519	6.1821	0.8297
ec0.5	0.5178	0.5525	4.3726	0.9736	0.3647	0.1831	35.0022	0.9138
ec0.7	0.6852	0.8549	4.4482	1.6890	0.6076	0.5100	30.9664	1.9185
ec1.0	0.7371	1.2044	3.6183	2.6951	0.6773	1.2093	15.2555	4.2824
ep0.3	0.3086	0.2810	1.5520	0.5689	0.4212	0.0498	6.1699	1.0998
ep0.5	0.3319	0.5657	3.5009	0.8116	0.2903	0.1459	20.3387	0.7985
ep0.7	0.3303	0.9435	4.0643	1.7916	0.2163	0.4263	22.3426	1.3600
ep1.0	0.5884	1.2618	3.2032	2.1096	0.4998	1.1315	10.6767	1.9860

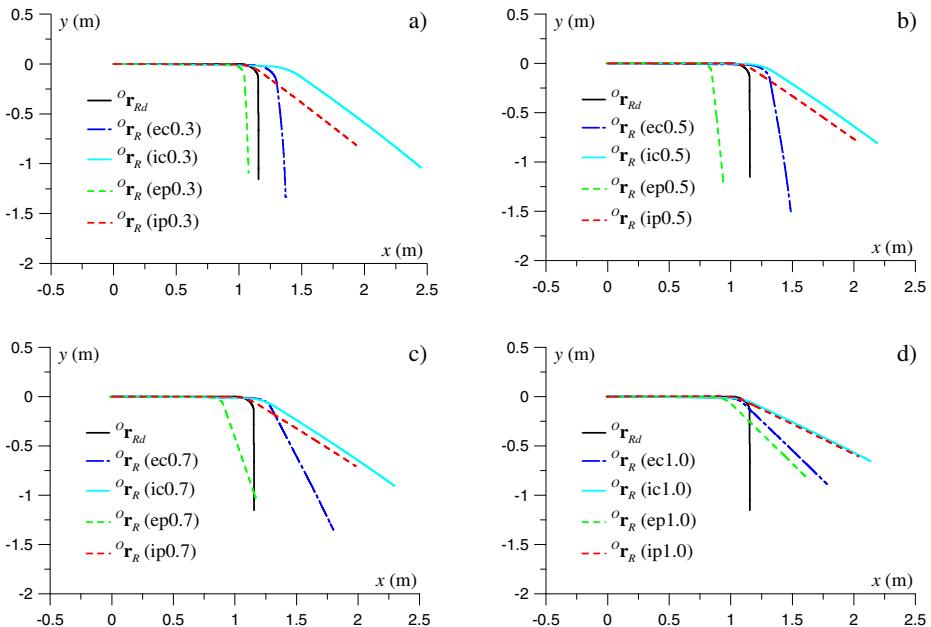


Fig. 6. Motion paths of characteristic point R of the robot for all experiment variants

Motion paths of characteristic point R of the robot for all analyzed variants of experiment are shown in Fig. 6.

The paths are grouped in subfigures with respect to values of desired velocity of the point R of the robot. Graphs show desired motion path denoted as ${}^o\mathbf{r}_{Rd}$ and actual paths for particular variants from Table 1 denoted as ${}^o\mathbf{r}_R$ with variant symbol from the table given in parentheses. The variant symbol contains encoded information about the version of control system, kind of ground and value of desired linear velocity used in experiment.

Fig. 7 shows time histories of desired and actual angular velocities of spinning for driven wheels, respectively $\dot{\theta} = [\dot{\theta}_3, \dot{\theta}_4]^T$ and $\dot{\theta}_d = [\dot{\theta}_{3d}, \dot{\theta}_{4d}]^T$ (marked in figure with red dashed lines), for selected experiment variants on PVC flooring. Presented results concern the lowest and the highest investigated linear velocities, that is, 0.3 m/s and 1.0 m/s respectively.

Fig. 6 clearly shows that the proposed external controller is effective in correcting the shape of actual trajectory. The error of angle tracking is growing with increasing desired linear velocity of robot motion. Main cause of this effect can be connected with limitations of power of driving motors which realize desired velocities of wheels with small error in case of motion with low linear velocities, but with large error when linear velocities are high (Fig. 7b and 7d). This effect should be further investigated in separate research.

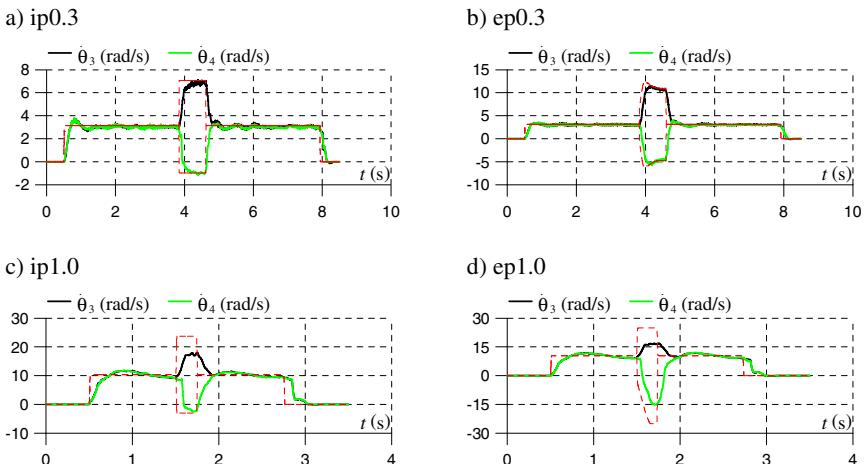


Fig. 7. Time histories of desired and actual angular velocities of spinning of driven wheels (values of desired velocities are marked in figure with red dashed lines)

Time histories of accelerations and velocities of characteristic point R of the robot are presented in Fig. 8. In the figure are shown only longitudinal components of the measured velocity and acceleration, in moving coordinate system $\{R\}$, which are the most important, because desired velocity was defined along longitudinal direction by the assumption (${}^R v_{Rd}^O = {}^R v_{Rd}^o$).

In Fig. 6 one may notice that in some experiments turning takes place earlier while in others later than it is marked by desired path shown by black solid line. The distance traveled by the robot is directly connected with its actual linear velocity shown in Fig. 8 and its results from longitudinal slip of the wheels.

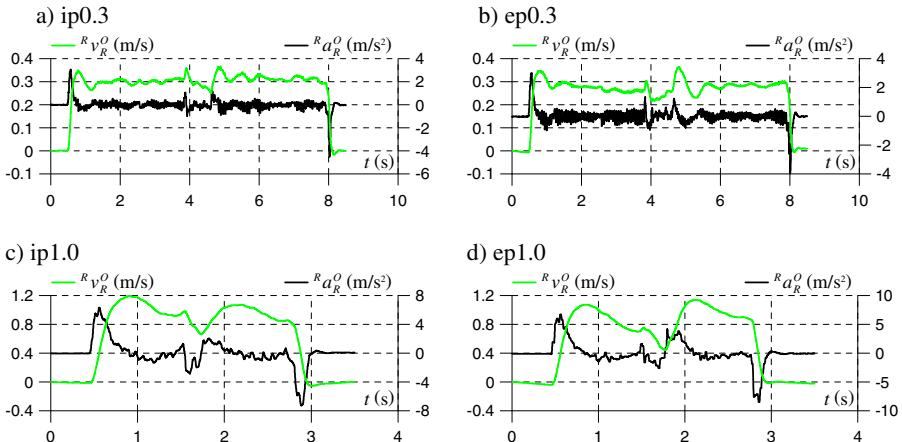


Fig. 8. Time histories of longitudinal velocities and accelerations of the characteristic point R of the robot, expressed in moving coordinate system $\{R\}$

Velocities ${}^R v_R^O$ are calculated on the basis of accelerations measured by on-board sensors. The accelerations are influenced by two main groups of factors. One group is associated with effects of realization of motion, like wheels' slip, which is the most relevant in the present case. However, the other group of factors includes all phenomena associated with inertial measurement technique, one of the most important being the effect of acceleration of gravity on horizontal components of measured acceleration.

Certain correction of this systematic effect was applied to the presented results, but the used method of correction might not be the most accurate. Improvement of accuracy of the method of inertial measurements might be one of the aims of future studies.

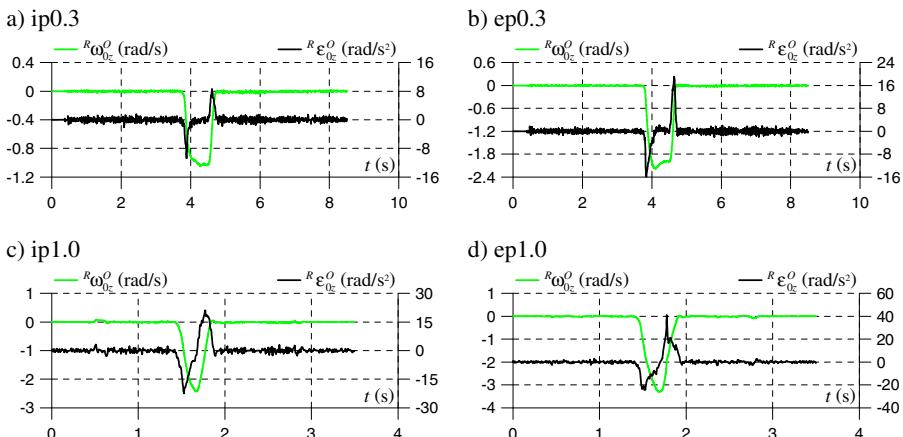


Fig. 9. Time histories of angular velocity and acceleration of yawing of robot's body for experiments on PVC flooring

In turn, in Fig. 9, the time histories of angular parameters of yawing of robot's body are shown whereas in Fig. 10 errors of robot's position and heading are presented for the analyzed experiment variants on PVC flooring.

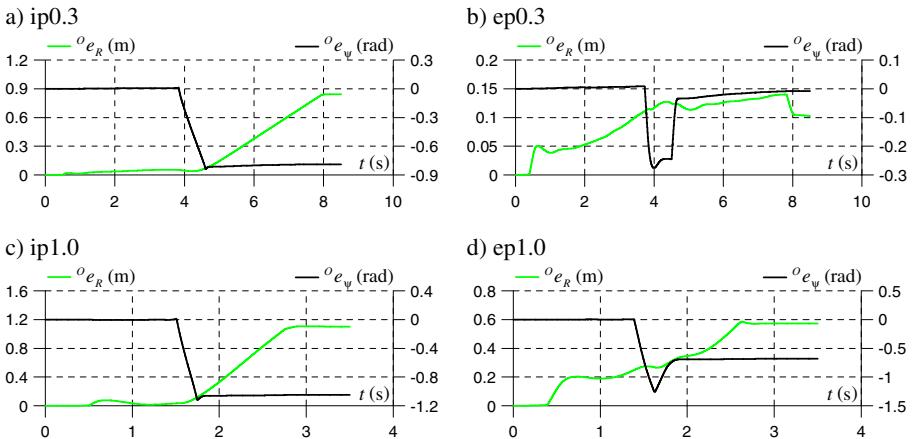


Fig. 10. Time histories of errors of robot's position and heading for experiments on PVC flooring

6 Conclusions and Future Works

Results of investigations of trajectory tracking control of four-wheeled PIAP SCOUT mobile robot in two configurations of control system were presented. In the first configuration robot was equipped only with internal wheels' velocity controller provided by the manufacturer. In the second configuration, to the internal controller was added new controller of yaw velocity of the mobile platform.

Most important conclusions of the work are summarized below:

- It is confirmed that trajectory tracking with only internal controller enabled and inputs based on kinematic relations is very inaccurate for all investigated experiment variants.
- It is found that introduction of the proposed simple linear controller (belonging to the class of PID controllers) of yaw velocity of the mobile platform significantly improves accuracy of motion realization.
- The improvement is most pronounced for lower linear velocities of motion and is degrading with increase of velocity. This effect may be attributed mainly to insufficient power of driving motors, which were not able to realize desired velocities of driving wheels for higher torques, but this issue requires further investigation.

Directions of future works will include:

- Implementation of more advanced structures of non-linear controllers, e.g. based on dynamics models of the robot, to compare with simple linear controller presented in this paper.

- Use of more data from IMU (Inertial Measurement Unit), for instance, linear accelerations to calculate linear velocities, which may be used as feedback to the controller.

Acknowledgements. The work has been realized as a part of the project entitled “Dynamics modeling of four-wheeled mobile robot and tracking control of its motion with limitation of wheels slip”. The project is financed from the means of National Science Centre of Poland granted on the basis of decision number DEC-2011/03/B/ST7/02532.

References

1. Giergiel, J., Kurc, K.: Mechatronics of the inspective robot. *Mechanics and Mechanical Engineering* 10(1), 56–73 (2006)
2. Trojnacki, M., Szymkarczyk, P., Andrzejuk, A.: Tendencies in the development of ground mobile robots (1). Review of mobile robots for special applications, *Pomiary Automatyka Robotyka* 2008(6), 11–14 (2008) (in Polish: Tendencje rozwoju mobilnych robotów lądowych (1). *Przegląd robotów mobilnych do zastosowań specjalnych*)
3. Wołoszczuk, A., Andrzejczak, M., Szymkarczyk, P.: Architecture of mobile robotics platform planned for intelligent robotic porter system–IRPS project. *Journal of Automation, Mobile Robotics & Intelligent Systems* 1(3), 59–63 (2007)
4. Kasprzyczak, L., Trenczek, S., Cader, M.: Robot for monitoring hazardous environments as a mechatronic product. *Journal of Automation, Mobile Robotics & Intelligent Systems* 6(4), 57–64 (2012)
5. Mobile robots for counter-terrorism (PIAP), <http://www.antiterrorism.eu>
6. Maalouf, E., Saad, M., Saliah, H.: A higher level path tracking controller for a four-wheel differentially steered mobile robot. *Robotics and Autonomous Systems* 54(1), 23–33 (2006)
7. Yi, J., Song, D., Zhang, J., Goodwin, Z.: Adaptive trajectory tracking control of skid-steered mobile robots. In: IEEE International Conference on Robotics and Automation, pp. 2605–2610 (2007)
8. Lucet, E., Grand, C., Bidaud, P.: Sliding-mode velocity and yaw control of a 4WD skid-steering mobile robot. In: Angeles, J., Boulet, B., Clark, J.J., Kövecses, J., Siddiqi, K. (eds.) *Brain, Body and Machine. AISc*, vol. 83, pp. 247–258. Springer, Heidelberg (2010)
9. Dąbek, P.: Evaluation of low-cost accelerometers for measurements of velocity of unmanned vehicles. *Pomiary Automatyka Robotyka* 2013(1), 102–113 (2013)

Universal Control System for Managing Multiple Unmanned Engineering Machines

Rafał Typiak

Military University of Technology, Warsaw, Poland
rtypiak@wat.edu.pl

Abstract. The following paper presents early stages of execution of a computer based system designed to allow for a single human operator to manage and control multiple unmanned engineering machines. It utilizes a concept of a universal control station toggled with a management server which tells it what should be the desired communication structure for each of the controlled machines. The paper highlights the requirements for both the universal control station in terms of structure as well as the management server in terms of functionality. Afterwards it describes the assumptions made during the initial stages of implementation. It showcases the machines selected for the integration and the current state of play in terms of control system's modules execution. The paper's summary describes plans for further development and desired functionality.

Keywords: control systems, unmanned platform sensory, ergonomics, multi machine control algorithms.

1 Introduction

Currently conducted peace keeping missions around the world require military forces to be in close proximity of civilians and civilian structures. This presents an ideal opportunity for any type of enemy force utilizing asymmetrical tactics, to bombard fortified positions from within urban populated areas. This poses a threat to soldiers conducting engineering missions both within their own bases as well as outside them. That is why, there is a visible trend of introducing unmanned machines for carrying out high threat level actions instead of sending in soldiers. However currently, there is a limited number of possible uses for these devices as they're dedicated solutions and real world engineering scenarios rarely require only one type of action to be performed. One possible solution would be to create a robust unmanned engineering machine which would be capable of completing engineering missions on its own; however, since these are high threat level actions, the possibility of losing such a machine is high. Another solution would be to utilize multiple, less expensive machines, which would complement each other's functionalities. This approach would decrease the possibility of losing all units in case of any unpredicted actions. This solution does have a drawback in a sense that currently each machine comes with its own control station (Fig. 1). As such, when trying to use two or more units, one needs two or more

operators. Not only that, but it's also crucial to allow them to communicate with each other. While during missions taking place inside military installations this does not seem to be a big issue, on-site missions in remote locations, where it's crucial to keep presence at minimum to provide adequate security, high number of personnel is not the right solution.



Fig. 1. Different control panels for unmanned engineering equipment: a) All All-purpose Remote Transport System's control panel (ARTS) (Source: <http://www.ara.com>), b) Bobcat's Robotic Controller Kit's control panel (Source: <https://www.qinetiq-na.com>)

This drawback has been a driving factor in creating a concept of a single, flexible control station which would allow a single person to take control on a number of unmanned engineering machines and use them in rotation to complete a complex engineering task. Because each machine's functionality can be divided into a mechanical and sensory, simple switching between machines proves inadequate, as taking control over a new machine prevents from using the last machine's sensory functionality. That is why the control station should be equipped with an onboard control system which would allow for control transfer while still retaining access to sensors from each machine used.

Using multiple machines poses also a problem of remote control and interference between devices using similar wireless solutions. In standard conditions this could possibly be a major problem, but with a single control station solution, this becomes a software issue instead of a hardware one.

There are a number of issues needed to be addressed in order to create a fully functional multi machine control station. Most importantly it needs to have a clearly defined functionality as this is the base for creating any working control stand. A target consumer needs to be defined in order to assess the work environment, possible powering solutions and mobility level. Next, there needs to be a clear logic defined which would concern the controlled machines and how they influence each other and the control station. With the knowledge gained from previous steps, one can extrapolate certain requirements for communication and control hardware and programming. Such aspects as available power supply limitations and presumed mobility allows for narrowing down possible hardware solutions and discarding those which do not meet presented requirements. The same relation takes place with machine control algorithms. Knowledge about how the operator has to be able to interact with selected

machines allows determining the number of hardware elements which need to be in use during normal working conditions. Obviously the idea of using multiple unmanned machines for completing a single, complex engineering task puts a huge focus on enabling the operator to use multiple sensors scattered across different hardware platforms (machines). This process needs to be done both efficiently and user friendly. Burying the operator under vast amount of data and letting him sort it out is not time-efficient, a criteria which is currently being introduced more openly on the market, however, this mostly involves control schemes, not data visualization. One such solution is using a game controlled to send steering commands to an unmanned machine (Fig. 2). This flattens the learning curve for new operators, which (as is assumed) mostly use these devices in their home entertainment systems.



Fig. 2. Complete Bobcat Robotic Controller Kit system with a gamepad controller (Source: <https://www.qinetiq-na.com>)

2 Universal Control Station's Functionality Requirements

Soldiers carrying out engineering tasks have access to logistical support which, most of the time, boils down to access to technical support vehicles. These vehicles then supply access to a wide range of powering solutions, from 12 V through 24 V DC current to 230 V AC. However, there needs to be a factor of mobility taken into account. It is assumed that the described control station should have the possibility to be mounted inside most (if not any) personal carriers available. Thus, the available voltage ranges are limited to 12 and 24 V DC installations. Should a single voltage value needed to be chosen, it is preferable to select the 12 V DC as it is also commonly found in civilian vehicles, which than could also serve as means of transport for the control station. The problem of powering every component from a single 12 V installation should not prove problematic, as a lot of currently available communication hardware is built in a way which allows powering from 9-36 V DC installations. The same goes for industrial and military grade visualization units and commonly available units are effectively powered from 12 V power adapters. Another argument for using the 12 V power source could be that some diesel-power units are equipped with both 230 V AC and 12 V DC power outputs.

As mentioned, a lot of effort should be put into creating the control station as mobile as possible. This would require creating a compact, but sturdy frame, preferably one that has an openwork structure to decrease mass. The idea is that the standalone version of the control station should comprise of three detachable sections: the operator's seat, the operating panel and the visualization panel. With these three parts functioning together but having the ability to be separated, it allows for using the same station within different work environments (Fig. 3). If space is not an issue, the whole set may be used. If the space is limited, the visualization part, or the operator's seat may be removed to allow for an in-vehicle placement. If, however, the space is highly limited, or in case of emergency, the operating panel should suffice to allow for basic control over every machine.

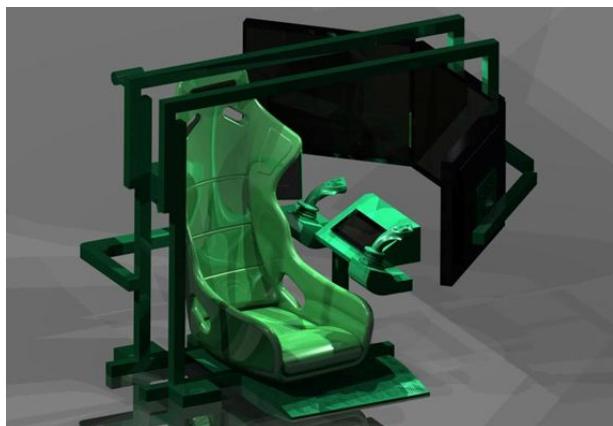


Fig. 3. A concept design of the universal control station with three sections: the operator's seat and controls, the operating panel and the visualization panel

3 Multiple Machine Control Algorithms

The problem of controlling multiple machines is being tackled by many research institutes; however most of their work concerns Unmanned Aerial Vehicles (UAVs) [1, 2]. The Military University of Technology has experience in strategies of use of remote control unmanned ground vehicles [3]. The presented solution is a set of rules and logic which would be implemented in order to allow a single operator to control one of several selected unmanned machines, with other units being on standby and supporting the task currently underway. A more detailed description of the logic itself may be found in [4].

The idea is that an operator-specialist, tasked with carrying out an engineering mission is given the mission parameters and selects, from any currently available units, the ones which he decides are going to be useful. The then registers them to his control station, effectively binding them to his control until released (Fig. 4). The control systems of these machines communicate with the control station, which not only tells

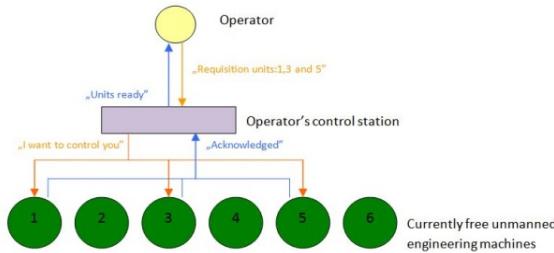


Fig. 4. A diagram describing the way for the operator to register desired unmanned engineering machines

them whose commands to obey, but also how many other machines are there in the grouping and sets an order in which the control protocol will be handled. Each machine would receive an ID by which it would communicate with the control station. After this, the process of configuration would be over and a normal work process would start.

Switching between machines would have an option to enforce a certain state upon the units currently not used. This would allow implying certain overall behaviorism upon the grouping. For example, it could force all the machines currently not used to go into idle state, which would, for example, decrease their engine's rpm and shut down any unnecessary systems to conserve power and decrease noise. This state would also decrease fuel consumption which would benefit the effective work time of the grouping.

4 Communications and Control Algorithms

Because there is a need to handle a lot of data traffic between the machines and the operator's station, it has been assumed that there is going to be a set of algorithms used on the hardware side to manage both control and sensory data.

The control station should be equipped with a set of two radio transmitters, working on the same frequency band, but on different channels. The lowest channel ("control channel") would be reserved for the currently controlled machine while the other one would be used for two way communications with the rest of the units. The main link would be used to transmit both control and sensory data with as high refresh rate as possible to allow as smooth control as possible. The second link would normally be used for siphoning data from sensors to the control station, but in case of switching the currently controlled machine it would send a command to the specific unit to become active. This way, the machine would reconfigure its transmitting unit to change onto the control channel. Additionally, the currently controlled machine would receive an order to turn inactive (Fig. 5).

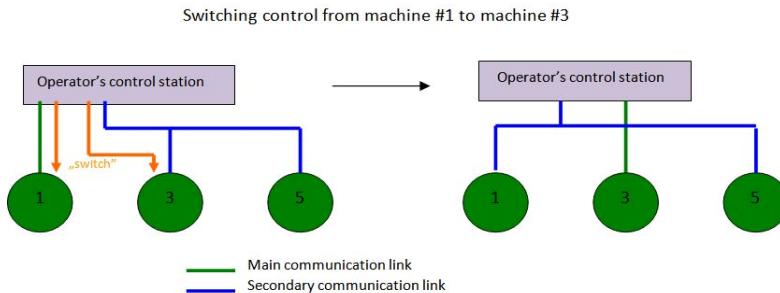


Fig. 5. A diagram describing the logic used for switching control over different unmanned engineering machines

It is obvious that any unmanned engineering machine working in the grouping would be required to undergo a certain adaptation process to enable it to communicate with the control station. However (assuming they all use the same control signal bus, like CAN-bus, etc.) the proposed solution would only require to change the currently used transmitting unit to an “intelligent” module working on a set frequency. The on-board logic of this unit would allow it to not only manage data transfer between the machine and the operator station but would also reconfigure itself when receiving a certain command. This way, there would not be any need to change the on-board control systems of these machines. Protocol handling would also be carried out by the newly attached unit as, because of several machines using the same frequency channel, an order will need to be introduced, resulting in data transfer frequency decrease. The unit would have the ability to emulate control data if needed to keep the machine in its desired state. Since the data traffic would use the same channel the transmission unit could track information transfers to see if other machines are currently sending data. If so, the unit would keep feeding the on-board control system with data corresponding to a safe state for the machine. If the traffic would cease and control with the operator’s station would be lost, the unit would have the possibility to execute a certain command (full shutdown, autonomy algorithm, etc.).

5 Sensor Data Visualization and Management

The described control station should be equipped with a set of visualization tools to help the operator easily obtain information he/she currently needs. Since the basic set of information used to generate control signals for the unmanned engineering machines comes from vision signals, the control station should be equipped with a set of monitors to allow natural access to these data. An analysis carried out by [5] has determined that a set of three monitors positioned in a way as to create a panoramic view is a good solution for these kinds of machines. They should be the main part of the control station’s visualization panel. Other being a set of two smaller monitors positioned underneath which would serve as additional sensory visualization units (Fig. 6).

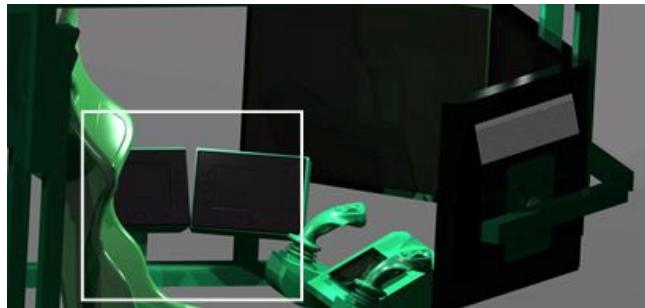


Fig. 6. The universal control station's additional monitors purposed with visualizing data acquired from different unmanned engineering machines

To allow easy access to data, there should be a system created to categorize data and a GUI to place that data in specific parts of the sensory monitors. That way, the operator would be able to switch between, for example thermal imaging and night vision with a push of a button. Additionally they may serve as means to display 3D maps of the environment [6, 7] This system could also interact with the control station's own system to ask for certain information to be sent. This could further decrease the traffic as data which would not be even displayed to the operator, would not be sent.

6 Control Station's Autonomy Algorithms

The idea of a single operator's station to serve as a means to controlling multiple unmanned engineering machines could also be used as means of introducing autonomous behaviors to machines not designed for such actions. Maintaining communication with each unit allows the control station to send each its individual control data. It would of course happen at a decreased rate, so the list of available autonomous behaviors would be limited, but it could prove invaluable when carrying out tasks required from all the machines at the same time (like reaching a certain destination). A trailing mechanic described in [8] could be used on each machine, with only a single unit having the whole sensory base needed. Others could enter certain modes, which could be executed with the need of only a small amount of additional information on their part. Currently two modes are being considered: shadow mode and follow mode.

In shadow mode, each unit would receive the same data (for example “move forward at 100% speed”) which would allow for a quick displacement of an entire grouping. The other one would be the follow mode, where certain data (presumably GPS) would be used to track each machine’s location. After selecting a formation, the system would prioritize units and set their desired location within this formation and execute commands to make them form up on each other. These calculations would probably be done on external hardware linked with the control station’s internal data bus.

7 Unmanned Engineering Machines Chosen for Integration

The concept of creating a universal control system which would allow for a single person to operate a number of unmanned engineering machines is a completely new approach to carrying out complex engineering tasks. Using multiple, specialized machines may prove more adequate in cases where space is limited and sending in a single, big machine is not a possibility. This strategy also allows for a more precise placement of different sensors which may aid the operator in his work, in comparison to every sensor being placed on a single platform. For the reasons stated above, three completely different unmanned ground units were selected (Fig. 7).



Fig. 7. Unmanned ground units and engineering machines selected for integration: a) Unmanned Kawasaki Mule, b) the “Dromader” UGV, c) Unmanned Engineering Machine “Marek”

First on is the fast moving, reconnaissance and sensor base Kawasaki Mule adapted specifically for the purpose of the developed control system. The second one is the proven “Dromader” UGV which was used in EDA research for electronic warfare and has the ability to manipulate relatively small objects. It’s capable of carrying out both inspection and small engineering missions. The last unit selected is the Unmanned Engineering Machine “Marek” which was specifically designed to carry out engineering tasks in hazardous areas.

8 System’s State of Play

Currently the system comprises of the control station and the server side as well as vision systems receivers and control system’s data transmitters. It allows for selecting any of the three units, which causes the system to start converting operator’s station input to a correct transmission protocol. Additionally, the build-in visualization unit changes to facilitate sensory data visualization of the selected unmanned unit. The HMI part of the system takes shape of an overlay, which is applied on top of the original control screens for each of the machines (Fig. 9). This way, there is no need to modify existing solutions while still retaining desired functionality.

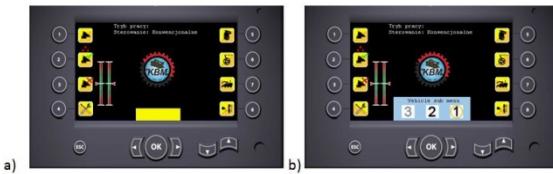


Fig. 8. The HMI element of the control system overlapping the original control screens:
a) original, b) modified

9 Summary

While the idea to control many machines by a single person may seem overwhelming, it needs to be stressed out that the operator is controlling a single machine at a time in a normal operation mode. In different modes, when the control is in fact sent to all the machines at once, the operator serves a role of a foreman, giving commands to a group. Managing relations between specific machines is to be done by the universal control system's subsystems.

As of now, the system only overlaps the “Marek” control screen, with other to be implemented. It needs to be noted that the remaining units either don't have any form of control screens developed (Kawasaki Mule) or have one developed for a different hardware platform (“Dromader”). Thus, there will need to be a lot of adaptation work carried out for the vehicles to be incorporated into the system.

References

1. Vidal, R.: Pursuit-Evasion Games with Unmanned Ground and Aerial Vehicles, Seul. IEEE (2001)
2. Howell, A.S., Hedrick, J.K.: Border patrol and surveillance missions using multiple unmanned air vehicles, Atlantis. IEEE (2004)
3. Typiak, A., Zienowicz, Z.: Utilization of remote controlled vehicle with hydrostatic driving system. In: 25th International Symposium on Automation and Robotics in Construction 2008, Lithuania (2008)
4. Typiak, R.: A design of a control system for a group of unmanned engineering machines. In: IARP RISE 2012, Warsaw (2012)
5. Bartnicki, A., Krogul, P., Przybysz, M.: Stanowisko zdalnego sterowania pojazdem bezzałogowym w zadaniach zmniejszenia zagrożenia wywołanego niekontrolowanym uwalnianiem substancji niebezpiecznych. Logistyka 2011(3) (2011) ISSN 1231-5478
6. Typiak, A., Gnatowski, M.: Map Building System for Unmanned Ground Vehicle. In: Conference on Mechatronic Systems, Mechanics and Materials 2011, Jastrzebia Gora (2011)
7. Typiak, A.: Use of Laser Rangefinder to detecting in surroundings of mobile robot the obstacles. In: 25th International Symposium on Automation and Robotics in Construction 2008, Lithuania (2008)
8. Dąbrowska, A., Typiak, A.: Nadążne sterowanie lekką platformą transportową w układzie teleoperatora, VI Międzynarodowa Konferencja Uzbrojeniowa, Waplewo (2006)

Motor Cascade Position Controllers for Service Oriented Manipulators

Tomasz Winiarski and Michał Walęcki

Warsaw University of Technology, Warsaw, Poland

tmwiniarski@gmail.com

<http://robotics.ia.pw.edu.pl>

Abstract. In the article a method is proposed to achieve some advantages of impedance control in the manipulators without direct torque controllers in joints, that is with indirect position–force control, position controlled joints and force–torque sensor in the wrist. The method is applicable to robots with backdrivable gears, where the external force exerted on the links causes the joints movement. The approach improves safety behaviour of a manipulator without the need for major changes in the hardware nor building a new robot, which is very important especially from the point of view of research systems.

Keywords: service robot, indirect force control, controller design.

1 Introduction

Improvements in the technology of both advanced sensors and control methods enable significant development of service robotics. Substantial number of robotic platforms have been built to investigate this field. From the mechanical point of view service robots have to contain at least a manipulation subsystem. More complex ones additionally possess mobile bases and active, multi-sensory heads. To deal with service oriented tasks, the impedance or position-force control of manipulators is needed [1, 2]. As an example the former approach is used in Justin robot [3], which is based on two LWR manipulators, and the latter in PR2 Willow Garage robot [4]. In terms of its overall performance Justin is a very successful impedance controlled robot. Unfortunately, it is rather costly, because of the elaborate construction of joints with customized torque sensors. As a result the alternative approach involving indirect position–force control determines the main trend in service robot arm design, e.g. [5], because it relies on a simpler, cheaper and more common mechanical construction. Impedance control in joints makes the unexpected contact of the manipulator with the environment (e.g. human) less dangerous than in the case of ordinary position controlled joints. This behavior is desirable for both service and social robots.

In the article a method is proposed to achieve some advantages of impedance control in the manipulators without direct torque controllers in joints, that is with indirect position–force control, position controlled joints and force–torque sensor in the wrist. The method is applicable to robots with backdrivable gears,

where the external force exerted on the links causes the joints to move. First the indirect force control scheme is described (sec. 2) with two variants of servomechanisms and direct current (DC) motor model. Then the cascade motor controller design method is presented to show how to add nonlinear joint compliance to the existing indirectly controlled robot (sec. 3). The approach improves safety behaviour of a manipulator without the need for major changes in the hardware or building a new robot, which is very important especially from the point of view of research systems. Then the experiments (sec. 4) illustrate the approach for both the classical unknown contour following position–force control benchmark and the obstacle–robot interaction case. The conclusions (sec. 5) summarize the work.

2 Indirect Force Control

Fig 1 presents the general structure of indirect force controller [6].

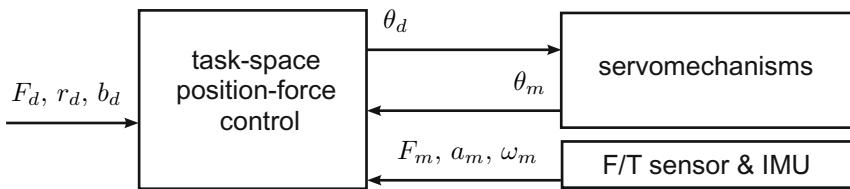


Fig. 1. Indirect force controller general structure, where: F_d is the vector of desired general force expressed in the task-space, r_d is the vector of desired position of the end-effector expressed in the same space, b_d is the vector of desired behaviors, θ_d is the vector of desired motors positions, θ_m is the vector of measured motors positions, F_m is the vector of measured general force expressed in the task-space, a_m is the vector of measured end–effector linear accelerations, ω_m is the vector of end–effector angular velocities

Usually, for the convenience of the operator or the system programmer, the manipulator task is specified in the task-space (sometimes called operational space) using three vectors: desired general force – F_d , desired pose – r_d and desired behaviour – b_d . Six directions can be distinguished in the task-space. In the simplest form the behaviour vector decides for which particular coordinates force control or position control is used. Both F_d and b_d consist of six elements each. The pose r_d consists of at least six elements (e.g. position and aggregated orientation axis-angle representation [1]).

The methods of indirect force controller development were investigated to obtain the desired behaviour of a position–force controller [6]. The controller uses general force measurements computed using wrist transducer data – F_m together with linear acceleration – a_m and angular velocity – ω_m [7] obtained from an

Inertial Measurement Unit (IMU). The exact methods of task specification are beyond the scope of this article [1]. The important property of position-force control is that it produces the desired motor position θ_d that can be directly sent to servomechanisms. Servomechanisms can be realized in two variants: direct (fig. 2(a)) and cascade (fig. 2(b)).

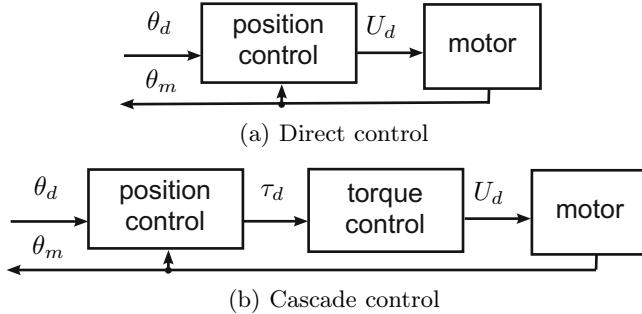


Fig. 2. Two types of servomechanisms, where: U_d is the desired voltage, τ_d is the desired torque

The classification is general and can be applied to both classic brushed direct current motors and modern brushless permanent magnet motors, nowadays commonly used in new manipulators. Direct position controller sets the desired voltage (in the case of the classic mechanically commutated DC motor) or desired amplitude of voltage waveforms (in the case of the brushless DC motor) – U_d to the motor power amplifier. By applying the desired voltage to the motor, the regulator influences directly the motor velocity and compensates the error resulting from the motor load [8]. In cascade control an internal torque control is introduced, where the desired torque – τ_d produced by the position controller is utilized to finally compute the desired voltage – U_d . The way the torque is controlled depends on the motor type. For brush motors the torque is roughly proportional to the current irrespective of speed. Torque control can be implemented as a PI feedback loop that computes the voltage to be applied to the motor. For brushless motors more sophisticated methods are used to achieve the desired torque, e.g. field oriented control [9]. Further on, a representative manipulator with harmonic, backdrivable gears and brushed direct current motors was utilized to illustrate the approach proposed in this work, hence the DC motor time domain model is introduced [10]:

$$U_d = iR + \frac{di}{dt}L + U_b \quad (1)$$

where: i is the current in motor circuit, R is the winding resistance, L is the winding inductance, U_b is the back electromotive force,

$$U_b = k_b * \omega \quad (2)$$

where: k_b is a proportional factor and ω is the motor shaft velocity. In DC motor the output torque τ can be computed from the following equation:

$$\tau = k_\tau i \quad (3)$$

where: k_τ is the proportional factor.

3 Cascade Controller Development Motivation and Procedure

Typical industrial robots are still classic rigid manipulators that implement a position control law [11]. Their position controllers (fig. 2(a)) work under the assumption of constrained position error. The controller main goal is to minimize the position error to accurately follow the desired trajectory. Any deviation from the desired trajectory is considered a system fault or overload [12]. Thus, for safety reasons, the internal robot state is monitored. Exceeding the position error limit or measured current limit implies an emergency stop and as a consequence motor power is switched off and joints mechanical brakes are activated. In some situations such robot behaviour can be dangerous, e.g. while the manipulator hits a person. For that reason the industrial manipulators work in well organized human-free areas, in opposition to service robots that should cooperate with humans in their unstructured environment.

In contrast to the position controllers, the torque controllers allow to constraint the motor or the joint torque while the manipulator is still operational, without the need to switch off the power nor need to activate the brakes. By using torque control, it is possible to constraint the force exerted on the environment (e.g. person) even if contact is not detected by the force sensor mounted in the end-effector. If additionally the manipulator is equipped with backdrivable gears, it is possible for the victim to push the manipulator backwards and break free. This behaviour is essential in service robotics. Torque controllers were typically utilized in direct control or impedance control of manipulators.

Further on the authors analyse the possibility of using motor torque controllers with constrained torque inside the indirect position force controllers. To achieve that, the new cascade structure is proposed as it is presented in fig. 3, for the case of DC motor propelling the joint. A significant feature of this approach is that it bases on an existing position controller, that previously worked in the basic controller structure in fig. 2(a). The conversion from structure presented in fig. 2(a) to fig. 3 can be executed individually for each motor of the manipulator. In the following the components of the final controller presented in fig. 3 are presented:

- **position control** - It is assumed that position controller already existed in the system and was developed according to methods presented eg. in [13] to work in the structure in fig. 2(a). This controller can be used in the new cascade structure (fig. 3), because the remaining parts of this structure were developed to form motor equivalent, that can be treated as motor itself with

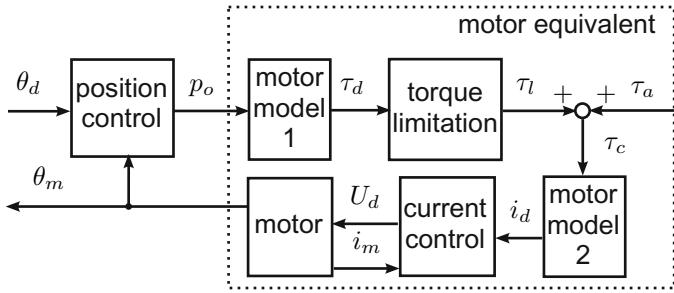


Fig. 3. Motor cascade controller structure, where: p_o is the output of the position controller, τ_l is the limited torque, τ_a is additional torque, τ_c is computed torque, i_d is desired current, i_m is measured current

torque constraint. If position controllers use an integrational component, an anti wind up is vital, otherwise the long-term position error, permissible in the new controller, can cause a significant overshoot or even instability,

- **motor model 1** - This component was introduced to allow torque limitation vital e.g. for service robotics. Initially the current is computed (4) basing on (1):

$$i = \frac{U_d - U_b - \frac{di}{dt}L}{R} \quad (4)$$

in practice it forms the following discrete implementation:

$$i_d^k = \frac{U_d^k - U_b^k - \frac{i_m^k - i_m^{k-1}}{\Delta T}L}{R} \quad (5)$$

where: k in the upper right index is discrete time instant, ΔT is the position controller time period and according to (2), U_b^k can be computed as follows:

$$U_b^k = k_b \omega^k = k_b \frac{\theta_d^k - \theta_d^{k-1}}{\Delta T} \quad (6)$$

finally τ_d^k is computed with accordance to formula (7).

$$\tau_d^k = k_\tau i^k \quad (7)$$

- **torque limitation** - The motor torque can be limited:

$$\tau_l^k = \begin{cases} \tau_u & \text{for } \tau_d^k > \tau_u \\ \tau_d^k & \text{for } \tau_b \leq \tau_d^k \leq \tau_u \\ \tau_b & \text{for } \tau_d^k < \tau_b \end{cases} \quad (8)$$

where: τ_u is the upper torque limit and τ_b is the bottom torque limit. This limits should be set slightly above the nominal torques for manipulator in motion.

- **torque sum** - The computed torque is the sum of both additional torque and limited torque:

$$\tau_c^k = \tau_l^k + \tau_a^k \quad (9)$$

This component is very important both for overall controller performance and ease of τ_u and τ_b calculation. The τ_a can be treated as the controller feedforward partial noise cancellation and should include at least static compensation for manipulator with typical load. In more elaborated structure partial dynamics compensation is desirable.

- **motor model 2** - Here the desired current is computed according to formula (7):

$$i_d^k = \frac{\tau_c^k}{k_\tau} \quad (10)$$

- **current control** - An adequate hardware structure should be provided, with motor current controllers [14].

4 Experiments

The experiments were conducted with modified IRb6 manipulator under supervision of MRROC++ robot programming framework [1]. The IRb6 manipulator is equipped with brush DC motors, hence the desired torque τ_d is proportionally related to the desired current i_d and the measured torque τ_m is proportionally related to the measured current i_m [8].

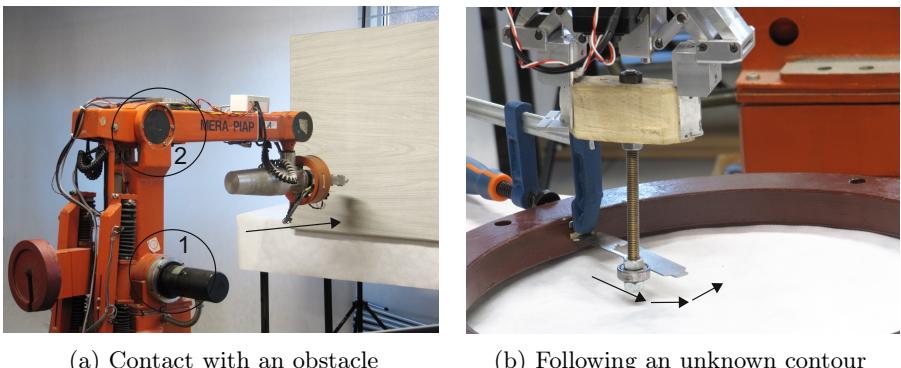
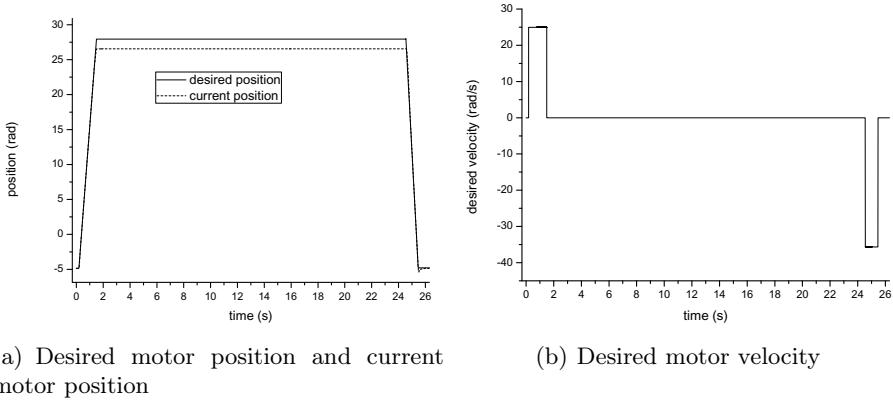


Fig. 4. Experimental test beds

Two following experiments were conducted to compare system behaviour for initial direct position control structure and proposed cascade controller. The first experiment (sec. 4.1) relies on the test bench with the manipulator end-effector shifting in horizontal plane by moving one of its motors (fig. 4(a)) and taking advantage of mechanical coupling between joints 2 and 3. The single motor was

chosen to illustrate the particular motor behaviour. Here the manipulator hits an obstacle. In this case no end-effector force feedback is used, hence this case is representative for the unexpected contact with the environment. In the second experiment the unknown contour following procedure is executed (sec. 4.2) as a common position-force control benchmark to compare the system behaviour with direct and cascade motor control. A few motors are in simultaneous motion and the manipulator behaviour from the point of view of its end-effector is presented.



(a) Desired motor position and current motor position (b) Desired motor velocity

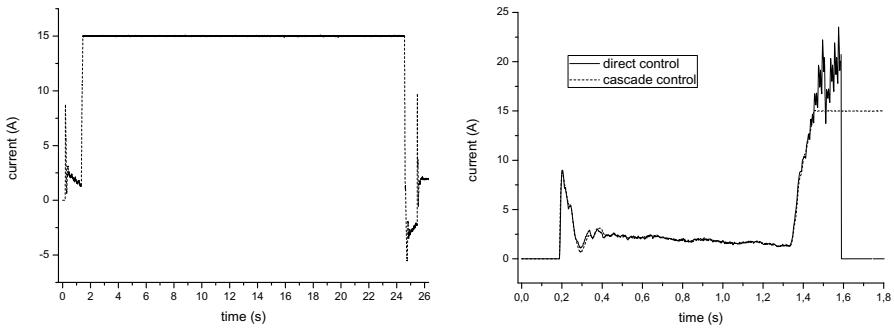
Fig. 5. Experiments with contact for cascade control

4.1 Unexpected Impact

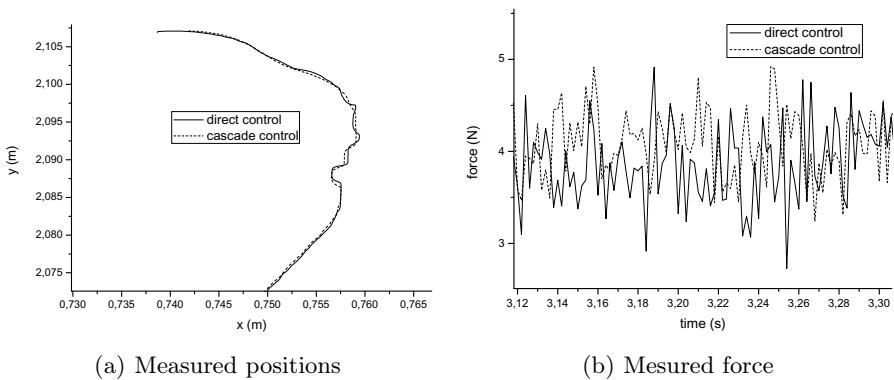
Let us consider the case where the manipulator hits the obstacle while it moves. Fig. 5 presents the current position, the desired position and the desired motor velocity for cascade controller. Initially (until $t \approx 1.4s$) the motor moves with the desired velocity, hence the desired position changes linearly. Then it hits the obstacle and the position error occurs. The point is that although the desired torque limit was reached (fig. 6) (equivalent $i_{max} = 15A$) for cascade controller, there was no emergency stop, hence in the end it was possible to set the velocity in opposite direction and manipulator moved backward. With the direct control in the same situation an overcurrent error caused the emergency stop and the power was cut (see $t = 1.4s$ in fig. 6(b)). It should be noted that peaks in the measured current plots are caused by the increases of desired velocity.

4.2 Following an Unknown Contour

The new motor cascade controller is currently utilized in all IRb6 manipulator applications executed in robotics laboratory at Warsaw University of Technology, i.e. haptic coupling, Rubik cube solving etc. Here one of the typical benchmarks is performed to compare system behavior for cascade and direct motor control.

**Fig. 6.** Current measured for experiments with contact**Table 1.** Basics statistics of the force measurements for the following of the same segment of contour

type of controller	nr. of samples	desired force value	average force value	standard deviation
direct	100	4	4.13	0.42
cascade	100	4	4.07	0.46

**Fig. 7.** Following of an unknown contour experimental results

Following an unknown contour [15, 16] constitutes the task of detecting the edge of the contour and staying in contact with it while the manipulator end-effector tip moves in a plane. The task is specified to maintain the desired contact force (in the case of this experiment $F_d = 4N$) in the direction orthogonal to the contour tangent and reach the desired velocity in the direction parallel to this contour. Fig. 4(b) presents the experimental test bed. The arrows show the segment of motion with the rapid direction change for the experimental results visualized in fig. 7. The plots show that the system behaviour for both controller

structures is almost identical, what is a desirable result. In fig. 7(b) the measured force in contour plane is visualised, for 100 samples of the same segment of motion both for direct and cascade control. Table 1 presents the basic statistics of the measurements.

5 Conclusions

The cascade structure of motor controller is the natural extension of direct position controller. The extended structure allowed to introduce some safety features, so far reserved for impedance controllers. The experiments performed for indirect force control showed that proposed motor control transition procedure leads to successful, final implementation. The future work will concentrate on modelling of manipulator to precisely compute additional torque τ_a . Other task is to utilize the approach to control various type of objects, e.g. active head [17].

Acknowledgment. This project was funded by the National Science Centre according to the decision number DEC-2012/05/D/ST6/03097. Tomasz Winiarski has been supported by the European Union in the framework of the European Social Fund through the Warsaw University of Technology Development Programme. The authors thank Michał Lisicki for his help in paper editing.

References

1. Zieliński, C., Winiarski, T.: Motion generation in the MRROC++ robot programming framework. *International Journal of Robotics Research* 29(4), 386–413 (2010)
2. Winiarski, T., Banachowicz, K.: Opening a door with a redundant impedance controlled robot. In: 9th Workshop on Robot Motion & Control (RoMoCo), pp. 221–226 (2013)
3. Dietrich, A., Wimbock, T., Albu-Schaffer, A.: Dynamic whole-body mobile manipulation with a torque controlled humanoid robot via impedance control laws. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3199–3206 (2011)
4. Meeussen, W., Wise, M., Glaser, S., Chitta, S., McGann, C., Mihelich, P., Marder-Eppstein, E., Muja, M., Eruhimov, V., Foote, T., et al.: Autonomous door opening and plugging in with a personal robot. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 729–736 (2010)
5. Schmid, A., Gorges, N., Goger, D., Worn, H.: Opening a door with a humanoid robot using multi-sensory tactile feedback. In: International Conference on Robotics and Automation (ICRA), pp. 285–291. IEEE (2008)
6. Winiarski, T., Woniak, A.: Indirect force control development procedure. *Robotica* 31, 465–478 (2013)
7. Winiarski, T., Banachowicz, K.: The acquisition system of general force of contact between robotic manipulator and the environment. *Pomiary Automatyka Robotyka* (2), 390–394 (2013) (in Polish); System akwizycji skorygowanej siły uogólnionej kontaktu robota manipulacyjnego z otoczeniem
8. Yeadon, W., Yeadon, A.: Handbook of Small Electric Motors. McGraw-Hill handbooks. McGraw-Hill Education (2001)

9. Drury, B.: The Control Techniques Drives and Controls Handbook, 2nd edn. Institution of Engineering and Technology, Stevenage (2009)
10. Kuo, B.C., Golnaraghi, M.F.: Automatic control systems, vol. 4. John Wiley & Sons, New York (2003)
11. Pires, J., Sa da Costa, J.: Position sensing and motor control in industrial robotics. In: Proceedings of the IEEE International Symposium on Industrial Electronics, ISIE 1997, vol. 3, pp. 866–871 (1997)
12. Herman, S.: Industrial Motor Control, 7th edn. Delmar Cengage Learning (2012)
13. Kurman, K.: Feedback Control: Theory and Design. Elsevier Science Inc., New York (1984)
14. Walęcki, M., Banachowicz, K., Winiarski, T.: Research oriented motor controllers for robotic applications. In: Kozłowski, K. (ed.) Robot Motion and Control 2011. LNCIS, vol. 422, pp. 193–203. Springer, Heidelberg (2012)
15. Ahmad, S., Lee, C.: Shape recovery from robot contour-tracking with force feedback. In: IEEE International Conference on Robotics and Automation, pp. 447–452 (1990)
16. Baeten, J., Verdonck, W., Bruyninckx, H., De Schutter, J.: Combining force control and visual servoing for planar contour following. Machine Intelligence and Robotic Control 2(2), 69–75 (2000)
17. Walęcki, M., Stefaczyk, M., Kornuta, T.: Control system of the active head of a service robot exemplified on visual servoing. In: 9th Workshop on Robot Motion and Control (RoMoCo), pp. 48–53 (2013)

Specification of Tasks in Terms of Object-Level Relations for a Two-Handed Robot

Cezary Zieliński and Tomasz Kornuta

Warsaw University of Technology, Institute of Control and Computation Engineering,
Nowowiejska 15/19, 00-665 Warsaw, Poland
{C.Zielinski,T.Kornuta}@elka.pw.edu.pl

Abstract. It is customary in artificial intelligence that object-level robot task planning is performed by stating sequences of situations (states of the environment) and that transitions between those states are executed by operations of an active agent. The paper presents a formal model of a robotic system in which operations treated as behaviours are defined in terms of transition functions. Those behaviours bring about the required environment state changes. The paper focuses on the transformation between the object-level plan expressed as a sequence of relations between objects into a low level sequences of robot behaviours. The theoretical considerations are illustrated on an example of a two-handed robot acquiring a jar from a closed cupboard, emptying the contents of the jar into a bowl and replacing the jar in the cupboard, i.e. a standard service task.

Keywords: robot tasks, task specification, task translation.

1 Introduction

Artificial intelligence usually uses preconditions and postconditions (effects) to specify an action in a certain environment [1]. Those conditions are formulated in terms of the required and produced state of that environment. The change of state of the environment is brought about by a certain agent acting in that environment. If the agent is a robot such an object-level specification has to be translated into manipulator-level actions pertaining to the control of devices constituting the robot (e.g. its actuators and sensors). The purpose of this paper is to show how the object-level specifications, well rooted in artificial intelligence, can be translated into low-level robot commands. The method will be presented on a non-trivial example of a two-handed robot possessing a mobile head opening a cupboard to retrieve a closed jar full of grains, subsequently emptying it into a bowl and finally replacing the closed jar in the cupboard.

Plenty of work pertaining to plan generation has been done since the first planners such as STRIPS [2] and ADL [3] were created. The major difference between those two was the assumption dealing with unmentioned literals: the former assumed them being false (the so called Closed World), while the latter assumed that they are unknown (Open World assumption). In 1998, inspired mainly by the two aforementioned planners, a language called PDDL (Planning

Domain Definition Language) [4, 5] was proposed. The key idea behind it was to separate the model of the planning problem into two parts: description of the domain and definition of the problem. Besides that PDDL was mainly supposed to serve as a common platform in the International Planning Competition (IPC). It turned out that during the last 15 years it has become a de facto planning standard. SHOP2 (Simple Hierarchical Ordered Planner 2) [6] and other hierarchical transition networks (HTNs)-based planners [7, 8] are examples of using PDDL. Recently a new proposal for plan description languages emerged – the CRAM plan language (CPL) [9]. The generation of a plan of actions leading to the goal is, however, only one facet of a problem. The other is how to translate the high-level actions into low-level commands executed by the robot. The most common approach is to define actions in terms of hierarchical, concurrent state machines [10]. One of the examples of a task-level executive system based on hierarchical concurrent state machines, which controls the overall behaviour of the system, that are currently available in Robot Operatin System (ROS) [11] is SMACH (State MACHine) [12], which replaced previously used TeleoReactive EXecutive (TREX) [13]. In this paper we seek the solution of a similar problem, whereas instead of state machines we propose utilization of transition functions for the description of behaviours of subsystems constituting the robot control system.

2 Universal Model of a Robotic System

2.1 General Inner Structure of an Embodied Agent

It is assumed that a robotic system is represented as a set of agents a_j , $j = 1, \dots, n_a$, where n_a is the number of agents in the system. Subscript j designates a particular agent. Those agents that have physical bodies are called embodied agents. This presentation focuses on embodied agents [14], but all other agents can be treated as special cases with no body, thus the presentation is general. An embodied agent a_j , or simply an agent from now, possesses effectors E_j , which influence the environment, receptors R_j (exteroceptors), which gather information from the environment, and a control system C_j that governs the actions of the agent. The exteroceptors of the agent a_j are numbered (or named), hence $R_{j,l}$, $l = 1, \dots, n_R$, and so are its effectors $E_{j,h}$, $h = 1, \dots, n_E$. Both the receptor readings and the effector commands undergo transformations into a form that is convenient from the point of view of control, hence the concepts of virtual receptors r_j and virtual effectors e_j appear. Thus the control system C_j is decomposed into: a number of virtual effectors $e_{j,n}$, $n = 1, \dots, n_e$, and virtual receptors $r_{j,k}$, $k = 1, \dots, n_r$, as well as a single control subsystem c_j (fig. 1a). Virtual receptors perform information aggregation, consisting in the composition of readings obtained from several exteroceptors or in the extraction of the required information from one, but a complex one. Moreover the readings obtained from the same exteroceptors $R_{j,l}$ may be processed in different ways, so many virtual receptors $r_{j,k}$ can be formed. Exteroceptor readings $R_{j,l}$ are processed by virtual receptors and subsequently transmitted to the control

subsystem c_j which generates commands influencing virtual effectors e_j , which in turn activate effectors E_j . The control subsystem c_j must be able to both re-configure exteroceptors R_j and influence the method how the virtual receptors r_j aggregate readings. It also must acquire proprioceptive data from the effectors. Hence the reverse connections with those subsystems also exist. An agent through its control subsystem is able to establish a two-way communication with other agents $a_{j'}$, $j \neq j'$.

Each of the enumerated components of the control subsystem uses communication buffers to transmit or receive information to/from the other components (fig. 1a). A consistent denotation method is used to designate both the components and their buffers. To make the description of such a system concise no distinction is being made between the denotation of a component buffer and its state – the context is sufficient. In the assumed notation a one-letter symbol located in the centre (i.e. E , R , e , r , c designates the component. To reference its subcomponents or to single out the state of this component at a certain instant of time we place extra indices around this central symbol. The left superscript designates the referenced buffer of the component by stating both the owner of the buffer and the component to which it is connected. The right superscript designates the time instant at which the state is being considered. The left subscript tells us whether this is an input (x) or an output (y) buffer and if there is no left subscript the internal memory of the subsystem is referred. The right subscript may be complex, where its elements are separated by commas. They refer to the ordinal numbers of: the agent, its component and subcomponent, or the ordinal number of the function. For instance ${}^x{}_c^i$ denotes the contents of the control subsystem input buffer of the agent a_j acquired from the virtual effector at instant i . In the case of a function (f) the left superscript designates the buffer in which the computed value will be placed. The right superscript represents its type.

2.2 General Subsystem Behaviour

The general work-cycle of any subsystem s , where $s \in \{c, e, r\}$, of any agent a_j is presented in fig. 1b. The functioning of a subsystem s requires a transition function processing taking the data contained in its input buffers ${}_x s_j$ and its internal memory ${}^s s_j$ as arguments to produce the output buffer values ${}_y s_j$ and update the memory ${}^s s_j$. Hence the subsystem behaviour can be described by a transition function ${}^s f_j$ defined as:

$$[{}^s s_j^{i+1}, {}_y s_j^{i+1}] := {}^s f_j({}^s s_j^i, {}_x s_j^i). \quad (1)$$

where i and $i+1$ are the consecutive discrete time stamps. Function (1) describes the evolution of the state of a subsystem s . A single function (1) would be too complex to define in a compact form, thus it is usually convenient to decompose it into a set of partial functions:

$$[{}^s s_j^{i+1}, {}_y s_j^{i+1}] := {}^s f_{j,u}({}^s s_j^i, {}_x s_j^i), \quad (2)$$

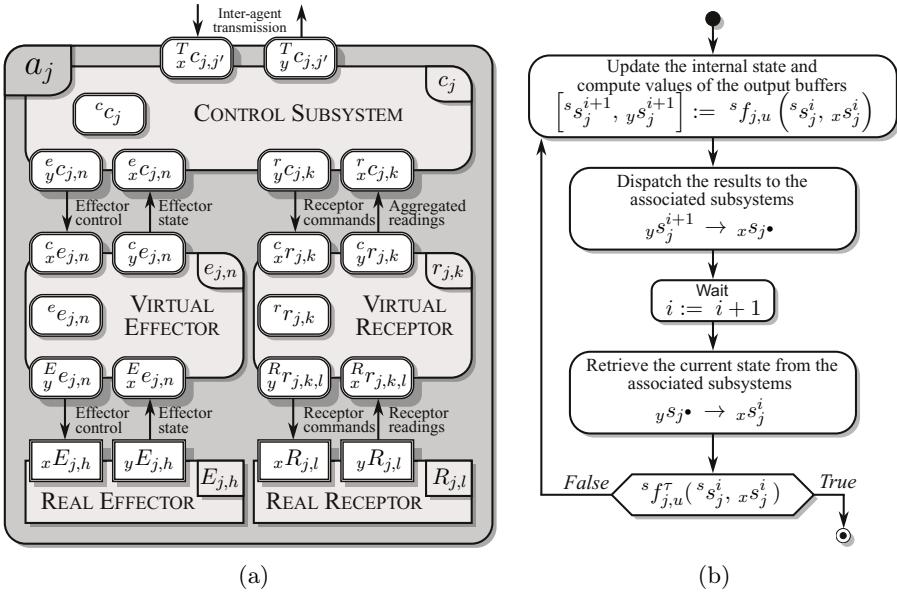


Fig. 1. (a) Internal structure of an agent a_j (b) General flow chart of a subsystem behaviour ${}^s B_{j,u}$

where $u = 0, \dots, n_{f_s}$. Capabilities of the agent arise from the multiplicity and diversity of the partial functions of its subsystems. Such a prescription requires rules of switching between different partial transition functions of a subsystem, thus two additional Boolean valued functions (predicates) are required:

- ${}^s f^\sigma$ defining the Initial Condition and
- ${}^s f^\tau$ representing the Terminal Condition.

The former selects the transition function for cyclic execution, while the latter determines when this cyclic execution should terminate. Hence a multi-step evolution of the subsystem in a form of a Behaviour $B_{j,u}$ defined as:

$${}^s B_{j,u} \triangleq {}^s B_{j,u} ({}^s f_{j,u}, {}^s f_{j,u}^\tau) \quad (3)$$

The execution pattern of such a behaviour is presented in fig. 1b. The s_j^\bullet , where $j^\bullet \in \{j, j'\}$, denotes all subsystems associated with s_j (in the case of Control Subsystem some of those subsystems even may not belong to the same agent, hence j' appears). The behaviours ${}^s B_{j,u}$ can be associated with the nodes of a graph and Initial Conditions with its arcs, thus a finite state automaton representation results fig. 2. ${}^s B_{j,0}$ is the default (idle) behaviour, activated when no other behaviour can be activated. Behaviour selection represented by a hexagonal block is executed as a stateless switch defined by the initial conditions ${}^s f_{j,u}^\sigma$.

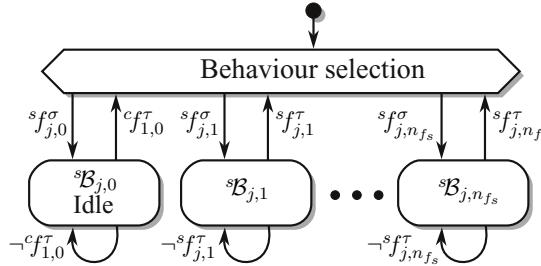


Fig. 2. State graph of the behaviour selection automaton

3 The Environment and Its Model

An embodied agent operates in a physical environment \mathcal{E} . To reason about the effects of the agent's actions in such an environments its model \mathcal{M} has to be created. This model will reside in the memory of the control subsystem of the agent, i.e. in ${}^c c_j$. The environment \mathcal{E} contains objects of interest \mathcal{O}_ν , where $\nu = 1, \dots, n_o$ and n_o is the number of objects of interest. Those objects are modelled as o_ν . The model of an object is created by specifying the set of its attributes \mathcal{D}_ν , where

$$\mathcal{D} = \bigcup_{\nu=1}^{n_o} \mathcal{D}_\nu \quad (4)$$

is the set of all attributes. The attributes describe for instance: the location of the object, the characteristic places on the object (e.g. bottom, top, handle), shape, colour, affordances. Some of them have constant values others are variable. The values of variable attributes are obtained from the virtual receptors through ${}^r c_{j,k}$ or from other agents through ${}^T x c_{j,j'}$. The current or future state of the environment, or in other words the situation that occurs or has to occur, is modelled as a set of object-level relations, i.e. relations between objects. Those can be binary, when two objects are the arguments of such a relation or unary, if a single object is the argument. In the former case, for instance, spacial relationships between objects can be defined, e.g. *book ON table*. In the latter case certain properties of an object are ascertained, e.g. its colour or shape. However object-level relations are of use only if they can be translated into the notions that the agent understands, i.e. are within its ontology. The agent operates on the values of attributes, so the object-level relations have to be defined in terms of mathematical relations between the attribute values. An example of such an approach was presented in [15, 16]. However the only mathematical relation used was the equality relation. Although this approach proved its utility it was a bit artificial, e.g. a book is on the table regardless of whether a certain coordinate frame affixed to the bottom of the book matches exactly a coordinate frame affixed to the top of the table or they are misaligned a bit. Such a misalignment is better expressed in terms of inequalities. Thus inequalities, regions or set inclusions would be of utility as well. This line of thought can be extended even

further. If the corner of the book protrudes from the table by, e.g. 1 cm, we would be inclined to treat the book as still on the table. However this requires fuzzy relations and definition of appropriate fuzzy sets with degree of membership functions. Nevertheless such a model captures only geometrical properties of the environment and neglects all other effects, e.g. gravity or friction. Unfortunately the exactness of the model is inversely proportional to its utility in automatic reasoning, so a certain tradeoff has to be found.

Relations in general are used for three purposes: 1) describing the state of the environment, 2) testing whether the state of the environment is as expected or 3) producing the required state of the environment. In the case of a unary relation pertaining to an object having an attribute representing the colour of that object, 1) we can state that the object is red, 2) we can ask whether it is red or 3) cause it to become red (paint it red).

For an embodied agent to execute actions in the environment it has to cause the occurrence of specific relations between objects. Objects do not move by themselves – they have to be translocated by the effectors, so to the model of the environment the model of the effectors, treated as objects having attributes, has to be added. Thus an effector represented as an object is modelled as ${}^E o_{j,h}$. In this way a single ontology covers both the agent and its environment. Hence situations can be described in terms of relations between objects o_ν and effectors ${}^E o_{j,h}$.

Of interest to us is the definition of an object-level relation \mathcal{R}_η , $\eta = 1, \dots, n_\varrho$, where n_ϱ is the number of considered relations. This paper focuses on binary relations ${}^b \mathcal{R}$. A binary relation ${}^b \mathcal{R}_\eta$ between objects o_κ and o_λ , i.e. $(o_\kappa \ {}^b \mathcal{R} \ o_\lambda)$ is defined in terms of attributes of those objects, i.e. attributes extracted from the sets \mathcal{D}_κ and \mathcal{D}_λ . For instance, if $d_{\kappa,\alpha} \in \mathcal{D}_\kappa$ and $d_{\lambda,\beta} \in \mathcal{D}_\lambda$, an equality based binary relation ${}^b \mathcal{R}_\eta$ can be defined as:

$$F_{\eta,1}(d_{\kappa,\alpha}) = F_{\eta,2}(d_{\lambda,\beta}) \quad (5)$$

where $F_{\eta,1}$ and $F_{\eta,2}$ are appropriately defined functions taking attributes as arguments. Obviously more complex equality based binary relations involving several attributes of both objects can be invented. Inequality based binary relation can be defined too, e.g.:

$$F_{\eta,2}(d_{\lambda,\beta}) \leq F_{\eta,1}(d_{\kappa,\alpha}) \leq F_{\eta,3}(d_{\lambda,\beta}) \quad (6)$$

Conversely this can be stated as $F_{\eta,1}(d_{\kappa,\alpha}) \in \mathcal{F}_\eta(d_{\lambda,\beta})$, if \mathcal{F}_η defines a set representing the required region. The mathematical definition of a unary relation is similar to (5) or (6), however here the value of the attribute is compared with a constant or has to be within a certain subset.

Either the current state of the environment or the expected one can be described in terms of relations defined by (5) or (6). Obviously in the case of the expected state, which implies an activity of the effectors causing it, such a relation requires at least that the attributes of ${}^E o_{j,h}$ will be involved in the definition of such a relation. The control subsystem of the agent will modify its state in such a way that the attributes of ${}^E o_{j,h}$ will change in such a way that the object-level

relation will be obtained. As usually the mentioned activities not only require the description of the required final state, but also the definition of the method of attaining it, the definition of the relation has to be extended by extra factors, e.g. definition of acceptable interaction forces in the process of producing the final outcome. However the described activities are defined in terms of transition function, thus the object-level task specification can be readily translated into the behaviours described in section 2.2.

4 An Example of Object-Level Specification

An exemplary task consists in a robot acquiring a jar from a closed cupboard, emptying the contents of the jar into a bowl and replacing the jar in the cupboard, i.e. a standard service task. The robot has two manipulators equipped with a three-fingered gripper each, and a pan tilt head with two cameras and two RGB-D sensors. The task is expressed in terms of the following relations:

1. Location of the cupboard is known,
2. Location of the door-handle is known,
3. Door-handle is grasped by the gripper: (*door-handle Grasped gripper*),
4. Cupboard door is open: (*door At open-pose*),
5. Location of the jar is known,
6. Jar is grasped: (*jar Grasped gripper*),
7. Jar is outside of the cupboard (*jar At outside-pose*),
8. Jar is open (*jar-lid At above-jar-pose*),
9. Location of the bowl is known,
10. Contents of the jar are in the bowl (*jar-contents In bowl*),
11. Jar is closed (*jar-lid Twisted jar*),
12. Jar is in the cupboard (*jar On cupboard-shelf*),
13. The door-handle is grasped by the gripper: (*door-handle Grasped gripper*),
14. Cupboard is shut: (*door At closed-pose*).

Assertions that a certain fact is known are specified in terms of unary relations, which in turn are associated with behaviours acquiring data, i.e. reading in the values of attributes through: ${}^r_c_{j,k}$, ${}^T_c_{j,j'}$ or ${}^e_c_{j,n}$. Several binary relations need to be defined: *In*, *On*, *At* and *Grasped*, *Twisted*. Here an infix notation has been used, but in AI prefix notation is preferred. It should be noted that in this simplified example some abstract objects, such as certain poses, have been introduced. In reality those should be defined as characteristic places of other objects. The scarcity of space lets us define just one of those relations. Let it be the *Twisted* relation. For that purpose we shall use (5) type of definition.

5 Translation of Object-Level Specification into Behaviours

The behaviours defined in section 2.2 are used as operations executed by the embodied agent (robot) and thus causing the required changes of state of the

environment. Hence object-level environment state changes can be reduced to behaviours of the embodied agent, defined in terms of transition functions and the buffers of the control subsystem. Let us analyse the execution of the object level relation (*jar-lid Twisted jar*). Both *jar* (O_{jar}) and *jar-lid* (O_{lid}) are objects of the environment. Their models are: o_{jar} and o_{lid} respectively. The sets \mathcal{D}_{jar} and \mathcal{D}_{lid} contain their attributes.

$$\mathcal{D}_{\text{jar}} = \{d_{\text{jar},\text{base}}, d_{\text{jar},\text{top}}, d_{\text{jar},\text{bottom}}, d_{\text{jar},\text{handle}}, d_{\text{jar},\text{twist}}, d_{\text{jar},\text{grasped}}, \dots\} \quad (7)$$

where $d_{\text{jar},\text{base}} = {}^0\mathcal{T}$ is a homogeneous matrix expressing the pose of the jar base coordinate frame JB with respect to the world reference frame 0, $d_{\text{jar},\text{top}} = {}^{JB}\mathcal{T}$ top of the jar with respect to its base, $d_{\text{jar},\text{bottom}} = {}_{Jb}\mathcal{T}$ bottom of the jar with respect to its base, $d_{\text{jar},\text{handle}} = {}^{JB}\mathcal{T}$ handle (grasp location) of the jar with respect to its base and $d_{\text{jar},\text{twist}}$ is the affordance of the jar, telling us that if we want to put the lid on the jar it has to be twisted. Similarly the attributes of the lid are grouped in the set

$$\mathcal{D}_{\text{lid}} = \{d_{\text{lid},\text{base}}, d_{\text{lid},\text{top}}, d_{\text{lid},\text{bottom}}, d_{\text{lid},\text{handle}}, d_{\text{lid},\text{twist}}, d_{\text{lid},\text{grasped}}, \dots\} \quad (8)$$

Now the relation *Twisted* can be defined using (5):

$$\begin{aligned} d_{\text{lid},\text{base}} * d_{\text{lid},\text{bottom}} &= d_{\text{jar},\text{base}} * d_{\text{jar},\text{top}} \\ d_{\text{lid},\text{twist}} &\in \mathcal{D}_{\text{lid}} \\ d_{\text{jar},\text{twist}} &\in \mathcal{D}_{\text{jar}} \\ d_{\text{lid},\text{grasped}} &= \text{yes} \\ d_{\text{jar},\text{grasped}} &= \text{yes} \end{aligned} \quad (9)$$

where $*$ is a homogeneous matrix multiplication operator. This definition tells us that it will occur if the bottom of the lid will be located where the top of the jar is. However both the lid and the jar need to have the ability to twist the lid on the jar. Moreover, both of them need to be grasped – otherwise they would not have the ability to move. Here, to simplify the explanation, concrete objects are used in the definitions. Obviously in reality we need to use object classes in the definitions. Otherwise we would need to define a plethora of relations for each pair of objects.

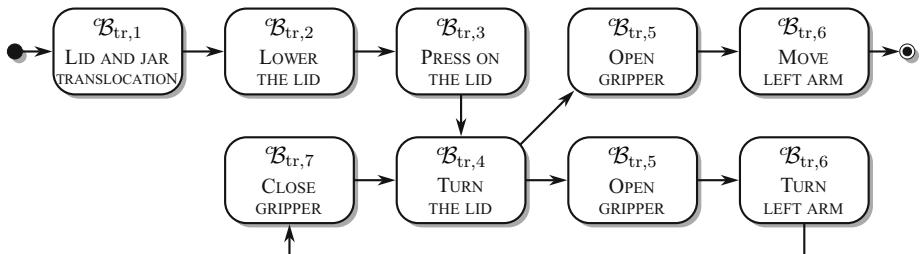


Fig. 3. Finite state automaton realizing the *Twisted* relation

As from the definition (9) the precondition for its execution requires that both the *jar* and *lid* have to be already grasped, so the execution of the *Twisted* relation between those two objects starts from bringing them together to the location where the lid can be twisted on the jar. The behaviours ${}^c\mathcal{B}_{tr,u}$ of the agent controlling the two-handed robot *tr*, are invoked by the finite state automaton presented in fig. 3. The sequence of behaviours is the following: ${}^c\mathcal{B}_{tr,1}$ – translocation of the lid by the left manipulator *lm* and the translocation of the jar to the location in which the twisting operation will be performed; ${}^c\mathcal{B}_{tr,2}$ – lowering the lid onto the top of the jar by executing a guarded motion; ${}^c\mathcal{B}_{tr,3}$ – pressing on the lid onto the jar using compliance control; ${}^c\mathcal{B}_{tr,4}$ – turning the lid using compliance control; ${}^c\mathcal{B}_{tr,5}$ – opening the gripper; ${}^c\mathcal{B}_{tr,6}$ – regrasp operation, i.e. reorientation (translocation) of the gripper of the left manipulator *lm* over the lid (pure pose control); ${}^c\mathcal{B}_{tr,7}$ – closing the gripper; ${}^c\mathcal{B}_{tr,5}$ – opening the gripper; ${}^c\mathcal{B}_{tr,6}$ – motion of the left manipulator *lm* from over the lid. Subsequence of behaviours: ${}^c\mathcal{B}_{tr,4}$, ${}^c\mathcal{B}_{tr,5}$, ${}^c\mathcal{B}_{tr,6}$ and ${}^c\mathcal{B}_{tr,7}$ is executed iteratively until the lid is tightly screwed on the jar. Behaviours ${}^c\mathcal{B}_{tr,5}$ and ${}^c\mathcal{B}_{tr,6}$ are used twice, however with different goal poses. The execution of each behaviour is governed by a transition function defining it and its termination is governed by a respective terminal condition. An example of a transition function for pure pose control is presented in, e.g. [17]) and that of guarded motion or compliance control in, e.g. [18]).

6 Conclusions

Currently the FSA associated with each of the object-level relations has to be generated by a programmer. However, once the list of the behaviours invoked by such FSAs will be established the FSAs will be generated automatically, taking into account just the mathematical relations between the values of the attributes of the objects being the arguments of that relation. Each of the behaviours is defined in terms of a transition function and a terminal condition, so control over all the elements of the system are exerted directly through them. The task is defined in terms of object-level relations, so the sequence of those is associated with triggering into action respective FSAs. As subsequences of the mentioned sequence can be invoked conditionally a hierarchically higher-level FSA is produced, thus hierachic FSAs have to be used. Nevertheless, the structure of the higher-level FSA is in a one-to-one correspondence with the task definition in terms of object-level relations, so the transformation is straight forward here.

Acknowledgement. This project was financially supported by the National Centre for Research and Development grant no. PBS1/A3/8/2012.

References

1. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River (1995)
2. Fikes, R.E., Nilsson, N.J.: STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial intelligence* 2(3), 189–208 (1972)

3. Sankar, S., Hayes, R.: ADL – an interface definition language for specifying and testing software. *ACM Sigplan Notices* 29, 13–21 (1994)
4. McDermott, D., Ghallab, M., Howe, A., Knoblock, C., Ram, A., Veloso, M., Weld, D., Wilkins, D.: PDDL – the planning domain definition language. Technical Report CVC TR98-003/DCS TR-1165, Yale Center for Computational Vision and Control (1998)
5. Gerevini, A.E., Haslum, P., Long, D., Saetti, A., Dimopoulos, Y.: Deterministic planning in the fifth international planning competition: Pddl3 and experimental evaluation of the planners. *Artificial Intelligence* 173(5), 619–668 (2009)
6. Nau, D.S., Au, T.C., Ilghami, O., Kuter, U., Murdock, J.W., Wu, D., Yaman, F.: SHOP2: An HTN planning system. *J. Artif. Intell. Res. (JAIR)* 20, 379–404 (2003)
7. Sohrabi, S., Baier, J.A., McIlraith, S.A.: HTN planning with preferences. In: Twenty-First International Joint Conference on Artificial Intelligence (2009)
8. Wolfe, J., Marthi, B., Russell, S.J.: Combined task and motion planning for mobile manipulation. In: ICAPS, pp. 254–258 (2010)
9. Beetz, M., Mosenlechner, L., Tenorth, M.: CRAM – a cognitive robot abstract machine for everyday manipulation in human environments. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1012–1017. IEEE (2010)
10. Nilsson, N.J.: A hierarchical robot planning and execution system. Technical report, Stanford Research Institute, Artificial Intelligence Center (1973)
11. Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., Berger, E., Wheeler, R., Ng, A.: ROS: an open-source Robot Operating System. In: Proceedings of the Open-Source Software workshop at the International Conference on Robotics and Automation (ICRA) (2009)
12. Bohren, J., Cousins, S.: The SMACH high-level executive. *IEEE Robotics & Automation Magazine* 17(4), 18–20 (2010)
13. McGann, C., Py, F., Rajan, K., Thomas, H., Henthorn, R., McEwen, R.: T-REX: A model-based architecture for auv control. In: 3rd Workshop on Planning and Plan Execution for Real-World Systems, vol. 2007 (2007)
14. Kornuta, T., Zieliński, C.: Robot control system design exemplified by multi-camera visual servoing. *Journal of Intelligent & Robotic Systems*, 1–25 (2013)
15. Zieliński, C.: TORBOL: An object level robot programming language. *Mechatronics* 1(4), 469–485 (1991)
16. Zieliński, C.: Description of semantics of robot programming languages. *Mechatronics* 2(2), 171–198 (1992)
17. Zieliński, C., Kornuta, T.: Generation of linear Cartesian trajectories for robots using industrial motion-controllers. In: 16th IEEE International Conference on Methods and Models in Automation and Robotics, MMAR 2011, pp. 62–67 (August 2011)
18. Zieliński, C., Winiarski, T.: Motion generation in the MRRROC++ robot programming framework. *International Journal of Robotics Research* 29(4), 386–413 (2010)

Using Integrated Vision Systems: Three Gears and Leap Motion, to Control a 3-finger Dexterous Gripper

Igor Zubrycki and Grzegorz Granosik

Lodz University of Technology

{igor.zubrycki,granosik}@p.lodz.pl

<http://robotyka.p.lodz.pl/>

Abstract. In this paper we have tested two vision based technologies as possible control interfaces for dexterous 3-finger gripper. Both qualitative analysis and quantitative comparison with sensor glove are presented. We also provide some ready to use solutions to directly control movements of the gripper and support operator in difficult manipulation tasks by applying gestures.

Keywords: vision system, Leap Motion, Three Gear, ROS.

1 Motivation

Manipulators with dexterous grippers can manipulate objects in constrained environments in ways that would be otherwise impossible with only 6-DOF manipulator and parallel-jaw grippers [1]. As robots become more and more used in areas less structured than typical industrial environments, the expectation that robots can manipulate objects in more dexterous ways is growing. However, co-ordinated motion of dexterous gripper is not trivial and, in fact, the whole system can be considered as a group of several manipulators - fingers, working in the common space.

In tasks like telemanipulation, programming or teaching robot the gripping actions, there is a common need for an intuitive control of the gripper. As we have explained in previous work, we can use sensor glove to successfully operate grippers in an intuitive way [2], [3]. However, this is not ideal solution as sensor gloves have some important disadvantages: they need to be calibrated as readings drift with time and temperature, they wear mechanically and one hand of the user is always occupied by the device.

Recently, several solutions for tracking hand motion using 3D vision systems became available; they have advantages that can be favorable when controlling dexterous gripper:

- based only on the vision information, they provide function to track fingers poses without any mechanical devices attached to hand (even markers)
- they return data on position and orientation of hands in space

- systems provide some motion and gesture information that can be additionally used in controlling behavior of manipulator or gripper

Based on these features we have examined two easily available solutions: Leap Motion (LM) sensor and Three Gears (3G) system, shown in Fig. 1. The gripper we have used in test was the 3-finger dexterous hand (Schunk Dexterous Hand).



Fig. 1. Three Gears (left) system and Leap Motion (center) vision systems, SHD gripper (right)

2 Basic Features of Leap Motion and Three Gears Vision Systems, and Their Comparison

Both Three Gears system and Leap Motion are integrated vision systems, that means they are not conventional 3D scanners providing information in the form of 3D points cloud but they are systems designed specifically for hands tracking and providing additional information about hand movements and gestures.

These are commercially available products with their main market being PC users seeking gestural interfaces to control various applications. Where did they come from? With the arrival of iPhones and Android devices, gestural interfaces became available for large group of people, proving that this kind of control can be very intuitive and comfortable for users [4]. As the original solution requires a touchscreen to function, users of normal desktops started looking for some substitute. Additionally, the big success of the kinect sensor proved that there is also a large market for 3D gestural interfaces.

Both tested systems use IR cameras and computer vision to determine the 3D shape of observed object, but as integrated solutions designed for hand tracking they provide additional functionalities, namely:

- Hands position and orientation. Systems automatically determine which part of the 3D scene is a hand, extract it and give information (position and orientation) about the center of the palm.

- Position and orientation for every visible finger.
- Recognition of some set of gestures.
- Additional high level information (e.g.: tracking accuracy, hand scale).

Although basic functionality of both systems is similar they differ significantly in application, comparison a few important features is provided in Table 1.

Table 1. Comparison of Three Gears and Leap Motion systems

Three Gears	Leap Motion
	Construction
Uses 3D cameras developed by Microsoft and PrimeSense sensor. Sensor uses Light Coding technology, where built in laser projector, projects a pattern on whole scene which is observed by IR camera. Readings are then decoded and sent to a computer as a 3D point cloud data and RGB video. The device was created to track the pose of whole body and because of that focal point is set for an according distance. All devices designed by PrimeSense can be used but Three Gears recommends using a short range sensor Carmine 1.09 (range 0.35 - 1.4 m). 3D scanner has to be set up 0.7 m above workspace on a camera stand.	Was designed specifically for hands tracking and can be placed on the desk just next to the keyboard. It is small in size: 76 x 30 x 17 mm and lightweight: 45 g. Leap Motion uses two cameras and 3 IR LEDs to improve lighting conditions. All calculations necessary to transform stereovision images into points cloud and further to objects positions are performed on the host computer. This can result in a high machine load: about 20% processor usage for Intel Core i3, 2.4GHz laptop.
Positioning precision 0.01 mm stated in specs and 0.2 measured in tests.	Positioning precision 1 mm stated in specs (for PrimeSense Carmine 1.09)
Hand Tracking API quaternion of palm rotation and vector of translation.	vector of direction and vector of translation. Speed and direction of hand motion
Finger Tracking API each joint identified by name with quaternion of rotation and vector of position relative to the base	direction and translation vectors of visible fingers without names
Gestures API static gestures of the hand using one or two hands (pressed, dragged, released, moved, simultaneously pressed etc.)	gestures based on hands motion (swipe, keyboard tap, wall tap, circle)

Table 1. (*continued*)

Additional API	
Tracking accuracy, hands scale, calibration information	ability to track long, straight objects (giving their direction and translation vector), calculating the radius of a ball fitting inside the palm (when imitating or catching something).
Operating systems and Main binary server has to be on Mac/Windows computer, library availability in JAVA, C++ also server publishes information in text protocol described on website, so any language that can parse it can be used (each message ends with carriage return)	Languages availability Windows/ Mac, full API and airspace (App store for Leap Motion programs) availability, Linux: beta version for developers of libraries and driver. WebSocket Json server, but with libraries for python, java, c++, javascript.
Robustness	
Basic 3D cloud acquisition done directly by Kinect device. Laser projector provides good cloud acquisition everywhere but direct sunlight/ around bright lights. Hand position tracking is reliable when hand is at least few cm above ground. Finger pose tracking reliability is highly dependent on hands orientation and works well only for small angles relative to sensor plane when fingers do not self-occlude. Gestures are well classified	Based on IR stereo vision camera with artificially lighted scene. Two modes that are switched based on the intensity of ambient light - normal and robust. Worst tracking observed for near switching light conditions. Stereovision based on edges detection with poor results for smooth skin - like hands in gloves. Finger tracking poor for large angles, fingers that reappear on the scene have different position on the list - difficult to track.
Popularity and support	
133 persons on Twitter, we have email contact with developers	26 894 persons on Twitter. Active developers forum and phone customer support.

3 Gripper Control

We have been working with the SDH gripper for a few months now, preparing convenient ROS-based control architecture [2] and testing various grips [3]. With its 7 DOF this three finger hand is really dexterous providing a few default grasps and independent control for all joints. However, intuitive operation of this device is challenging. We have started our development with using sensor glove as the operators interface, now we are proposing vision systems as possible substitution or improvement because:

- it offers non-contact control, operators hand is not occupied permanently (which might be important for some tasks, e.g. SWAT operations)

- information of the position and orientation of the entire hand can be used to control manipulator while fingers pose can be used to control gripper
- no need to calibrate
- some hand gestures based on the movement or position of hands relative to each other can be detected and used

However, vision systems have their own limitations, mainly instability and self-occlusion of objects. The latter problem is important for a single view schemes - like in case of Three Gears and Leap Motion. Taking into account both advantages and limitations we have tested two possible scenarios of controlling the SDH gripper: direct and indirect that we are explaining in the following sections.

3.1 Direct Control

Integrated vision systems can be used for direct control of grippers. In this case information about the position and orientation of fingers is translated into particular pose of the gripper. If the SDH is considered the motion of 5 human fingers has to be transformed to the joint position of 7 DOFs. We have tested different ways of direct mapping, taking into account specific features of two devices, namely:

- 3G provides angle values for all phalanges and therefore we can calculate (using some constant ratio) angle positions for each joint of SDH. We can choose three particular human fingers to follow or more sophisticated mapping.
- LM provides information about position of the fingertip and its orientation and therefore, we have to solve inverse kinematics problem for the grippers finger to calculate appropriate joint angles.
- LM has also special function to track position and orientation of elongated objects (visible in the scene) - based on this feature we have built mock-up gripper (reproducing Willow Garage experiment) which position/orientation can be followed by grippers position/orientation. The most important limitations we have observed are: occlusion, erroneous pose detection and API malfunctions.

3.2 Indirect Control - Using Gestures and Higher Level Information for Gripper Control

Recent success of devices such as tablets or smartphones proves that gestural control can be intuitive and ergonomic for users. Previous works in robot control using gestures focused on important task of gesture recognition from flat 2D pictures as well as providing a set of robust gestures that could be used in many environments [5], [6].

Integrated vision systems use 3D data in gesture recognition, and provide additional information about speed, direction and position of hands or fingers when the gesture was done. This can give users additional control, without enlarging the set of gestures necessary to control the gripper [7].

Both LM and 3G systems can recognize a number of gestures, although these are gestures of different type. Leap Motion recognizes gestures based on hand

movement and finger movement, similarly to smartphones or tablets. API can process such motions as: swipe, rotate, wall tap (finger moving in direction of imagined wall), keyboard tap (finger moving up or down). For each gesture programmer has also additional information: about speed of motion, its direction and other data specific to gesture. There are also libraries for learning new gestures basing on a frame motion or a hand pose. Leap Trainer provides tools for various learning algorithms e.g. Geometric Template Matching or Neural Networks.

For the Three Gears system, gestures are similar to those done with an ordinary mouse, but using motion of the whole hand (as opposed to finger movement recognized by Leap Motion). System differentiates: pinch gesture, drag, release and pinch with both hands. Provided in the same time are information about positions of both hands, which enable implementing a bigger spectrum of gestures, like rotate with two hands instead of two fingers.

Control Modes Using Gestures. We can propose three approaches to utilize gestures:

- Discrete type gestures with force control: in this method user can choose different grip type using gestures. Each grip type provides different finger orientation and different expected forces. The gripping action ends when expected force structure is acquired or a force threshold is exceeded.
- Continuous movement gestures: in this mode, gesture controls the degree of gripping. Recognized gesture is translated into continuous movement of gripper, that ends with the end of gesture.
- Discrete type gestures that have a helper function in the control system: i.e. they can change an operation mode of the gripper.

The advantages of using gestures are of several types: they can be implemented without the need of calibration, recognizing hand gestures and tracking hand is much easier than tracking particular finger movement so it can be done robustly.

However, gestural interface for gripper has one important limitation: the user need to remember functions of all movements. In case of direct control hand motion translates directly into gripper motion. Therefore, the user has a natural method to learn the control just by observing the movement, and for well-tuned parameters interaction can be fully intuitive.

In case of gestural control, the user has to remember entire mapping as it is based on symbolic and not direct connection [8]. Also, gestures being recognized by a vision system are done in the air, without screen that can provide additional information as in case of tablets or smartphones.

As gestural control is new field there is ongoing effort to create a universal vocabulary of gestures that could be used in human-machine interaction. In case of interaction with robots, it is not yet standardized, and we can see many different gestures used by researchers for even simple commands like stop [5], [9]. Fortunately, there are some procedures for designing gestural interfaces [10] and also some universal principles of designing any kind of human machine interface

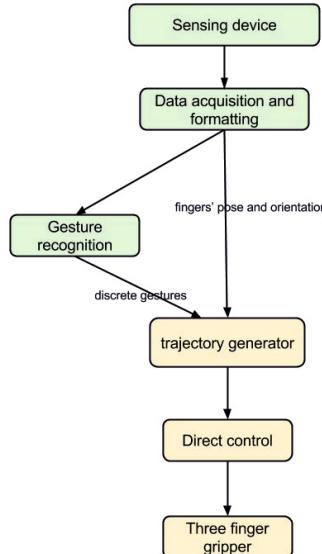


Fig. 2. Control structure based on ROS

such as visibility, feedback, consistency, non-destructive operations, scalability and reliability [17].

As gripping task requires full attention from the operator, we suggest the following four rules that make feasible controlling of the gripper using gestures:

1. use relatively small number of gestures to make the recall task easier, especially in stressful situations [10], [11],
2. give feedback when gesture is recognized as well as the command is successful, this will help the operator whose focus will be on the task performed by the gripper - we recommend using signals different than visual such as: haptic (e.g. vibrations) or audio [6], [12], [13],
3. provide user with simulation environment or training mode of operation,
4. use only gestures that are recognized robustly. This means for example setting wide range for gesture parameters (e.g. swipe can be performed with various speeds). Use other feature to avoid erroneous recognition, e.g. position of gesture can be used to discriminate intentional or unintentional action as it is very well sensed by both vision systems.

4 Control System

For the purpose to control a 3-finger dexterous gripper and test both vision systems we have created a control structure based on the Robot Operating System (ROS), as shown in Fig. 2. In this system we are using universal pose or gesture



Fig. 3. Operation scheme of the Three Gears system

commands that could be generated by either vision systems, sensors glove or combination of these. We have tested the efficiency, robustness and usability of different solutions in controlling 3-finger gripper.

Data coming from a sensing device, such as 3D scanner or sensors glove is formatted and sent, as hands pose and gestural commands to the trajectory generator node. Gesture commands can change state of device and hand pose can be used for direct control of the gripper. Direct control node realizes the reference trajectory (produced by previous node) by sending velocity commands to the gripper, with position feedback directly from SDH hand. It also transfers to the system all information about the current state of the gripper.

To prepare this controller in the ROS we had to write a few new and to modify some ROS packages, i.e. ROS HandKinect package. Some implementation problems were associated with the limitation of the Three Gears system only to Windows or Apple platforms. Therefore, this equipment is running on the separate Windows machine as a TCP/IP server that publishes a text based protocol. On the Linux platform with ROS middleware we have created special wrapper package to receive and transform these data, as schematically shown in Fig. 3. Our package converts data stream into ROS topics describing flexion angles and recognized gestures, as well as converts rotation quaternions and position vectors described in absolute coordinate system into a series of transform frames that could be used for forward kinematics and visualization, as presented in Fig. 4

Application of the Leap Motion system in ROS required modification of the Rosleapmotion package. It contains a series of python nodes that read the LM protocol and translate it to ROS topics. For our purpose it was necessary to update the code for the newest Leap Motion API and expand it by additional functions providing gestural data info.

5 Practical Applications and Tests

We have tested two integrated vision systems in the following modes:

- direct control of the SDH gripper
- calibration of the sensor glove

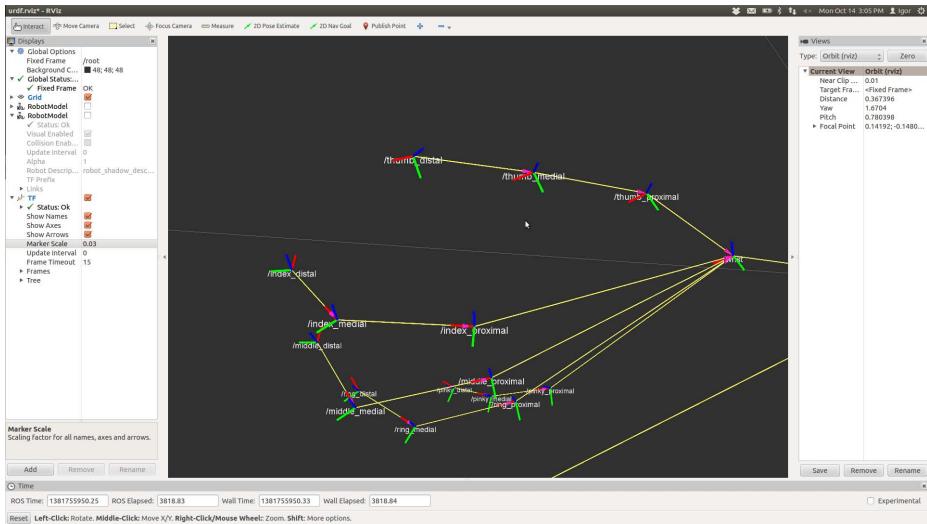


Fig. 4. Visualization of the hands posture in Rviz based on the transformation frames

- hybrid system composed of the direct control of the SDH gripper and discrete switching of states by gestures recognized by the Three Gears system

Direct Gripper Control. In the direct control mode, joint angle data was translated into the gripper movement. As noted in the Section 3.1 fingers pose recognition is very unstable and dependent on the orientation of the whole hand (possibility of the self-occlusion), and therefore sometimes gripper movements were chaotic.

Sensor Glove Calibration. In our previous research [2] we have used sensors glove as a sensor device. It has many attractive features such as no self-occlusion or no dependence on lightning conditions. However, flexion sensors used in it change their parameters in time, as well as they wear out mechanically. Resistance of these sensors can vary over 30% between batches and therefore sensors glove requires calibration.

The Three Gears system with HandKinect node provides information about flexion of each joint of the finger, the same values can be measured by the flex sensor mounted in our sensors glove we have collected these data and showed in Fig. 5. Although there is a large number of outliers, there is also a linear correlation between these readings. Relation was calculated using a least squares method. The best fit was obtained when each finger was flexing slowly and separately. This minimizes an error for vision system as well as gives enough data for robust estimation.

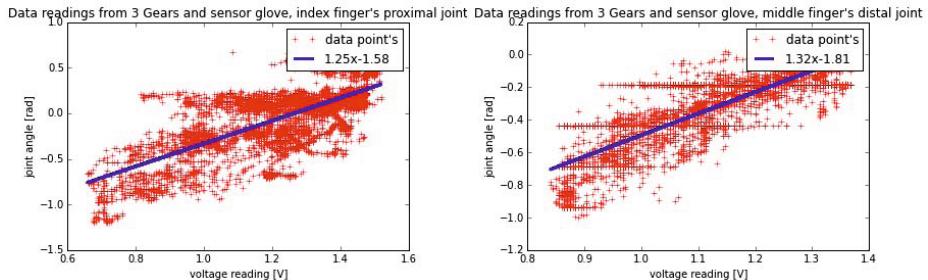


Fig. 5. Data readings from 3 Gears and sensor glove for index finger's proximal joint and middle finger's distal joint

Hybrid System. The biggest problem with using sensor glove to control a gripper is the fact that one hand is occupied all the time. Its pose is translated into the gripper movement, therefore, this hand cannot be used to other actions like pushing buttons or gestures. Moreover, when the operator is precisely controlling some manipulation task he/she should have eyes always focused on target action and as a consequence using the keyboard or control buttons has to be avoided. Interesting solution for this problem might be the use of gestures. They can be recognized by a vision system, they can be done in the air, almost anywhere on the scene, without touching anything, without finding accurate spot, and even without watching.

We have tested the following scenario of controlling the gripper with the sensors glove on one hand, and with special actions provided by gestures made with a second hand. Gestures recognized by the vision system can switch between three states:

1. sensor glove controls gripper directly,
2. sensors glove controls model (so called shadow hand [2]) and movement of real gripper is confirmed,
3. stop gripper in the current pose.

Tested vision systems provide recognition of different gestures, therefore, we have used two sets for the same switching commands, as described in Table 2.

Table 2. Set of gestures to activate different operation modes

Action /Gestures	Three Gears	Leap Motion
Activate gripper control	Single pinch	Keyboard tap
Activate model control	Double pinch	Circle
Stop motion	Single pinch again	Horizontal swipe



Fig. 6. Students testing our control system with 3D vision

We have tested the proposed system with several students seeing it for the first time (see Fig. 6). After only single description of the task and control methods they succeeded to manipulate the gripper and use all gestures. We have observed that users needed some kind of feedback about the gesture interpretation and confirmation of an action see our postulates of successful control with gestures collected in Section 3.2. This is particularly important when gestures are removed from the device [12] or the device is changed to another one with different set of gestures. In our system we have used sounds generated from ROS node with pygames program.

6 Conclusion

In this paper we have described two commercially available integrated vision systems, both giving accurate information about hands position and orientation. We have analyzed their potential role in controlling the dexterous gripper and proposed adequate control system architecture. We are also providing some experimental results with direct and indirect control using Three Gears and Leap Motion.

Acknowledgment. Research was partially supported by the National Centre for Research and Development under grant No. PBS1/A3/8/2012.

References

1. Onda, H., Suehiro, T.: Motion planning and design of a dexterous gripper-graspability, manipulability, and virtual gripper. In: Proc. IEEE/RSJ Int. Conference on. Intelligent Robots and Systems, vol. 1 (1998)
2. Zubrycki, I., Granosik, G.: Test setup for multi-finger gripper control based on robot operating system (ROS). In: Proc. of 9th Int. Workshop on Robot Motion and Control, Wsowo, Poland, July 3-5, pp. 135–140 (2013)

3. Zubrycki, I., Granosik, G.: Grip recognition and control of 3-finger gripper with sensor glove. In: Proc. of Int. Conference on Robotics and Artificial Intelligence. Problems and Perspective (RAIPAP 2013), Brest, Belarus, November 4-6 (2013) (to be published)
4. Lomas, N.G.: 1.2 Billion Smartphones, Tablets To Be Bought Worldwid. In: 2013; 821 Million This Year: 70
5. Pietrasik, M., Arychta, D.: Multimedial methods for control of robots. In: Postpy Robotyki (part 1), Warszawa, pp. 331–338 (2006) (in Polish)
6. Zhai, S., et al.: Foundational Issues in Touch-Screen Stroke Gesture Design-An Integrative Review. Foundations and Trends in Human-Computer Interaction 5(2), 97–205 (2012)
7. Dhawale, P., Masoodian, M., Rogers, B.: Bare-hand 3D gesture input to interactive systems. In: Proc. of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction: Design Centered HCI, pp. 25–32 (July 6, 2006)
8. Norman, D.A., Nielsen, J.: Gestural interfaces: a step backward in usability. Interactions 17(5), 46–49 (2010)
9. Ruttum, M., Parikh, S.P.: Can robots recognize common Marine gestures? In: IEEE 42nd Southeastern Symposium on System Theory (SSST) (2010)
10. Nielsen, M., et al.: A procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction. In: Proc. of the 5th International Gesture Workshop, pp. 1–12 (March 2003)
11. Nacenta, M.A., Kamber, Y., Qiang, Y., Kristensson, P.O.: Memorability of Pre-designed and User-defined Gesture Sets. In: Proc.s of the 31st Annual ACM SIGCHI Conference on Human factors in Computing Systems, CHI 2013 (2013)
12. Wigdor, D., Wixon, D.: Brave NUI world: designing natural user interfaces for touch and gesture, pp. 81–95. Elsevier (2011)
13. Williamson, J., Murray-Smith, R.: Audio feedback for gesture recognition. Technical Report TR-2002-127, Dept. Computing Science, University of Glasgow (2002)
14. Weichert, F., et al.: Analysis of the Accuracy and Robustness of the Leap Motion Controller. Sensors 13(5), 6380–6393 (2013)
15. <https://forums.leapmotion.com/forum/general-discussion/general-discussion-forum/130-processing-where-it-is-computed>
16. <https://www.sparkfun.com/datasheets/Sensors/Flex/flex22.pdf>
17. Norman, D.A.: Natural user interfaces are not natural. Interactions 17(3), 6–10 (2010)
18. Hasanuzzaman, M., et al.: Real-time vision-based gesture recognition for human robot interaction. In: IEEE International Conference on Robotics and Biomimetics, ROBIO 2004, pp. 413–418 (August 22, 2004)

Part III

Measuring Techniques and Systems

Influence of the Environment on Operation of Checkweigher in Industrial Conditions

Piotr Bazydło¹, Michał Urbański¹, Marcin Kamiński¹, and Roman Szewczyk²

¹ Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland

{pbazydlo, murbanski, mkaminski}@piap.pl

² Institute of Metrology and Biomedical Engineering,

Warsaw University of Technology, Warsaw, Poland

szewczyk@mchtr.pw.edu.pl

Abstract. This paper presents impact of selected environmental factors on operation of specified checkweigher created by the one of worldwide leaders in the field of dynamic weighting RADWAG. The research was carried out in three different industrial facilities, in order to ensure that dependence between different environments and measurements. The analysis of the results allowed to elements of the entire disruption, which were described in the material. In addition, the paper contains guideline for effective measurement of environmental conditions.

Keywords: checkweigher, dynamic weighting, electronic weighting module, industrial conditions.

1 Introduction

The increasing requirements and standards imposed on industry results in growing interest in the field of automatic measurement systems [1]. That trend leads to steadily increased activity in the researches aimed at different parts of production processes, for example dynamic weighting [2]. Checkweighers appear to be devices, which possibilities and properties have great influence on reliability of production line. Weighting objects during their movement, can be usually spotted at the end of line, especially in food, pharmacy or chemical industry in order to determine if weight of the product is suitable within given range. However, weighing during the action of conveyor belt can be easily disrupted by the number of external and internal factors depending on construction of checkweigher, what turns out to be major drawback of automatic mass measurement. That problem sets handicap on implementation of industrial application in large factories, where checkweigher is exposed to harsh conditions.

The purpose of this paper is to collect data about environment in industrial conditions and connect them with operation of electronic weighting module, in order to specify the range of parameters, in which voltage signal is holding a specific accuracy. It will help to fill the gap, which results from the current concern about internal disruptions of the system, such as torque ripple of the brushless direct-current motors.

As a result, it may be an indicator of acceptable conditions for automatic mass measurement based on specified checkweigher and indicator of occurring unacceptable measurements deviation.

2 Researched Checkweigher and Developed Measurement System for Studying Environment Conditions

Dynamic weight measurement is performed by electronic weighting module connected with short length conveyor belt. This module is fixed with construction made from stainless steel. Described checkweigher is supposed to move belts with the maximum speed of 1.5 meters per second and provide measurements with accuracy of 0.1 gram for weights from 1 gram to 750 gram.

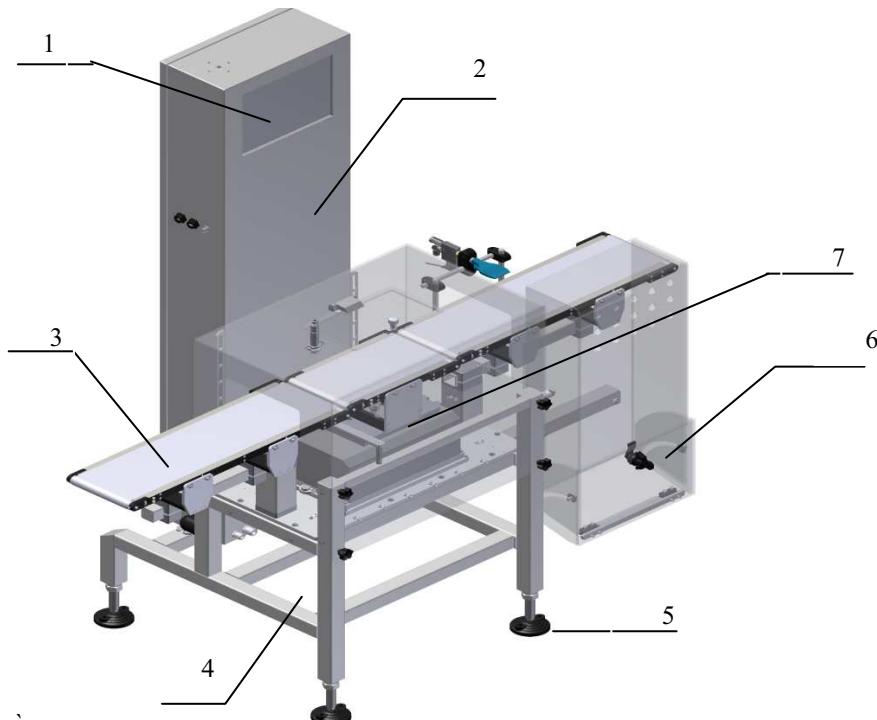


Fig. 1. Construction of researched checkweigher: 1 – HMI screen, 2 – control cabinet, 3 – conveyor belt, 4 – construction, 5 – suppressing elements, 6 – container, 7 – weighting module

Due to the presence of the mass, weighting pan changes position of the coil what increases level of voltage signal, which is then processed to obtain desired measurement. The more extreme conditions, the harder it is to filter out disruptions for dynamically changed measuring signal.

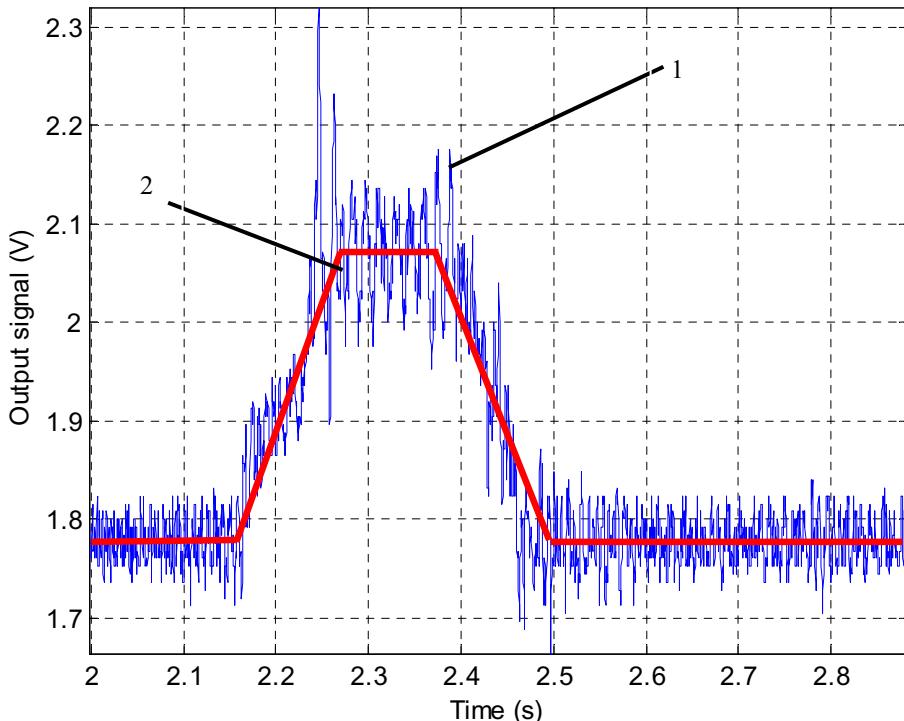


Fig. 2. Signal during measurement of 785gram with speed of 1.5 m/s before signal processing:
1 – read signal, 2 – theoretical measurement of mass

In the case of electronic weighting module, sources of external disruptions may be: temperature, humidity, vibrations, electrostatic charges, air movements, constant and alternating magnetic field [2]. According to that, set of measurement instruments has been developed and used for the investigation concerning their effect on voltage signal, which indicates weight.

The measuring system assembled for the task consist of five separable subsystems, as is presented in Figure 3. For temperature and humidity measurement the SENSIRON SHT71 [3] sensor was used. This sensor works under control of NXP LPC1115 [4] ARM microcontroller, with I²C – similar interface. The accuracy of SHT71 for temperature and humidity measurement is $\pm 0.4^{\circ}\text{C}$ and $\pm 3\%\text{RH}$ respectively, for supply voltage of 3.3 V. Measurement results are saved to a text file located on a SD Card.

Communication between LPC1115 microcontroller and SD Card is provided by a SPI interface [5]. The temperature and humidity measuring system is powered by two 9V batteries, which give an operating time of up to 14 hours. The measurements are taken every second, which equals to a sampling rate of 1 Sample/s. Results are then processed using a PC class computer.

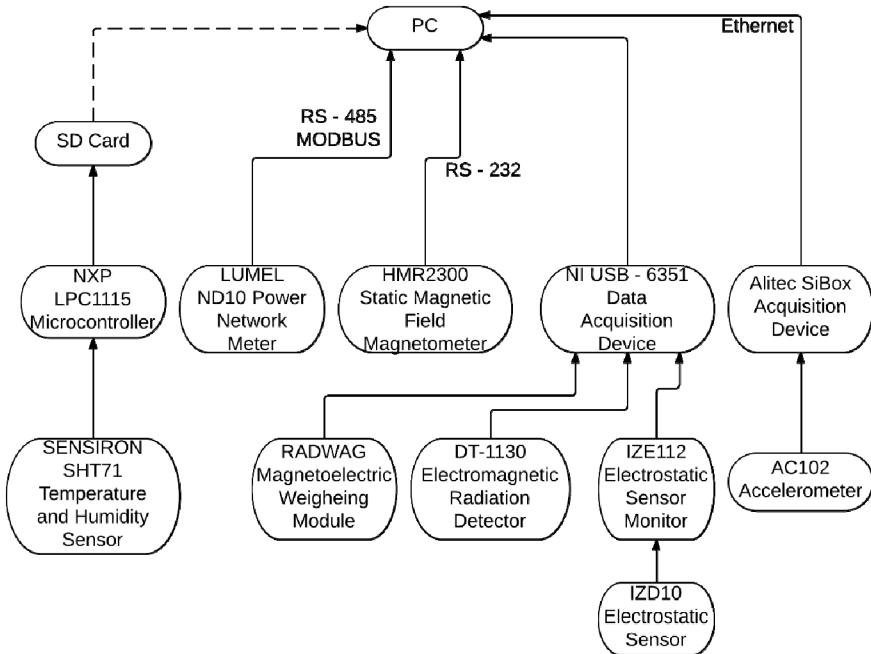


Fig. 3. Schematic block diagram of given measurement setup

For power network parameters a LUMEL ND10 Power Network Meter [6] was used. Measured parameters of a power supply line were: Voltage (L-N), Frequency, Total Harmonic Distortion (THD). Accuracy of LUMEL ND10 PWM is presented in Table 1.

Table 1. Accuracy of Measurement for LUMEL ND10

Parameter	Basic Error
Voltage (L-N)	$\pm 0.2\%$ mV
Frequency	$\pm 0.2\%$ mV
Total Harmonic Distortion (THD)	$\pm 5\%$

Results of measurements were read from device's internal Data Registers by a dedicated LabVIEW application. Communication was arranged over a RS-485 interface, using MODBUS protocol. Sampling rate for this measurement was 1 Sample/s, and the values were saved to a text file.

HMR2300 Magnetometer [7] was used for a measurement of a strength and direction of a static magnetic field. Magnetometer's three digital outputs indicated X, Y and Z vector components of a magnetic field. The values were read by a LabVIEW application over a RS-232 interface. Sampling rate for a HMR2300 measurements was set to 10 Samples/s. Resulting values presented in gauss unit (Gs) were saved to a text file located in a LabVIEW application directory.

Measurements of alternating magnetic field and electrostatic charges, as well as the voltage signal from weighting module were collected using National Instrument USB-6351 Data Acquisition Device set up to work under LabVIEW. For electrostatic charges an IZD10 Electrostatic Sensor [8] with IZE112 Electrostatic Sensor Monitor was used. Electrostatic sensor was placed over weighting module conveyor belt, with a 50 mm clearance, which allowed it to operate at ± 20 kV range. Sampling rate for electrostatic charge measurement was set to 1 samples/s.

Measurement of alternating magnetic field was conducted with a DT-1130 Electromagnetic Radiation Detector. For the purpose of given task voltage signal was isolated from the device and then processed to set measuring rate to $0 \div 100$ nT within 50 Hz – 2000 MHz frequency range.

Voltage signal from electromagnetic weighting module was collected as a reference signal, indicating module's responses to external disruptions. These, if processed, results in direct weight indication. Sampling rate for this measurement was set to 800 Samples/s, to ensure, that each signal deviation would be noticeable.

Each of those signals were collected with a NI USB – 6351 Data Acquisition Device, using separate channels, with respect to ground (RSE Mode). Measurements were saved to separate text files, located in a LabVIEW application directory.

For vibrations measurement up to three AC102 Accelerometers were used, with a sensitivity of 100 mV/g, an ± 50 g range (peak). Signal collection was provided by a alitec® SiBox Data Acquisition Device communicating with PC class computer with an Ethernet connection. Data was processed using alitec® ATC_Daq software and MATLAB dedicated drivers for data acquisition, which resulted with acceleration and vibration spectrum measurement. The acceleration was measured with a Sampling rate of 8192 samples/s.

All measurement data was processed and presented using MATLAB Environment for numerical computation and visualization. The summary of sampling frequencies for each measuring subsystem is presented in Table 2.

Table 2. Summary of sampling rates for measurement

Measurement	Sampling Rate
Temperature and Humidity	1 Sample/s
Power Network Parameters	1 Sample/s
Static Magnetic Field	10 Samples/s
Alternating Magnetic Field	1 Sample/s
Electrostatic Charge	1 Sample/s
Weighting Module Voltage Signal	800 Samples/s
Vibrations	8192 Samples/s

3 Influence of Selected Conditions on Dynamic Weighting.

In order to keep track of changes occurring during exposure on harsh conditions, the linear approximation (1) of sensing characteristic of checkweigher has been developed:

$$m = \frac{U_o[V] - U_m[V]}{0.00035} [\text{gram}] \quad (1)$$

where m – measured mass, U_o – voltage output selected for measurement, U_m – mean voltage of unloaded checkweigher.

Modules reserved for dynamic weighting, work with low resolution of 0,1 gram. According to that, most of the factors which disrupt measurement with higher accuracy, such as 1 μ gram for static weighting, have no significant influence on operation of checkweigher (assuming that device has been built in a proper way). That leads to narrowing sources of unacceptable deviations to vibrations and air movements. All voltage signals coming from the checkweigher has been measured before signal processing, because of searching for raw sources of disruptions.

3.1 Air Movements

The main sources of air movements disrupting operation of electronic weighting module are: movements caused by the presence of human and heating, ventilation, and air conditioning systems (HVAC).

To illustrate the correlation between measurement and disruptions, the source of constant air flow has been used. It has been set in front of checkweigher in distance of 0.5m with working speed of 1.8 m/s.

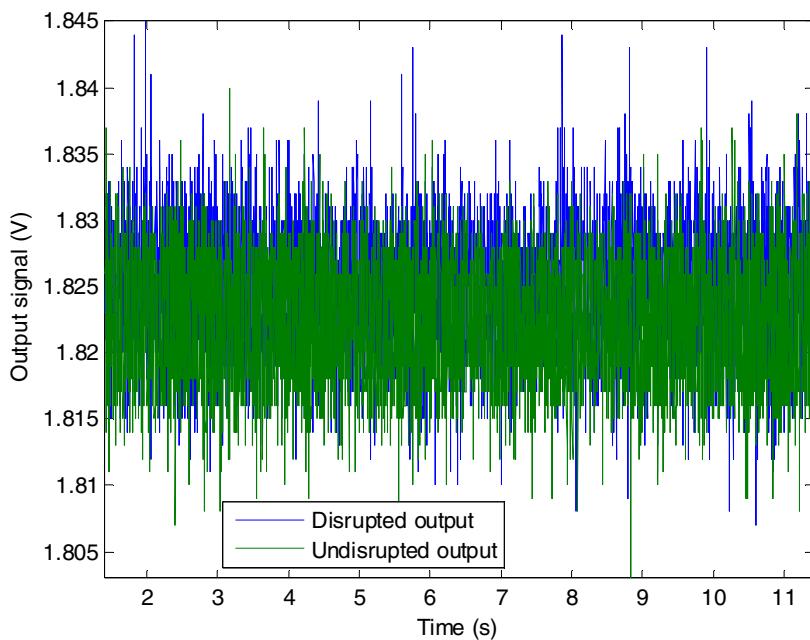


Fig. 4. Influence of 1,8 m/s air movement on output signal

According to figure 4, signal for unloaded device is 1.822 V, while for noised measurement that value equals 1.825 V. That difference causes almost 9 gram of deviation.

Table 3. Influence of specified air speed (in front of device) and their influence on value of output voltage

Module signal	Air Movement Velocity			
	0 m/s	1.2 m/s	1.4 m/s	1.8 m/s
Max voltage (V)	1.840	1.842	1.842	1.845
Avg voltage (V)	1.822	1.824	1.825	1.825
Min voltage (V)	1.803	1.806	1.808	1.807

Such values as shown in Table 3. are possible to reach even by the usual acts of human near the application. Characteristic of the changes will not be constant but it can lead to the increased deviation of single measurement.

3.2 Influence of Vibrations

The main sources of vibrations are: working machines, human interference and unforeseen accidents such as objects hitting the floor [9]. Depending on the construction of checkweigher and its ability to suppress vibrations in different frequencies, the

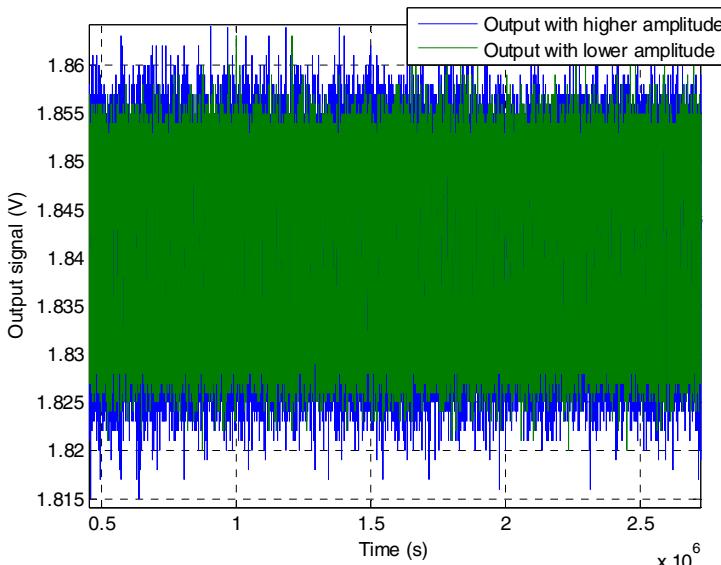


Fig. 5. Voltage output in two different cases of constant vibrations

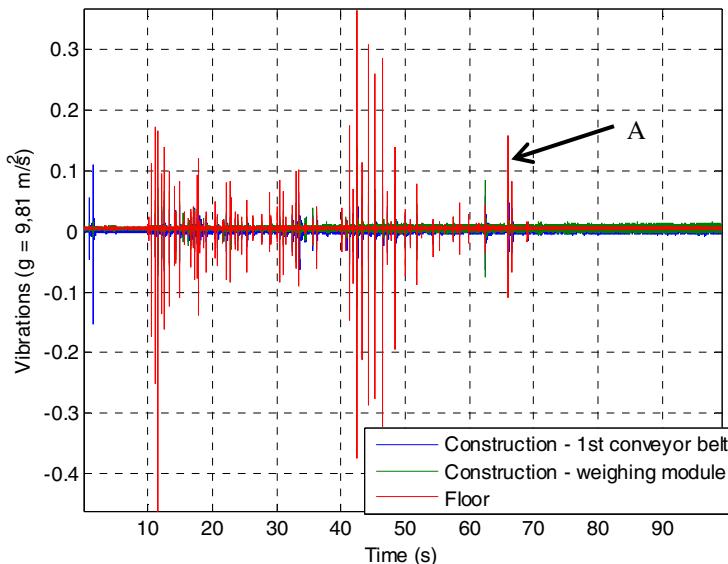


Fig. 6. Vibrations in three points during increased human activity. A marks researched disruption

correlation between measurement and specified disruption will be different. The tested case of the device is resistant to constant vibrations reaching 0.02 g, with dominating frequencies of 50, 183 and 512 Hz. Value of the 0.02 g was the biggest constant vibration measured in three different industrial environments.

Studies have shown, that different values of constant vibrations has changed the amplitude of output, while mean output voltage stays at the same level (Fig. 5). In that case, implemented adaptive filter is able to process the signal in order to select mean value of the output sufficiently for the accuracy of 0.1 gram.

The most problematic factors are single vibrations induced mainly by the presence of human (for example walking or dropping objects). Depending on the structure of the floor, walking man can cause vibrations reaching even 0.35 g.

The set of two graphs (Figure 6 and 7), along with close-up of marked sections presents correlation between output voltage and vibrations caused by walking people. The difference between selected pairs of output/disruptions results from frequency of vibrations and positioning of the accelerometer. In reference to this research, walking human can provide deviation of 0.25 V. After signal processing, such a single pulse can cause deviation reaching 15 grams. It can be serious problem when checkweigher is located near the routine path of the employees.

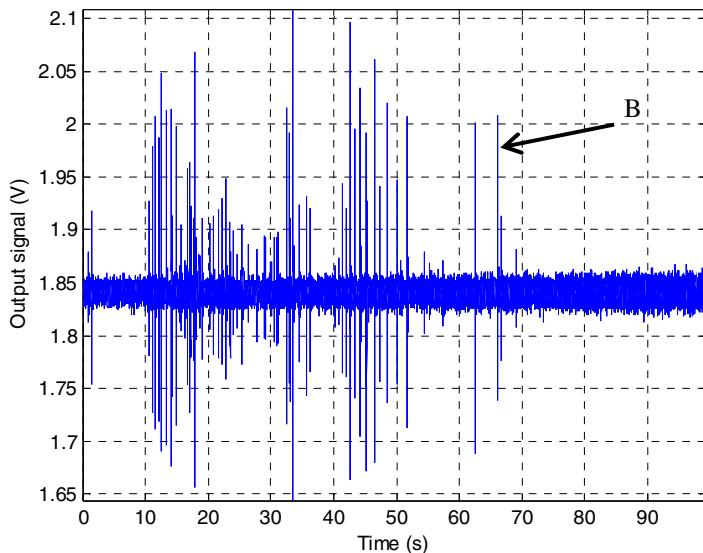


Fig. 7. Output voltage during increased human activity. **B** marks researched response to disruption

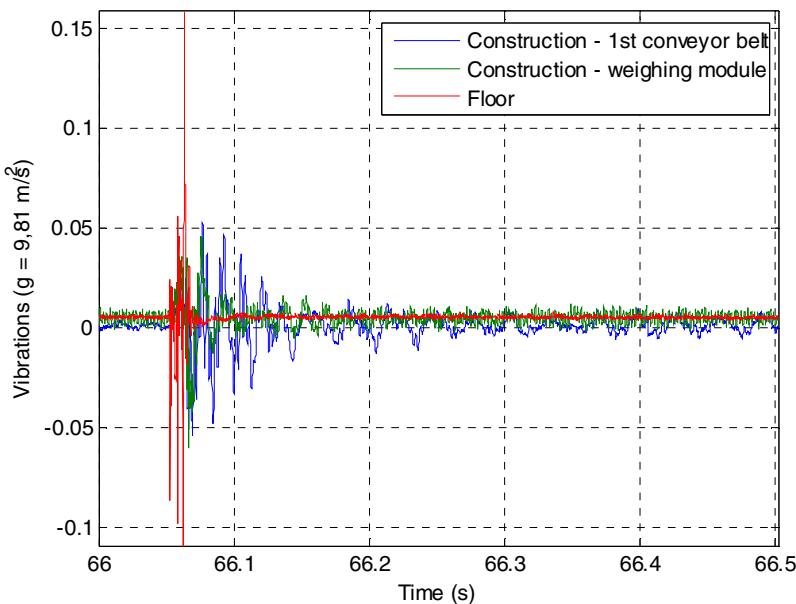


Fig. 8. Close-up of section A marked on Figure 6

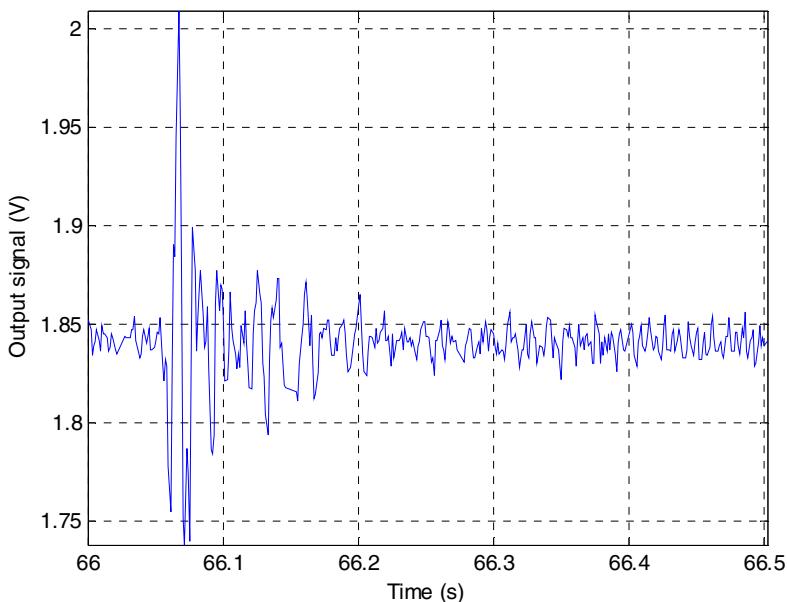


Fig. 9. Close-up of section B marked on Figure 7

According to the Fig. 9. voltage signal of the checkweigher needs almost 0.2 second to stabilise after a single vibration reaching 0.16g. This is partly caused by the stabilisation time of construction. In relation to figure 8., conveyor belt needs more time to suppress vibration than the part with module fixed to it. Assuming that it is possible to weigh five objects during a second, that case will probably disrupt only one measurement. The longest noticed time of output stabilisation for shock during the research was 0,25 seconds for 0,28g.

4 Summary

Results of the measurement have shown a correlation between air movements and vibrations, and checkweigher's voltage signal. Therefore those disruptions might have a significant influence on checkweigher's weight indication. Presented research sets base for further analysis of disruption sources, disruptions itself, and checkweigher's response to given disruptions. These analysis could provide deeper knowledge of checkweigher's signal processing. With that knowledge a set of improvements can be provided for checkweigher solution, in order to filter or prevent those disruptions from having an influence on weight indication.

We would like to thank RADWAG Balances and Scales Company for providing us with prototype Checkweigher solution, their cooperation and knowledge exchange on the field of dynamic weighing.

Acknowledgment. This work was partially supported by The National Centre of Research and Development (Poland).

References

1. Yam, K.L.: Encyclopedia of Packaging Technology. John Wiley & Sons (2009) ISBN 978-0-470-08704-6
2. Szmyrka-Grzebyk, A., Merlone, A., Falkiewicz, K., Grudniewicz, E., Migala, K.: Metrology for Pressure, Temperature, Humidity and Airspeed in the Atmosphere. Journal of Automation, Mobile Robotics & Intelligent Systems 6(3), 56–60 (2012)
3. SENSIRON SHT7x Datasheet, http://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/Humidity/Sensirion_Humidity_SHT7x_Datasheet_V5.pdf
4. NXP LPC111x Product Datasheet,
http://www.nxp.com/documents/data_sheet/LPC111X.pdf
5. Kozien, K., Mysinski, W.: Implementacja sprzętowa protokołu PS/2 w układzie FPGA XILINX Spartan III XC3S1000. Pomiary Automatyka Robotyka (3), 81–83 (2010) (in Polish)
6. LUMEL ND10 Power Network Meter Datasheet,
http://www.lumel.com.pl/download/ZZZ4L2x1bWVsL2VuL2R1ZmF1bHRfbXVsG1saXN0YV9wbG1rb3cudjAvNDk1/nd1019__23.07.12.pdf
7. HMR 2300 Magnetometer Datasheet,
<http://www51.honeywell.com/aero/common/documents/myaerospacecatalog-documents/Missiles-Munitions/HMR2300.pdf>
8. IZD10 Electrostatic Sensor Datasheet, http://www.smc.co.nz/files/products/Accessories1273124228_1.pdf
9. Goliński, J.: Wibroizolacja maszyn i urządzeń, Warszawa. Wydawnictwo Naukowo-Techniczne, pp. 83–204 (1979) (in Polish) ISBN 83-204-0067-8

Method for Limitation of Disturbances in Measurement Data in 3D Laser Profilometry

Piotr Czajka¹, Wojciech Mizak¹, Jacek Galas², Adam Czyżewski²,
Maciej Kochanowski², Dariusz Litwin², and Maciej Socjusz²

¹ Institute for Sustainable Technologies – National Research Institute
6/10 Pulaskiego str., Radom, Poland
piotr.czajka@itee.radom.pl

² Institute of Applied Optics, 18 Kamionkowska str., Warsaw, Poland

Abstract. The paper presents a 2D laser profilometer with the triangulation head. The system has been equipped with a robust algorithm for effective elimination of measurement artefacts typical for triangulation systems. The approach consists in multiple scanning of the object surface for various angular positions, followed by a process of fusion of these data images. The location of the rotation axis of the object has been calculated with the use of the specially developed calibration gauge which is crucial for a precise data assembling procedure. The paper covers the sequence of data processing, the design of the profilometer as a whole, system control and a collection of algorithms verified experimentally.

Keywords: optical profilometry, 2D laser triangulation technique, shadow effect, surface geometry reconstruction.

1 Introduction

The technique that utilizes the laser triangulation sensor is one of the most popular approaches to contactless distance measurements [1, 2]. Because of a relatively small measurement spot (typically at the level of a few μm) [3], which results in high selectivity, the method can be applied to surface geometry reconstruction (optical profilometry). The measurement head (in its standard 1D version) utilizes a point-based scanning method. The most important limitation of the method is a long time of a surface profile measurement. In order to accelerate the scanning process, the 2D laser triangulation is used where a full surface profile line is measured instead of a sequence of points. Thanks to this approach a fast 2D surface scanning is possible. Although the method is much faster, in surface topography measurements it suffers from large areas of shadows, where collecting reliable data is not possible. The disturbances may affect both detection and laser beam projection processes. In the first case a detector "cannot see" the illuminated area, whereas in the second case the illuminating beam does not reach the region of interest [4, 5]. In both cases the measurements are not possible. In this paper we propose a method, which diminishes the influence of the above-mentioned disturbances. The technique consists in multiple scanning of

the object for some selected angular orientations and followed an intelligent fusion of collected images, simultaneously rejecting "bad" regions.

2 2D Laser Triangulation Technique

In the 2D laser triangulation technique a light line is projected on the surface being measured. The light reflected from the object surface is imaged at the CCD or CMOS photodetector array by means of a specially designed optical system at an angle to the laser beam (Fig. 1). The distance change between the head and the surface results in a signal shift on the sensor. The distance between the surface and the head is calculated by taking into account the position of the incoming signal at the sensor and the angle between the axis of the detector and the laser beam and distance between them [2].

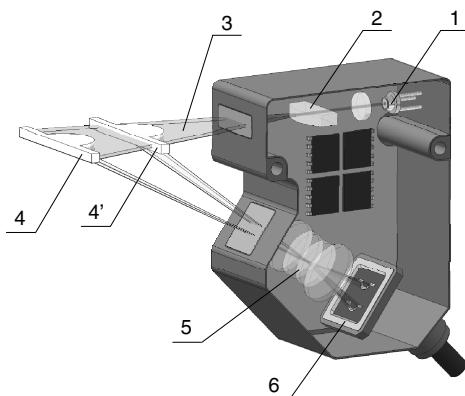


Fig. 1. The principle of the measurement using the 2D laser triangulation technique: 1- laser diode, 2- set of cylindrical lenses, 3- laser beam, 4- object surface at two positions, 5- lens system, 6- CCD or CMOS detector

During a line scanning process the system collects a set of points, that corresponds to the light line projected on the object surface [5].

One of the important advantages of the 2D triangulation process is the significant increase in scanning speed compared to point-based approaches. Commercially available heads offers the resolution at the level of 600 to 1300 points [3, 6, 7].

3 Sources of Disturbances in the Triangulation Methodology

When the laser triangulation technique is used there are several disturbances resulted mainly from the optical nature of the approach and the surface characteristic of the object [8, 9]. Among them there are:

- the laser spot overshadow due to a complex shape of the object;
- limitation of the reflected line illumination due to the incidence angle of the laser beam;

- uncertainty of the measurement at the edges of the object and in the area close to them;
- disturbance of the measurements resulted from the surface structure;
- disturbance caused by the reflective coefficient of the surface;
- disturbances caused by other optical phenomena like absorption and diffusion;
- laser spot blurring introduced by the motion of the setup.

The shadow and edge effects are the most troublesome phenomena. The former manifest themselves as large areas not illuminated by a laser beam. This occurs because complex walls and edges of the object overshadow inner surfaces. Also a light beam illuminating edges is a potential source of reflections and interreflections, which results in difficult to predict errors in the detector readouts [5].

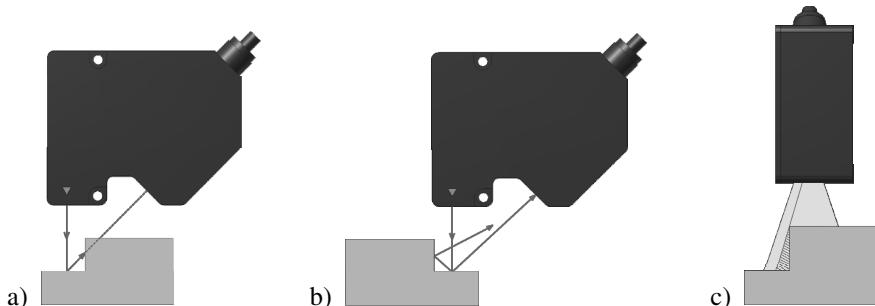


Fig. 2. Main sources of disturbances in laser triangulation technique [5, 10]: a) shadow effect, b) reflection and interreflection effects, c) dead zone effect

The shadow effect occurs when an edge or wall of the object prevents the detector from seeing the reflected light (Fig. 2.a) [1, 5, 8, 9, 10]. When the surfaces in question are characterized by a high reflective coefficient detector readouts can be disturbed by reflections and interreflections. The latter are those produced by light reflected in sequence from more than one surface (Fig. 2.b) [2, 5, 9, 10]. The dead zone effect takes place in 2D laser triangulation heads (Fig. 2.c). The effect enlarges with the growing distance between the measurement point and the centre of the scanning laser line [5, 10]. The edge of the object prevent the surface to be efficiently illuminated by the laser line.

4 Scanning Technique to Avoid Measurement Disturbances

The 2D measurement head enables a profile measurement along the projected laser line. If the object dimension does not exceed the head range it is necessary to use only one linear actuator. However, when the object is larger than the head range and it is necessary to maintain high accuracy of the process, it is necessary to perform measurements in parts, and then connect separately collected data into one coherent image. This requires two linear stages - configured in a cross. The scanning method in the X-Y reference system is widely applied in many profilometer constructions

[5, 8, 11, 12]. However, to reduce the measurement disturbances mentioned before, a new method – involving multiple scanning – has been developed and tested.

The solution consist in introduction of an auxiliary rotation stage. The stage performs rotation along Z axis (Fig. 3.a.). The configuration enables scanning for different angular positions of the object (Fig. 3.b.).

Following the result of the rotation in reference to the head position, data errors occur in subsequent measurements in deferent areas of the object surface. This allows developing an efficient reconstruction algorithm. It depends on overlapping data images from different angular positions, after removing erroneous data from disturbed areas. Since the sets of data complement one another the final result produces continuous surface of high accuracy [5].

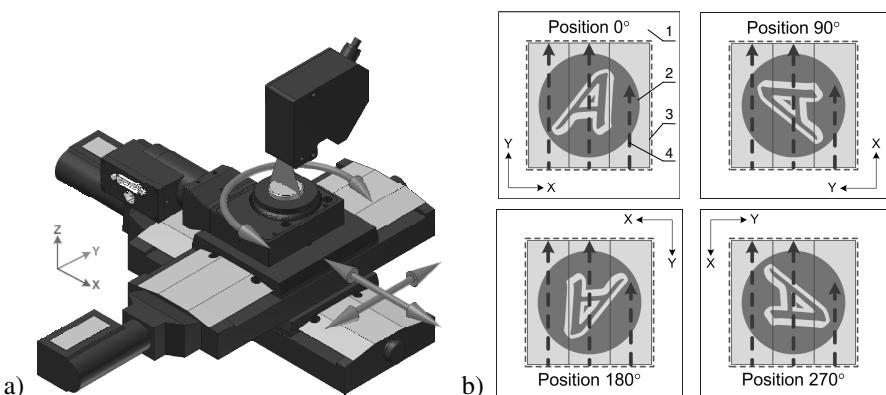


Fig. 3. Disturbance eliminating technique: a) profilometer configuration, b) scanning for different angular positions: 1- working region, 2- object, 3- scanning area, 4- scanning trajectory

The main advantage of the technique is the significant reduction of measurement artefacts connected to the shadow and reflecting effects. The only disadvantage is 4 times longer measurement procedure compared to a single scanning process. However, when 2D heads are applied increased time is less significant.

5 Data Fusion of Different Position of the Rotary Stage

The data collected for different angular position of the object have to be precisely mutually correlated in order to eliminate the shadow effects.

Some initial tests have revealed that the algorithm based on several independent angular measurements and followed by fitting them using the correlation function has not ensured the head accuracy.

The key point in the algorithm is the precise determination of the object rotation centre in the Cartesian X-Y coordinate system. The X-Y system is defined by the axes of the translation stages. A bearing ball, presented in Fig. 4, has been chosen as a calibration object for the rotation centre finding algorithm. Such object can be correctly measured without producing any aforementioned artefacts. The ball is placed arbitrarily in the measurement area.

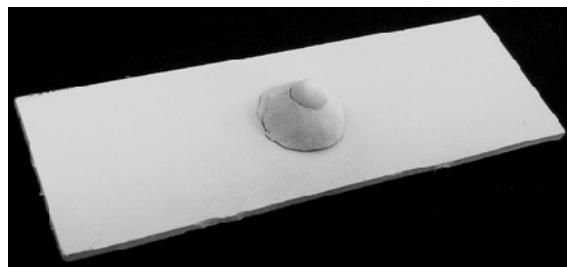


Fig. 4. Image of the reference object

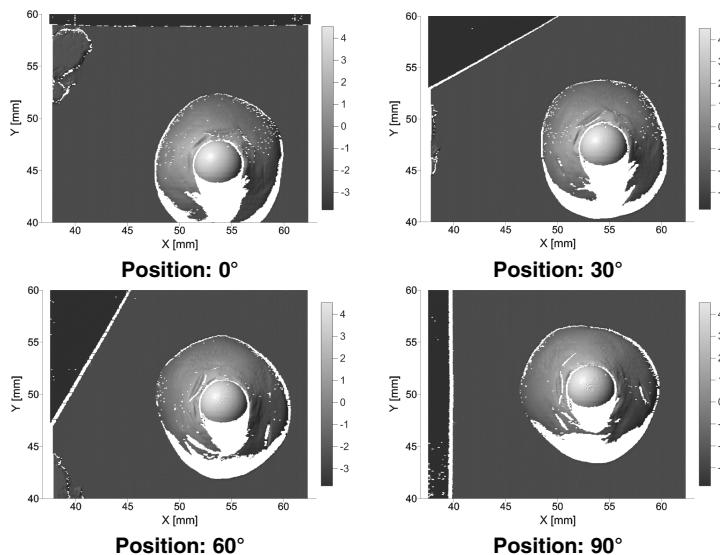


Fig. 5. Data images of the calibration object (a bearing ball) at four angular positions

Then the profilometer scans the ball for several angular positions – typically with the step equal to 30° . Fig. 5 presents four scans of the calibration object for 4 object angular positions. The position of the ball is calculated in the X-Y system for each angular position.

The centre of rotation is calculated according to a least-squares circle fitting algorithm. Fig. 6 presents the result of the least square approximation of bearing ball centre position for 12 scans in various angular orientations. When the centre of the rotation stage is defined, its coordinates can be easily used in rotation transformations for the artefact elimination process, which is described in the next section.

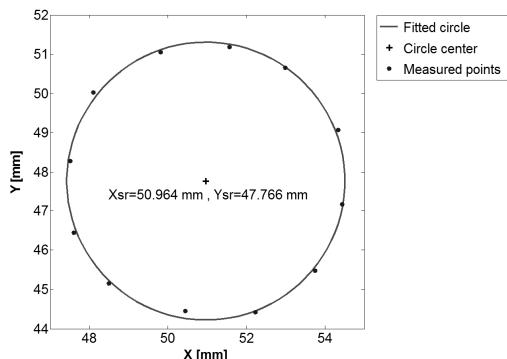


Fig. 6. Circle least square approximation visualisation

6 Measurement Data Processing Algorithms

The controller of the laser head, delivered by the manufacturer, has been equipped with its own collection of processing algorithms. However, for the purpose of this work the authors of the paper have not used them. All the presented data are based on authors' original concepts. Measurement process sequence is as follows:

- The laser head controller performs angular position correction according to the initial head orientation in the profilometer mount;
- The profilometer form a data grid in its XYZ coordinate system by collecting data from translation stages, rotary stage and laser head;
- The profilometer software performs mathematical transformations of the collected data, so that they can be put in the same grid pattern;
- Selective fusion of the measurement data from all scans into the final data array is performed.

The multiply scanning measurement allows the profilometer not only to measure places, which are beyond the reach of the head in a single scanning process, but also increases efficiency of processing algorithms, by providing more data. In each point within the measurement grid there are four values from laser head to process. If the sensor cannot return a reliable value in one angular position, it is very likely it will in another position. In addition, in most of the measurement points 4 values are available, which reduces uncertainty of measurement by calculating the mean value.

During the tests it has been discovered that the shadow effects may cause errors also in the close neighbourhood of the shadow line (Near Shadow Effect). To avoid influence of this phenomenon on the final results, in each point where four values have been read by the head, the outlier is discarded and the final result is calculated as the mean of the remaining three. Fig. 7 presents the results of data processing by the Near Shadow Effect elimination algorithm.

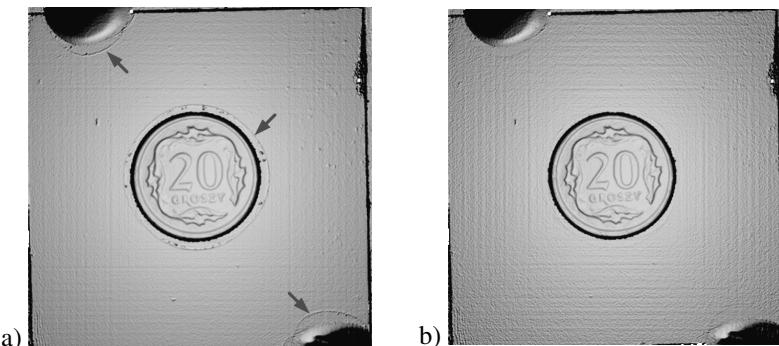


Fig. 7. Results of the near shadow effect elimination algorithm: a) without filtering b) with filtering

Preliminary tests have demonstrated that the algorithm of assembling measurement data from different positions of the rotation stage into basic coordinates system is accurate enough for the profilometer purposes. It means that the head accuracy declared by its manufacturer has been at least maintained despite of the data assembling process. Simultaneously, typical triangulation-origin artefacts have been eliminated or significantly decreased. The algorithm is designed in a way that in all positions data are read directly to the specially arranged array without interpolation.

7 Measurement System Design

The developed system consists of two main elements: the 3D laser profilometer (Fig. 8.a.) and control unit with a specialized software installed (Fig. 8.b.).

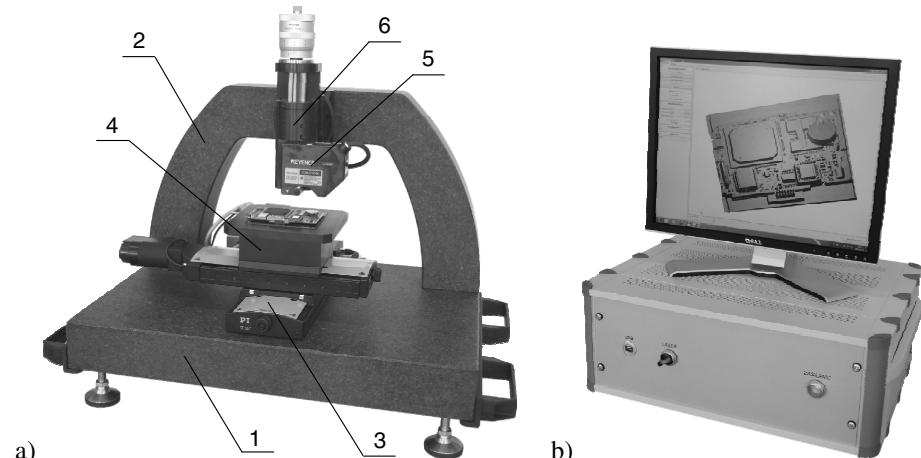


Fig. 8. System for surface topography measurement: a) 3D laser profilometer design, b) general view of the control system: 1- granite base, 2- head mount, 3- X-Y linear stages, 4- rotating stage, 5- measurement head, 6- head adjustment module.

The profilometer has a modular design. It consists of 3 main parts: the granite base with the head mount, X-Y linear stages with rotation stage, measurement head, equipped with adjustment module [5].

The module of the linear stages enables movement of the measured object in the X-Y directions. The rotation stage performs rotation of the object along Z axis. The parameters of the stages can be found in Table 1. [5, 14].

Table 1. Specification of the actuators [5, 14]

Parameter	Value	
Type	linear	rotating
Range	100 mm	360°
Resolution	0.125 µm	32 µrad
One directional repeatability	0.2 µm	50 µrad
Two directional repeatability	2 µm	-
Screw pitch / Transmission ratio worm gear	0.5 mm	50:1
Maximum speed	15 mm/s	90°/s

The measurement unit consists of the laser head, positioner and the mount. The latter is placed on a granite arch. A precise adjustment of the head is performed using the positioner. The specification of the head are presented in Table 2.

Table 2. Specification of the measurement head [5, 10]

Parameter	Value
Range in Z axis	20 mm
Range in X axis	20 ÷ 25 mm
Resolution in Z axis	1 µm
Resolution in X axis	800 points (33 µm pitch)
Repeatability in Z axis	1 µm
Repeatability in X axis	5 µm
Nonlinearity in Z axis	± 0.1 % measurement range

An industrial computer controls the system. Communication between the computer and other devices is realized via USB interface. The linear stages and rotation stage are controlled by three single-channel control units connected via a network. Thanks to such solution all three actuators can be controlled using a single USB port [5].

Rough data incoming from the laser triangulation head are transferred to the head control unit, which is placed in the system controller housing. After initial processing the data are sent to the PC via the USB port.

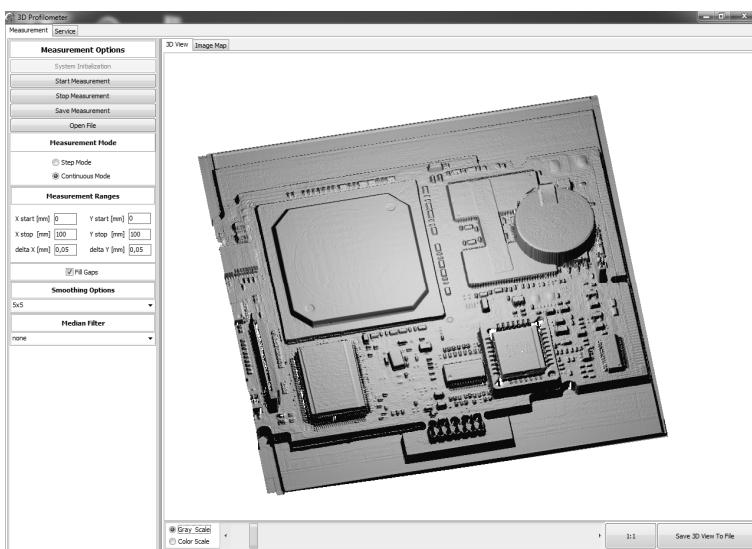


Fig. 9. User interface – section "Measurement"

The profilometer software has been installed on the PC. Its user interface has two main tabs: Measurement and Service. The Measurement section is designed for the user, the view of which is demonstrated in Fig. 9.

The measurement section is divided into 2 parts. The components designed for performing measurements are placed on the left hand side of the screen. After switching on the system the user has to initialize it pressing "System initialization" button. Next the values of the fields: X start, X stop, delta X, Y start, Y stop, delta Y should be inserted. They characterize measurement region. When the above is done, the user can start measuring by clicking the button "Start Measurement". The options concerning data processing are placed below the measurement fields. "Fill Gaps" options allows the user to interpolate missing data.

The software enables context filtering of the collected data using a low-pass smoothing filter or a median filter [5, 13]. On the right-hand side of the interface there is a visualization screen. When the measurement process is finished it is possible to see a 3D plot of the object under study by clicking "3D View".

The section "Service" allows independent control of each system components including translation and rotation stages (Fig. 10). The user can manoeuvre the stages utilizing simple navigation panels localized on the left-hand side of the interface. The navigation utilities allow the user to move the stage X (along the axis parallel to the measurement axis of the head), stage Y (perpendicularly to the measurement axis of the head) and the rotary stage.

Each panel includes the following buttons: „<<”, „<<”, „0”, „>>”, „>>l”. These can be used to move the translation stages to the limits, centre position, or advance by a specified distance ("Jump"). The laser head panel is localized on the right-hand side of the interface. It displays the plot of the currently collected surface profile [5, 13].

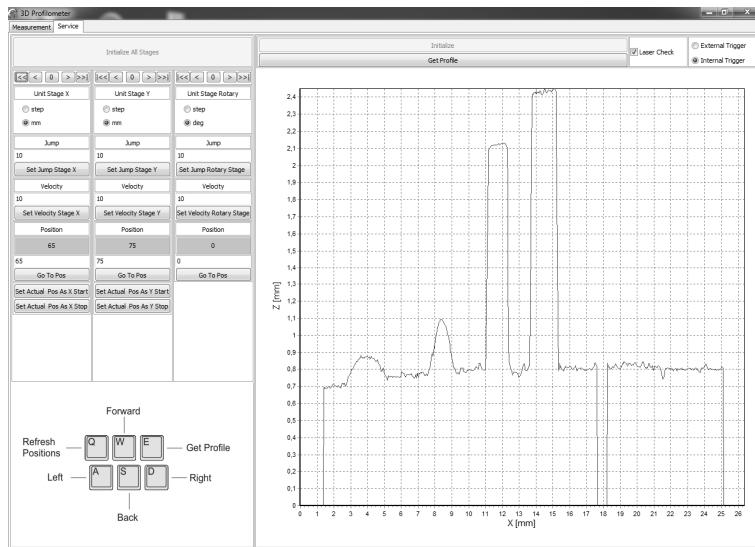


Fig. 10. Bookmark "Service"

8 Sample Surface Scans

The surface of a minicomputer module ("System On Module" - SOM type) has been used as a test object (Fig. 11).



Fig. 11. An electronic circuit during the measurement process

The system has recorded partial (at different angular positions of the object) and final results. The scanning results are demonstrated as top views of the object. The data images easily visualize discontinuity of regions, where measurements could not be possible (white areas) Fig. 12.

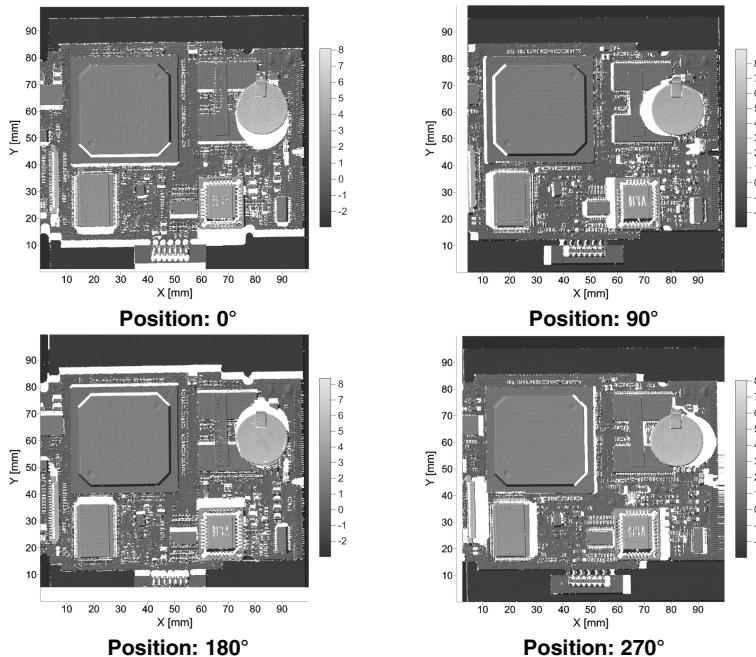


Fig. 12. Data images of a minicomputer module ("System On Module" – SOM type) at four different angular positions

The scanning area has covered the entire measurement range of the translation stages XY, which is 100 mm with the scanning step equal to 50 μm . The final results of the data fusion of four scans are presented in Fig. 13.

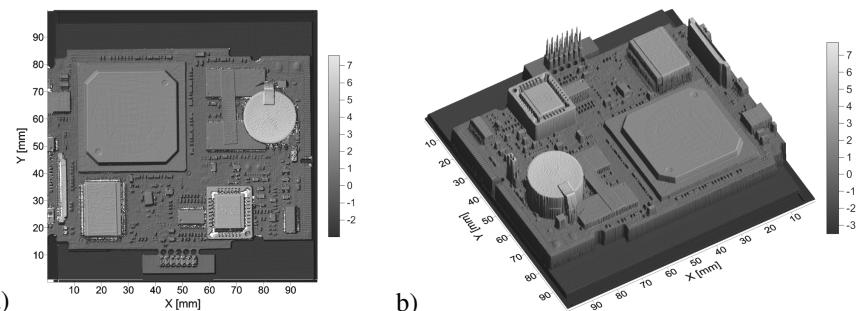


Fig. 13. Final view of the surface of a SOM module, combined of 4 partial measurements: a) top view, b) 3D view with "invisible" points interpolated

Despite of the significant reduction of the regions of uncertainty, there are still areas where the profilometer cannot find correct values, especially - e.g.: holes of the small diameters. The user has two options: to leave as they are or eliminate by interpolating the correct values [5, 13].

9 Conclusion

A method for eliminating artefacts in a triangular laser system dedicated to surface reconstruction has been presented in this paper. The technique involves multiple scanning of the object under study at a sequence of different angular positions in respect to the scanning direction of the laser head. The procedure allows effective surface characterization, simultaneously eliminating or significantly reducing typical disturbance effects of a triangular measurement systems. A robust control software developed by the authors allows precise location of the rotation axis and combining partial data sets recorded for 4 angular positions of the object. The main advantage of the software is its ability to correct location of the data points on line, during scanning. The data analysis and fusion are based on simple and effective algorithms, which saves the processing time. The system as a whole has been tested successfully using objects such as: coins, shells and circuits boards. The results clearly indicate that the device can be applied in both industrial and research laboratories focused on precise measurements and quality control related issues.

Acknowledgement. Scientific work executed within the Strategic Programme “Innovative Systems of Technical Support for Sustainable Development of Economy” within Innovative Economy Operational Programme.

References

1. Kupiec, M.: Optical - contact technique for coordinate measurements, Cracow University of Technology Digital Library, Cracow (2007) (in Polish)
2. Reiner, J., Stankiewicz, M.: Elimination of image geometrical disturbances in an laser triangulation technique. Pomiary Automatyka Kontrola 56(1), 54–57 (2010) (in Polish)
3. Keyence Datasheet, <http://www.keyence.eu>
4. Kulak, A.: A laser measurement system for quality control applications. Problemy Eksplatacji (2), 115–124 (2004) (in Polish)
5. Czajka, P. (ed.): The development of 3D laser profilometry, Final report, ITeE-PIB Radom (2012) (in Polish)
6. LMI Technologies DataSheet, <http://www.lmi3d.com>
7. Micro-Epsilon Datasheet, <http://www.micro-epsilon.com>
8. Giesko, T. (ed.): Testing and research instrumentation for surface characterization using laser profilometry, Final report, ITeE-PIB Radom (2006) (in Polish)
9. Bradshaw, G.: Non-contact surface geometry measurement techniques. Technical Report TCD-CS (1999)
10. High-accuracy 2D Laser Displacement Sensor, LJ-G Series User’s Manual, Keyence (2010)
11. NanoFocus AG Datasheet, <http://www.nanofocus.com>
12. Nanovea Datasheet, <http://www.nanovea.com>
13. Galas, J. (ed.): Development of software modules for reconfigurational two-axes measurement manipulator, Final report, INOS Warszawa (2012) (in Polish)
14. Physik Instrumente Datasheet, <http://www.physikinstrumente.com>

Preisach Based Model for Predicting of Functional Characteristic of Fluxgate Sensors and Inductive Components

Piotr Frydrych¹ and Roman Szewczyk²

¹ Institute of Metrology and Biomedical Engineering, Warsaw University of Technology
Warsaw, Poland

p.frydrych@mchtr.pw.edu.pl

² Industrial Research Institute for Automation and Measurements, Warsaw, Poland
rszewczyk@piap.pl

Abstract. This paper presents model of magnetic characteristics of cores of fluxgate sensors and of inductive components for electronic devices. It is based on extended Preisach model. The model consists of two parts. First describes structure of material by coercivity distribution matrix, what enables to recreate previous material domain structure. Second part calculates high frequency losses. The model is very useful for prediction high frequency asymmetrical hysteresis loops.

Keywords: amorphous alloys, magnetic hysteresis, soft magnetic materials , fluxgate sensor, ferrites, modelling.

1 Introduction

Nowadays in many electronic devices magnetic cores are excited by high frequency asymmetrical signals. They have both high expectations of sensitivity and minimalisation of losses. For those features higher accuracy of the simulation of output characteristic is required. Examples of these areas are fluxgate sensors. They are very applicable in many fields, such as navigation, nondestructive testing [1], safety and security solutions, current sensors [2] and pulse power conversion devices, far more effective and flexible especially for direct currents [3, 4]. Parameters like coercion, saturation, permeability and remanence are often used but they are not sufficient. Especially small and asymmetric loops for different materials with the same values of those parameters may strongly differ. Model can be used instead of prototypes testing to predict characteristics, what shortens time necessary to introduce the product to the market and lowers production costs.

2 Experimental Setup

In this research two different cores were tested. First was single layer frame shaped sensor core made of as-quenched Fe₈₀B₁₁Si₉ amorphous ribbon [5, 6] second was ring

shaped bulk $\text{Mn}_{0.70}\text{Zn}_{0.24}\text{Fe}_{2.06}\text{O}_4$ ferrite core for power applications [7, 13-15]. Ribbon core was placed in fluxgate sensor in Foerster configuration. Inner diameter of frame was 25 mm and outer 35 mm. Each branch of frame shaped core had 25 sensing and 15 excitation turns. Coils configuration in sensing circuit were changeable, so it is possible to measure average magnetic hysteresis loops of the core placed in fluxgate sensor. The ferrite core was frame shaped with dimensions 70 mm high, 30 mm wide, and 15 mm thick, it had a bulk density $\rho = 4.8 \cdot 10^6 \text{ g/m}^3$.

3 Idea of the Model

Model is divided into two parts (fig. 1). The first one is Preisach based domain distribution model of magnetic characteristics of the core. Model parameters are optimized for representative static hysteresis loops, which are connected only with material properties. It represents also the state of material, which is an effect of previous magnetization changes. Second part estimates high frequency losses in the material. It is connected with field's changes speed, core dimensions and shape. Therefore losses field is calculated, which influences effective magnetic field in the material. This approach was chosen to simplify material testing process before modelling.

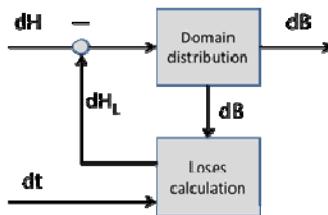


Fig. 1. Model scheme

4 Domain Distribution Model

Preisach model implies statistic distribution of domain coercion field. It is useful for modeling magnetic characteristics of amorphous materials, because of its correlation with microstructural disorder [8]. For ferrites magnetic characteristic usually Jiles-Atherton model is used [9]. Even though the link between the model and magnetization mechanisms in bulk materials is not so obvious, presented Preisach model is effective from phenomenological point of view in modelling characteristics of materials such as ferrites [10, 11]. In this model Preisach matrix is used (fig. 2). It represents distribution of coercivity operators in material. Operators can switch between positive and negative.

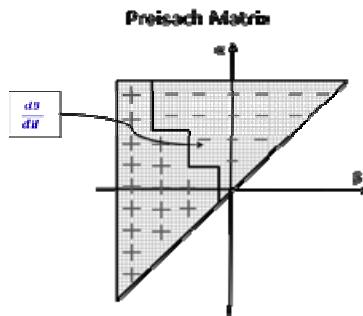


Fig. 2. Preisach matrix idea

For increasing field all operators which have lower α value than actual field value switch up, whereas for decreasing field all operators with higher β value switch down. Domain distribution can be recorded by using operator matrix. Every cell of the matrix represents differential region in Preisach plane, which dimensions depend on required accuracy of the model. Differential region value is proportional to amount of domains which switch for that range of magnetic field and is connected with magnetic permeability. Therefore it is possible to record full domain structure, which is effect of historical magnetic field values and current magnetic field. This ability is needed especially in asymmetrical and minor loops. Two dimensional normal distribution was assumed (fig. 3).

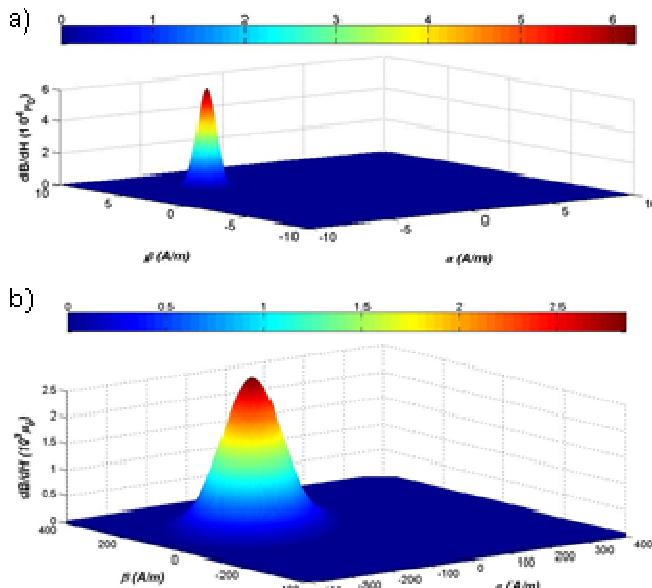


Fig. 2. Preisach plane with distribution of domain corecivity for a) as-quenched Fe80B11Si9 b) Mn_{0.70}Zn_{0.24}Fe_{2.06}O₄ ferrite

Center of distribution has to be on symmetry axis, because of symmetrical coercivity of material. Thanks that there are only four distribution parameters: A – scaling factor, m – mean value of domain coercivity field , sc – coercivity variance and si – interaction variance (1).

$$p(\alpha, \beta) = \frac{1}{2\pi \cdot s_\alpha \cdot s_\beta} \exp \left(-\frac{(\alpha_0 - m_\alpha)^2}{2s_\alpha^2} - \frac{(\beta_0 - m_\beta)^2}{2s_\beta^2} \right) \quad (1)$$

Model parameters were optimised for static loops. For amorphous alloy they were measured for multilayer rolled tape instead of single layer frame shaped core, which was used in fluxgate sensor. The reason was too weak signal for frequency lower than 1 Hz. In optimisation process evolution strategy ($\mu+\lambda$) was used [12]. Modelling results are shown on figure 3 and 4. Achieved R2 Pearson coefficient was over 0.98 for amorphous alloy and over 0.92 for ferrite.

5 Domain Distribution Model

Domain coercivity field distribution does not change with frequency. To model losses effective field H_E is calculated. Losses field H_L is correlated with change of internal magnetisation in time. It takes in to account dimensions of the core, material structure and conductivity. External field is therefore sum of effective field H_E and losses (2). For given field value and its change in time losses can be calculated, thanks to computed Preisach matrix which represents induction change for external field value. This approach results from Lenz's law. According to it an induced electromotive force always gives rise to a current which produced magnetic field opposite to the original change in magnetic flux. After computing the distribution for static loops losses factors for different mechanisms can be calculated (3) comparing static hysteresis loop with higher frequencies loops. Two losses mechanisms can be taken into account: eddy current losses (4) and excess losses (5). As it is assumed for domain coercivity field distribution does not change with frequency, so the differences between static and high frequency loops are the effect of lower effective field H_E .

$$H_E = H - H_L \quad (1)$$

$$H_L = H_{ed} + H_{ex} \quad (2)$$

$$H_{ed} \sim \frac{dB}{dt} \quad (3)$$

$$H_{ex} \sim \sqrt{\frac{dB}{dt}} \quad (4)$$

Losses factors computed for hysteresis loops can be used for effective field calculation of other signals. The way of calculating presented above simplifies modeling.

6 Functional Characteristics Modeling

Magnetic rate dependent model with ability to approximate losses in the core is necessary to predict induction in core for magnetic given field. This is important especially because fluxgate sensors work for frequencies higher than 1 kHz. Magnetic field has to be calculated for windings and core dimensions. Output signal can be calculated in the same way. In Foerster configuration signals from two cores are added in electric circuit. Therefore predicted fluxgate output signal is achieved, what is presented on figure 5. In this case it can be observed, that there are some imperfections, like oscilloscope offset, influence of filtration, noise and others, which are difficult to predict.

Modeling of magnetic characteristics of power ferrite can be used to predict losses in the core. Usually producers inform about losses values for particular core. In fact they are measured for symmetrical loops, what is not appropriate for many power conversion applications. For example in resonant converters amplitude of magnetic field inducted in the core is between zero and near to saturation of the core. Losses calculated using model show significant differences (fig. 6).

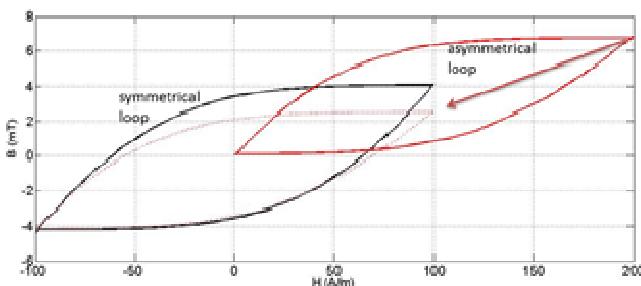


Fig. 3. Comparison of symmetrical and asymmetrical loops simulated using developed model for $\text{Mn}_{0.70}\text{Zn}_{0.24}\text{Fe}_{2.06}\text{O}_4$ ferrite

7 Conclusions

Applied Preisach matrix and effective field calculation method simplifies magnetic characteristics modelling. Designed Preisach based model is very sufficient to optimisation of fluxgate parameters and magnetic cores of electronic devices. It is significant, that model parameters optimised for particular core can be appropriate for other cores made from the same material as it was in case of multilayer rolled core and single layer frame shaped core. Therefore model optimised for material can be generalised for prediction of characteristic of different devices with different shapes of cores made of that material. Effective field calculation can be also useful for further losses research.

Acknowledgment. This work has been supported by the European Union in the framework of European Social Fund through the Warsaw University of Technology Development Programme.

Calculations for the modelling were made in the Interdisciplinary Centre for Mathematical and Computational Modelling of Warsaw University, within grant G36-10.

References

1. Bao, S., Gong, S.F.: Magnetomechanical behavior for assessment of fatigue process in ferromagnetic steel. *J. Appl. Phys.* 112, 113902 (2012)
2. Yang, X., Zhang, B., Wang, Y., Zhao, Z., Yan, W.: The optimization of dual-core closed-loop fluxgate technology in precision current sensor. *J. Appl. Phys.* 111, 07E722 (2012)
3. Hasse, L., Szewczyk, R., Smulko, J., Bienkowski, A., Frydrych, P., Kolano-Burian, A.: Low Current Transformer Utilizing Co-Based Amorphous Alloys. *Trans. Magn.* 48, 1493–1496 (2012)
4. Zhang, J., Li, P., Wen, Y., He, W., Yang, A., Lu, C., Qiu, J., Wen, J., Yang, J., Zhu, Y., Yu, M.: High-resolution current sensor utilizing nanocrystalline alloy and magnetoelectric laminate composite. *Rev. Sci. Instrum.* 83 (November 2012)
5. Frydrych, P., Szewczyk, R., Salach, J., Trzcinka, K.: Two- Axis Miniature Fluxgate Sensors. *Trans. Magn.* 48, 1485–1488 (2012)
6. Lei, C., Lei, J., Yang, Z., Zhou, Y.: Improved micro fluxgate sensor with double-layer Fe-based amorphous core. *Microsyst. Technol.* 19, 167–172 (2013)
7. Bieńkowski, A., Szewczyk, R.: The Dependence of the Magnetoelastic Properties of Zn–Mn Ferrites on Their Magnetocrystalline Properties. *Phys. Stat. Sol.* 3, 825–828 (2002)
8. Andrei, P., Stancu, A., Caltun, O.: Hysteresis in particulate recording media. Experiment and simulation with Preisach and Jiles–Atherton models. *J. Appl. Phys.* 206, 160–164 (1999)
9. Bieńkowski, A., Rożniatowski, K., Szewczyk, R.: Effects of stress and its dependence on microstructure in Mn–Zn ferrite for power applications. *J. Magn. Magn. Mater.* 254–255, 547–549 (2003)
10. Lu, H.Y., Zhu, J.G., Ron Hui, S.Y.: Measurement and Modeling of Thermal Effects on Magnetic Hysteresis of Soft Ferrites. *Trans. Magn.* 43(11) (November 2007)
11. Frydrych, P., Szewczyk, R.: Application of Preisach model for modeling ferrite-based cores of sensors. *J. Electr. Eng.* 61, 73–76 (2013)
12. Szewczyk, R.: Extension of the model of the magnetic characteristics of anisotropic metallic glasses. *D: Appl. Phys.* 40, 4109 (2007)
13. Szewczyk, R.: Modelling of the magnetic and magnetostrictive properties of high permeability Mn-Zn ferrites. *J. of Physics* 67(6), 1165–1171 (2006)
14. Szewczyk, R., Bienkowski, A.: Magnetoelastic Villari effect in high-permeability Mn-Zn ferrites and modeling of this effect. *J. of Magnetism and Magnetic Materials* 254, SI284–SI286 (2003)
15. Salach, J., Bieńkowski, A., Szewczyk, R., Jackiewicz, D., Frydrych, P., Kolano-Burian, A.: Utilizing of magnetic amorphous alloys in magnetoelastic tensile stress sensors. *Pomiary Automatyka Robotyka* 16(2), 556–560 (2012) (in Polish)

Automated Measurement Systems for Meters of Heat

Tadeusz Goszczyński

Industrial Institute of Automation and Measurement PIAP, Warsaw, Poland
tgoszczyński@piap.pl

Abstract. Heat meters are often used in buildings with central heating systems for financial settlements with residents of apartments. The meter of heat is a set of 3 sensors and a calculator. For producers it is very important to have appropriate instruments and procedures for validation of different type of heat meters. In this paper procedures for validation using automated systems are proposed and analysed taking into account its efficiency.

Keywords: heat meters, sensors, automated systems, efficiency, validation tests.

1 Introduction

Heat meters are instruments intended for measuring the heat that, in a heat-exchange circuit, is absorbed by a heat conveying liquid. These heat meters are used for financial settlements between heat supplier and residents of apartments in buildings with central heating systems and are submitted for control by legal metrology services. The meters shall comply with the requirements formulated in standards [1] and [2]. For manufacturing of heat meters the main limitation in productivity lies in process of their validation. Validation is the most time consuming process in production and therefore optimizing and automation of it is critical for producers of these instruments. Selecting the most appropriate modular validation system and procedure can offer enormous benefits in terms of system quality, cost reduction and manufacturing productivity. This paper proposes a possible approach to support decision-making in the area for heat meter validation systems and a methodology of selecting instruments for modular systems considering technological efficiency and economic approach. The next sections are organized as follows: Sections 2 gives the theoretical framework of heat meter validation, Section 3 considers configurations of system for measurements, Section 4 proposes new method of maximum error calculation. Section 5 presents calculation of time for validation of compact meters and calculation of time for validation of combined meters. The last Section provides some conclusions and gives suggestions for future work.

2 Theoretical Framework

Heat meters are made of few sub-assemblies. Flow sensor at either the flow or return of a heat-exchange circuit. Temperature sensor pair which senses the temperatures of

the heat conveying liquid at the flow and return of a heat-exchange circuit. Calculator using signals from the flow sensor and the temperature sensors, calculates the quantity of heat absorbed.

Flow sensor is always verified on the special separate water stand. There are two main types of heat meter's construction: compact and combine.

- Compact meter calculator with inseparable temperature sensor pair is validated on one special stand.
- Combined meter validation needs separate stand for calculator and separate stand and procedure for sensor pair.

While validating the temperature sensor pair, both sensors must be inserted into a temperature controlled bath and their resistance should be measured at three temperature points. The standard [1, 9] requires calculations evaluating the worst-case error of the pair (maximum ratio of error to maximum permissible error – MPE) in full measurement range of the heat meter. This worst-case error must be calculated precisely, therefore, the graphical methods normally used are not acceptable. The standard method requires sensor pair calculations for every point of measurements area given by producer. For example, if the gradient of 0.1°C is used for the calculation it results in more than 10^6 calculations of quadratic equations. Even if it takes only 2–3 minutes for every pair, it is irritating for personnel and lowers efficiency. The proposed here method requires calculating a much smaller number of equations. At first, the parameters (t , Δt) of extremum points are determined. In the next step the error values in those points only are calculated, and the worst-case error is selected.

When validating the calculator with inseparable temperature sensor pair, the two temperature sensors should be immersed in two thermostatic baths. Within each of the three temperature ranges given in [1] accuracy of heat calculation should be examined. Heat meter calculates actual values of heat power and counts energy in time, using algorithms given in [1]. Heat energy Q measured by heat meter is a difference of energy of heat conveying liquid on the input of the heat exchanger Q_{in} and on the output of it Q_{out} . This energy measured in time from 0 to t_1 can be presented as:

$$Q = \int_0^{t_1} [t_1 C_w(t_1) \rho(t_1) v_1(t_1) - t_2 C_w(t_2) \rho(t_2) v_2(t_2)] d\tau \quad (1)$$

where: t_1 – temperature of water on input, t_2 – temperature of water on output (return temperature), $C_w(t_1)$ – heat capacity on input, $C_w(t_2)$ – heat capacity on output, $\rho(t_1)$ – water density on input, $\rho(t_2)$ – water density on output, $v_1(t_1)$ – volume flow rate of water on input, $v_2(t_2)$ – volume flow rate of water on output, τ – time.

In process of validation a heat meter shall be tested in three points of temperature difference $\Delta t = (t_1 - t_2)$:

1. for return temperature 40°C to 70°C :

$$\Delta t_{min} \leq \Delta t \leq 1,2 \Delta t_{min} \quad (2)$$

2. for return temperature from 40 °C to 70 °C:

$$10 \text{ } ^\circ\text{C} \leq \Delta t \leq 20 \text{ } ^\circ\text{C} \quad (3)$$

3. for:

$$\Delta t_{\max} - 5 \text{ } ^\circ\text{C} \leq \Delta t \leq \Delta t_{\max}. \quad (4)$$

It is known [3] that two coefficients $k_1(t_1, t_2)$ and $k_2(t_1, t_2)$, which are functions of t_1 and t_2 , can be presented as:

$$[t_1 - t_2] k_2(t_1, t_2) v_2(t_2) = t_1 Cw(t_1) \rho(t_1) v_1(t_1) - t_2 Cw(t_2) \rho(t_2) v_2(t_2) \quad (5)$$

$$[t_1 - t_2] k_1(t_1, t_2) v_1(t_1) = t_1 Cw(t_1) \rho(t_1) v_1(t_1) - t_2 Cw(t_2) \rho(t_2) v_2(t_2) \quad (6)$$

After applying it to (1) we have two equations: one, for measurement of volume of water on output of heating system

$$Q = \int_0^{\tau_1} [t_1 - t_2] k_2(t_1, t_2) v_2(t_2) d\tau \quad (7)$$

and second, for measurement of volume of water on input

$$Q = \int_0^{\tau_1} [t_1 - t_2] k_1(t_1, t_2) v_1(t_1) d\tau \quad (8)$$

The validation system is destined for measurement of heat meters accuracy according to one of these equations.

3 Configurations of Measurement System for Compact Meters

The use of modular structure architecture of the validation system has many advantages [4]. Therefore selection of every apparatus is one of the first steps. It was checked that in thermostatic baths of middle range one reference sensor and up to 8 heat meter sensors could be immersed. For measurement of the two standard reference sensors resistance one precision resistance meter can be used with IEEE-488 bus for communication with computer. The 8-channel multiplexer connected to the standard reference sensors and to precision resistance meter with 4-wire technique of measurement should be used. Another multiplexer for connection of heat meter's pulse outputs to pulse counters should have number of channels appropriate to number of heat meters in a group [5]. Heat meters working in verification mode need to have a group of pulses received on their inputs in order to generate on its outputs number of pulses proportional to quantity of heat measured. Calculators could be physically connected to the system using the multiplexer. Then in the same phase of measurement few heat meters can be connected to pulse generator and few pulse counters. In every such phase temperature in the two baths is measured using reference sensors connected one after another to the same resistance meter. In this way

uncertainty due to temperature stability in the baths is minimized. The measurement can be made twice: before and after generating pulses to heat meter and average value taken into account.

Since every heat meter must be checked at 3 points of its temperature characteristic and since stabilizing temperature of bath takes time, it is recommended to realize measurements for groups of heat meters after stabilizing temperature in baths. Measurement data are stored in a database and after every change of temperature in baths the procedure of measurement for all groups of meters under test is realized.

4 Calculation of Error for Combined Meters

Tegeler et al. [6] present a mathematical model for the calculation of the temperature difference uncertainty of calibration of paired temperature sensors. It provides requirements for apparatus used for the calibration but cannot be used for calculation of maximum error of the controlled pairs.

For combined meters the stand for checking pairs of temperature sensors measures resistance of 2 sensors for every meter in 3 temperature values and calculates parameters of resistance/temperature relation of every temperature sensor. But the value of maximum error of temperature difference calculated by the heat meter must be defined in next step. The error value should be chosen as a maximum error value calculated in full range of temperature and temperature difference given for this type of heat meter. Calculating all error values using step 0.1 K for Δt and t gives more than few hundred thousands consecutive calculations. To shorten time of heat meter validation Goszczyński in [7] proposed a method for finding first a region of the characteristic for every heat meter where the maximum error is placed.

Standard EN 60751 [2] defines resistance/temperature relation as

$$r = R_0(1 + At + Bt^2) \quad (9)$$

and standard values: $R_0; A; B$.

The measured by stand resistance values are used in the system of three equations to calculate the three constants $R_{01}; A_1; B_1$ for one sensor and $R_{02}; A_2; B_2$ for the second sensor.

This method of calculation is mandatory for the calculation of error to comply with the requirements formulated in the standard [1]. But it need not be done for every point in full temperature range and full temperature difference range if the analysis of extremum points location is carried out.

Korytkowski et al. [8], used resistance difference value instead of temperature difference, as an output signal of the temperature sensor pair and calculated the relative error:

$$e_r = \frac{\Delta r_{sp} - \Delta r_{st}}{\Delta r_{st}} \quad (10)$$

where

$$\Delta r_{sp} = R_{01} \left(1 + A_1 t_1 + B_1 t_1^2\right) - R_{02} \left(1 + A_2 t_2 + B_2 t_2^2\right) \quad (11)$$

and

$$\Delta r_{st} = R_0 \left(1 + At_1 + Bt_1^2\right) - R_0 \left(1 + At_2 + Bt_2^2\right) \quad (12)$$

and from these

$$\begin{aligned} \Delta r_{sp} - \Delta r_{st} = & (R_{01} - R_{02}) + (R_{01}A_1 - R_{02}A_2)t + (R_{01}A_1 - R_0A)\Delta t + (R_{01}B_1 - R_{02}B_2)t^2 + \\ & (R_{01}B_1 - R_0B)2t\Delta t + (R_{01}B_1 - R_0B)\Delta t^2 \end{aligned} \quad (13)$$

$$\text{where } \Delta t = t_1 - t_2 \text{ and } t = t_2 \quad (14)$$

These equations can be used in numerical methods from [7] of calculating maximum error using resistance instead of temperature.

Every manufacturer of the heat meter defines for it: t_{min} , t_{max} , Δt_{min} and Δt_{max} .

Expressions in clauses in Eq. (13) are substituted by w_i to w_5

$$R_{01} - R_{02} = w_1; \quad R_{01}A_1 - R_{02}A_2 = w_2; \quad R_{01}A_1 - R_0A = w_3;$$

$$R_{01}B_1 - R_{02}B_2 = w_4; \quad R_{01}B_1 - R_0B = w_5 \quad (15)$$

Consequently, the equation for further analysis has a form

$$\Delta r_{sp} - \Delta r_{st} = w_1 + w_2t + w_3\Delta t + w_4t^2 + 2w_5t\Delta t + w_5\Delta t^2 \quad (16)$$

and ratio of a relative error divided by maximum permissible error

$$e_{rr} = \frac{e_r}{|E_r|} \quad (17)$$

where E_r is the maximum permissible error (MPE)

$$|E_{rMPE}| = 0,5 \frac{\Delta r_{st} + 6 \cdot 0.385 \Delta t_{min}}{\Delta r_{st}} \quad (18)$$

substituting Eqs. (10), (12), (16) and (18) to Eq. (17) we obtain

$$e_{rr} = \frac{2w_1 + 2w_2t + 2w_3\Delta t + 2w_4t^2 + 4w_5t\Delta t + 2w_5\Delta t^2}{6\Delta r_{min} + R_0 \left(1 + At_1 + Bt_1^2\right) - R_0 \left(1 + At_2 + Bt_2^2\right)} \quad (19)$$

In order to find points of extremum of equation (18) the derivatives over temperature and over temperature difference are determined and compared to zero.

For temperature the roots, (extremum points) are:

$$t_{1,2} = \frac{-4w_4(6\Delta r_{\min} + R_0 A \Delta t + R_0 B \Delta t^2) \pm \sqrt{del1}}{8w_4 R_0 B \Delta t} \quad (20)$$

where

$$\begin{aligned} del1 = & 16w_4^2(6\Delta r_{\min} + R_0 A \Delta t + R_0 B \Delta t^2)^2 - \\ & 16w_4 R_0 B \Delta t \left((2w_2 + 4w_5 \Delta t)(6\Delta r_{\min} + R_0 A \Delta t + R_0 B \Delta t^2) - 4R_0 B \Delta t (w_1 + w_3 \Delta t + w_5 \Delta t^2) \right) \end{aligned} \quad (21)$$

For temperature difference:

$$\Delta t_{1,2} = \frac{-4w_5 6\Delta r_{\min} + 4R_0 B(w_1 + w_2 t + w_4 t^2) \pm \sqrt{del2}}{2(2w_5(R_0 A + 2R_0 B t) - 2R_0 B(w_3 + 2w_5 t))} \quad (22)$$

where

$$\begin{aligned} del2 = & (4w_5 6\Delta r_{\min} - 4R_0 B(w_1 + w_2 t + w_4 t^2))^2 - 4(2w_5(R_0 A + 2R_0 B t) - \\ & 2R_0 B(w_3 + 2w_5 t))(2(w_3 + 2w_5 t)6\Delta r_{\min} - 2(w_1 + w_2 t + w_4 t^2)(R_0 A + 2R_0 B t)) \end{aligned} \quad (23)$$

These equations can be used to determine points of maximum or minimum error values on horizontal lines and vertical lines limiting the range of heat meter's measurements. Then the error values in search for the maximum value can be calculated according to standard method but only for less than 10 pairs of parameters values of Δt and t .

5 Calculation of Time Needed for Validation

Heat meters have two main types of construction: combined and compact. As was mentioned earlier there are different procedures for the two types.

Procedures for Compact Meters

The producer can choose from validating one meter after another or in groups of meters. Measurement data are stored in a database and used for calculation of errors after finishing all measurements.

Total time of validation is a sum of time needed for following operations:

- connections and time for stabilizing temperature of immersed sensors - t_1
- bath temperature control - t_2
- pulses generation - t_3
- test pulses counting - t_4

- temperature measurements in baths - t_5
- change of temperature in baths - t_6
- disconnections of heat meters - t_7

In Table 1 are presented values of time needed for the operations when validating different models of compact heat meters (named here from A to I).

Table 1. Values of time needed for the operations when validating different models of compact heat meters (time in seconds)

	t_1	t_2	t_3	t_4	t_5	t_6	t_7
A	60	10	20	20	10	1800	30
B	60	10	20	180	10	1800	30
C	60	10	20	300	10	1800	30
D	60	10	20	20	10	3600	30
E	60	10	20	300	10	3600	30
F	120	10	40	20	20	3600	30
G	120	10	40	20	20	1800	30
H	120	10	20	10	10	3600	60
I	180	20	60	300	20	3600	120

* Types from A to I in the table represent different models of heat meters

Time needed for validation for the following methods:
one after another

$$t_{\text{compact1}} = t_1 + t_8 + t_6 + t_8 + t_6 + t_8 + t_7 \quad (24)$$

automatically in 1 group of 8 items multiplexed (4 x to 2 pulse counters) manner

$$t_{\text{compact2}} = (t_1 + 4t_8 + t_6 + 4t_8 + t_6 + 4t_8 + 8t_7) / 8 \quad (25)$$

automatically in 10 groups of 8 items connected to 8 pulse counters (no multiplexer) manner

$$t_{\text{compact3}} = (10t_9 + t_6 + 10t_9 + t_6 + 10t_9) / 80 \quad (26)$$

where:

$$t_8 = t_2 + t_3 + t_4 + t_5 \quad (27)$$

$$t_9 = t_1 + t_8 + 8t_7 \quad (28)$$

In Table 2 are presented values of time needed for the validation procedures when validating different models of compact heat meters (named here from A to I) using the 3 presented here methods.

Table 2. Values of time needed for the validation procedures when validating different models of compact heat meters using the 3 presented here methods (time in seconds, recalculated for one meter).

	t_{compact1}	t_{compact2}	t_{compact3}
A	3930	577,5	180
B	4570	817,5	240
C	5050	997,5	285
D	7530	1027,5	225
E	8650	1447,5	330
F	7710	1080	258,75
G	4110	630	213,75
H	7580	1050	333,75
I	9100	1642,5	667,5

* Types from A to I in the table represent different models of heat meters

For combined meters the following processes are needed:

1) On the sensors pair validation stand

- Immersing a group of sensors in the bath and waiting for stabilization of temperature
- Measuring the resistances of the immersed sensors
- Calculating maximum error of every sensor pair

2) On the calculator validation stand

- Simulating resistances of temperature sensors in 3 points of characteristics
- Calculating errors in the points.

For temperature sensors pairs in combine meters following operations (and time for it) are needed:

- connections and waiting for stabilizing temperature of immersed sensors - t_{11}
- bath temperature measurement - t_{12}
- multiplexing - t_{13}
- sensor resistance measurement - t_{14}
- temperature measurement in baths - t_{15}
- change of temperature in baths - t_{16}
- disconnections of heat meters - t_{17}
- calculation of maximum error for one pair - t_{18}
- calculation of error for one pair using method of maximum search - t_{19}

Checking calculator error on separate stand:

- with automatic connection of resistors simulating sensors in 3 measurement points - t_{20}
- hand operated - t_{21}

In Table 3 are presented values of time needed for the operations when validating typical combine heat meter.

Table 3. Values of time needed for the operations when validating typical combine heat meter (time in seconds)

t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}
60	10	20	20	10	1800	30	180	20	30	180

Time needed for validation for the following methods:

- one after another

$$t_{\text{combined1}} = t_{11} + t_{12} + t_{13} + t_{14} + t_{15} + 3t_{16} + t_{17} + t_{18} + t_{21} \quad (29)$$

- automatically in group of 6 items with hand operated calculator validation

$$t_{\text{combined2}} = [8(t_{11} + t_{12} + t_{13} + t_{14} + t_{15} + t_{17} + t_{18} + t_{21}) + 3t_{16}] / 8 + t_{21} \quad (30)$$

- automatically in group of 8 items without maximum error search method

$$t_{\text{combined2}} = [8(t_{11} + t_{12} + t_{13} + t_{14} + t_{15} + t_{17} + t_{18}) + 3t_{16}] / 8 + t_{20} \quad (31)$$

- automatically in group of 8 items with maximum error search method

$$t_{\text{combined2}} = [8(t_{11} + t_{12} + t_{13} + t_{14} + t_{15} + t_{17} + t_{19}) + 3t_{16}] / 8 + t_{20} \quad (32)$$

Table 4. Values of time needed for validation (time in seconds, recalculated for one meter)

$t_{\text{combined1}}$	$t_{\text{combined2}}$	$t_{\text{combined3}}$	$t_{\text{combined4}}$
4110	960	810	650

6 Conclusions

In present paper solutions that are emerging in validating heat meters and their impact on efficiency were analyzed. Understanding these issues and their mathematical derivations and algorithms are essential for later works on this matter. The author proposed methods of analysis for calculation of errors needed for validation of heat meters. The author have undertaken the experimental trials, which preliminarily prove the presented approach being very promising. In the next step analysis of automatic connections and disconnections of heat meters from the system are planned.

References

1. European Standard EN 1434-5:1997/A1: 2002 (E) Heat meters – Part5: Initial verification tests
2. European Standard EN 60751 (IEC Publication 751) Industrial platinum resistance thermometer sensors

3. Organisation Internationale De Metrologie Legale. International Recommendation. Heat meters. OIML R 75. Edition 1988 (E). Paris (1988)
4. Goszczyński, T., Korytkowski, J.: System for validation of hybrid heat meters. Pomiary Automatyka Robotyka 9, 11–12 (2006) (in Polish)
5. Goszczyński, T.: System for compact heat meters measurements.: PL-205079 – Patent Office RP
6. Tegeler, E., Heyer, D., Siebert, B.: Uncertainty of the calibration of paired temperature sensors for heat meters. In: Tempmeko 2007 Conf., Lake Louise, Canada (2007),
<http://www.tempmeko2007.org>
7. Goszczyński, T.: Error Determination for Heat Meter Validation. Heat Transfer Engineering 31(1), 83–89 (2010)
8. Korytkowski, J., Goszczyński, T., Jachczyk, E.: Methods of computer testing system for examining the accuracy of temperature probes of heat meter. In: Automation 1998 Conf., Warsaw, Poland (1998) (in Polish)
9. Jachczyk, E.: Around MID Directive - harmonized standard PN-EN 1434 “Heat meters”. Pomiary Automatyka Robotyka 13(2), 40–44 (2009) (in Polish)

Influence of Stresses on Magnetic B-H Characteristics of X30Cr13 Corrosion Resisting Martensitic Steel

Dorota Jackiewicz¹, Roman Szewczyk²,
Jacek Salach², Adam Bieńkowski², and Maciej Kachniarz¹

¹ Industrial Research Institute for Automation and Measurements, Warsaw, Poland
d.jackiewicz@mchtr.pw.edu.pl, kachniarz.maciej@gmail.com

² Institute of Metrology and Biomedical Engineering, Warsaw University of Technology
Warsaw, Poland

{r.szewczyk,j.salach,a.bienkowski}@mchtr.pw.edu.pl

Abstract. Paper presents the frame-shaped cores based methodology of testing of magnetoelastic characteristics of energetic steels such as X30Cr13 steel, subjected to tensile stresses. In presented method, the magnetic circuit of the sample is closed. For this reason, the results of magnetoelastic investigation are independent of the shape of the sample. To validate the proposed method, the influence of tensile stresses on B(H) hysteresis loop of X30Cr13 martensitic corrosion resistant steel was carried out. On the base of these result, clear criteria for non-destructive assessment of mechanical stresses in the material were determined.

Keywords: magnetoelastic effect, corrosion resistant steel, stress assessment.

1 Introduction

Under the influence of mechanical stresses, the shape of $B(H)$ magnetic hysteresis loop changes significantly for both crystalline [1] and amorphous [2] magnetic materials. This effect is commonly known as magnetoelastic effect [3, 16]. In spite of the fact that changes of flux density B under stresses are most significant for high permeability amorphous and nanocrystalline alloys [4, 5, 6], the magnetoelastic effect is also observed in steels [7, 8, 9]. As a result, measurements of stress-induced changes of shape of hysteresis loop of steel, creates the possibility of non-destructive assessment of mechanical stresses generated during the use mechanical component.

Magnetoelastic-effect based non-destructive testing of corrosion resisting steels can be considered only for the martensitic steels, such as X30Cr13 steel. Austenitic steels are non-magnetic, however, under specified mechanical stresses austenite may be converted to martensite [10, 11, 12]. This phenomenon is also very interesting from the non-destructive testing point of view, but its principle is different than magnetoelastic effect in martensitic steels.

Martensitic corrosion resisting X30Cr13 steel exhibits good resistance to the corrosive effect of salt water, weather and various corrosion agents. Moreover, this steel

exhibit good resistance to high temperature oxidation up to approximately 750°C, which makes it suitable for development of energetic turbines elements. However, the lack of standardized method of measurements of magnetoelastic characteristics of energetic steels as well as lack of information about detailed magnetoelastic characteristics of specific types of energetic steels is the significant barrier in industrial application of magnetoelastic effect for non-destructive testing of materials. This paper is trying to fill both of these gaps.

2 Methodology of Measurements

For effective measurements of magnetoelastic characteristics of soft magnetic materials, the methodology of measurements has to fulfill following requirements:

- magnetic circuit of the tested sample has to be closed. If magnetic circuit is open (e.g. strip samples are used) demagnetization energy appears [3]. For this reason, if magnetic circuit is open, the results of measurements are strongly dependent on the shape of the sample. Moreover, open magnetic circuit reduces permeability and stress sensitivity of the sample,
- possibility of winding of the sample by magnetizing and sensing windings has to be provided. This condition is especially important during the testing of magnetoelastic properties of steel-based elements. Such materials exhibit significant coercive force and requires large number of turns of magnetizing coil,
- uniform distribution of stresses has to be assured during the magnetoelastic testing. If stress distribution is non-uniform or both compressive and tensile stresses appear, possibility of physical interpretation of experimental results is limited. In the worst case, when both tensile and compressive stresses appear [13], stress sensitivity of the sample may be misleading.

To fulfill above requirements during the magnetoelastic tests, special methodology utilizing the frame-shaped samples was developed. Physical dimensions of the martensitic corrosion resisting X30Cr13 steel based sample are presented in figure 1a.

In proposed method, both sensing and magnetizing windings are made on the columns of the sample. Furthermore, magnetizing and sensing windings should be made on both columns as it is presented in figure 1a. To minimize demagnetization effects, sensing winding should be made under the magnetizing winding.

To analyse the distribution of flux density in proposed frame-shaped sample, the finite elements based simulation was performed. The results of this simulation are presented in figure 1b. It should be indicated, that nearly uniform distribution of flux density B in the columns of the sample was observed. This phenomenon is very useful from the point of view of further interpretation of experimental results.

Figure 2a presents the mechanical force reversing system enabling generation of uniform tensile stresses in frame shaped sample. With the use of proposed system, the tensile stresses may be generated by oil press providing compressive force. The schematic diagram of oil press with proposed mechanical force reversing system is presented in figure 2b, whereas photography of whole system is presented in fig. 2c.

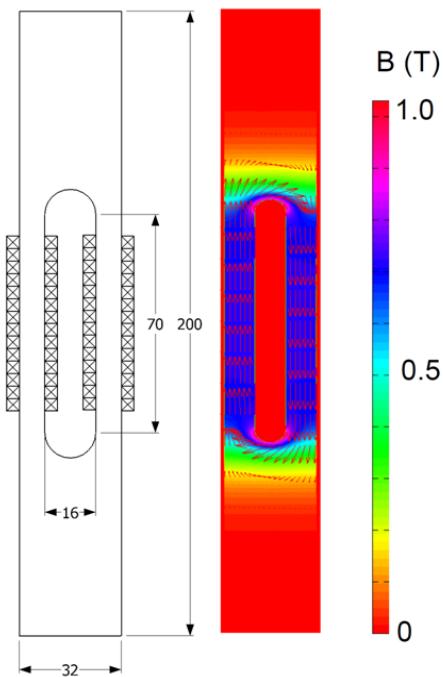


Fig. 1. Frame-shaped sample: a) physical dimensions of the X30Cr13 steel, b) results of finite elements method based simulations of flux density B distribution in frame-shaped sample with constant magnetic amplitude permeability

It should be indicated, that the main elements of mechanical force reversing system are made of non-magnetic materials such as aluminum. As a result, influence of large elements of mechanical system on the results of magnetomechanical measurements may be neglected.

Figure 3 presents the schematic block diagram of computer-controlled experimental system for magnetoelastic properties of construction steel testing. Personal computer equipped in type NI-PCI-6221 National Instruments data acquisition card is driving BOP36-6 voltage to current converter produced by KEPCO. Magnetizing winding of the tested sample is driven by the output of this voltage to current converter to provide controlled waveform of driving current.

Signal from sensing winding is provided to the input of Lakeshore 480 fluxmeter. As a result, the real-time information about flux density B in the sample is given to the analog input of data acquisition card. Therefore $B(H)$ hysteresis loop may be determined.

During the tests, the compressive force F is provided to the mechanical force reversing system by the oil press. This oil press is manually controlled by the operator. However, construction of oil press gives possibility of keeping constant value of compressive force, which may be also measured by measurement of the pressure in the oil chamber of oil press.

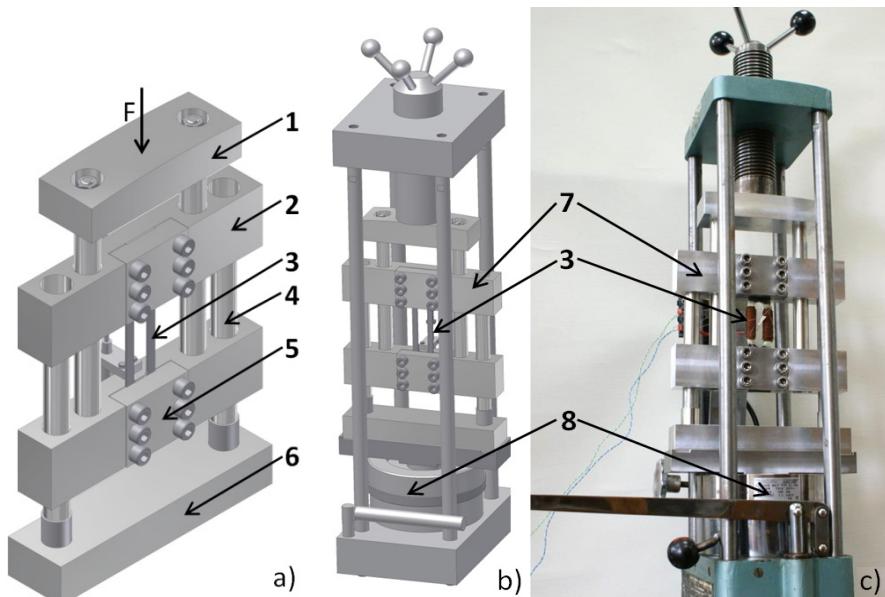


Fig. 2. Mechanical setup for testing the magnetic and magnetoelastic properties of frame-shaped samples: F – compressive force, a) the mechanical force reversing system: 1 – upper bar, 2 – moving bar, 3 – tested frame-shaped sample, 4 – cylindrical columns, 5 – sample holder, 6 – base of the device, b) schematic diagram of oil hydraulic press, c) photography of mechanical setup for testing the magnetic and magnetoelastic properties of frame-shaped samples: 7 – reversing system, 8 – force sensor

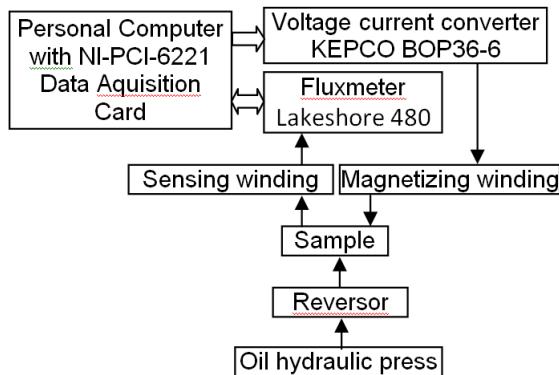


Fig. 3. The schematic block diagram of computer-controlled experimental system for magnetoelastic properties of construction steel testing

3 Experimental Results

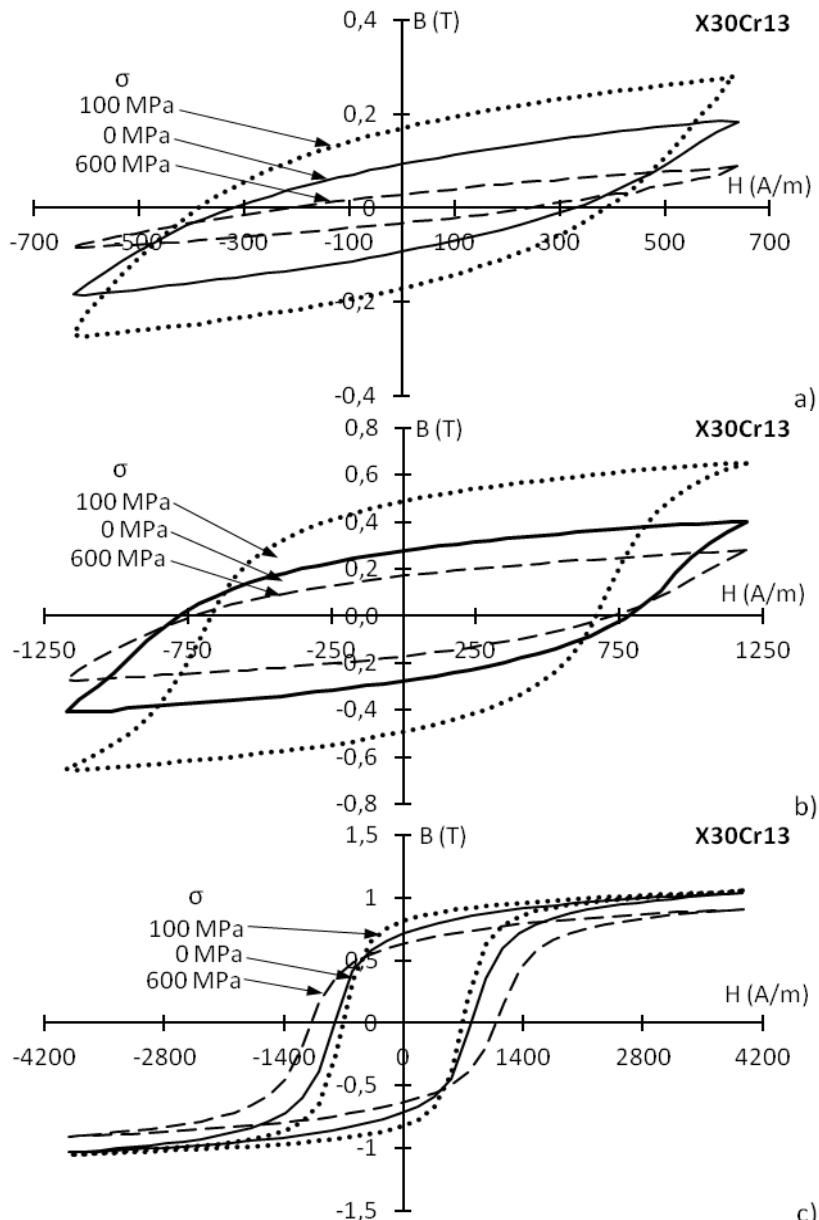


Fig. 4. The tensile stresses dependence of magnetic $B(H)$ characteristics of frame-shaped samples made of X30Cr13 steel, for the three amplitudes H_m of magnetizing field: a) $H_m = 640 \text{ A/m}$, b) $H_m = 1200 \text{ A/m}$, c) $H_m = 4000 \text{ A/m}$

The influence of tensile stresses up to 600 MPa on the shape of $B(H)$ hysteresis loop of frame-shaped samples made of the martensitic corrosion resistant X30Cr13 steel is presented in figure 4. This influence was tested for diversified values of the amplitude of magnetizing field H_m : 640 A/m, 1200 A/m and 4000 A/m respectively. It should be highlighted, that under the compressive stresses σ , the maximal value of flux density B achieved for given amplitude of magnetizing field H_m first increase and then decrease. Similar effect is connected with the total field of hysteresis loop, determining total losses during the full cycle of magnetization of the sample. These losses first increase and, for tensile stresses σ exceed 100 MPa, significantly decrease.

Figure 5 presents the influence of tensile stresses σ on the value of maximal flux density B in the tested frame-shaped core achieved under the amplitude H_m of magnetizing field. These magnetoelastic characteristics clearly indicate, that under tensile stresses σ , maximal flux density B characteristic is non-monotonous. Extreme on the $B(\sigma)_{Hm}$ characteristic is called magnetoelastic Villari [14, 15] point and for martensitic corrosion resisting steel X30Cr13 was reached for σ equal to about 150 MPa.

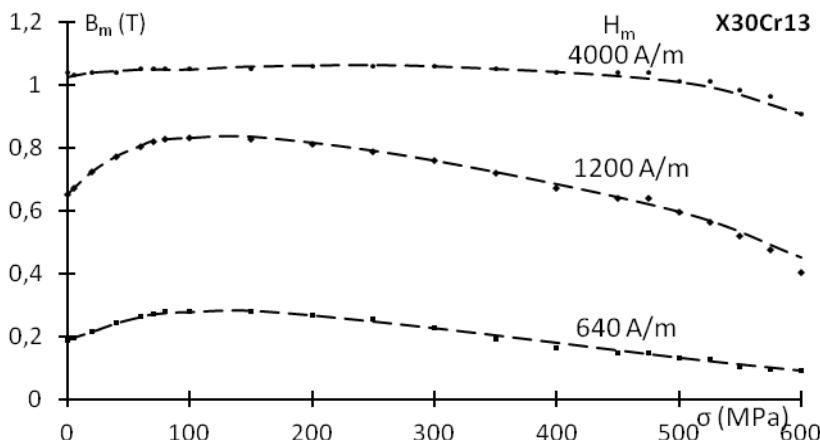


Fig. 5. Magnetoelastic $B(\sigma)_{Hm}$ characteristics of frame-shaped sample made of X30Cr13 steel under the tensile stresses. Sample was tested for different values of amplitude of magnetizing field H_m . Magnetoelastic Villari point may be observed for stresses about 150 MPa

It should be noted, that magnetoelastic Villari point is better visible for lower values of amplitude of magnetizing field H_m . This phenomenon is connected with participation of magnetoelastic energy and magnetization energy in the total free energy of the material of tested sample. For lower values of the amplitude of magnetizing field H_m , participation of magnetoelastic energy is much higher and phenomena connected with magnetoelastic effect are clearly visible.

From practical point of view, appearance of magnetoelastic Villari point may be used for determination of clear criteria for non-destructive test. For given value of the amplitude of magnetizing field H_m , e.g. 1200 A/m, the increase of flux density B above its initial value indicates, that sample is subjected to acceptable values of tensile stresses σ . On the other hand, if maximal value of flux density B achieved for

amplitude of magnetizing field H_m equal 1200 A/m is lower than its initial value, the clear signal is given, that stresses in the sample are near the region of plastic deformation or rupture of the element. Such criteria enable novel approach to utilization of magnetoelastic effects in non-destructive testing of construction elements made of the martensitic corrosion resistant X30Cr13 steel.

4 Conclusions

Presented method of magnetoelastic testing of frame-shaped samples made of corrosion resistant, martensitic construction steels, opens the possibility of filling the gap connected with the lack of information about their magnetoelastic characteristics. Due to the fact, that magnetic circuit of the tested samples is closed, the measured magnetoelastic characteristics are independent of the shape of the sample. As a result, the non-destructive test oriented methodology of determination of magnetoelastic characteristics of the construction steels was created.

Presented results indicate, that magnetoelastic characteristics of martensitic corrosion resistant X30Cr13 steel under tensile stresses σ , are non-monotonous. Extreme on the $B(\sigma)_{H_m}$ characteristic (so called magnetoelastic Villari point) appears for tensile stresses σ equal to about 150 MPa.

The fact, that magnetoelastic Villari point appears on magnetoelastic characteristics of X30Cr13 steel may be utilized for determination of the clear criteria for non-destructive tests. For amplitude of magnetizing field H_m equal 1200 A/m, the increase of flux density B above its initial value indicates, that sample is subjected to acceptable values of tensile stresses σ . If maximal value of flux density B achieved for this amplitude of magnetizing field H_m is lower than value for unstressed sample, the clear signal is given, that stresses in the sample are near the region of plastic deformation or rupture of the element. It should be highlighted, that presented criteria enable practical utilization of magnetoelastic effects in non-destructive testing of energetic turbines elements made of the martensitic corrosion resistant X30Cr13 steel.

Acknowledgment. This work was partially supported by The National Centre of Research and Development (Poland) within grant no. PBS1/B4/6/2012.

References

1. Aphrodite, K., Evangelos, H.: Stress Dependent Magnetization and Vector Preisach Modeling in Low Carbon Steels. *IEEE Transactions on Magnetics* 48, 1433–1436 (2012)
2. Stoklosa, Z., Rasek, J., Kwapiński, P., Badura, G., Haneczok, G., Pająk, L., Lelątko, J., Kolano-Burian, A.: Magnetic, Electrical and Plastic Properties of Fe76Nb2Si13B9, Fe75Ag1Nb2Si13B9 and Fe75Cu1Nb2Si13B9 Amorphous Alloys. *Journal of Alloys and Compounds* 509, 9050–9054 (2011)
3. Jiles, D.C.: *Introduction to Magnetism and Magnetic Materials*. Chapman&Hall, London (1998)

4. Bieńkowski, A., Szewczyk, R.: The Possibility of Utilizing the High Permeability Magnetic Materials in Construction of Magnetoelastic Stress and Force Sensors. *Sensors and Actuators A* 113, 270–276 (2004)
5. Mamalis, A.G., Hristoforou, E.: On the Magnetic and Magnetoelastic Uniformity Measurements on Fe₇₈Si₇B₁₅ Amorphous Ribbons and Wires. *Materials Science Forum* 670, 87–91 (2010)
6. Hlenschi, C., Corodeanu, S., Chiriac, H.: Magnetoelastic Sensors for the Detections of Pulse Waves. *IEEE Transactions on Magnetics* 49, 117–119 (2013)
7. Sablik, M.J., Augustyniak, B., Chmielewski, M.: Modeling Biaxial Stress Effects on Magnetic Hysteresis in Steel with the Field and Stress Axes Noncoaxial. *Journal of Applied Physics* 85, 4391 (1999)
8. Wang, G.D., Wang, M.L., Zhao, Y., Chen, Y., Sun, B.N.: Application of magnetoelastic stress sensors in large steel cables. *Smart Structures and Systems* 2, 155–169 (2006)
9. Song, Y.S., Ding, Y.L.: Fatigue monitoring and analysis of orthotropic steel deck considering traffic volume and ambient temperature. *Science China-Technological Sciences* 56, 1758–1766 (2013)
10. Solomon, N., Solomon, I.: Deformation Induced Martensite in AISI 316 Stainless Steel. *Revista de Metalurgia* 46, 121–128 (2010)
11. Ueji, R., Takagi, Y., Tsuchida, N., Shinagawa, K., Tanaka, Y., Mizuguchi, T.: Crystallographic Orientation Dependence of ε Martensite Transformation During Tensile Deformation of Polycrystalline 30% Mn Austenitic Steel. *Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing* 576, 14–20 (2013)
12. Hilkhuizen, P., Geijsselaers, H., Bor, T., Perdahcioglu, E., Boogaard, A., van den Akkerman, R.: Strain Direction Dependency of Martensitic Transformation in Austenitic Stainless Steels: The Effect of Gamma-Texture. *Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing* 573, 100–105 (2013)
13. Mohri, K., Sudoh, E.: New Extensometer Using Amorphous Magnetostrictive Ribbon Wound Cores. *IEEE Transactions on Magnetics* 17, 1317–1319 (1981)
14. Bieńkowski, A., Szewczyk, R., Salach, J., Kolano-Burian, A.: The Magnetoelastic Villari Effect in Fe₂₅Ni₅₅Si₁₀B₁₀ Amorphous Alloy Subjected to Thermal Treatment. *Reviews on Advanced Materials Science* 18, 561–564 (2008)
15. Bienkowski, A., Szewczyk, R., Kolano, R.: Influence of thermal treatment on magnetoelastic Villari effect in Fe₇₈Si₁₃B₉ amorphous alloy. In: 11th International Conference on Rapidly Quenched and Metastable Materials, *Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing*, vol. 375, pp. SI: 1024–SI: 1026 (2004)
16. Salach, J., Bieńkowski, A., Szewczyk, R., Jackiewicz, D., Frydrych, P., Kolano-Burian, A.: Utilizing of magnetic amorphous alloys in magnetoelastic tensile stress sensors. *Pomiary Automatyka Robotyka* 16(2), 556–560 (2012) (in Polish)

Anode Current Control in the Microwave Heating Equipment

Marek Kuna-Broniowski and Piotr Makarski

University of Life Sciences, Department of Technology Fundamentals, Lublin, Poland
marek.kunabroniowski@gmail.com

Abstract. The correct power density of microwave radiation is decisive as regards the quality of the raw material undergoing thermal processing. It is of particular significance especially in case of materials sensitive to overheating. Therefore, the power of falling microwave radiation should be controlled using a stepless method and within broad limits.

Keywords: magnetron power control, microwave cooker, thermal processing.

1 Introduction

The microwave heating equipment has numerous applications for heating a wide range of products owing to such properties as: heating the object in the entire volume, high power efficiency, and process speed. Special benefits of microwave heating are observed in case of drying, where thanks to the appropriate selection of exposure parameters we can achieve proper formation of capillaries without closing them on the surface. Owing to that the drying process is quick allowing free evacuation of steam from the inside of the object to the outside without destroying material structure [5]. However, in order to ensure the optimum course of the process it is necessary to provide the heating power to the object in a precise and smooth way – depending on the object mass as well as the required drying characteristics [7, 4]. Otherwise the advantageous properties of microwave heating will be destroyed.

In the laboratory experiments a typical microwave cooker is used for this purpose due to a low price as well as that it can be used as a microwave drier without altering the mechanical construction (it is equipped with a rotary plate where samples can be placed as well as forced air circulation used to remove released steam). However, this method has disadvantages which limit its use in a significant way, in particular in case of small sampled or sensitive biological materials. Following a number of tests it was demonstrated that the power generated by the magnetron is usually too high because thermal decomposition of samples takes place. What is more, most frequently used control method i.e. setting a lower power on the cooker panel does not bring desired results either. It is caused by the fact that the control of medium power microwaves involves cyclical switching on and off the magnetron where the on-time is several seconds. This time is long enough to allow some materials to undergo thermal decomposition or to burst as a result of excessive steam pressure that is generated

inside them during the operation of the magnetron at full capacity. In order to eliminate the above faults of the typical microwave heating equipment there is a need to limit the magnetron power as well as to ensure its smooth power control. According to the authors study, with the magnetron power exceeding 300 W each sample underwent destruction of organic material. The problem of the magnetron power control is not new and there are a few typical solutions to this problem that are presented below.

2 Commonly Used Control Methods of the Microwave Heating Equipment

The following methods are used to adjust the amount of microwave energy falling on the raw material during thermal processing:

- In multi-magnetron systems each magnetron or a pair of magnetrons are supplied from separate power supplies. The adjustment of electromagnetic radiation power is achieved by switching on and off individual magnetrons (or their pairs).
- In single-magnetron systems or with a pair of magnetrons supplied from a transformer with balanced winding, the adjustment to the average power of microwaves is achieved by cyclical switching on and off of the transformer power supply (with the duty cycle independent from the desired average power of microwaves). The time of individual commutating cycles ranges from a few to a few dozen seconds.
- In the systems with a few power supplies working in parallel for an individual magnetron, the power control is carried out by switching on or off individual power supplies.

Despite the simplicity and reliability of these magnetron supply systems they have one significant drawback, namely that the power control is effected in a stepwise manner (multi-magnetron systems or/and power supplies working in parallel) or it is an on-off type (single-magnetron systems or composed of a pair of them, in the system with a transformer with balanced winding).

Nevertheless, none of the above methods ensures smooth power control, its lack in these systems may not allow thermal processing of materials sensitive to overheating.

Another way is inclusion cascading in the path of waveguide – the adjustable microwave attenuator. Due to the high power output of the magnetron (about 1 kW), power control radiation in this way, it becomes little useful [1, 2]. The relatively high power dissipated in the attenuator poses serious constructional problems and troubles with effective heat dissipation [6]. Also other ways used to suppress power, such as variable resistance based on PIN diode or FET transistors, for this power range – cannot be realized. Only properly designed attenuator with waveguide operating below cutoff frequency, or attenuator with a lossy wedge are able to work with such a large powers. However, using the attenuators, the power control is mechanically adjusted, by changing the position of the relevant elements. In practice, this prevents the automatic execution of process.

3 Smooth Power Control of Microwave Heating Systems

The system of the smooth power control of the magnetron presented in this study originated as a result of research on the method of microwave drying of small portions of organic materials. What the authors originally did was to supply an anode transformer with reduced voltage from the control autotransformer while maintaining the rated voltage of the cathode heater current. However, the disadvantage of this method was that the generated microwave power depended on voltage fluctuations in the mains supply and required frequent adjustments effected by the change to the position of the autotransformer knob. This hampered the automation of the control process and apart from that due to the size it was not possible to build-in the autotransformer to the inside of the cooker. Therefore, the authors designed a smooth control magnetron power system in the form of additional modules of small sizes that can be embedded in the structure of a typical microwave cooker.

This solution uses an original magnetron anode supply system, a special controller designed by the author and a separate magnetron cathode heater current transformer. The controller operates based on a thyristor-controlled phase regulator with current feedback. Owing to that the average value (for each half period) of the anode power supply current (i.e. the average value of the generated microwave power as well) remains constant disregarding the voltage oscillations in the mains and the smooth control of the value of this current is carried out by the voltage change at the control input. Such a way of the magnetron power control is the subject of the patent application by the authors.

3.1 Magnetron Supply in the Microwave Heating Systems

In the commonly used microwave heating systems the anode circuits of magnetrons are supplied using step-up transformers with a balanced output (dual-magnetron system – Fig. 2) or transformers with an unbalanced output and a rectifier in the voltage doubler (single-magnetron system – Fig. 1).

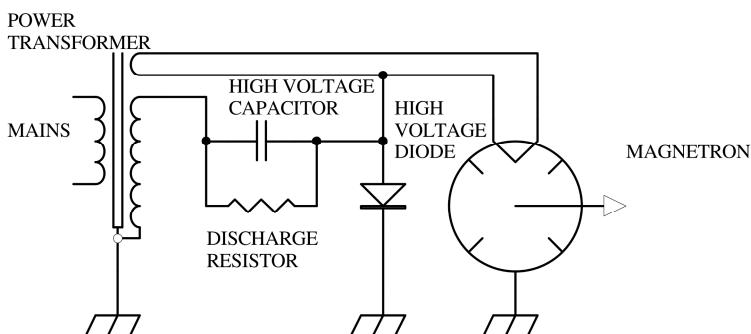


Fig. 1. Transformers magnetron supply with an unbalanced output and a rectifier in the voltage doubler

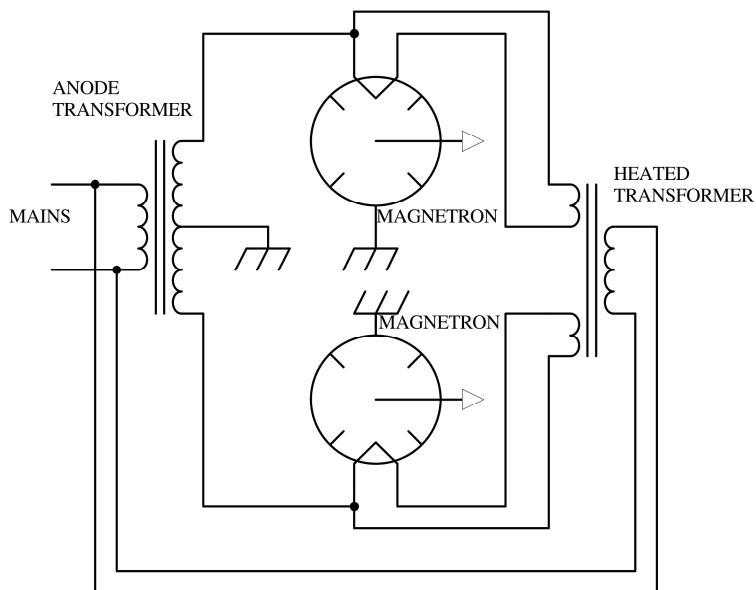


Fig. 2. Dual magnetron transformers power supply

The current and voltage characteristic of the magnetron shows a small motional resistance at the voltage drop of a few kV [3]. An example of characteristic for 2M218 magnetron used in microwave cookers, determined by the authors is presented in Fig. 3.

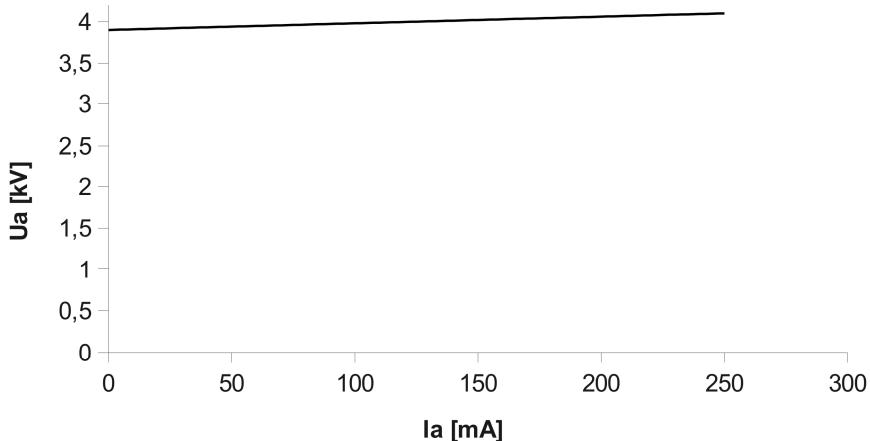


Fig. 3. Current and voltage characteristic of 2M218 magnetron (authors' own research)

Within the range of 0 – 250 mA of anode currents the anode voltage changes in a linear way within the range 3.9 – 4.1 kV. Therefore, the relative change of voltage is only 4.88%, whereas the dynamic resistance of the magnetron is 800Ω . In order to

ensure proper supply to such a receiver we need a power supply with a soft voltage current characteristic. It is accomplished by the use of transformers with a high leakage reactance or by including a series reactor from the mains side. For example, transformer – type YHAI-800NTC used in microwave cookers has a short-circuit voltage of 37% (according to the research conducted by the authors).

3.2 Smooth Power Control of the Magnetron

The smooth power adjustment system of the magnetron was designed and constructed by the authors operates based on the phase control of the anode transformer. However, unlike classical phase controllers regulating the mean or root mean square value of voltage, this system maintains the set mean current value (the switching angle of the thyristor switch is being adjusted on an ongoing basis in order to ensure that the half-cycle current value is constant).

The block diagram is presented in Fig. 4.

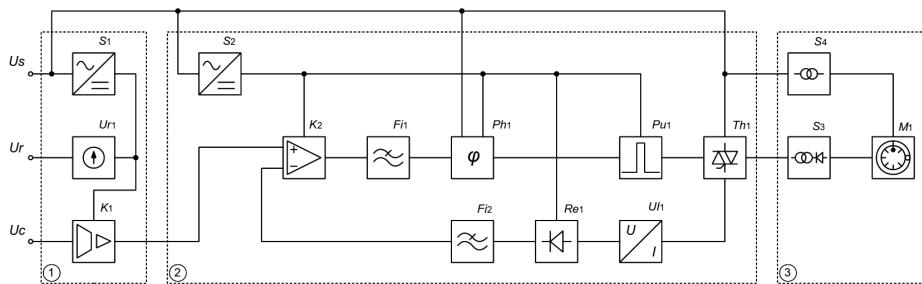


Fig. 4. Block diagram of the smooth power control of the magnetron

The system consists of 3 basic units: galvanically isolated control circuit (1), phase controller with the current feedback (2) and a magnetron power supply system (3). Unit (1) contains an isolation amplifier K_1 and a source of reference voltage U_{r1} , supplied from power supply S_1 . Thanks to that the control voltage input U_c and the reference voltage output U_r are galvanically isolated from the mains inputs U_s . The input power adjustment of the magnetron is carried out by the change to the input voltage U_c within the range of 0 – 10V. If a potentiometer is used for this purpose, the reference voltage output U_r of 10V can be used. The phase controller unit (2) cooperating with the power supply system of the magnetron (3) has an external characteristic resembling that of the current source. The phase detector Ph_1 via the gate pulse generator Pu_1 controls the thyristor switch Th_1 . The release angle of this switch depends on the control voltage at the phase detector input. This voltage is an output voltage of the differential amplifier K_2 filtered by the low-pass filter F_{l1} . The current-voltage transducer UI_1 generates the voltage signal directly proportional to the instantaneous value of the load current. This signal is rectified in the active rectifier Re_1 and undergoes filtration in filter F_{l2} releasing a constant – voltage component directly proportional to the half-cycle value of the load current. This voltage appears at the inverting input of the amplifier K_2 and constitutes a signal

compensating the changes to the mean value of the load current. The reference signal is voltage at the non-inverting input of the amplifier. The anode power supply S3, supplying the magnetron M1, from the input side is connected to the output of the thyristor switch Th1. The cathode filament circuit of the magnetron is supplied from a separate power supply S4 and not from a separate anode winding as it normally occurs in microwave cookers. It is necessary as the effective voltage value at the anode transformer input can change within broad limits depending on the set value of the anode current. Dependence of the anode current of the magnetron I_0 in the control voltage function U_s .

The nonlinear course of the characteristic (Fig. 5) results mainly from the fact that the current feedback signal comes from the primary winding current of the anode transformer and not from the secondary winding current (such a solution is simpler because there is no problem of galvanic isolation at the voltage of a few kilovolts).

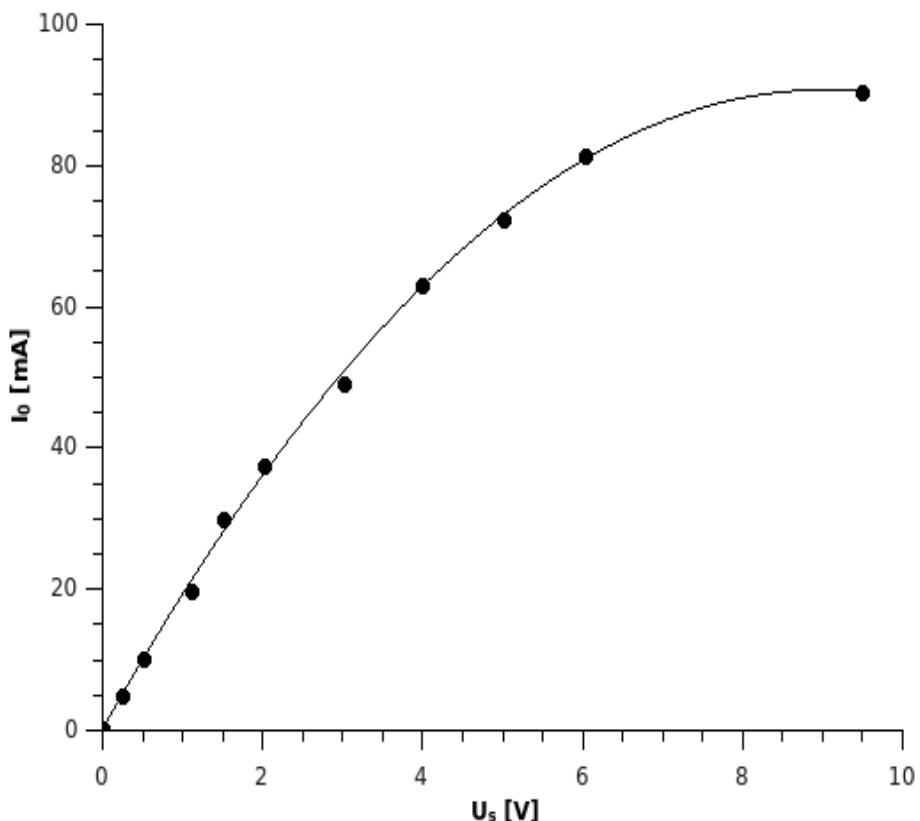


Fig. 5. Control characteristic of the smooth power adjustment of the magnetron

The average half-cycle value of the current flowing through the primary winding is not in direct proportion to the analogical value of the secondary winding current value (i.e. also to the average value of the magnetron current) due to a relatively high value

and nonlinear dependence in the no-load current function of the transformer. Under operating conditions this nonlinearity has no significance, it can be taken into account while scaling control elements.



Fig. 6. Anode current control installed in the microwave heating equipment

4 Conclusions

The presented power control system of the magnetron was embedded into a typical microwave cooker and has been used for over a year. It ensured, among others, precise drying of biological materials and trouble-free processing of food materials which basically cannot be processed in the microwave cooker. The smooth control of the heating power is possible for very small levels ranging from 1-2 W to 1 000 W. The power control can be smooth according to the set heating curve, which is of particular importance for the drying systems, as well as stepwise, involving fast increase and equally fast decrease of power.

References

1. Dobrowolski, J.: Technika wielkich częstotliwości, OWPW – 1 wyd, 170–172 (2001)
2. Galwas, B.: Miernictwo mikrofalowe, WKiŁ – 1 wyd, 87–92 (1985)

3. Hauser, J.: Podstawy elektrotermicznego przetwarzania energii, Zakład Wydawniczy K. Domke – 2 wyd. popr, 334–335 (1996)
4. Jałoszyński, K., Figiel, A.: Drying kinetics and antioxidant activity of oregano. *Acta Agrophysica* 11(1), 81–90 (2008)
5. Marzec, A., Zadrożna, M.: Wpływ mocy mikrofal na jakość suszu jabłkowego. *Acta Agrophysica* 12(2), 457–467 (2008)
6. Thomas, H.E.: Techniki i urządzenia mikrofalowe, WNT – 1 wyd, 143–150, 192 (1978)
7. Warchałewski, J.R., Dolińska, R.: Analiza mikroskopowa ziarna pszenicy dwu pokoleń wyhodowanych z nasion ogrzanych mikrofalami. *Acta Agrophysica* 10(3), 727–737 (2007)

The High-Resolution Camera in Estimation of the Position of the Hydraulic Valve Spool

Piotr Owczarek¹, Dominik Rybarczyk¹, Jarosław Gośliński²,
and Adam Owczarkowski²

¹ Institute of Mechanical Technology,

² Institute of Control and Information Engineering,

Poznan University of Technology,

ul. Piotrowo 3, 60-965 Poznań, Poland

{piotr.owczarek,dominik.rybarczyk}@put.poznan.pl,

{jaroslaw.a.goslinski,

adam.j.owczarkowski}@doctorate.put.poznan.pl

Abstract. In the paper, the vision system for position measurement of the slider of proportional valve is presented. The valve spool's slider was connected with a synchronous motor. The position was detected by industrial camera with a resolution of 1.6x1.2 kpixels. The research focused on the correctness of detection of the slider position. The use of the lens with the control aperture allowed to separate background from the object of research. The study confirms that it is possible to use vision system for measuring position of actuators in any hydraulic device. Future study will be narrowed to a measurement of a hydraulic cylinder pistons displacement in an application of a electro-hydraulic manipulator. This can significantly reduce the cost of mechanical sensors as well as shorten the overall time.

Keywords: vision system, proportional valve spool, synchronous motor.

1 Introduction

Precision of the hydraulic servo drives positioning depends on the hydraulic amplifier parameters, the feedback system and the control system. Traditional measurement systems mounted in the hydraulic cylinder or valve are very expensive and can bring a potential problem with the assembly. Therefore it is important to search for new and inexpensive ways of control methods. The authors propose the use of the vision feedback in a electro-hydraulic proportional valve, controlled by low-power synchronous motor.

In the literature, several different applications of visions based control system can be found. The development of a position tracking system for a hand-held tool based on the low-cost sensors and vision system was described in [1]. In [2] the vision based control systems for the low frequency material testing was presented. Authors stressed the importance of low frame rate and the delay of image processes. These are the main reasons to degrade the control performance. Also it was mentioned that these

factors had no significant influence in the case of slow changing process. The mechanical sensor was replaced by CCD camera. In different work [3] implementation of a vision system in a micro aerial vehicle to image processing and assist an autopilot system was described. In the article [4] three problems were described such as uncertain visual measurements, different sampling rates and compensation of the sensor delay. The camera was mounted on end-effector of 5-DOF parallel hydraulic manipulator. Vision system based on the generalized Hough transform to find the position and the orientation of hydraulic components was described [5]. The vision system sent information about position of product to controller of mechanical arm.

In the article authors propose use of a vision system for measurement of position of a slider of valve spool. Next the vision system will be used in order to measure the length (displacement) of hydraulic cylinder piston in electro-hydraulic manipulator.

2 Mathematical Background of Image Processing

A position estimation was done with use of the gradient edge detection method. This is a popular method in image processing and it is given by equation:

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \quad (1)$$

Detection of edge in axis x (fig. 1.a):

$$\nabla f = \left[\frac{\partial f}{\partial x}, 0 \right] \quad (2)$$

Detection of edge in axis y (fig. 1.b):

$$\nabla f = \left[0, \frac{\partial f}{\partial y} \right] \quad (3)$$

Also it is possible to compute the gradient direction (fig. 1.c):

$$\theta = \tan^{-1} \left(\frac{\partial f}{\partial x} / \frac{\partial f}{\partial y} \right) \quad (4)$$

The edge strength is given by the gradient magnitude:

$$\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \quad (4)$$



Fig. 1. The direction of the gradient

The value of edge strength parameter is crucial. If this value is too big, the detection of edge will be impossible. On the other hand, if this value is too small, all edges will be detected, that can be cause by noisy signal.

3 The Experimental Stand

The test stand consists of a proportional valve with synchronous motor, an industrial controller and a vision system. The vision system was based on high resolution camera. The resolution of camera was 1600x1200 pixels. The focal length of the lens was 16 mm. This camera was equipped with high-performance processor and real-time operating system. Application was written in NI Vision Builder [6]. The camera observes the valve spool movement area (see fig. 1).

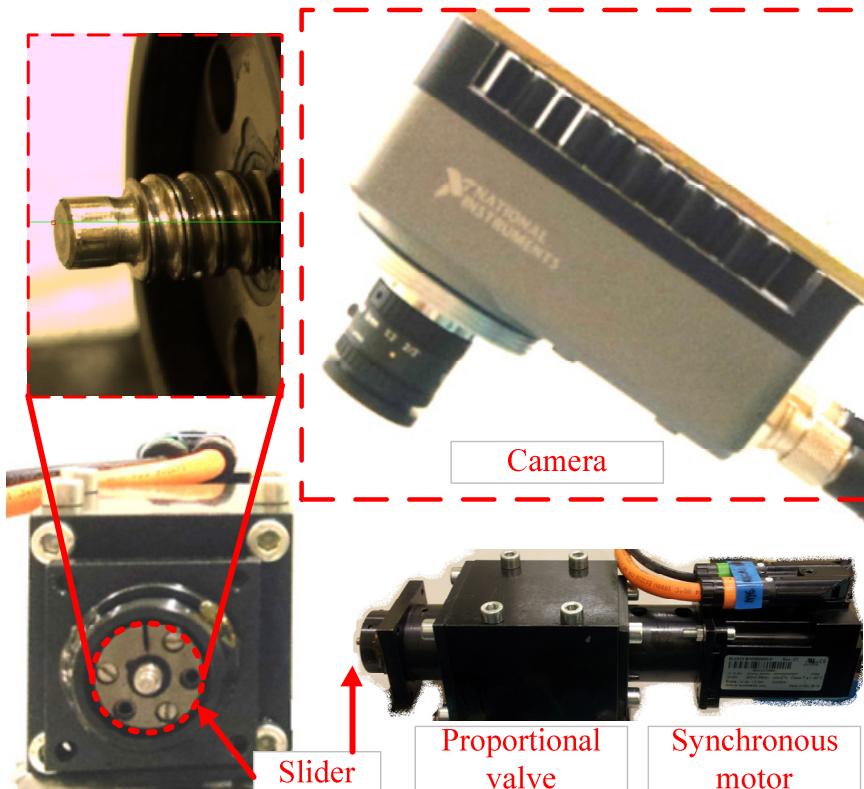


Fig. 2. Test stand view

During the experiment the slider's position of the proportional valve spool was measured. The valve's slider was moved by the synchronous motor, type PMSM (permanent-magnet synchronous motor). The connection between the slider and the motor was realized by coupling bellows. Rotation of the motor shaft is change to linear movement of the slider by ball-gear. As it can be seen in fig. 3, the movement of the valve spool's slider allows to open or close valve gaps.

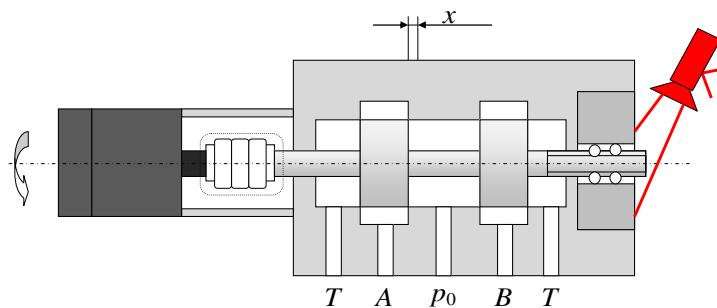


Fig. 3. Valve construction schema

4 Experimental Results

The aim of the experiment was to verify the estimation's precision of valve spool's slider position, realized by vision system. The spool was moved in range of -2.5 mm to 2.5 mm. The research was performed for various distances between the camera lens and object (for 175 mm and 65 mm) (fig. 4, 5).

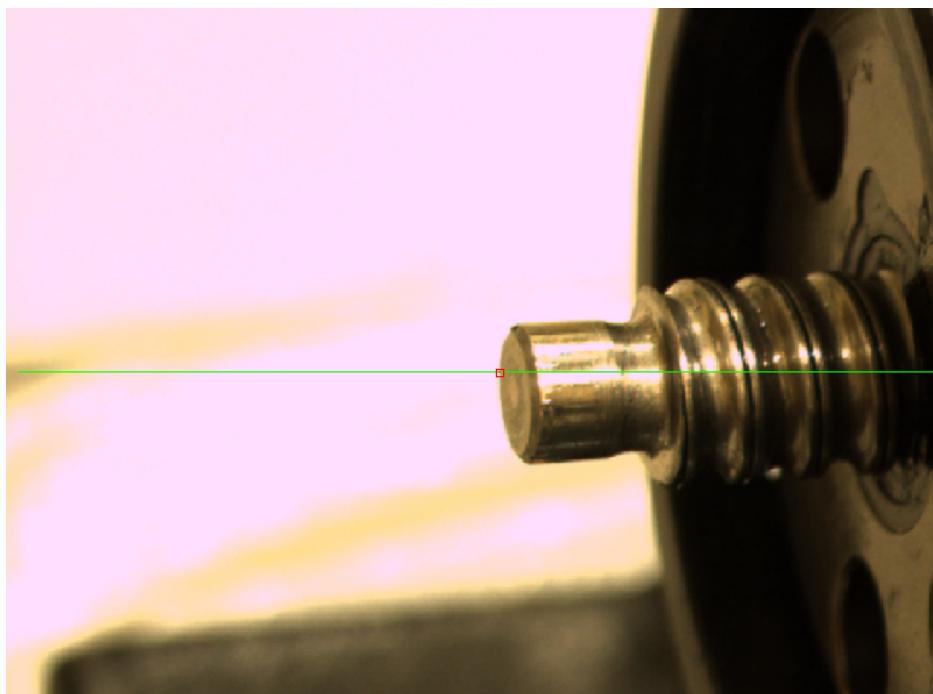


Fig. 4. Image from the camera (distance 175 mm)

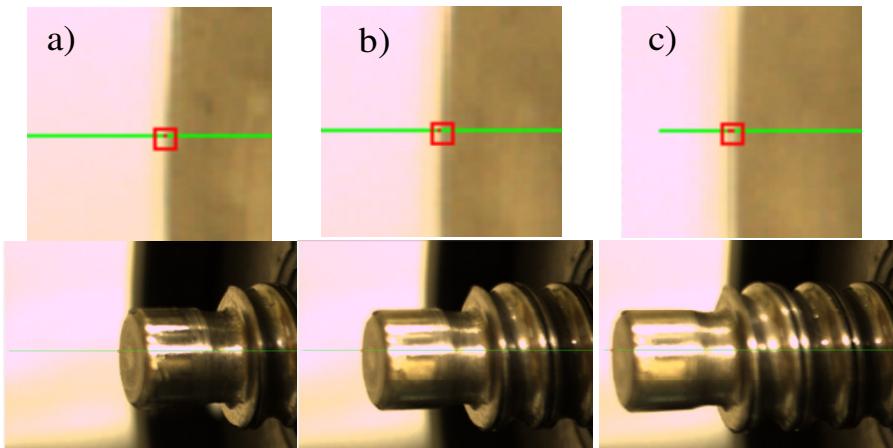


Fig. 5. View of screw from the vision system (distance 65 mm)

Figure 5 shows the maximum and minimum positions of the valve spool (fig. 5.a, 5.c), and the center position (fig. 5.b) in which the valve is closed. The edge detection in the image was realized by line on axis x, which is marked by green color. Also on this line red square was placed. In this square, pixels, which shows the place of edge detection are marked with red points. It was possible, because a differences between brightness of the background color and the valve spool color were big. That was caused by the modification of an aperture of lens. In the experiment, many tests of edge detection for different positions of the valve were done. During the tests, the repeatability of measurements was also examined. In tab. 1 the selected results were shown.

Table 1. The experiment results

Position [mm]	Vision system [pixel]		
	1	2	3
-2,5	36	36	36
-2	58	58	59
-1,5	80	80	81
-1	103	103	103
-0,5	125	126	126
0	148	148	148
0,5	170	170	170
1	192	192	192
1,5	215	215	215
2	237	237	237
2,5	259	259	259

In rows are the results of axis x from the vision system. Tests of positions are similar. That results proved the correctness of detection method. In figure 6 the results from tab. 1. was visualized. The linearity of waveforms confirms that the estimation method was noise-resistant.

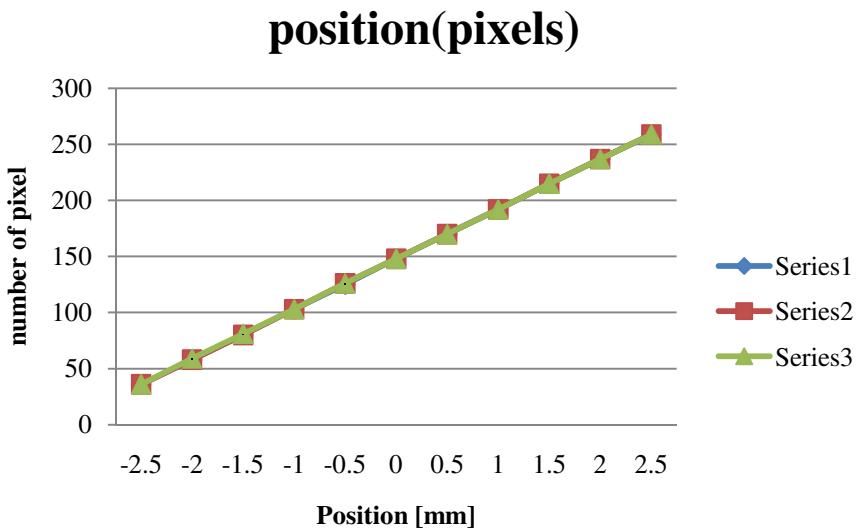


Fig. 6. Position of valve and pixels form vision system

In the future the presented system will be used to measure all pistons positions of the electro-hydraulic manipulator (fig. 7).

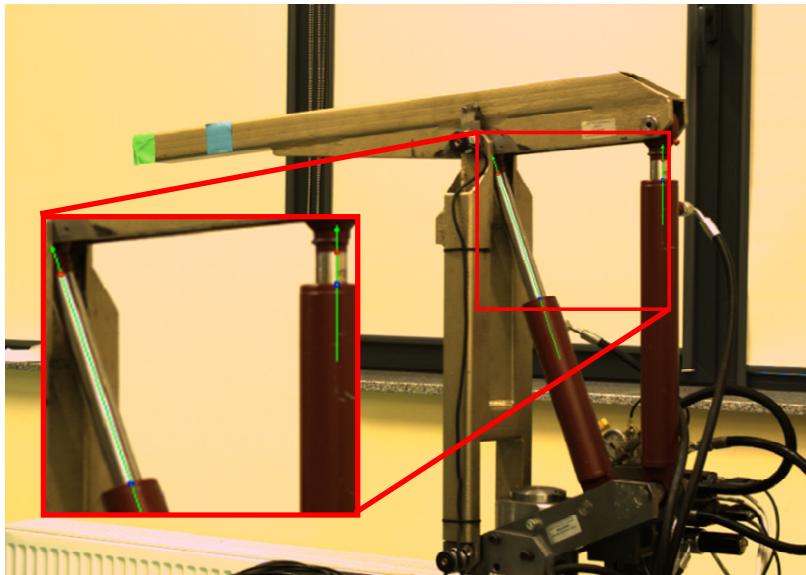


Fig. 7. View of the electro-hydraulic manipulator

On the manipulator two lines were marked. These lines show the direction of edge detection during the measurement of the length of pistons. Information about length of piston can be used to compute the angle in joints of the manipulator. Future research will be focused on control of the manipulator only with vision feedback.

In figure 8 and 9, results of the strength edge of the cylinder with long and short piston were shown. The vertical yellow lines represent points where the edge is detected properly. Horizontal blue lines present the threshold, in which edges cannot be properly classified.



Fig. 8. View of the edge strength profile from the vision system (long piston)

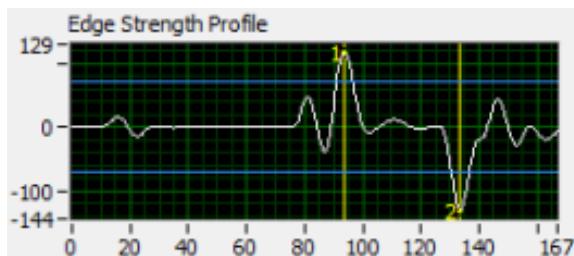


Fig. 9. View of the edge strength profile from the vision system (short piston)

5 Conclusion

The recognition system works as required. Use of the presented system for the manipulator with many actuators can reduce the cost of the measurement system. Future tests will be performed on a real object, with running drive system. The usage of the vision system for measuring the position of the individual cylinders in the joints can be very important for many application, because the mechanical assembly with standard sensors measuring system can be very difficult, costly, time consuming, or even impossible.

References

1. Parnian, N., Won, S.P., Golnaraghi, F.: Position sensing using integration of a vision system and inertial sensors. In: The 34th Annual Conference of IEEE Industrial Electronics, IECON (2008)

2. Ray-Hwa, W., Min-Siou, L., Wang, Y.T.: A vision based low-frequency electro-hydraulic fatigue testing machine. In: Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT) (2011)
3. Edwards, B., Archibald, J., Fife, W., Lee, D.J.: A Vision System for Precision MAV Targeted Landing. In: International Symposium on Computational Intelligence in Robotics and Automation, CIRA 2007 (2007)
4. Alkkioimaki, O., Kyrki, V., Kalviainen, H.: Yong Liu, Handroos, H., Challenges of Vision for Real-Time Sensor Based Control. In: Canadian Conference on Computer and Robot Vision, CRV 2008 (2008)
5. Pace, S.: Fast localisation of mechanical components using Hough transform. In: Proceedings of the International Conference on Image Analysis and Processing (1999)
6. National Instruments, <http://www.ni.com>

Sensitivity and Offset Voltage Testing in the Hall-Effect Sensors Made of Graphene

Oleg Petruk¹, Roman Szewczyk², Tymoteusz Ciuk³, Włodzimierz Strupiński³,
Jacek Salach², Michał Nowicki², Iwona Pasternak³, Wojciech Winiarski¹,
and Krzysztof Trzcinka¹

¹ Industrial Research Institute for Automation and Measurements PIAP,
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
opetruck@piap.pl

² Warsaw University of Technology, Institute of Metrology and Biomedical Engineering,
sw. A. Boboli 8, 02-525 Warsaw, Poland
szewczyk@mchtr.pw.edu.pl

³ Institute of Electronic Materials Technology, Wólczyńska 133, 01-919 Warsaw, Poland
włodek.strupinski@itme.edu.pl

Abstract. Paper presents the results of the hall effect testing in the graphene structures. Special hall effect structures were designed and build, using large graphene sheets. Laboratory testing stand was developed to test sensitivity and offset voltage in hall effect structures under external magnetic field. Characteristics of investigated structures were measured, including such impacting factors as structure size, external magnetic field strength, temperature and time.

Keywords: Hall effect, graphene, magnetic sensor.

1 Introduction

The Hall effect is the production of a voltage difference (the Hall voltage) across an electrical conductor, transverse to an electric current in the conductor and a magnetic field perpendicular to the current [1]. The Hall coefficient is defined as the ratio of the induced electric field to the product of the current density and the applied magnetic field. It is a characteristic of the material from which the conductor is made, since its value depends on the type, number, and properties of the charge carriers that constitute the current. A Hall effect sensor is a transducer that varies its output voltage in response to a magnetic field [2].

Graphene is a 2-dimensional, crystalline allotrope of carbon. In graphene, carbon atoms are densely packed in a regular sp^2 -bonded atomic-scale hexagonal pattern. Graphene can be described as a one-atom thick layer of graphite [3, 4]. It is the basic structural element of other allotropes, including graphite, charcoal, carbon nanotubes and fullerenes. It is exceptional conductor of electricity and heat. Experimental results from transport measurements show that graphene has a remarkably high electron mobility at room temperature, with reported values in excess of $15,000 \text{ cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$. Additionally, the symmetry of the experimentally measured conductance indicates that hole and electron mobilities should be nearly the same. Scattering by the acoustic

phonons of graphene intrinsically limits room temperature mobility to $200,000 \text{ cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$ at a carrier density of 10^{12} cm^{-2} . The corresponding resistivity of the graphene sheet would be $10^{-6} \cdot \text{cm}$. This is less than the resistivity of silver, the lowest known at room temperature.

Therefore, graphene with its exceptionally high theoretical and measured electron mobility and conductivity, is expected to be a good candidate for hall effect sensors construction [5].

2 Principles of Operation of a Hall Effect Sensor

The Hall effect is due to the nature of the current in a conductor. Current consists of the movement of charge carriers, typically electrons, holes, ions etc. When a magnetic field is present, that is not parallel to the direction of motion of moving charges, these charges experience the Lorentz force [1]. When such a magnetic field is absent, the charges follow approximately straight paths between collisions with impurities, phonons, etc. However, when a magnetic field with a perpendicular component is applied, their paths between collisions are curved so that moving charges accumulate on one face of the material. This leaves equal and opposite charges exposed on the other face, where there is a scarcity of mobile charges. The result is an asymmetric distribution of charge density across the Hall element that is perpendicular to both the ‘line of sight’ path and the applied magnetic field. The separation of charge establishes an electric field that opposes the migration of further charge, so a steady electrical potential is established for as long as the charge is flowing.

For a simple metal where there is only one type of charge carrier, that is the electrons, the Hall voltage V_H is given [2]:

$$V_H = -\frac{I \cdot B}{n \cdot t \cdot e} \quad (1)$$

where: I – current across the plate length, B – perpendicular magnetic field, t – thickness of the plate, e – elementary charge, n – charge carrier density of the carrier electrons. From practical reasons, the Hall coefficient R_H is defined as [2]:

$$R_H = \frac{E_y}{j_x \cdot B} = \frac{V_H \cdot t}{I \cdot B} = -\frac{1}{n \cdot e} \quad (2)$$

where: j_x – current density of the carrier electrons (in x direction), E_y – induced electric field (in y direction).

As a result, the Hall effect may be very useful as a means to measure either the carrier density or the magnetic field.

3 Developed Hall Sensor Structures and Testing Setup

To measure the sensitivity and offset voltage in the hall effect graphene structures, special structures were designed and build, using large graphene sheets produced on SiC during the epitaxial process.

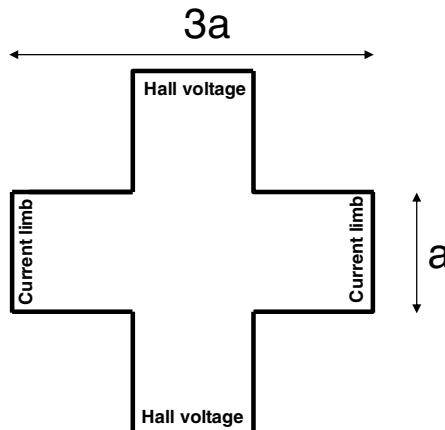


Fig. 1. Schematic diagram of Hall sensor with cross shape

Monolayer graphene was grown using the Chemical Vapor Deposition (CVD) method on the Si face of semi-insulating on-axis 4H-SiC(0001) substrates in a standard hot-wall CVD Aixtron VP508 reactor. Prior to the growth, in situ etching of the SiC surface was carried out in hydrogen atmosphere. The epitaxial CVD growth of graphene was realized under dynamic flow conditions that simultaneously inhibit Si sublimation and promote the mass transport of propane molecules to SiC substrate [3]. The as grown samples were characterized by Hall effect measurements in Van der Pauw geometry (0.55T Ecopia HMS-3000 setup) with the four golden probes placed in the corners of the 10 mm × 10 mm substrates.

Table 1. Transport properties of samples.

Sample No.	graphene concentration [cm ⁻²]	graphene mobility [cm ² /Vs]	graph resistance [Ω/sq]
1979_4H_27	4E12	1300	1200
1979_4H_85	5E12	1000	1100
1979_4H_83	6E12	1000	1000
1980_4H_43	6E12	940	1100
1980_4H_64	8E12	740	1000

Tab.1 illustrates transport properties of the five samples. The substrates were subsequently photolithographically patterned to form five graphene Hall bars on each. 20nm_Ti/80nm_Au ohmic contacts (200 μm × 200 μm) were e-beam deposited. The Hall bars took the form of symmetrical, equal-arm crosses made of 5 squares with length equal a , as it is presented in Fig. 1. The a dimension spanned the range between

200 μm and 1000 μm , with 200 μm interval. Hall voltage was measured on one pair of contacts, whereas current flow was implemented between the other two opposite contacts. Photograph of developed Hall effect sensor with graphene structures is presented in Fig. 2.

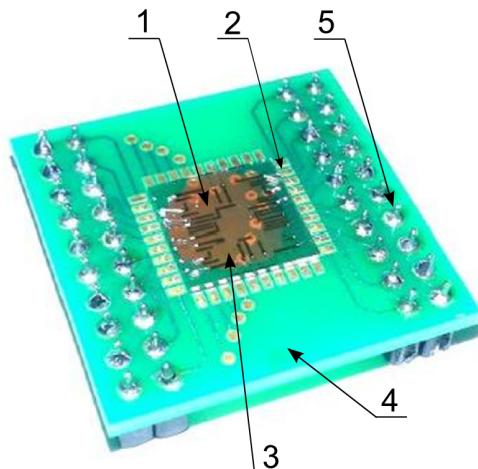


Fig. 2. Photograph of the developed Hall effect sensor structure: 1 – cross made of graphene (transparent, not visible), 2 – bonding, 3 – SiC wafer, 4 – PCB board, 5 – pins with 100 mils raster

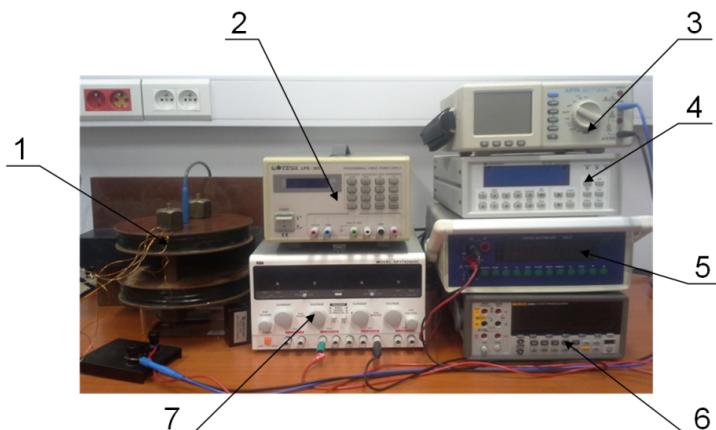


Fig. 3. Photograph of the testing setup for Hall effect sensor structures: 1 – DC magnetic field (Helmholtz coil), 2 – Hall effect structure power supply (Motech), 3 – Helmholtz coil power supply measurement (Appa multimeter), 4 – Magnetic field graphene measurement (Lake-shore), 5 – Hall effect structure power supply current measurement (Metrol), 6 – Hall effect structure Hall voltage measurement (Fluke multimeter), 7 – Helmholtz power supply (NDN)

For the testing of functional parameters of Hall effect structures in the laboratory conditions, special measuring setup utilizing high precision Helmholtz coil was developed. Helmholtz coils were powered from regulated power supply to enable changing of magnetic field strength acting upon the investigated structure in the range of ± 10 mT. Current in the Helmholtz coils was measured by reference Appa 207 multimeter. Moreover Lakeshore 455 gaussmeter was used to measure this field directly for rough verification.

The Hall effect structure was powered directly by the Motech laboratory power supply, and the current was measured simultaneously by the Metrol DM-22 multimeter. The bias current was 1,25 mA. The Hall voltage was measured by precision Fluke 8808A voltmeter. Photograph of developed testing setup is presented in Fig. 3.

4 Results of Testing

Results of the tests of developed Hall effect structures on testing setup are presented in Figures 4–7. It should be highlighted, that nearly linear dependence of Hall-effect sensor output signal versus measured magnetic field was observed in the whole measured range for most structure types.

Linear dependence fit is also presented in figures 4–7. It may be observed, that developed Hall effect sensors exhibit highly linear characteristics. The offset voltage was compensated.

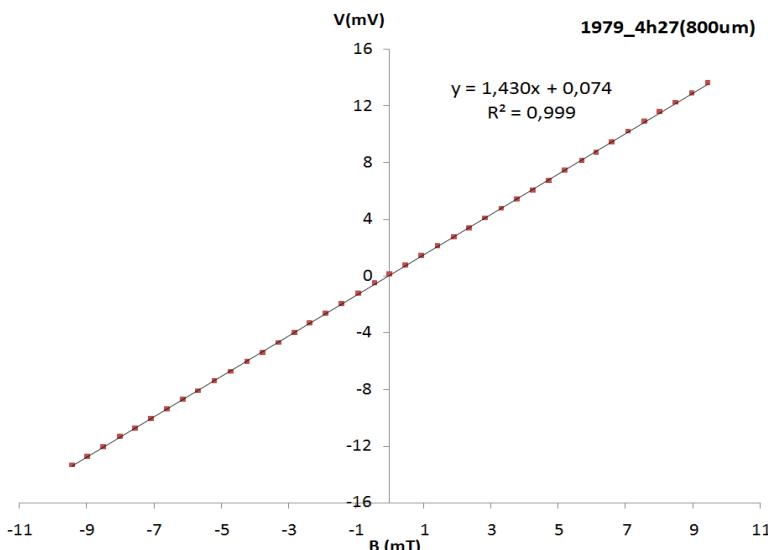


Fig. 4. Experimentally measured characteristic of developed graphene Hall effect structure (3a equal to 800 μ m, $n=4\times 10^{12}$ cm $^{-2}$, $\mu=1300$ cm 2 /Vs, $R_S=1200$ Ω /sq)

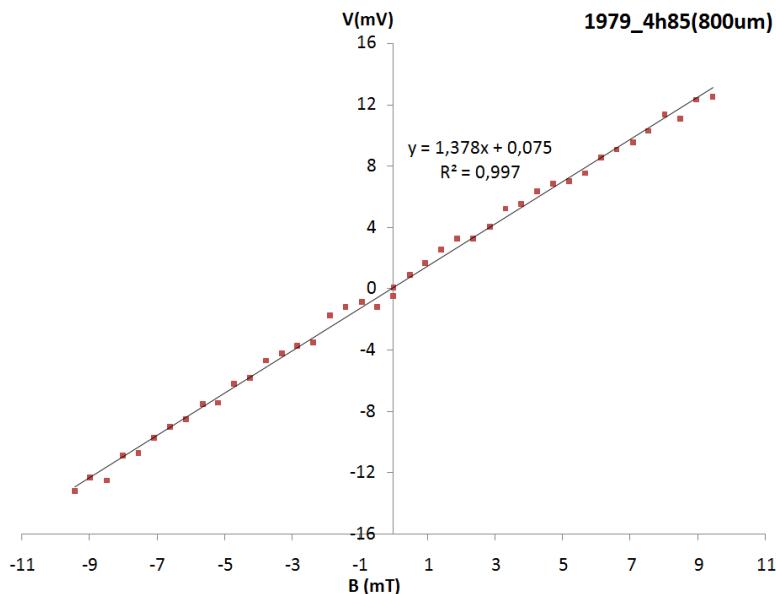


Fig. 5. Experimentally measured characteristic of developed graphene Hall effect structure (3a) equal to 800 μm , $n=5\times10^{12}\text{cm}^{-2}$, $\mu=1000\text{cm}^2/\text{Vs}$, $R_S=1100\Omega/\text{sq}$)

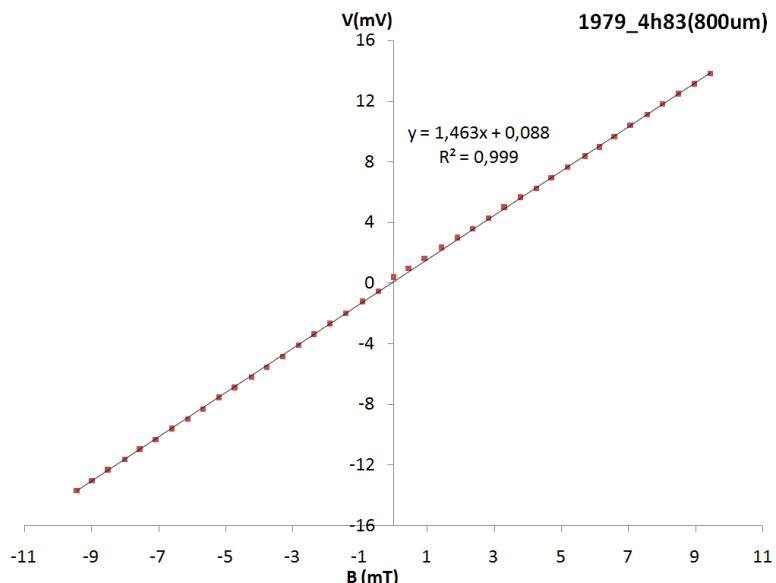


Fig. 6. Experimentally measured characteristic of developed graphene Hall effect structure (3a) equal to 800 μm , $n=6\times10^{12}\text{cm}^{-2}$, $\mu=1000\text{cm}^2/\text{Vs}$, $R_S=1000\Omega/\text{sq}$)

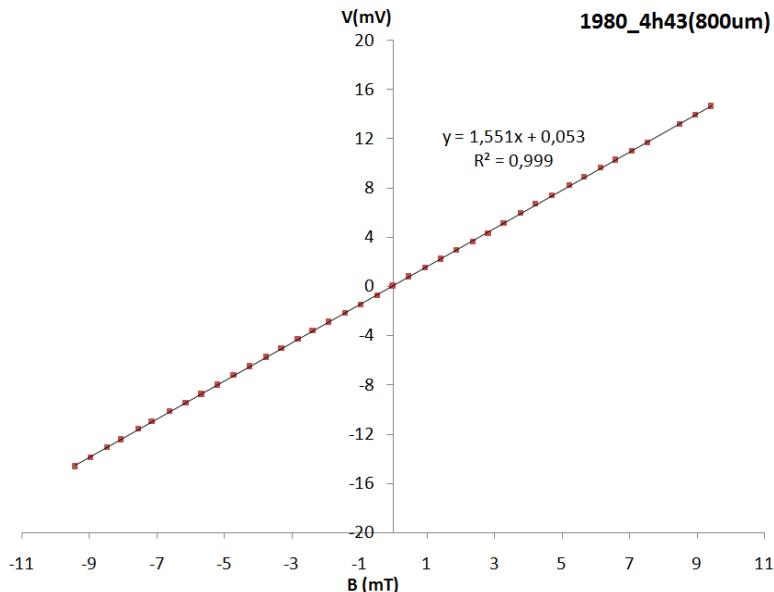


Fig. 7. Experimentally measured characteristic of developed graphene Hall effect structure (3a equal to 800 μ m, $n=6\times10^{12}\text{cm}^{-2}$, $\mu=940\text{cm}^2/\text{Vs}$, $R_S=1100\Omega/\text{sq}$)

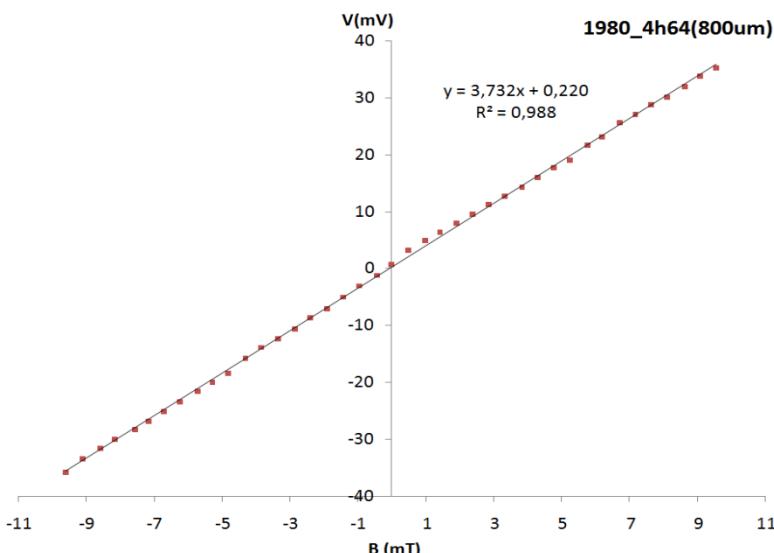


Fig. 8. Experimentally measured characteristic of developed graphene Hall effect structure (3a equal to 800 μ m, $n=8\times10^{12}\text{cm}^{-2}$, $\mu=740\text{cm}^2/\text{Vs}$, $R_S=1000\Omega/\text{sq}$)

Linear fit of the function $f(x) = k \cdot x + b$ were made by the least square method. High linearity of the sensor was confirmed by the R-square parameter exceeding 0.99 for most of the tested structure variants.

The Table 2 shows the measured functional parameters of tested graphene based Hall sensors. The most sensitive variant is highlighted.

Table 2. Calculated functional parameters of tested Hall-effect structures. (Transport properties of each sample are presented in Tab. 1)

Hall effect structure	sensitivity mV/mT	offset voltage mV
1979_4h27(200um)	0,564	544,9
1979_4h27(300um)	0,892	540,6
1979_4h27(400um)	0,600	278,0
1979_4h27(500um)	0,659	106,0
1979_4h27(800um)	1,430	409,6
1979_4h85(200um)	0,512	398,9
1979_4h85(300um)	0,400	411,9
1979_4h85(400um)	0,500	236,5
1979_4h85(500um)	0,563	177,4
1979_4h85(800um)	1,378	319,5
1979_4h83(200um)	0,739	261,2
1979_4h83(300um)	0,629	298,6
1979_4h83(500um)	0,372	119,3
1979_4h83(800um)	1,463	411,2
1980_4h43(600um)	1,168	510,1
1980_4h43(700um)	1,138	476,4
1980_4h43(800um)	1,168	510,6
1980_4h43(900um)	1,578	224,9
1980_4h43(600um)	1,551	331,9
1980_4h64(800um)	3,732	1190
1980_4h64(700um)	3,670	1872
1980_4h53(1000um)	2,152	877,7
1980_4h53(800um)	2,516	1100,0
1980_4h53(700um)	2,232	773,6
1980_4h53(600um)	0,464	402,0

From practical point of view, in the case of Hall sensors, the most important are temperature drift of offset voltage as well as its time stability. Fig. 9 shows the temperature dependence of offset voltage for tested graphene based Hall sensor. Significant temperature drift, in the range of the Hall effect voltage for 10 mT, can be clearly seen.

Fig. 10 shows the offset voltage time drift for graphene based Hall sensor. In this case time drift for 100 minutes, in the range of the Hall effect voltage for 10 mT was also observed.

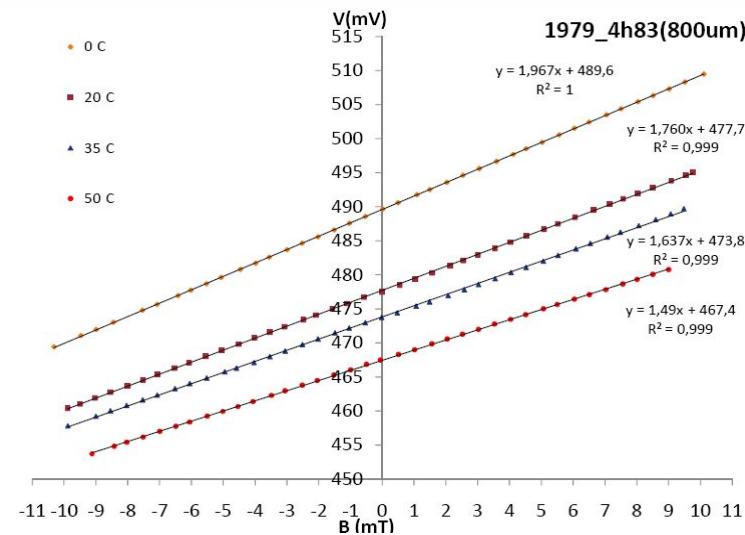


Fig. 9. Operating temperature drift of offset voltage of developed graphene based Hall sensor ($n=6\times10^{12}\text{cm}^{-2}$, $\mu=1000\text{cm}^2/\text{Vs}$, $R_S=1000\Omega/\text{sq}$)

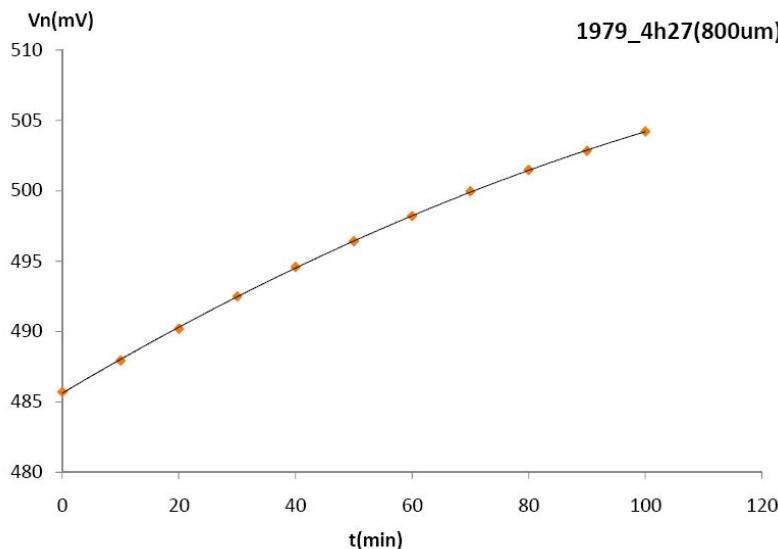


Fig. 10. Time drift of offset voltage of developed graphene based Hall sensor ($n=4\times10^{12}\text{cm}^{-2}$, $\mu=1300\text{cm}^2/\text{Vs}$, $R_S=1200\Omega/\text{sq}$)

5 Conclusion

Experimental results presented in the paper confirm, that Hall effect structure utilizing graphene material exhibit highly linear characteristics as well as high magnetic field

sensitivity. Moreover, this sensitivity is diversified from the point of view of transport properties of samples, which creates further possibility of optimisation of Hall sensors production process.

However, experimental tests indicated significant drift of sensors offset voltage caused by both temperature and time. This drift is significant limitation from the point of view of practical application of developed graphene based Hall effect sensors. As a result, developed sensors have to be optimised from the point of view of minimisation of these drifts as well as drift compensation mechanisms should be implemented in signal processing.

Acknowledgments. This work was partially supported by The National Center for Research and Development within GRAFTECH Program.

References

1. Ripka, P.: Magnetic Sensors and Magnetometers. Artech, Boston (2001)
2. Ramsden, E.: Hall-Effect Sensors, 2nd edn. Theory and Application, Newnes (2006)
3. Strupinski, W., Grodecki, K., Wysmolek, A., et al.: Graphene Epitaxy by Chemical Vapor Deposition on SiC. *Nano Lett.* 11(4), 1786–1791 (2011)
4. Krupka, J., Strupinski, W.: Measurements of the sheet resistance and conductivity of thin epitaxial graphene and SiC films. *Applied Phys. Letts.* 96, 082101 (2010)
5. Xu, H., Zhang, Z., Shi, R., Liu, H., Wang, Z., Wang, S., Peng, L.M.: Batch-fabricated high-performance graphene Hall elements. *Nature Sci. Rep.* 3, 1207 (2013)

Digitally Controlled Current Transformer with Hall Sensor

Oleg Petruk¹, Roman Szewczyk², Jacek Salach², and Michał Nowicki²

¹ Industrial Research Institute for Automation and Measurements PIAP,
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
opetrak@piap.pl

² Warsaw University of Technology, Institute of Metrology and Biomedical Engineering,
sw. A. Boboli 8, 02-525 Warsaw, Poland
szewczyk@mchtr.pw.edu.pl

Abstract. Paper presents the digitally controlled high resolution DC current transformer enabling contactless measurements of DC current. To increase the sensitivity of current transformer, high permeability, amorphous alloy based ring-shaped core was used. Moreover, analyses of magnetic field distribution in the measuring gap were carried out. Magnetic field in this gap was measured by the hall sensor connected with the high resolution sigma-delta analog to digital converter. As a result, current up to 100 A may be measured with uncertainty of about 1A.

Keywords: current transformer, Hall sensor.

1 Introduction

Current transformers are commonly used in industrial measurements of large alternating currents [1]. In the case of alternating currents magnetic cores with linear characteristics should be used [2]. In such a case, the current transformer guarantee galvanic separation of measuring circuit as well as provides voltage simply proportional to the current value.

However, in some industrial applications DC current should be measured. Such measurements are especially sophisticated if high potentials on the conductor appears. Such measuring conditions are common during energetic elements degradation monitoring, where current measurements are required [3]. In that case, to enable galvanic separation between conductor and DC current measuring circuit, the gapped current transformer should be used. Then, value of the flux density in the gap should be measured. It should be indicated, that gap in the core caused linearization of its characteristic due to the demagnetization effect [4]. Moreover, application of the high permeability magnetic materials, such as amorphous alloys [5, 6], increases sensitivity of DC current transformer and enables measurements of lower values of DC current.

For measurements of flux density in the transformer's gap, both hall effect sensors as well as miniaturized flux gate sensors were used [1]. However, miniature fluxgate

sensors are still expensive and require sophisticated filtration of second harmonics during analogue signal processing. As a result, in DC current transformers hall effect sensors are commonly used. Such hall effect sensors, with sensing element made of InAs and InSb[1] or even graphene [7] exhibit sufficient sensitivity for all industrial applications.

Paper presents the technical solution of advanced DC current transformer utilizing both high permeability amorphous alloy core as well as high sensitivity hall sensor connected with high-resolution analogue to digital converter. In presented solution the result of measured current may be transmitted to microcontroller via RS-232 interface for further digital processing. As a result, presented digitally controlled DC current transformer may be the base for development of intelligent measuring systems for specialized applications in energetic elements degradation monitoring.

2 Principles of Operation of DC Current Transformer

According to the Ampere's law, flow of electric current I causes appearance of magnetic field H and, consequently, magnetic flux density B in material with relative permeability μ , as it is presented in figure 1 and quantitatively described by the equation 1.

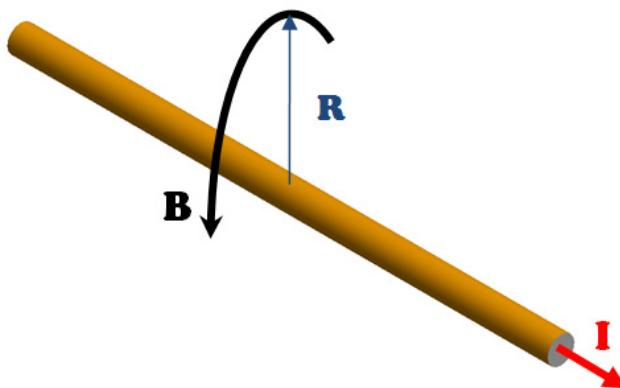


Fig. 1. Schematic diagram presenting magnetic flux density around the conductor: I – electrical current, B – magnetic flux density, R – distance

$$B = \mu\mu_0 H = \mu\mu_0 \frac{I}{2\pi R} \quad (1)$$

where: μ_0 is magnetic constant.

During the operation of DC current transformer, magnetic flux density B_{in} in the core causes significant increase of magnetic flux density B_{ex} in the gap, as it is presented in the figure 2. Due to the appearance of gapped core, the high sensitivity of the current transformer may be achieved.

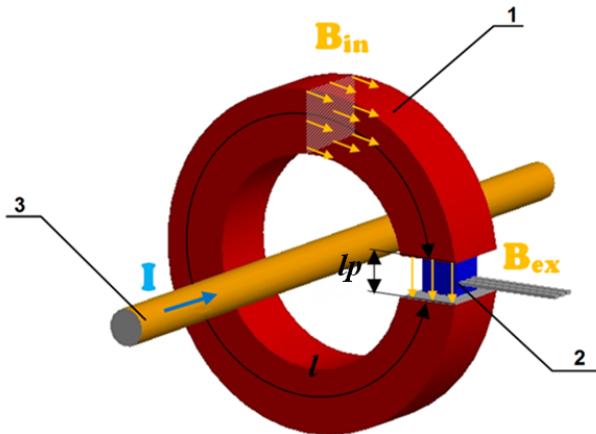


Fig. 2. Schematic diagram of DC current transformer: 1 – magnetic core with the gap, 2 – Hall sensor for B_{ex} measurements, 3 – wire with measured current I

Value of the flux density B_{ex} in the gap may be approximated by the equation 2 [8]:

$$B_{ex} = \frac{I}{l_p \mu_0} - \frac{l - l_p}{l_p} \frac{\mu}{\mu_0} B_{in} \quad (2)$$

where I is the current in the conductor, l – magnetic length of the core, l_p – length of the gap and μ – permeability of the core. It should be indicated, that for high values of permeability m of the core, equation 2 may be simplified to:

$$B_{ex} = - \frac{l - l_p}{l_p} \frac{\mu}{\mu_0} B_{in} \quad (3)$$

where B_{in} is proportional to the measured current I .

3 Technical Solutions of Hall Current Transformer

Schematic block diagram of developed digitally controlled current transformer is presented in figure 3. This solution consist of ring-shaped gapped core, high sensitivity Hall-effect sensor, sigma-delta analogue to digital converter microcontroller NXP1115 with ARM Cortex M0 architecture as well as RS-232 interface. To reduce influence of external distortions on sensor operation as well as to separate grounds of computer and sensor, transceptor barrier was applied.

In presented solution ring-shaped gapped core made of FeSiB amorphous alloy (2605SA1 produced by Metglass Co.) was used. Such amorphous alloy is produced by rapid cooling of liquid phase with cooling speed exceeding 10^6 K/s. As a result there is no crystalline phase in the material structure, whereas ferromagnetic properties appear due to the short range local order. As a result, amorphous magnetic alloys,

such as 2605SA1, exhibit high relative permeability (exceeding 10^5) and significant mechanical stress strength [9]. Thus, such core is especially suitable for development of gapped current transformers. Ring-shaped core of developed transformer had outside diameter 16 mm, inside diameter 10 mm and 5 mm height. Gap in the core was 1 mm wide.

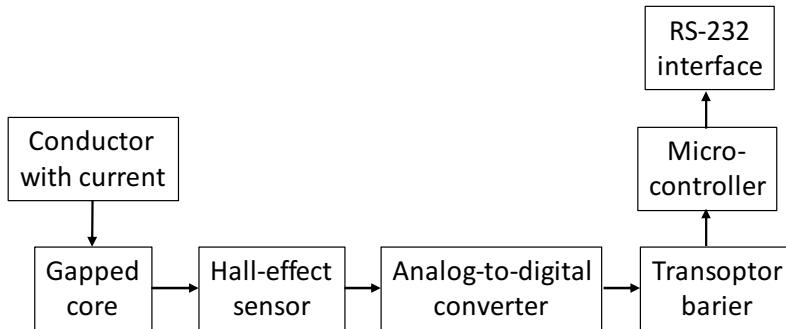


Fig. 3. Schematic diagram of digitally controlled current transformer

Spatial distribution of the flux density in the core of developed current transformer was calculated with the use of finite element method (FEM). Figure 4 presents this distribution caused by 1 A current flow. In this case average flux density in the gap was equal 1.27 mT.

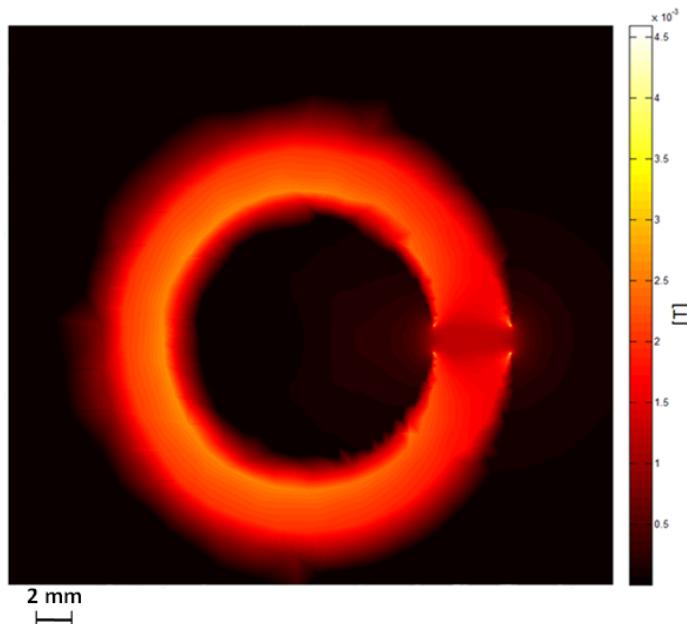


Fig. 4. Simulation of the spatial distribution of flux density B in the gapped core and in the gap (measured current equal 1 A)

As a sensor of magnetic field in the current transformer gap, advanced Hall-effect sensor TLE4990 [10] (produced by Infineon Technologies) was used. The most important parameters of this sensor are presented in the table 1.

Table 1. Most important parameters of TLE4990 Hall-effect sensor [10] produced by Infineon Technologies

Parameter name	Unit	Value
Magnetic field range	(mT)	204
Sensitivity	(mV/mT)	15
Resolution	(T)	60
Output noise	(mV)	1,2

Developed current transformer with electronic transducer is presented in figure 5. Analogue signal from Hall-effect sensor (2) in the gap of current transformer (1) is measured by sigma-delta AD7191 analogue to digital converter (3) produced by Analog Devices. Then, it is transmitted through the transceptor barrier (4) to the LPC1115 microcontroller and then, via RS-232 interface to the personal computer.

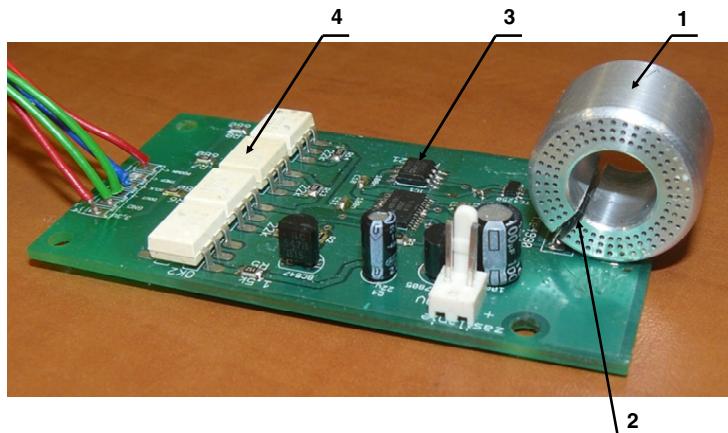


Fig. 5. View of developed current transformer: 1 – gapped core made of amorphous alloy (in aluminum casing), 2 – gap with Hall-effect sensor, 3 – sigma-delta analogue to digital converter, 4 – transceptor barrier

4 Testing Setup for Developed Current Transformer

Testing of functional parameters of current transformers maybe difficult in the laboratory conditions due to the fact, that such sensors operate in the range of currents up to 100 A. For this reason, special measuring setup utilizing welding transformer was developed. Welding transformer was powered from regulated autotransformer to enable changing of current in its short-circuited secondary winding. Current in secondary winding was measured by reference clamp multimeter BM 357 with about 1 % accuracy.

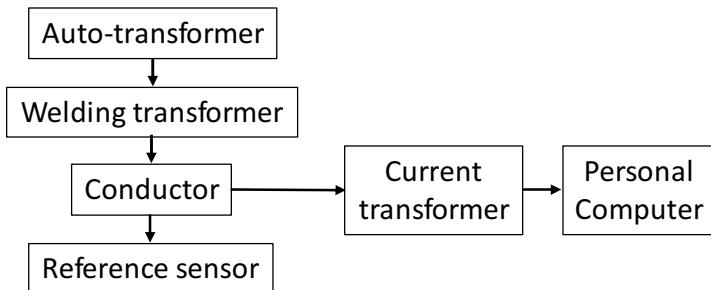


Fig. 6. Schematic block diagram of the testing setup for current transformers

5 Results of Testing

Results of tests of developed current transformer on testing setup are presented in figure 7. It should be highlighted, that nearly linear dependence of Hall-effect sensor output signal versus measured current was observed up to 100 A.

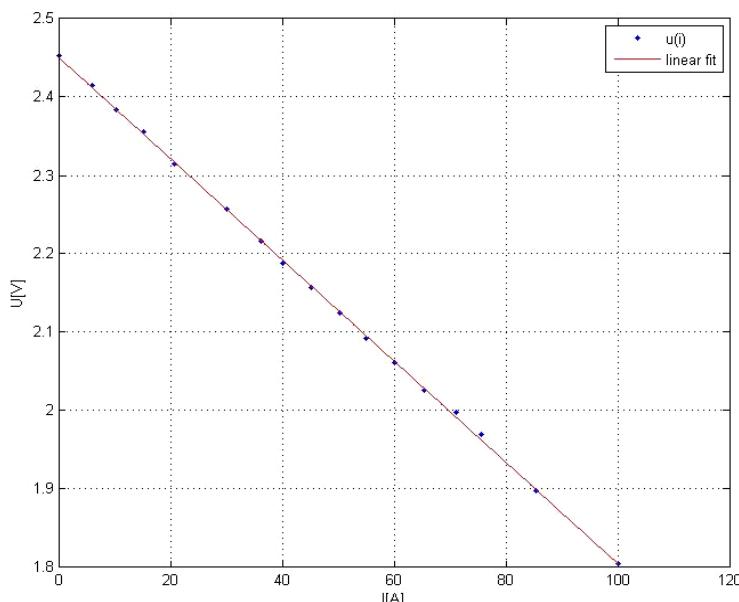


Fig. 7. Experimentally measured characteristic of developed DC current transformer with its linear fit

Linear fit of the function $f(x) = k \cdot x + b$, where $k = -0.0065 \text{ V/A}$ and $b = 2.46 \text{ V}$ was made by the least square method. High linearity of the sensor is confirmed by the R-square parameter exceeding 0.99. The most important parameters of linear fit are presented in table 2.

Table 2. The accuracy of the linear fit

Parameter	unit	value
SSE ¹	(V)	0.00014
R-square	(V)	0.9998
Adjusted R-square	(V)	0.9997
RMSE ²	(V)	0.00306

6 Conclusion

Experimental results presented in the paper confirm, that current transformer utilizing high permeability amorphous alloys exhibit highly linear characteristics as well as sufficient sensitivity for most industrial applications.

Moreover, high resolution sigma-delta analogue to digital converter enable application of other types of Hall-effect sensors, such as newly developed sensors with graphene-based Hall sensing elements. As a result, presented solution opens new possibilities of further practical applications of different types of magnetic field sensors.

Acknowledgments. This work was partially supported by The National Center for Research and Development within GRAFTECH Program.

References

1. Ripka, P.: Magnetic Sensors and Magnetometers. Artech, Boston (2001)
2. <http://www.magnetec.de/en/nanopermr-products/metering-products/>
3. Salach, J., Hasse, L., Szewczyk, R., Smulko, J., Bienkowski, A., Frydrych, P., Kolano-Burian, A.: Low Current Transformer Utilizing Co-Based Amorphous Alloys. IEEE Trans. Magn. 48, 1493–1496 (2012)
4. Jiles, D.C.: Introduction to Magnetism and Magnetic Materials. Chapman&Hall, London (1998)
5. Meydan, T.: Application of amorphous materials to sensors. J. Magn. Magn. Mater. 133, 525 (1994)
6. Kulik, T., Ferenc, J., Kolano-Burian, A.: Magnetically soft nanomaterials for high-temperature applications. Journal of Alloys and Compounds 434, 623–627 (2007)
7. Shen, T., Gu, J.J., Xu, M., Wu, Y.Q., Bolen, M.L., Capano, M.A., Engel, L.W., Ye, P.D.: Observation of quantum-Hall effect in gated epitaxial graphene grown on SiC (0001). Applied Physics Letters 95, 17210 (2009)
8. Nałęcz, M., Jaworski, J.M.: Magnetic measurements, WNT (1968)
9. http://metglas.com/products/magnetic_materials/2605sa1.asp
10. <http://www.infineon.com>

¹ SSE – sum of squared errors of linearization.

² RMSE – root mean squared error.

Measuring Station for Testing of Graphene Flow Sensors

Marcin Safinowski¹, Wojciech Winiarski¹, Kamil Domański¹, Oleg Petruk¹,
Szymon Dąbrowski¹, Roman Szewczyk², and Krzysztof Trzcinka¹

¹ Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland
*{msafinowski,wwiniarski,kdomanski,opetruk,
sdabrowski,rszewczyk,ktrzcinka}@piap.pl}*

² Warsaw University of Technology, Institute of Metrology and Biomedical Engineering
szewczyk@mchtr.pw.edu.pl

Abstract. This article describes the construction and operation principle of a measuring station for testing of prototype graphene flow sensors. The measuring station enables for checking an impact of different parameters, such as volume flow value, flowing liquid concentration and liquid temperature on value of generated electric charge on the graphene's surface. This paper also presents basic information about developed transducer which converts signals from graphene sensor. The essential part of the text is chapter about tests of the station. It contains information essential for depicting of real conditions during measurements of generated electric charge on the graphene's surface. Reliability of the carried out research was checked by estimation of uncertainty of measurement equipment.

Keywords: graphene, sensor, measurement, converter.

1 Introduction

From the results of studies conducted in various laboratories worldwide show that the flow of water with the chloride ions in the vicinity of graphene structure causes the generation of electrical charge on the surface of graphene. The resulting potential difference is dependent on the fluid flow rate and the concentration of chloride ions. This phenomenon can be used for the construction of the flowmeter.

Within the frameworks of the project “Graphene based, active flowsensors” were conducted research on the development of methods for producing graphene flow sensors. Just to be able to study the properties of prototype graphene flow sensors, it was necessary to build a testing station. The main task of the station was to carry out research through which information regarding the impact of the flow rate, the concentration of chloride ion, and temperature on the voltage generated on the surface of graphene was obtained. The measuring station was also used to tests aimed at optimizing the geometric construction of the sensors and the process for their preparation. Research conducted on this measuring station should to allow for the development of a prototype sensor, check his work on various parameters, and possible to check the finished product.

2 Construction of a Measuring Station

The measuring station is used to study the flow in order to determine, what the flow speed, the concentration of chloride ions, temperature, locations of the electrodes, and the shape of the sensor have an impact for the voltage generated on the surface of graphene.

Systems of control and station measurements provide making the test objectively, what is achieved by eliminating the possibility of subjective impact of maintenance: given and measured values, assessment of results, and creating records documenting the tests performed. Objective tests performance was carried out by automation and computerization of the measuring station. The flow in the station is forced by gravity flow solution. This solution was used to eliminate the ripple and interference generated by the pumps.

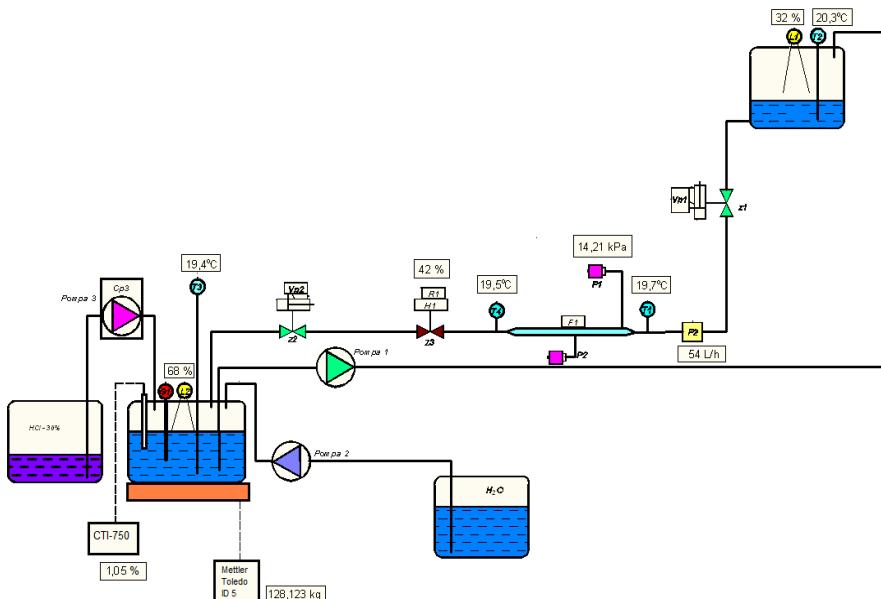


Fig. 1. A general schema of a measuring station for testing of graphene flow sensors

Main elements of a measuring station for testing of graphene flow sensors:

1. Temperature sensor T2 – Pt 100 type 816 Czaki
2. Top tank level sensor UT18-750 SELS
3. Top tank ELBI CB 200L
4. Cut-off valve Z1 – OMAL 101 SR 30
5. Control unit
6. Flowmeter MAG 6000 SITRANS F M MAG1100 DN6 – SIEMENS
7. Tank for distilled water ELBI CB 200L

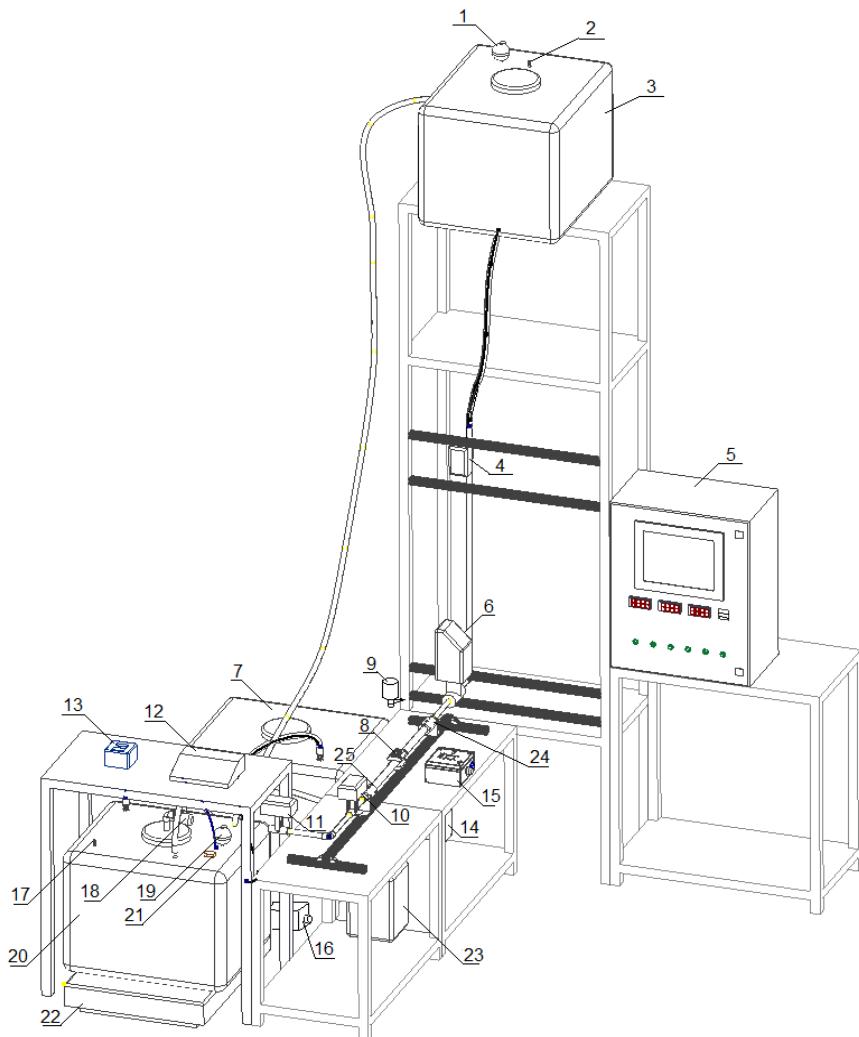


Fig. 2a. A measuring station for testing of graphene flow sensors

1. Tested graphene flow sensor
2. Pressure sensor PC-28 – Aplisens
3. Flow regulation valve Z3 – VALPES serii ER PLUS
4. Cut-off valve Z2 – SAFI RC210DA KSSA 0-1C F05 CM14/45
5. Scales terminal METTLER TOLEDO ID 5
6. Conductometer terminal JUMO CTI-750
7. P3 pump – for acid or sodium chloride solution VERDER M500
8. Converter for graphene sensor
9. P2 – pump for distilled water OMNIGEN WZ 250
10. Bottom tank level sensor UT18-750 SELS

11. Mixer MP12V20W
12. Temperature sensor T3 - Pt 100 typ 816 Czaki
13. Solution tank ELBI CB 200L
14. Heater UKQ - NUGA
15. ScalesMETTLER TOLEDO type KCC 150 RP T 01 51
16. Tank for sodium chloride and other concentrated substances
17. Temperature sensor T1 - Pt 100 typ 371 Czaki
18. Temperature sensor T4 - Pt 100 typ 371 Czaki



Fig. 2b. A measuring station for testing of graphene flow sensors

In addition, the measuring station includes compressed air system needed to power the actuators, and auxiliary control systems and measurements do not have a direct impact on the course of the tests. The station is also equipped with a system capable for changing the direction of liquid flow in slotted circuit in the graphene sensor's space from horizontal to vertical.

Because of opportunity to use hydrochloric acid for tests, our stand is equipped with safety guards and installation rapid run-off of solution from the top tank.

The position can be operated in an automatic cycle, wherein the control of the devices is implemented by a controller, and the effect of control and measurement data are displayed on the screen in a manual cycle, and in which the control of the devices is done by manual adjustment of the value of using SCADA interface.

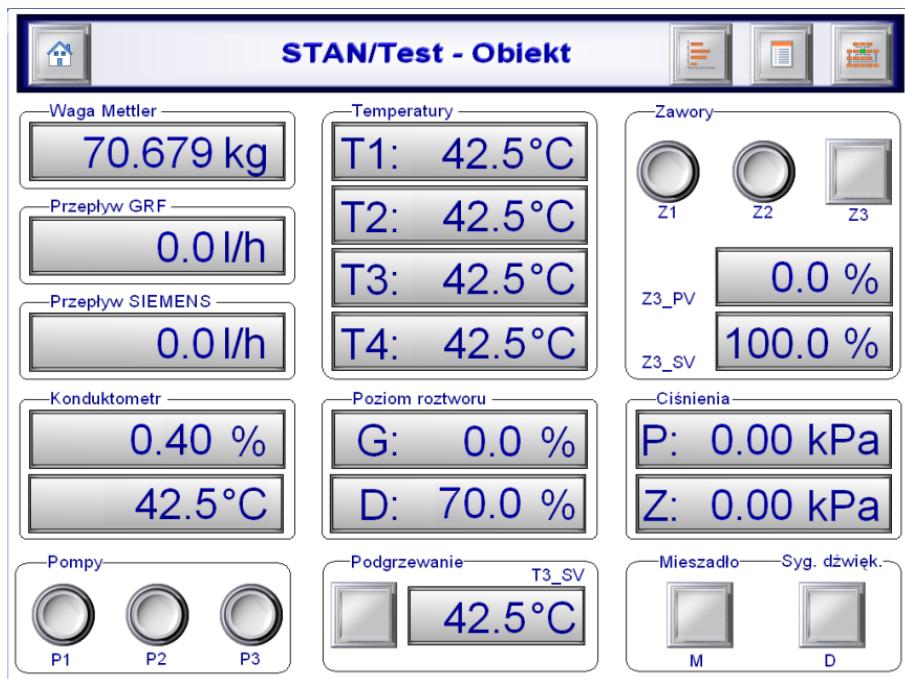


Fig. 3. Manual control panel

The flow tests procedure can be divided into a number of stages. Before starting the tests, one needs to define the parameters for a test, i.e. the concentration of solution, the volume flow, the temperature of the solution, etc. During the first stage a solution of the desired concentration is prepared. Programmer weighs an appropriate amount of water by pumping it to the scales tank, and then appropriate amount of acid or other active substance is weighed. In the next step, the solution is stirred and its concentration is measured.

If the concentration is not correct, an appropriate amount of one of the components is added. Then, the solution is heated to the required temperature. In a next step, the prepared solution is pumped to the upper tank. Then, the system is filled with the solution and venting step is executed. When the system is vented, one can start measurement step. In this stage, the appropriate flow volume is extorted, and all parameters – such as the volume flow from the flow, the signals from the graphene converter, mass flow, temperature in the four measurement points, pressure, etc. are stored on a computer disk. At the same time graphs of selected values are created. On completion of the tests, in the last stage, the liquid is being removed from the system.

Functions of station operating control, as well as measurements, are performer by PLC – OMRON SYSMAC CJ2M controller, implemented in the station. The controller is connected to an industrial computer with a touch monitor Advantech TPC – 1570 HC1AE. For the computer a software SCADA iFIX ver.5.5 PL for 150 points is installed. It provides current visualization of the process, and enables communication between the controller and the operator performing the tests.

In addition, with SCADA software works program Proficy Historian for SCADA – ver. 4.5 (for 32 points), which enables archiving measurement data, making the analyzes, diagrams and charts. The program groups measurements signals in way which makes analysis of interrelation between the values measured very clear.

At the same time, it is possible to create your own groups of measuring signals – if it turns out that it is necessary to find correlations among others, non-grouped measuring signals. The measuring station can be operated in an automatic cycle, wherein the control of devices is performed by a controller, and the effect of control and measurement data are displayed on the screen; as well in a manual cycle, in which the control of the devices is done by manual adjustment of the SCADA screen.

3 Data Acquisition and Transmission Device for Project FlowGraf

3.1 The Circuit

The purpose of the device is to interface a graphene flow sensor with a research workstation. Combined these two shall provide the functionality necessary to conduct research on the sensor.

The main control element of the device is an LPC1769 ARM MCU. The MCU controls communication interfaces to the computer station: a 100 Mbps Ethernet device for bidirectional data transfer with a future research computer application, an RS-232 industrial serial interface providing data to SCADA control panel through a PLC, and an UART serial interface with a serial-to-USB converter for debugging and testing purposes. Programming of the MCU is accomplished through SWD interface, similar to JTAG.

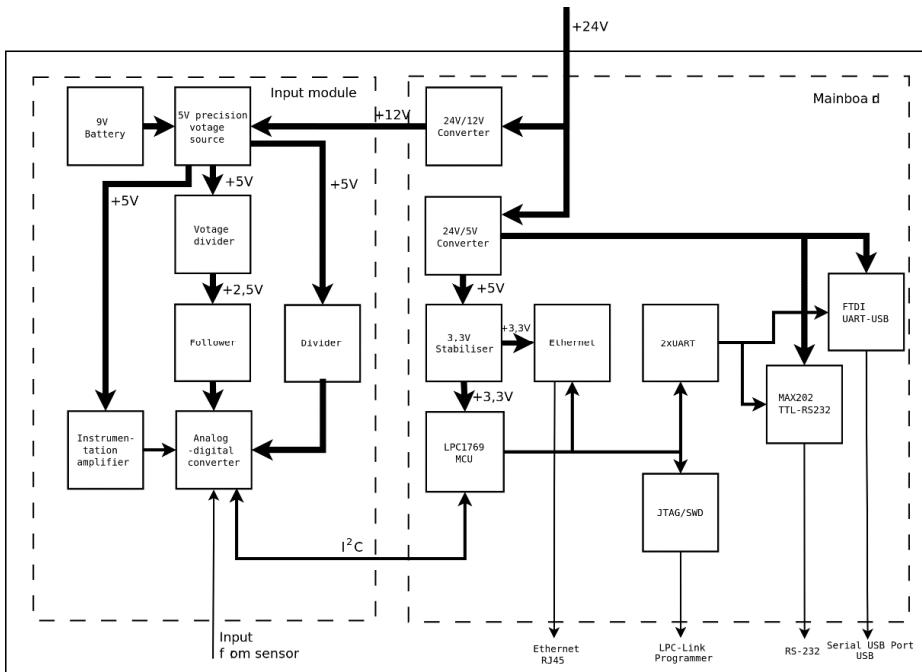


Fig. 4. Simplified schematic of the device

The MCU acquires its data from an MCP3425 analog-to-digital converter over an I²C bus. The converter samples the signal with 16-bit resolution at up to 15 samples per second and allows for a ± 2.048 V differential input range.

3.2 Input Module

Given the yet unconfirmed electrical character of the graphene sensor, the device has been designed to sport a modular design with quickly interchangeable input stage modules thus allowing for easy changes of input stage topology. The input stage is connected to the main board with four wires: two for analog power supply rails and two for the I²C bus.

The currently employed input stage uses an AD620 instrumentation amplifier with 10 G Ω input impedance to amplify a differential voltage signal 50 to 500 times and add a 2.5 V bias to the output which is then compared by the differential ADC to the same 2.5 V voltage. An LT1019 precision reference voltage source provides a stable power supply for the ADC and amplifier through a unity follower, as well as the bias voltage through a manually calibrated divider and another follower.

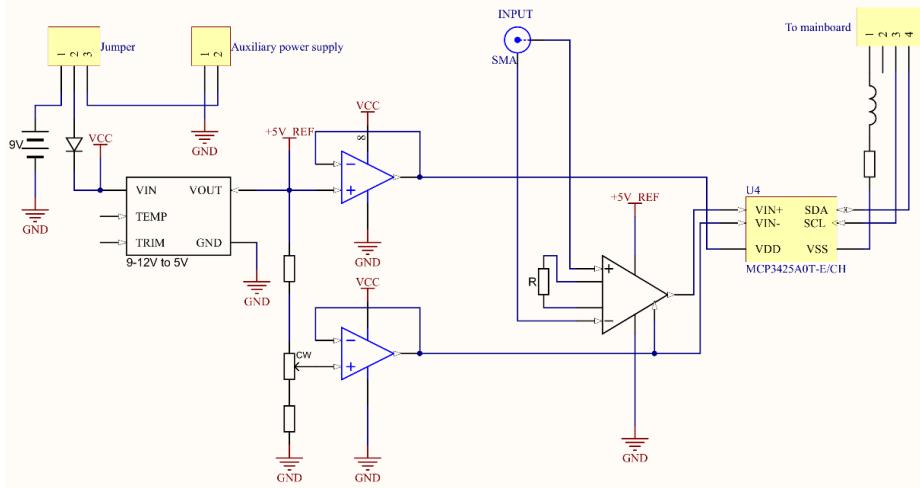


Fig. 5. Input module used in preliminary device tests

3.3 Embedded Software

The code running on the device has been written in C and utilities CMSIS library for the LPC1769 MCU, which provides a standard interface with conforming to a standard by ARM Holdings. This allows not only for seamless transition to another ARM platform but also for easy integration for various utility libraries available for the ARM architecture, for example lwIP library providing a TCP/IP stack for Ethernet communications.

After bootup the software executes and initialisation procedure consisting of the following actions:

- UART interface parameter setup (19200 baud, two stop bits, even parity),
- selection of pins for UART (as each of the interfaces can be connected to more than one pin pair),
- system clock configuration,
- I2C interface initialization,
- TCP/IP stack initialisation with a static IP address,
- finding the ADC's address on I²C bus and setting the converter's configuration register.

Subsequently the main loop begins. Data is read from the ADC and multiplied by a configured factor resulting in flow stream volume value which is then immediately sent over both UART interfaces if UART receive handler activated a flag signalling data request from PLC. The flag is afterwards disabled. Then the time elapsed from last periodic transmission is evaluated and if it is greater than a predefined interval then data is transmitted over both UARTs, as well as over Ethernet in a UDP packet broadcast on port 2500.

4 Analysis of Measurement Uncertainty

4.1 The Uncertainty of Measurement of the Flow Rate

The measurement of the flow rate is done by two methods. The first method is the continuous measurement with electromagnetic flowmeter SITRANS FM MAG 6000 MAG1100. Measurement error of the volume flow, declared by the manufacturer is: $e_q = \pm 0.2\%$ in the range of 10 to 100% of measurement range. The uncertainty of measurement of volume flow is calculated using a B type method. We assume rectangular expansion, therefore the type B uncertainty is calculated from the formula:

$$u(q) = \frac{e_q}{\sqrt{3}} \quad (1)$$

We obtain uncertainty of mass measurement equal to $u(q) = 0.12\%$. For the calculation of the expanded uncertainty we accept extension factor $k = 2$. Finally, the expanded uncertainty amounts $U(q) = 0.24\%$. For measurement range $0 \div 520 \text{ dm}^3/\text{h}$ the expanded uncertainty is $1.3 \text{ dm}^3/\text{h}$.

The second method of volume flow measurement is measuring mass flow over time, which is converted to volume flow. In this measurement is used for weight METTLER TOLEDO Type KCC 150 RP T 01 51 with measurement range $0 \div 150 \text{ kg}$. Resolution of the scales measurement is 0.001 kg , while the measurement error declared by the manufacturer is: $e_m = \pm 0.01 \text{ kg}$.

Uncertainty of mass measurement is calculated using the B type method. We assume rectangular expansion, therefore the type B uncertainty is calculated from the formula:

$$u(m) = \frac{e_m}{\sqrt{3}} \quad (2)$$

We obtain uncertainty of mass measurement equal to $u(m) = 0.0058 \text{ kg}$. For the calculation of the expanded uncertainty we accept extension factor $k = 2$. Finally, the expanded uncertainty amounts $U(m) = 0.012 \text{ kg}$. Mass flow is determined on by measuring the mass which flows at a given time. It is described by this pattern:

$$Q_m = \frac{m}{t} \quad (3)$$

m – measured out liquid mass – kg

t – time of liquid measurement – s

Q_m – mass flow – kg/s

A complex uncertainty of measuring the mass flow is given by:

$$u(Q_m) = \sqrt{\left(\frac{1}{t}\right)^2 e_m^2 + \left(-\frac{\bar{m}}{t^2}\right)^2 e_t^2} \quad (4)$$

Measurement uncertainty increases with decreasing measurement time. It was assumed that the minimum measurement time will be $\bar{t} = 200\text{s}$. For maximum mass flow $Q_m = 360 \text{ kg/h}$, the weight of which will be metered is $\bar{m} = 20 \text{ kg}$. For the values above complex uncertainty was $u(Q_m) = 0.42 \text{ kg/h}$.

For the calculation of the expanded uncertainty assume factor $k = 2$. Finally, the expanded uncertainty for the mass flow $Q_m = 360 \text{ kg/h}$ was $U(Q_m) = 0.84 \text{ kg/h}$.

Measurement uncertainty in this method depends on the amount by weight of the solution, which we take. The greater the amount of weight, the more accurate measurement, however, increasing the weight extends the research.

4.2 Uncertainty of Temperature Measurement

The temperature is measured using temperature sensors Pt 100 type 371 and Czaki 816. Measurement error of temperature sensors declared by the manufacturer is: $e_t = \pm 0.3^\circ\text{C}$.

The uncertainty of the temperature measurement is calculated using type B method. We assume rectangular expansion, therefore the B type uncertainty is calculated from the formula:

$$u(t) = \frac{e_t}{\sqrt{3}} \quad (5)$$

We obtain measurement uncertainty equal to $u(t) = 0.18^\circ\text{C}$. The uncertainty of the temperature measurement is also affected by the temperature sensors module. Error of the signal processing from the temperature sensors, declared by the manufacturer of the module is: $e_p = \pm 0.3\%$. For the predetermined measurement range equal to $0 \div 100^\circ\text{C}$, measurement uncertainty is $e_p = \pm 0.3^\circ\text{C}$.

The uncertainty of the signal conversion is calculated using type B method. We assume rectangular expansion, therefore the B type uncertainty is calculated from the formula:

$$u(p) = \frac{e_p}{\sqrt{3}} \quad (6)$$

We obtain uncertainty of temperature signal processing equal to $u(p) = 0.18^\circ\text{C}$.

Complex uncertainty equals to:

$$u(tp) = \sqrt{u(t)^2 + u(p)^2} \quad (7)$$

Therefore, complex uncertainty was $u(tp) = 0.25^\circ\text{C}$. For the calculation of the expanded uncertainty assume factor $k = 2$. Finally, the expanded uncertainty was $U(tp) = 0.5^\circ\text{C}$.

4.3 Uncertainty of Measuring the Concentration of Ions

Ion concentration was measured using a conductometer JUMO CTI-750. The conductometer measured conductance, which computed the concentrations of the solution.

Error of conductance measurement, declared by the manufacturer of the module is: $e_c = \pm 0.5\%$. Uncertainty of mass measurement solution we calculate B type method. We assume rectangular expansion, therefore B type uncertainty we calculate using a formula as follows:

$$u(c) = \frac{e_c}{\sqrt{3}} \quad (8)$$

We obtain conductance measurement uncertainty equal to $u(k_{\%}) = 0.29\%$.

For conductometer's operating range 0-500 mS/cm measurement uncertainty is $u(k) = 1.5 \text{ mS/cm}$.

Converting it to a hydrochloric acid concentration we obtain $u(\text{HCl}) = 0.015\%$. Converting it to a sodium chloride concentration we obtain $u(\text{NaCl}) = 0.05\%$.

The uncertainty of measurement is also affected by the temperature at which the measurement is made. For a concentration of about 2% HCl influence of temperature in the range $10^{\circ}\text{C} \div 65^{\circ}\text{C}$ is described by equation (9). The equation was obtained by approximations of characteristics points used to compensate for the temperature influence on the concentration of HCl solution.

$$C_{Tk} = -0,698 \ln(T) + 4.2377 \quad (9)$$

C_{Tk} – concentration of hydrochloric acid, depending on temperature,

T – solution temperature,

$$u(C_{Tk}) = \frac{0.7}{T} e_T \quad (10)$$

$u(C_{Tk})$ – uncertainty of hydrochloric acid concentration measurement caused by temperature measurement error,

e_T – temperature measurement error.

Measuring station has temperature measurement error $e_T = 0,25^{\circ}\text{C}$. Maximum uncertainty of hydrochloric acid concentration measurement occurs for the lower temperatures. It was assumed that the lowest working temperature is $+10^{\circ}\text{C}$.

In this case, the uncertainty of measurement of the temperature is $u(C_{Tk}) = 0.018\%$.

The complex uncertainty of measurement of hydrochloric acid is described by the formula:

$$u(Ck) = \sqrt{u(\text{HCl})^2 + u(CTk)^2} \quad (11)$$

Complex uncertainty is $u(Ck) = 0,024\%$. For the calculation of the expanded uncertainty assume factor = 2. Finally, expanded uncertainty of concentration of hydrochloric acid measurement equals to $U(Ck) = 0.048\%$.

For a concentration of about 2% NaCl influence of temperature in the range $10^{\circ}\text{C} \div 65^{\circ}\text{C}$ is described by equation (12). The equation was obtained by approximations of characteristics points used to compensate for the temperature influence on the concentration of NaCl solution:

$$C_{Ts} = 0,0006 T^2 - 0,0778 T + 3,642 \quad (12)$$

C_{Ts} – the concentration of sodium chloride, depending on temperature,

T – solution temperature,

$$u(CTs) = (0.0012T - 0.0778)e_T \quad (13)$$

$u(CTs)$ – uncertainty of concentration measurement caused by temperature measurement error,

e_T – temperature measurement error.

The measurement station has temperature measurement error $e_T = 0,25^\circ\text{C}$. Maximum uncertainty of sodium chloride concentration measurement occurs for the lower temperatures. It was assumed that the lowest working temperature is $+10^\circ\text{C}$. In this case, the uncertainty of measurement of the temperature is $u(C_{Ts})=0,025\%$.

The complex uncertainty of measurement of concentration is described by the formula:

$$u(Cs) = \sqrt{u(NaCl)^2 + u(CTs)^2} \quad (14)$$

The complex uncertainty of measurement of concentration equals to $u(Cs) = 0.056\%$. For the calculation of the extended uncertainty we assume expansion factor $k = 2$. Finally, extended of measurement of the concentration of sodium chloride is $U(Cs) = 0.12\%$.

5 Tests of the Measuring Station

5.1 Tests of the Flow

A series of flow tests, under which specified the uncertainty of station measurement using A method, was performed.

Table 1. The results of the measurement uncertainty of the volume flow measurement for slotted line of the station

Q_s (dm ³ /h)	Q_w (dm ³ /h)	difference in measurement $Q_s - Q_w$ (dm ³ /h)	$U(Q_s)$ (dm ³ /h)	$U(Q_s \%)$ (%)	$U(Q_w)$ (dm ³ /h)	$U(Q_w \%)$ (%)
390.096	390.23	0.134	0.68	0.17	0.71	0.18
328.783	329.1	0.317	0.79	0.24	0.67	0.20
253.576	254.03	0.454	0.48	0.19	0.46	0.18
148.056	148.05	-0.006	0.41	0.28	0.37	0.24
25.14	24.27	-0.874	0.17	0.68	0.09	0.37

Volume flow was measured by flowmeter SIEMENS MAG 6000 SITRANS F M MAG1100 DN6. This volume flow was measured by weight. This was done by measuring the weight of the liquid supplied to the scales tank in a known period of time, and then counting the volume flow, taking into consideration the density of the liquid and temperature.

A series of 15 measurements in 5 points was performed. Table 1 presents mean values of the measurements. The standard uncertainty of readout set at a trust level equal to 95%.

Symbols from Table 1 are described below:

Q_s – volume flow measured with flowmeter SIEMENS MAG 6000 SITRANS F M MAG1100 DN6,

Q_w – volume flow measured with scales METTLER TOLEDO Typ KCC 150 RPT0151,

$U(Q_s)$ – standard uncertainty of readout of volume flow measurement measured with flowmeter SIEMENS MAG 6000 SITRANS F M MAG1100 DN6,

$U(Q_s \%)$ – standard uncertainty of readout of volume flow measurement measured with flowmeter SIEMENS MAG 6000 SITRANS F M MAG1100 DN6, related to result of measurement,

$U(Q_w)$ – standard uncertainty of readout of volume flow measurement measured with scales METTLER TOLEDO type KCC 150 RPT0151,

$U(Q_w \%)$ – standard uncertainty of readout of volume flow measurement measured with scales METTLER TOLEDO type KCC 150, related to result of measurement.

5.2 Tests of the Concentration Regulation

On the station tests were performed to control the concentration of hydrochloric acid. They consisted in checking the accuracy of the control system, which produces a solution having a predetermined concentration. Maximum average difference between the reference value and the resulting concentration was 0.04% HCl.

The standard uncertainty equal to $u(C_p) = 0,057\%$ at a trust level 95% was obtained. Relative extended uncertainty of measurement readout was $U(C_p) = 1,1\%$ for measurement range $0 \div 5\%$.

5.3 Tests of the Temperature Regulation

With which the measurement uncertainty, the station measures the temperature of the liquid, was checked. Indications of temperature sensors of the station with the standard thermometer were compared. The individual values of the uncertainty of the readout of measurement are correlated in table 2.

Table 2. Uncertainties of the temperature measurement in the individual points of the measuring station's installation

Uncertainty of the readout of the temperature in the top tank T2	0.18°C
Uncertainty of the readout of the temperature measurement before the graphene sensor T1	0.15°C
Uncertainty of the readout of the temperature in the bottom tank T3	0.15°C
Uncertainty of the readout of the temperature behind the the graphene sensor T4	0.17°C

Indications differences in the individual points of the station of the hydraulic system fluid circulation were checked for 11 temperature values.

When heated solution flows, temperature T_2 (measured in the top tank) decreases in relative to temperature T_w in the bottom tank, as a result of the heat exchange with the environment, according to the dependence:

$$\Delta T_{w-2} = 0.0256e^{0.0494T_w} (\text{°C}) \quad (15)$$

The dependence was obtained by approximation of measurement results. It is valid in liquid temperature range $+15^{\circ}\text{C} \div +65^{\circ}\text{C}$, and for the environment temperature $T_0 = +20^{\circ}\text{C}$. Measurement uncertainty was $\pm 0.18^{\circ}\text{C}$.

When heated solution flows, temperature T_1 , measured before the graphene sensor, decreases in relative to temperature T_2 in the top tank, according to the dependence:

$$\Delta T_{2-1} = 0.0003T_2^2 - 0.0115T_2 + 0.1793(\text{ }^{\circ}\text{C}) \quad (16)$$

The dependence was obtained by approximation of measurement results. It is valid in liquid temperature range $+15^{\circ}\text{C} \div +65^{\circ}\text{C}$, and for the environment temperature $T_0 = +20^{\circ}\text{C}$. Measurement uncertainty was $\pm 0.18^{\circ}\text{C}$.

Temperature T_4 , measured behind the graphene sensor, decreases in relative to temperature T_1 (before the graphbene sensor), according to the dependence:

$$\Delta T_{1-4} = 0.0002T_1^2 - 0.0101T_1 + 0.128(\text{ }^{\circ}\text{C}) \quad (17)$$

The dependence was obtained by approximation of measurement results. It is valid in liquid temperature range $+15^{\circ}\text{C} \div +65^{\circ}\text{C}$, and for the environment temperature $T_0 = +20^{\circ}\text{C}$. Measurement uncertainty was $\pm 0.18^{\circ}\text{C}$.

6 Conclusion

Applied for the station measurement systems allow for objective and accurate execution of tests. The station is able to work in automatic mode, what reduces the formation of measurement errors caused by man. The station has been developed for research in hydrochloric acid medium, but also gives you the opportunity to work in solutions containing other ions.

Construction and operation of the station ensures both the safety of persons doing the tests, as well as instruments and devices installed on it.

The station allows for the operating parameters defined below:

- Range of volume flow measurement $0 \div 520 \text{ dm}^3/\text{h} \pm 0.7 \text{ dm}^3/\text{h}$
- Solution temperature $+10 \div +65 \text{ }^{\circ}\text{C} \pm 0.2^{\circ}\text{C}$
- Concentration of the solution HCl $0 \div 5 \% \pm 0.06\%$
- Concentration of the solution NaCl $0 \div 15 \% \pm 0.12\%$
- Solution mass $50 \div 150 \text{ kg} \pm 0.01\text{kg}$.

It is possible to extend the parameters of the station, if during testing the grapheme conductor it turns out that it is necessary to conduct research in other conditions. Actions aimed at reducing the measurement uncertainty of the station will be running.

The station allows to perform research, which may result in the launch of a new generation of flowmeters. Features of graphene theoretically allow for the construction of highly sensitive flowmeters with small dimensions, insensitive to operational environment, whose main advantage is the absence of an external power source. Innovation and versatility graphene flow sensors can be used in various industries, medicine and everyday life. Flow convectors examined on this station are expected to be used firstly in heatmeters households.

Acknowledgment. This work was partially supported by The National Centre of Research and Development (Poland) within GrafTech Program.

References

1. Guide to Expression of Uncertainty in Measurement, ISO 1995, Switzerland. Polish Trans. By GUS, Warsaw 999 (1995)
2. Arendarski, J.: Niepewność pomiarów. Oficyna Wydawnicza Politechniki Warszawskiej (2006) (in Polish)
3. Newaz, A.K.M., Markov, D.A., Prasai, D., Bolotin, K.I.: Graphene Transistor as a Probe for Streaming Potential
4. He, R.X., Lin, P., Liu, Z.K., Zhu, H.W., Zhao, X.Z., Chan, H.L.W., Yan, F.: Solution-Gated Graphene Field Effect Transistors Integrated in Microfluidic Systems and Used for Flow Velocity Detection. *Nano Letters* (2012)
5. Dhiman, P., Yavari, F., Mi, X., Gullapalli, H., Pulickel, Y.S., Ajavan., M., Koratkar, N.: Harvesting Energy from Water Flow over Graphene
6. Heinz, G.: Technika pomiarów przepływu wody i ścieków Seidel-Przywecki (1999)
7. Turkowski, M.: Przemysłowe sensory i przetworniki pomiarowe Oficyna Wydawnicza Politechniki Warszawskiej (2002)
8. Jun, S.-W., Yasutomi, K.: Linearized settling error calibration for a pipeline A/D converter using non-slewing amplifiers. *JAMRIS* 4 (2009)
9. Bartyś, M.: Smart volume flow rate transmitter. *Pomiary Automatyka Robotyka* 14(11), 74–78 (2010) (in Polish)

Influence of Tensile Force on Magnetic Properties of Amorphous Fe₈₀B₁₁Si₉ Alloys in Different States of Thermal Relaxation

Jacek Salach¹, Dorota Jackiewicz², Adam Bieńkowski¹,
Michał Nowicki¹, and Magdalena Gruszecka¹

¹ Institute of Metrology and Biomedical Engineering, Warsaw University of Technology,
Warsaw, Poland

{j.salach,a.bienkowski,m.nowicki}@mchtr.pw.edu.pl

² Industrial Research Institute for Automation and Measurements, Warsaw, Poland
d.jackiewicz@mchtr.pw.edu.pl

Abstract. The paper presents the method of measurements of the tensile stresses dependence of the magnetic characteristics of the ring-shaped cores made of Fe₈₀B₁₁Si₉ amorphous alloy in as quenched and annealed state. The results of investigation on influence of tensile stresses on magnetic characteristics of those cores have been done. First core was in as-quenched state, whereas others were annealed in 350°C for one hour, annealed in 355°C for one hour, and annealed in 360°C for one hour. Presented results confirm the high magnetoelastic sensitivity of Fe₈₀B₁₁Si₉ alloy in as-quenched and annealed states.

Keywords: amorphous magnetic alloys; magnetoelastic effect; tensile stresses; thermal relaxation.

1 Introduction

Knowledge about the magnetoelastic properties of soft amorphous alloys is very important from both practical and theoretical points of view. Thermodynamically reverse effect connected with the influence of the external tensile stresses on the magnetic properties of this alloys, so called magnetoelastic Villari effect [2, 4, 5], has also significant, technical consequences.

From the other hand knowledge about the influence of stresses on the magnetic properties of magnetic materials, despite the many years of development, is still low. There are many studies on the effects of compressive stresses on both the ceramic magnetic materials [7], as well as amorphous magnetic materials [2, 6, 9, 10–13]. Effect of the shear stresses which is a composite effect of the compressive and tensile stresses acting at 45 degrees to the direction of the magnetic circuit is also known from the literature [8]. However, there is a lack of knowledge about the effects of tensile stresses on the magnetic properties of magnetic materials, both crystalline and amorphous materials. The paper is designed to fill this gap, presenting the method to obtain a uniform state of stresses in the material as well as obtaining a closed

magnetic circuit. The tested material was also subjected to thermal relaxation aimed at demonstrating the applicability of amorphous materials as the core tension sensors.

2 Methodology of Investigation

Investigation was carried out on four ring-shaped cores made of $\text{Fe}_{80}\text{B}_{11}\text{Si}_9$ amorphous alloy. Cores were wound of amorphous ribbon to ring shaped cores. Each core had outside diameter 32 mm, inside diameter 25 mm and height 10 mm. First core was in as-quenched state, second was annealed in 350°C for one hour in protective atmosphere, third was annealed in 355°C for one hour and fourth annealed in 360°C for one hour [3].

The tensile stresses σ from the external force F are applied in a perpendicular direction to the direction of the magnetizing field H in the core. The main problem with this method is winding the core. This problem was solved by means of the device shown schematically in figure 1.

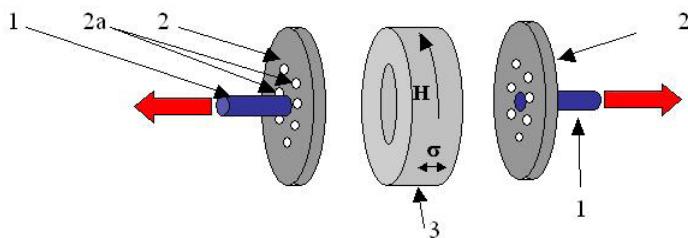


Fig. 1. Schematic diagram of device for application of uniform tensile stresses σ to the ring-shaped core: 1 – rod, 2 – nonmagnetic backings, 2a – holes for the magnetizing and measuring windings, 3 – amorphous ring shaped core

The amorphous ring shaped core (3) is fixed between two special backings (2) made of non-magnetic material (brass). In the backings have been made holes (2a), allowing for winding of the core. The tensile forces are applied to the rods (1).

The presented methodology of the application of the tensile stress on the magnetic characteristics of amorphous ring-shaped cores allows for a uniform stress distribution.

Mechanical setup in which investigation was carried out is presented in figure 2. Adjustable screw tensile stress generator (5) acts on the reference force sensor (2) coupled with electronic processing (3) and the tested core (1). The amorphous ring shaped core is wound by magnetizing and measuring windings.

Influence of tensile stresses on magnetic characteristics was measured according to the following procedure: 1 – applying the tensile force F , 2 – obtaining the stresses in the core, 3 – demagnetization of the core, 4 – determination of the $B(\sigma)_{H_m}$ characteristics based on the measured characteristics of $B(H)_\sigma$.

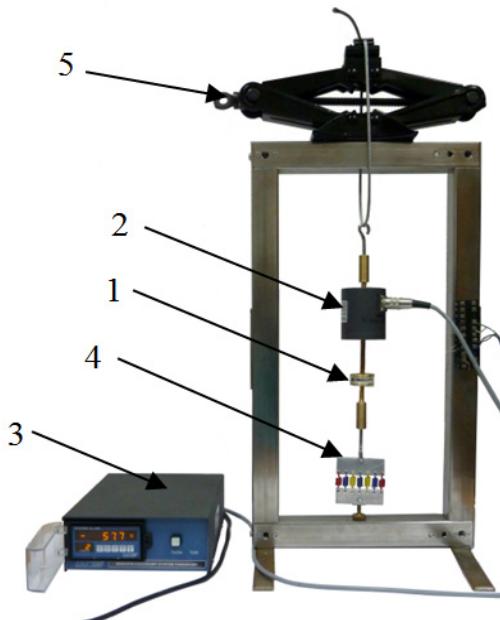


Fig. 2. Mechanical setup for application of tensile stresses σ in the ring-shaped core: 1 – amorphous ring shaped core, 2 – verification sensor, 3 – electronic processing staff of sensor, 4 – set of spring, 5 – force generator

Tensile stresses obtained in the cores during the measurements were from 0 MPa to 3 MPa at 0,25 MPa intervals.

3 Results

Figure 3 shows the influence of tensile stresses σ on the shape of $B(H)_\sigma$ hysteresis loops of Fe₈₀B₁₁Si₉ amorphous alloy. Under the tensile stresses σ up to 3 MPa, value of flux density B significantly decreased, and the value of coercive field H_c does not change.

The magnetoelastic $B(\sigma)_{H_m}$ characteristics of cores made of Fe₈₀B₁₁Si₉ amorphous alloy is presented in figure 4. Under the influence of tensile stresses, the value of maximal flux density B decreases. Figure 5 shows the influence of tensile stresses σ on the magnetic permeability. Based on these results the magnetoelastic sensitivity can be determined. Sensitivity was calculated as the relative change in the magnetic permeability of the sample for the full load range of 0 to 3 MPa. For core in as-quenched state, the magnetoelastic sensitivity is equal to 58,11%. For core annealed in 350°C for one hour, the magnetoelastic sensitivity is equal to 63,36%, for core annealed in 355°C for one hour magnetoelastic sensitivity is equal to 56,86%. And for core annealed in 360°C for one hour, the magnetoelastic sensitivity is equal to 60,28%.

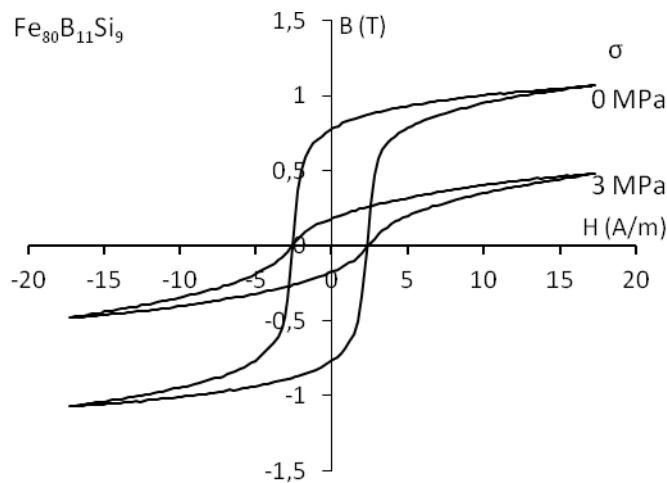


Fig. 3a. Influence of tensile stresses σ on the shape of hysteresis loop $B(H)_\sigma$ of tested cores in as quenched state

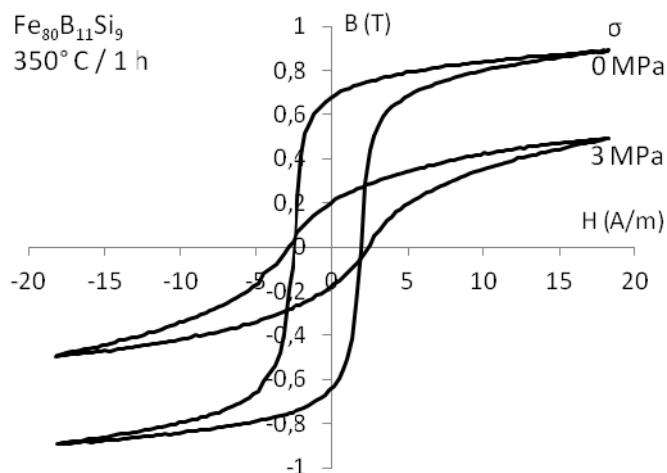


Fig. 3b. Influence of tensile stresses σ on the shape of hysteresis loop $B(H)_\sigma$ of tested cores after annealing in 350°C for 1 hour

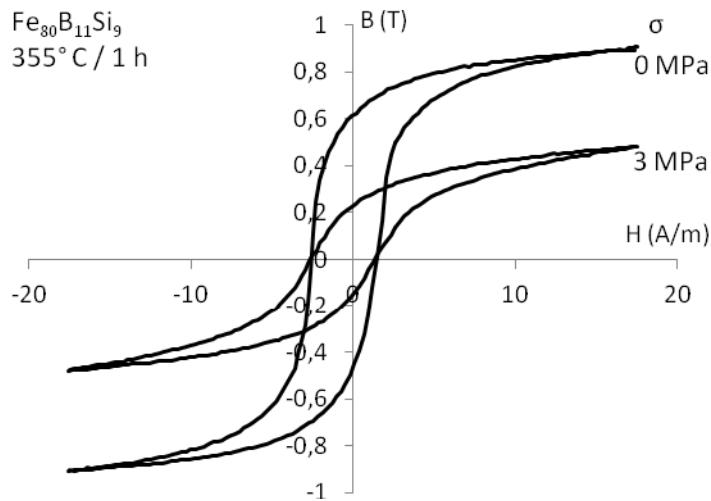


Fig. 3c. Influence of tensile stresses σ on the shape of hysteresis loop $B(H)_\sigma$ of tested cores after annealing in 355 °C for 1 hour

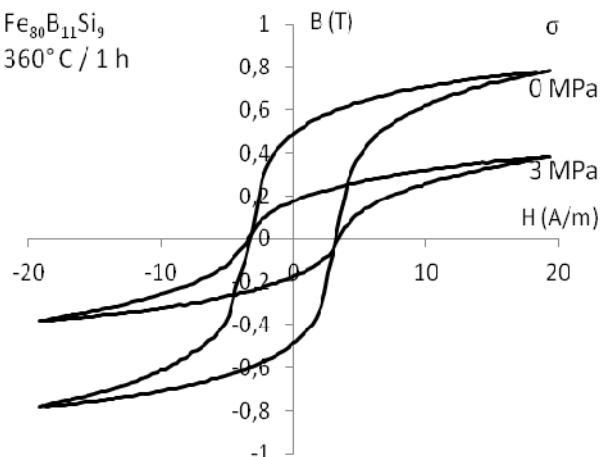


Fig. 3d. Influence of tensile stresses σ on the shape of hysteresis loop $B(H)_\sigma$ of tested cores after annealing in 360°C for 1 hour

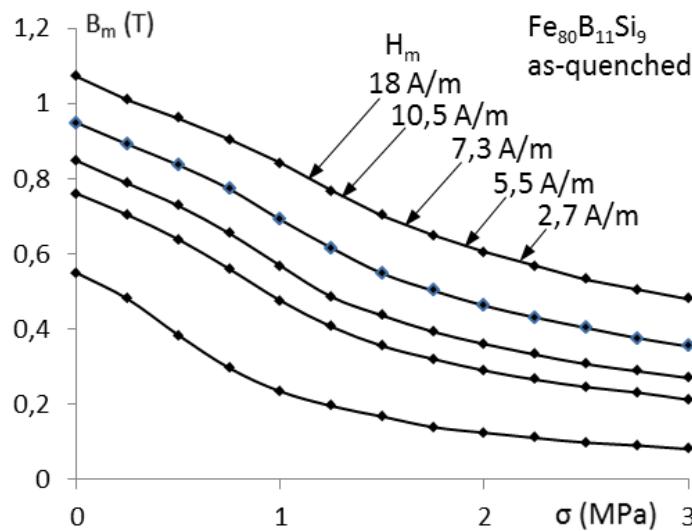


Fig. 4a. Magnetoelastic $B(\sigma)_{H_m}$ characteristics of tested cores made of $Fe_{80}B_{11}Si_9$ amorphous alloys in as quenched state

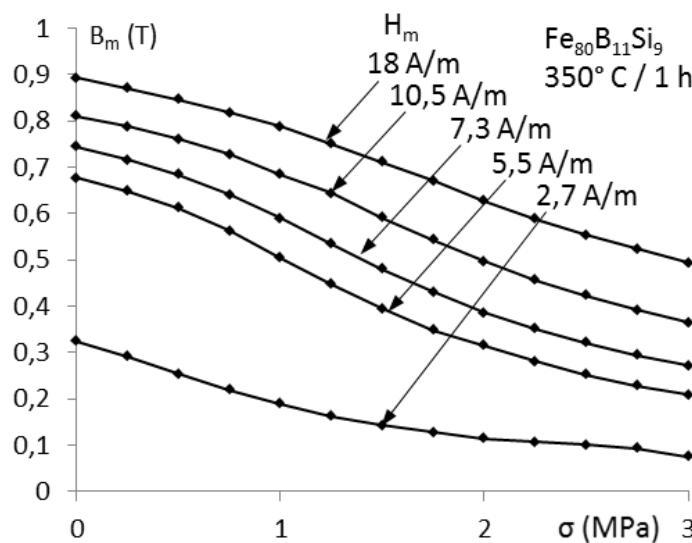


Fig. 4b. Magnetoelastic $B(\sigma)_{H_m}$ characteristics of tested cores made of $Fe_{80}B_{11}Si_9$ amorphous alloys after annealing in $350^\circ C$ for 1 hour

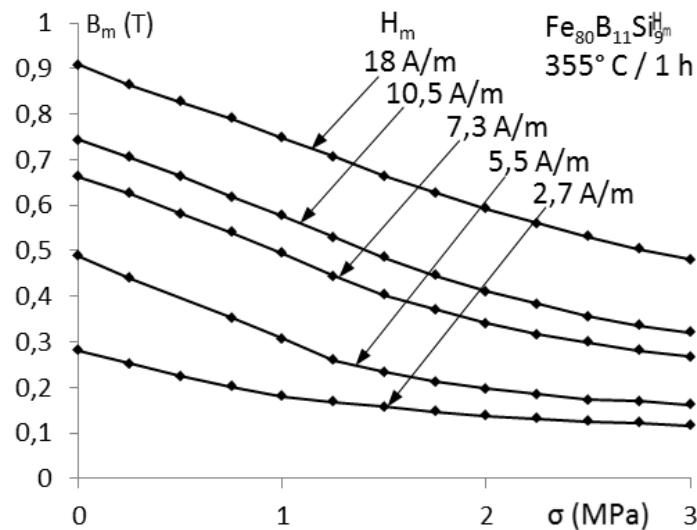


Fig. 4c. Magnetoelastic $B(\sigma)_{H_m}$ characteristics of tested cores made of $Fe_{80}B_{11}Si_9$ amorphous alloys after annealing in $355^{\circ}C$ for 1 hour

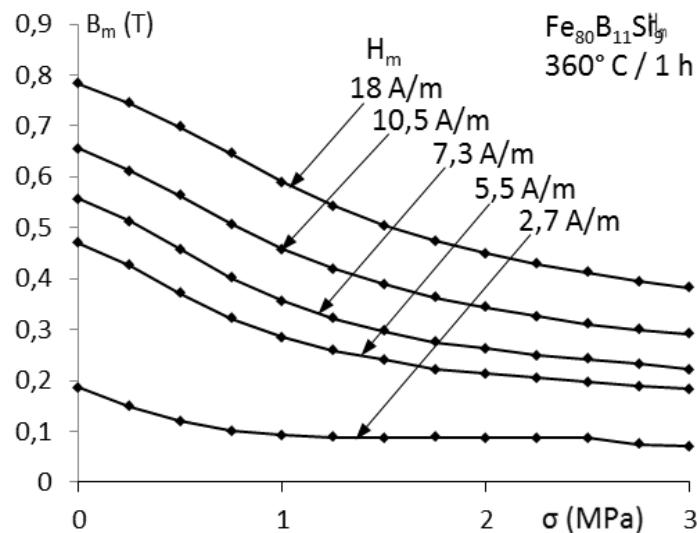


Fig. 4d. Magnetoelastic $B(\sigma)_{H_m}$ characteristics of tested cores made of $Fe_{80}B_{11}Si_9$ amorphous alloys after annealing in $360^{\circ}C$ for 1 hour

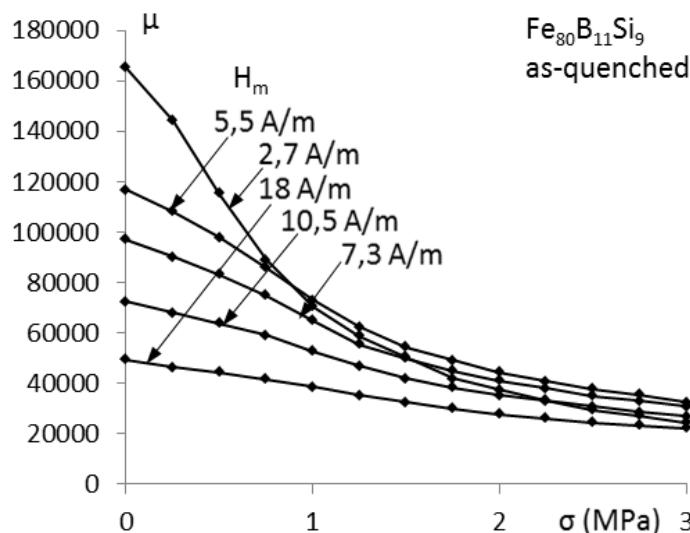


Fig. 5a. Influence of tensile stresses σ on the magnetic permeability of tested cores made of $Fe_{80}B_{11}Si_9$ amorphous alloys in as quenched state

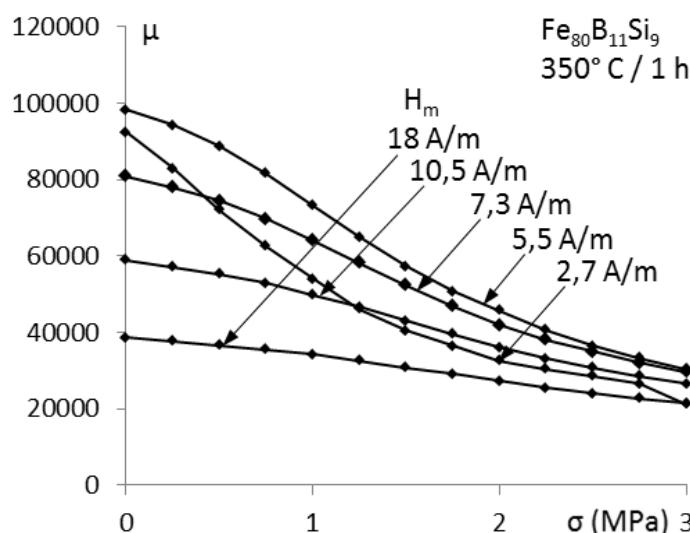


Fig. 5b. Influence of tensile stresses σ on the magnetic permeability of tested cores made of $Fe_{80}B_{11}Si_9$ amorphous alloys after annealing in $350^\circ C$ for 1 hour

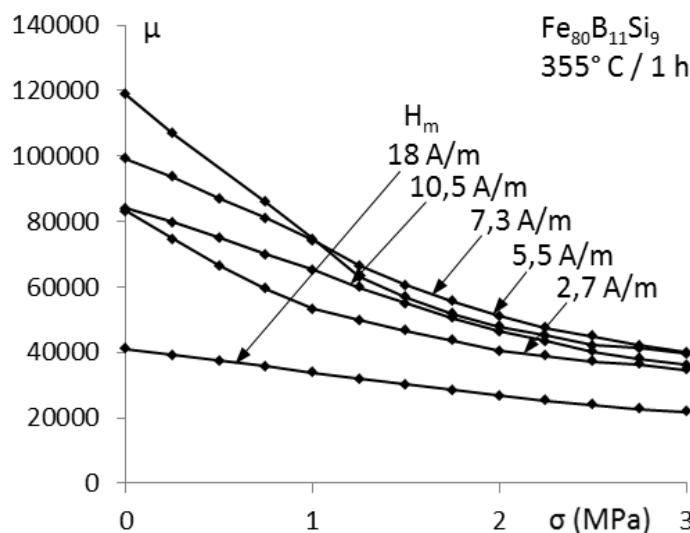


Fig. 5c. Influence of tensile stresses σ on the magnetic permeability of tested cores made of Fe₈₀B₁₁Si₉ amorphous alloys after annealing in 355°C for 1 hour

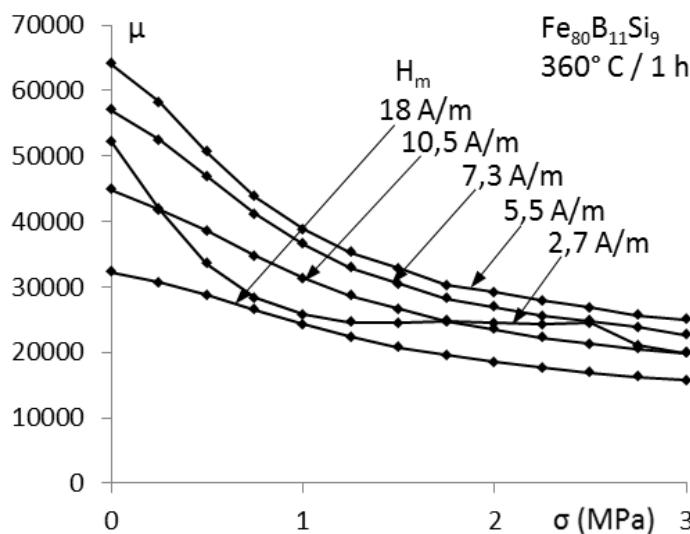


Fig. 5d. Influence of tensile stresses σ on the magnetic permeability of tested cores made of Fe₈₀B₁₁Si₉ amorphous alloys after annealing in 360°C for 1 hour

4 Conclusions

The measurement method presented in this paper is adequate for the measurements of the tensile stresses dependence of the magnetic characteristics of the ring-shaped cores made of Fe₈₀B₁₁Si₉ amorphous alloy in as quenched and annealed state.

It was observed, that the magnetoelastic sensitivity changes with the thermal annealing of the amorphous alloy sample. The highest value was obtained after annealing in 350°C for 1 hour. However, for the remaining cores, the magnetoelastic sensitivity was also very high.

Presented results confirm, that the amorphous Fe₈₀B₁₁Si₉ alloy can be used in the development of magnetoelastic sensors of the tensile stresses σ . Due to the high magnetoelastic sensitivity such sensors can be used in industrial applications, mechatronic systems, civil engineering [1].

Acknowledgment. This work was partially supported by The National Centre of Research and Development (Poland) within grant no. PBS1/B4/6/2012.

References

1. Frydrych, P., Szewczyk, R., Salach, J., Trzcinka, K.: Two-Axis, Miniature Fluxgate Sensors. *IEEE Trans. Magn.* 48, 1485–1488 (2012)
2. Salach, J., Szewczyk, R., Bieńkowski, A., Frydrych, P.: Methodology of testing the magnetoelastic characteristics of ring-shaped cores under uniform compressive and tensile stresses. *J. Electr. Eng.* 61, 93–95 (2010)
3. Szewczyk, R., Svec, S.P., Svec, P., Salach, J., Jackiewicz, D., Bieńkowski, A., Hosko, J., Kamiński, M., Winiarski, W.: Thermal annealing of soft magnetic materials and measurements of its magnetoelastic properties. *Measurements Automation Robotics* 2, 513–518 (2013)
4. Kolano-Burian, A., Varga, L.K., Kolano, R., Kulik, T., Szynowski, J.: High-frequency soft magnetic properties of Finemet modified with Co. *J. Magn. Magn. Mater.* 316, e820–e822 (2007)
5. O’Handley, R.: Modern magnetic materials – principles and applications. John Wiley & Sons (2000)
6. Shi, Y., Zang, Y.: Application of the nanocrystalline alloy in lubrication oil pressure- measuring for auto engines. *Adv. Mat. Res.* 661, 7–10 (2013)
7. Bieńkowski, A., Rozniatowski, K., Szewczyk, R.: Effects of stress and its dependence on microstructure in Mn-Zn ferrite for power applications. *J. Magn. Magn. Mater.* 255, 547–549 (2003)
8. Salach, J., Bieńkowski, A., Szewczyk, R.: The ring-shaped magnetoelastic torque sensors utilizing soft amorphous magnetic materials. *Journal of Magnetism and Magnetic Materials* 0304-8853 316, e607–e609 (2007)
9. Liu, K.-H., Lu, Z.-C., Liu, T.-C., Li, D.-R.: Magnetoelastic anisotropy of FeSiB glass-coated amorphous microwires. *Chinese Phys. Lett.* 30, 017501 (2013)
10. Salach, J., Bieńkowski, A., Szewczyk, R., Jackiewicz, D., Fydrych, P., Kolano-Burian, A.: Utilizing of magnetic amorphous alloys in magnetoelastic tensile stress sensors. *Pomiary Automatyka Robotyka* 16(2), 556–560 (2012) (in Polish)

11. Bienkowski, A., Szewczyk, R., Kolano, R.: Influence of thermal treatment on magnetoelectric Villari effect in Fe₇₈Si₁₃B₉ amorphous alloy. *Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing* 375, SI, 1024–1026 (2004)
12. Bienkowski, A., Szewczyk, R.: New possibility of utilizing amorphous ring cores as stress sensor. *Physica Status Solid A-Applied Research* 189(3), 787–790 (2002)
13. Szewczyk, R.: Modelling of the magnetic and magnetostrictive properties of high permeability Mn-Zn ferrites. *J. of Physics* 67(6), 1165–1171 (2006)

Application of X-ray Fluorescence to Determine Qualitative Parameters of Coal

Waldemar Sobierajski, Marek Kryca, and Artur Kozłowski

EMAG Institute of Innovative Technologies, Katowice, Poland

{Waldemar.Sobierajski,Marek.Kryca,Artur.Kozlowski}@emag.pl

Abstract. EMAG Institute has specialized in working out the devices for qualitative control of coal. In most cases, they are radiometer appliances taking advantage of absorption or dispersion of gamma radiation of artificial radioactive source or natural radioactivity (ALFA and RODOS Ash Meters). Such methods not always ensure a required accuracy of measurement in case of mineralogical changes of a tested material. In 1980s a PYLOX Meter was worked out in EMAG that makes use of X-ray fluorescence for determining contents of non-combustible solids in mine dust from zones which protect against dust explosions. It was recognised that it is the only method which could allow to measure the quality of hard coal of a very diversified mineralogical contents. Therefore, it was decided to research aimed to determine the usability of PYLOX Meter for measurements of quality of coals from different coal mines and the possibility of improvement of metrological parameters by using semiconductor detectors. Research was conducted on samples originated from Uzbekistan and „Piast” Coal Mine. On the basis of received outcomes, it was stated that accurate determination of qualitative parameters of hard coal in the size grade of 0–3 mm is possible. The accuracy of order of 0,02 % A_a (ash) and 0,002 % S_t (sulphur) was obtained.

Keywords: Qualitative Parameters Control, Radiometr Measurement Devices, Ash Meter, X-ray Fluorescence.

1 Introduction

One of the disciplines that EMAG Institute carries out is a control of coal quality. Within the confines of it, EMAG has worked out a range of devices used both for continuous measurements and lab-technological ones. In majority, there are the appliances taking advantage of absorption or scattering phenomenon by a tested material of gamma radiation originated from artificial radioactive sources or natural radioactivity. Appliances based on this rule are ash meters: ALFA, GAMMA NATURA 2, LAWON and RODOS (non-isotopic one, offered in explosion-proof version as well).

The basic limitation of listed ash meters is sensitivity to changes of mineralogical contents of the material. In practice, the measurement of mixtures quality obtained from different coal mines and even coal beds may be impossible. In case of meters

using natural radioactivity, an additional limitation is the level of radiation determining minimum measurement time, disturbing the impact of background radiation and occurrence of dependences between the radiation intensity and a determined property.

In 1980s in EMAG, it was worked out the PYLOX Meter of Non-combustible Residue that uses X-ray fluorescence. Basing on the experiences gained during its working out and implementation, it was decided to determine the requirements for a new generation of appliances that could be applied for coals of a distinct origin as well.

2 PYLOX Meter

POLOX Meter has been provided for, first of all, as a laboratory device for a precise determination of non-combustible residue in mine dust. A measurement method applied in PYLOX Meter takes advantage of X-ray fluorescence. On the basis of activated characteristic radiation in a sample (K lines of calcium and iron and dispersion range) the contents of non-combustible residue was determined.

Originally, PYLOX Meter consisted of: a measuring head containing a detector, activated radiation source, a container for a tested sample and a computer with multichannel analyser. As a result of research papers, it was decided that processing-calculating system will be placed outside the computer forming an independent meter with a head. Whereas, the computer will fulfil the function of visualisation, data and measurement results archiving. Appearance of PYLOX Meter was shown in Fig. 1, and in Fig. 2 there was presented a counting geometry.



Fig. 1. Appearance of PYLOX Meter

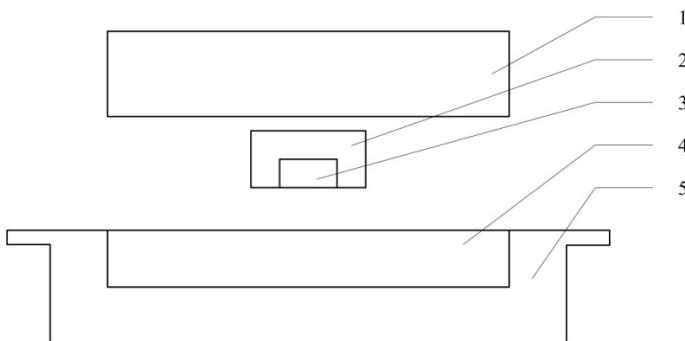


Fig. 2. Counting geometry of PYLOX Meter: 1 – Detector, 2 – Source container, 3 – Radiation Source, 4 – Sample, 5 – Measuring Vessel

The function of detector is fulfilled by an argon proportional counter. As an activated radioactive source it was used the Pu-238 one.

Originally, it was assumed that in order to gain a satisfactory measurement accuracy, a sample should be in an air moisture free analytical state, freely heaped and a surplus swept. The shape of appliance for sweeping the surplus of a material ensures simultaneously a good flattening of a sample's surface.

Obtaining the measurements accuracy for real conditions amounted to below 0,5 % of non-combustible residue contents. Next, it was decided to spread the application range of PYLOX Meter for measuring coal quality. Measured parameters were the contents of ash and sulphur. For analytical samples the accuracies have been obtained better than 0,3 % A_a and 0,05 % S_t . Later, it was stated that a quite precise determination of contents of non-combustible residue in mine dusts in raw state is possible, i.e. of graining of 0–1 mm, what significantly shortened the preparation time of a sample – it was enough to lead it to air moisture free state. During the utilisation, it was stated that in some cases, PYLOX ensures a sufficient accuracy of measurements in the event of disturbances occurrence in the form of indefinite impurities of smoke dusts. It let one suppose that the most proper method used for determining qualitative parameters of coals of various origin can be X-ray fluorescence.

3 Comparative Measurements

In order to obtain a comparative material, first of all, the measurements have been conducted by means of PYLOX Meter. The samples of hard coal from Uzbekistan and “Piast” Coal Mine were the tested material. Uzbek samples have been got in order to perform a calibration of Gamma-Natura 2 Ash Meter. On the samples there was tested an usability of absorptive method applied in LAWON Ash Meter as well. The results gained by the methods used in such devices were also taken into consideration to assess the outcomes received by X-ray fluorescence method.

The measurements have been conducted for two principal cases: a sample of a small graining and in the air moisture free state, and in the analytical air moisture free state.

3.1 Measurements Taking Advantage of Proportional Counter

Applied in PYLOX Meter proportional counter filled with argon is characterised by a relatively high active flank. Effective scope of activity amounts to from below 3 keV to 20 keV, and resolving power approx. 1 keV for 5,9 keV. Mainly, for the sake of resolving power, it is not treated as a perspective solution for new devices.

A representative sample spectrum of hard coal gained by means of a proportional counter was presented in Fig. 3.

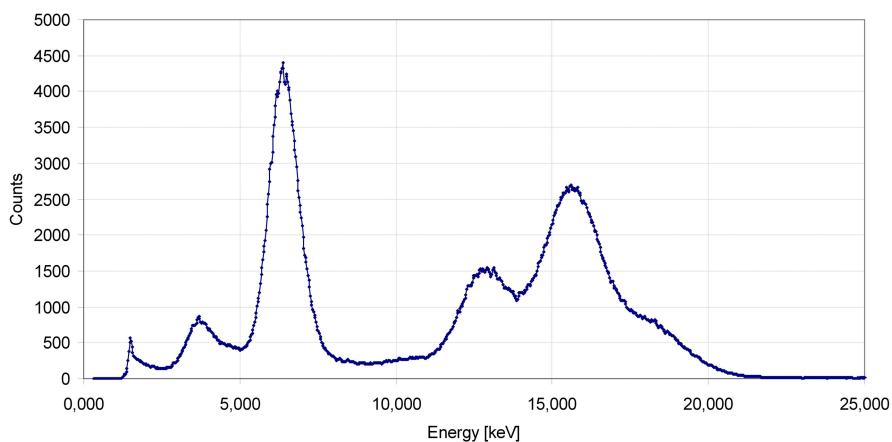


Fig. 3. Spectrum of activated radiation of a hard coal sample gained by means of a proportional counter

Outcomes of measurements for hard coal samples from Uzbekistan in relation to laboratory determinations were included in Tables 1 and 2.

Table 1. Outcomes of measurements of hard coal samples from Uzbekistan. The sample are in the air moisture free state. The measurements conducted by means of a proportional counter.

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{tlab} [%]	S _t [%]	S _{tlab} -S _t [%]
uz_1	42.51	43.02	-0.51	1.78	1.46	0.34
uz_2	40.83	41.23	-0.40	1.01	1.28	-0.27
uz_3	55.06	54.28	0.78	1.27	1.11	0.16
uz_4	54.32	54.54	-0.22	1.51	1.66	-0.15
uz_5	44.87	45.22	-0.35	1.04	1.17	-0.13
uz_6	34.85	34.15	0.70	1.68	1.64	0.04
	s.d.= 0,92			s.d.= 0.35		

Table 2. Outcomes of measurements of hard coal samples from Uzbekistan. The sample are in the air moisture free state. The measurements conducted by means of a proportional counter.

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{lab} [%]	S _t [%]	S _{lab} -S _t [%]
uz_1	42.51	43.92	-1.41	1.78	1.70	0.08
uz_2	40.83	40.81	0.02	1.01	1.01	-0.001
uz_3	55.06	54.02	1.04	1.27	1.29	-0.02
uz_4	54.32	54.31	0.008	1.51	1.54	-0.03
uz_5	44.87	45.48	-0.61	1.04	1.02	0.02
uz_6	34.85	33.90	0.96	1.68	1.73	-0.05
	s.d.= 0.84			s.d.= 0.07		

Denotations accuracy, also for samples in the analytical state were similar to ones gained in „Gama Natura” Such bad results can be justified by a fact that the samples originated from different coal mines, as it was turned out. However, LAWON Ash Meter coped with it pretty well, despite using less-sophisticated measurement method.

3.2 Measurements with the Application of Semiconductor Detector

Semiconductor detectors are characterised by much better parameters than proportional counters. For comparative measurements, it was selected a Si-PIN Detector. In comparison with a proportional counter, it possesses a relatively small active flank (13 mm²). The range of recorded energies is similar to the case of a proportional counter and amounts to from 2 to 30 keV, but the resolving power is much better – 150 eV.

Counter geometry applied in measurements by semiconductor detector was presented in Fig. 4. The spectrum of activated radiation of one of the samples of hard coal from Uzbekistan was presented in Fig. 5.

For the sake of a significantly less active flank, the measurement time was longer than in case of a proportional counter and amounted to 1000 s. The measurements outcomes were presented in Tables 3 and 4.

The accuracy obtained by means of a semiconductor detector is much better than in case of the application of a proportional counter. However, values of 0,76 % A_a and 22 % S_t (raw state 0–3 mm) and 0,54 % A_a and 0,08 % S_t (analytical state) should be considered insufficient.

During the previous research for the sample uz_1, as a result of conducting chemical determinations, considerably different outcomes from themselves within the range of 26 %, 42,50 % A_a have been obtained. An assumption has occurred that a sample is tightly heterogeneous. Therefore, it was decided to repeat the test but without sample uz_1. The results were presented in Tables 5 and 6.

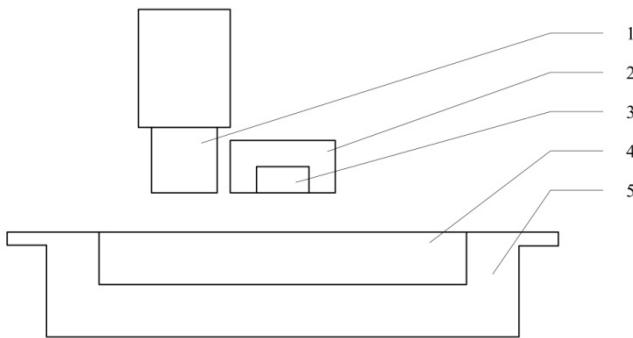


Fig. 4. Counter geometry applied for measurements with the application of semiconductor detector: 1 – Detector, 2 – Source Container, 3 – Radiation Source, 4 – Sample, 5 – Measuring Vessel

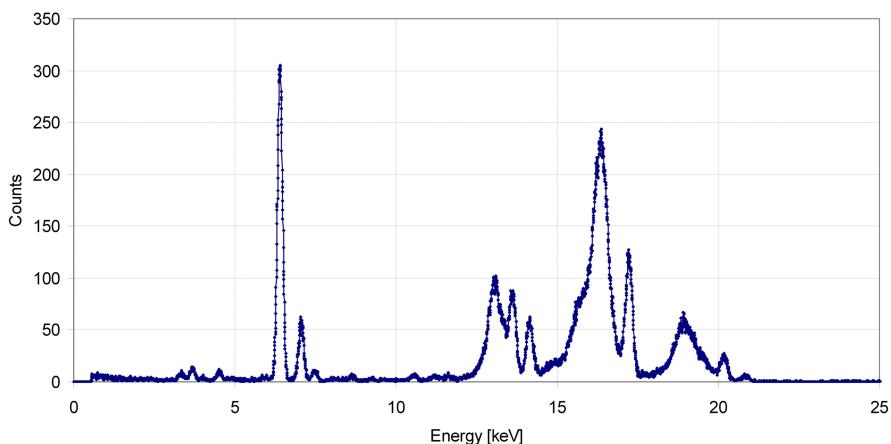


Fig. 5. Spectrum of activated radiation of a hard coal sample from Uzbekistan obtained by means of a Si-PIN semiconductor detector

Table 3. Outcomes of measurements of hard coal samples from Uzbekistan. The sample are in the air moisture free state. The measurements conducted by means of a semiconductor detector.

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{tlab} [%]	S _t [%]	S _{tlab} -S _t [%]
uz_1	42.51	42.69	-0.18	1.78	1.80	-0.02
uz_2	40.83	40.56	0.27	1.01	0.94	0.072
uz_3	55.06	55.26	-0.20	1.27	1.26	0.01
uz_4	54.32	53.68	0.64	1.51	1.41	0.10
uz_5	44.87	45.62	-0.75	1.04	1.25	-0.21
uz_6	34.85	34.63	0.22	1.68	1.63	0.05
	s.d.= 0.76			s.d.= 0.16		

Table 4. Outcomes of measurements of hard coal samples from Uzbekistan. The sample are in the analytical air moisture free state. The measurements conducted by means of a semiconductor detector.

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{tlab} [%]	S _t [%]	S _{tlab} -S _t [%]
uz_1	42.51	43.11	-0.60	1.78	1.69	0.089
uz_2	40.83	40.75	0.085	1.01	1.03	-0.015
uz_3	55.06	54.94	0.12	1.27	1.26	0.013
uz_4	54.32	54.11	0.21	1.51	1.57	-0.056
uz_5	44.87	45.06	-0.19	1.04	1.03	0.015
uz_6	34.85	34.48	0.37	1.68	1.73	-0.045
	s.d.= 0.54			s.d.= 0.08		

Table 5. Outcomes of measurements of hard coal samples from Uzbekistan. The sample are in the air moisture free state. The measurements conducted by means of a semiconductor detector. Sample uz_1 was excluded from the calibration

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{tlab} [%]	S _t [%]	S _{tlab} -S _t [%]
uz_2	40.83	40.84	-0.0063	1.01	1.012	-0.002
uz_3	55.06	54.99	0.0671	1.27	1.248	0.022
uz_4	54.32	54.37	-0.0537	1.51	1.527	-0.017
uz_5	44.87	44.90	-0.0263	1.04	1.049	-0.009
uz_6	34.85	34.83	0.0193	1.68	1.674	0.006
	s.d.= 0.09			s.d.= 0.03		

Table 6. Outcomes of measurements of hard coal samples from Uzbekistan. The sample are in the analytical air moisture free state. The measurements conducted by means of a semiconductor detector. Sample uz_1 was excluded from the calibration.

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{tlab} [%]	S _t [%]	S _{tlab} -S _t [%]
uz_2	40.83	40.83	-0.00	1.01	1.010	-0.000
uz_3	55.06	55.05	0.012	1.27	1.269	0.001
uz_4	54.32	54.33	-0.010	1.51	1.511	-0.001
uz_5	44.87	44.87	-0.003	1.04	1.040	-0.000
uz_6	34.85	34.85	0.003	1.68	1.680	0.000
	s.d.= 0.02			s.d.= 0.002		

Very well accuracies that were obtained after excluding sample uz_1 would confirm an assumption of its strong heterogeneity. Finally, it was decided to conduct another series of measurements, in which each of the samples was measured three times and as a result accepting an average value. The measurements were performed only for an analytical state. The results were presented in Fig. 7.

Table 7. Outcomes of measurements of hard coal samples from Uzbekistan. The sample are in the analytical air moisture free state. Each of the samples were measured three times. The measurements conducted by means of a semiconductor detector.

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{tlab} [%]	S _t [%]	S _{tlab} -S _t [%]
uz_1	42.51	42.47	0.04	1.78	1.776	0.004
uz_2	40.83	40.81	0.02	1.01	1.005	0.005
uz_3	55.06	55.10	-0.04	1.27	1.267	0.003
uz_4	54.32	54.30	0.04	1.51	1.510	-0.000
uz_5	44.87	44.87	-0.00	1.04	1.046	-0.006
uz_6	34.85	34.89	-0.04	1.68	1.685	-0.005
s.d.= 0.05				s.d.= 0.008		

Obtained results are a little bit worse than the previous ones, but still they should be regarded as very well and taking into consideration all of the samples.

Then, the measurements for hard coal samples originated from “Piast” Coal Mine have been conducted. The results of the measurements for the samples in the analytical state were presented in Table 8.

Table 8. Outcomes of measurements of hard coal samples from „Piast” Coal Mine. The samples are in the analytical air moisture free state. The measurements conducted by means of a semiconductor detector.

Sample	A _{alab} [%]	A _a [%]	A _{alab} -A _a [%]	S _{tlab} [%]	S _t [%]	S _{tlab} -S _t [%]
piast1	17.81	17.84	-0.03	0.893	0.890	0.003
piast2	11.51	11.54	-0.03	1.004	1.003	0.001
piast3	26.95	26.95	-0.01	0.865	0.876	-0.011
piast4	17.31	17.30	0.01	0.919	0.911	0.008
piast5	18.73	18.72	0.01	0.877	0.879	-0.002
piast6	22.07	22.01	0.06	0.962	0.963	-0.001
piast7	36.03	36.05	-0.02	0.695	0.693	0.002
s.d.= 0.05				s.d.= 0.008		

Obtained results were comparable to gained ones for coal samples from Uzbekistan. However, equation coefficients differed so much than it was not possible to create a common equation for samples from „Piast” Coal Mine and Uzbekistan.

4 Conclusions

The application of X-ray fluorescence enables to obtain better accuracies of measurements than for the methods based on dispersion or absorption, especially in case of small samples graining. However, it requires the application of a detector of significantly better resolving power than scintillation detectors or even proportional counters. In order to minimize the impact of samples heterogeneity, it is beneficial to apply detectors of possibly high active flanks and utilization per some measurements by

different sample orientation in respect of the detector. Because the active area of semiconductor detectors is relatively not large ($30\text{-}100\text{ mm}^2$), the measurement time should be a few times longer than in case of the application of a proportional counter and an radiation source of the same activity.

For the samples of hard coal that differ significantly from the mineralogical contents, even through the application of semiconductor detectors of a large resolving power, it is very difficult to determine a common dependence allowing to define qualitative parameters with a satisfactory accuracy.

References

1. Dziunikowski, B.: Radiometric methods of chemistry analysis. Wydawnictwa Naukowo-Techniczne, Warszawa (1991) (in Polish)
2. Sikora, T., Czerw, B.: Control-measuring devices and control systems for coal preparation plants in the activities of EMAG. Development and current state. Mechanizacja i Automatyzacja Górnictwa 5(412) (2005)

Measurement and Control System of the Plasmatron Plasma Reactor for Recovery of Metals from Printed Circuit Board Waste

Jakub Szałatkiewicz, Roman Szewczyk, Eugeniusz Budny,
Tadeusz Missala, and Wojciech Winiarski

Industrial Research Institute for Automation and Measurements PIAP
{jszalatkiewicz,rszewczyk}@piap.pl

Abstract. The article presents the measurement and control system of plasmatron plasma reactor for utilization and recovery of metals from waste of printed circuit boards built in the Industrial Research Institute for Automation and Measurements in Poland. The plasma reactor is the key component of a test stand developed to conduct research on processing and recovery of metals from used electric and electronic equipment, especially from electronic printed circuit boards. The system allows, recording multiparametric measurements realized in real time and time-correlated, visualization of current values of measurements and states by means of SCADA software. Besides main functions of data acquisition and visualization, the system allows local and remote control of operation of reactor subsystems. Analysis of measurement chain is presented for the key parameter of reactor work, which is temperature measurement inside the reactor. Standard uncertainty of measurement of the temperature is evaluated and in this case equal to 5 °C.

Keywords: recycling, electronic waste utilization, plasma technology, temperature measurements.

1 Introduction

Electronic printed circuit boards occur in most electric and electronic devices, which become hazardous waste after device stops serving its purpose and is disposed of by the user. Their mass production leads to significant use of natural resources in form of non-renewable deposits of metals, including precious metals, which are essential for their production. Metals exploited for production of devices remain in waste in the used electric and electronic equipment, which can be their valuable source. For this reason, during processing of waste of electric and electronic equipment (WEEE) it is possible at the same time to: recover (for further reuse) non-renewable resources like metals, and neutralize negative influence of this hazardous waste on environment.

It is estimated that in the 27 European Union countries, mass of produced WEEE was at the level of 8.3–9.1 mln Mg (tons) in 2005, out of which amount only 25% was gathered and processed, whereas the remaining 75% was not registered [1, 2]. The

growing mass of gathered waste and steady increase of supply of new devices to the market leads to necessity of development of new technologies and methods of processing of WEEE, aimed at recovery of resources they contain and at their neutralization. This need is especially important in Poland, because of existence of organizational and technological gap between our country and Western European countries, and also due to law regulations on the EU scale concerning WEEE processing (amendment of 2002/96/EC WEEE directive from 2012 [3]).

Traditional simple technologies of WEEE processing allow effective processing of majority of mass of WEEE devices. However, after such processes as manual dismantling and mechanical processing by milling, certain groups of waste remain and their handling requires specialized technologies and processes. Those groups include electronic printed circuit boards and the subscreen fraction from waste milling and screen analysis, with large content of milled printed circuit boards. At present, in Europe, there exist a few plants processing waste of printed circuit boards, where this waste is treated by means of pyrometallurgical technologies. In Poland installations like that are absent. Waste of electronic printed circuit boards are only part of charge in those technologies because of their composition and its influence on the conducted process. Lack of comprehensive processes that allow complete neutralization and recovery of metals from waste of electronic printed circuit boards was motivation to taking up research and development project, founded by NCBiR, aimed at investigations of applicability of plasma process for their processing.

2 Plasma-Based Technology of Utilization of Electronic Printed Circuit Boards

In the Industrial Research Institute for Automation and Measurements a test stand was developed for plasma-based processing of waste of electronic printed circuit boards from used electric and electronic equipment for the purpose of recovery of metals contained in this waste. A general view of the stand is presented in Fig.1. The key component of the stand is the plamsatron plasma reactor, in which three sources of plasma in form of plasmatrons located at every 120 degrees around reactor circumference are applied. The test stand is equipped with control and measurement apparatus and auxiliary devices necessary for support of reactor work, data acquisition and carrying out research.

The developed high-temperature plasma reactor is key component of process line for realization of research on high-temperature technology of processing and neutralization of selected electric and electronic waste (WEEE), allowing recovery metals including precious ones [4]. A block diagram of the process line is presented in Fig. 2.

The developed test stand allows realization of wide range of research on technology of processing of selected electric and electronic waste for the purpose of recovery of metals and simultaneous neutralization of this waste. The processing relies on technological steps presented in diagram shown in Fig. 2. Segregated waste is fed by means of automatic feeder of waste portions into the plasma reactor, where under action of plasma streams they undergo incineration and melting. Organic substances

contained in the waste are incinerated at plasma temperature, and products of incineration are then cooled to recover heat [5]. Metals and non-metals in liquid form pour into mould, where they solidify and can be reused. Fig. 3 presents cross-section of reactor's chamber with shown: route of waste in the system – orange arrows, plasma stream – black arrow, and direction of stream of exhaust gases – red arrow.



Fig. 1. General view of the test stand. 1. Plasma reactor, 2. Plasmatron, 3. Molten product collector, 4. Chimney with exhaust gases fan, 5. Belt conveyor of waste portions with linear storage, 6. Plasmatrons power supply module, 7. PLC cabinet – control and data acquisition, 8. Feeder of waste portions to the reactor

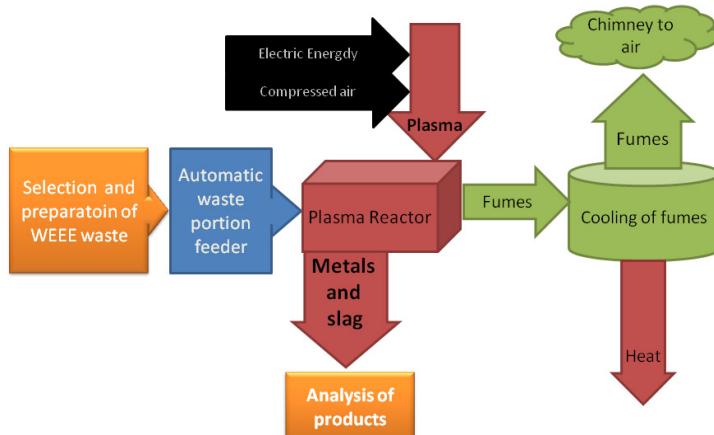


Fig. 2. Block diagram of process line for research on high-temperature technology of processing and utilization of selected electric and electronic waste (WEEE) by means of plasma reactor, allowing recovery of precious metals and rare earth metals

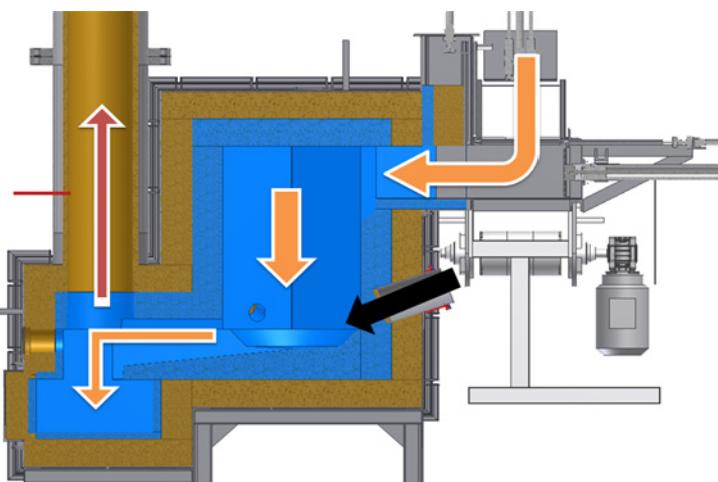


Fig. 3. Cross-section of plasma reactor with shown route of flow of waste and molten products (orange arrows, from top to down), exhaust gases (red arrow) and plasma (black arrow)

2.1 Plasma Reactor

Plasma reactor is made of three layers: refractory concrete, thermal insulation and metal mantle. Reactor chamber has hexagonal shape. This design allows keeping temperature of the order of 1500–1600 °C in its volume. In turn, within zone of direct action of plasmatrons, temperature outnumbers many times the temperature measured in reactor volume, but because of difficulties of measurement of temperatures above 2000 °C its measurement is not applied at the moment. Sources of heat in the reactor are three electric arc plasmatrons of 20 kW power each. Plasmatrons generate plasma stream made from compressed air, which flows into the bottom part of the reactor chamber where the waste is being incinerated and melted.

3 Measurement and Control System of Plasma Reactor

The plasma reactor for utilization of waste of printed circuit boards is equipped with measurement and control system that allows real-time measurement, processing, recording and visualization of multiple physical quantities. This system consists of sensors of various physical quantities (temperature, pressure, flow, electric current and voltage), analog-to-digital converters of signals, a PLC controller and SCADA (supervisory control and data acquisition) software, where the process is visualized.

Main function of the system is acquisition and visualization of measurement data, and control of reactor and auxiliary devices, to allow stable operation of the whole system and safety of both system users and system components.

The developed measurement and control system, apart from acquisition function has also operational function. It was realized based on SCADA software and PLC

controller, which signals control particular operational units such as: actuators of automated system of loading of waste portions, regulation of amount of cooling water, fan frequency inverter, switching on/off particular components like plasmatrons, pumps, etc.

The heart of the system is Mitsubishi Melsec-Q PLC controller equipped with digital 24 V input (16 pcs.) and output (32 pcs.) modules, analog 4–20 mA input (24 pcs.) and output (6 pcs.) modules, and thermocouple input modules (24 pcs.). The PLC controller integrates all measurement signals and transmits them via Ethernet to PC computer, where SCADA software is installed. SCADA performs three principal functions. The first function is communication between user and the measurement and control system (HMI) – this allows user to issue commands to be executed by the PLC controller and read current values of measurements and parameters gathered by PLC (visualization). Figure 4 presents the control and visualization window of parameters of the test stand. Moreover, the software allows recording the time histories of measured values and desired (imposed) output values with common time basis. The last function of SCADA is conducting current calculations based on gathered data to present aggregated results, e.g. showing amount of heat flowing away with exhaust gases or realization of operation programs as in case of automation of loading of waste portions using 4 pneumatic actuators.

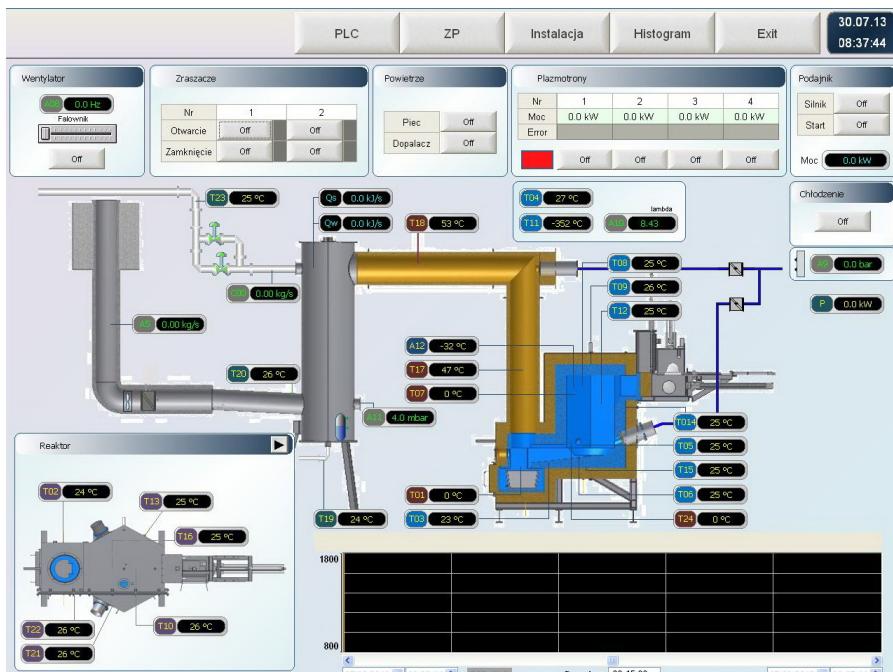


Fig. 4. View of screen of control and visualization of data of control computer (SCADA)

3.1 Analysis of Measurement Chain for Measurement of Temperature Inside Reactor Chamber

Temperature inside plasma reactor chamber is measured using type B (Pt Rh) thermocouple in ceramic sheath. The tip (hot-junction) of thermocouple is located inside the reactor chamber at the distance of 30 mm from the wall, at the height of 350 mm over reactor bottom. The sensor is on symmetry axis of the wall segment – its location is shown in Fig. 5.

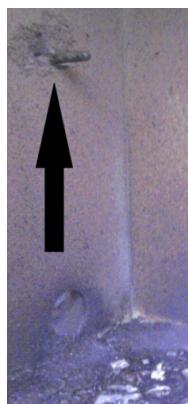


Fig. 5. Location of the point of temperature measurement inside plasma reactor

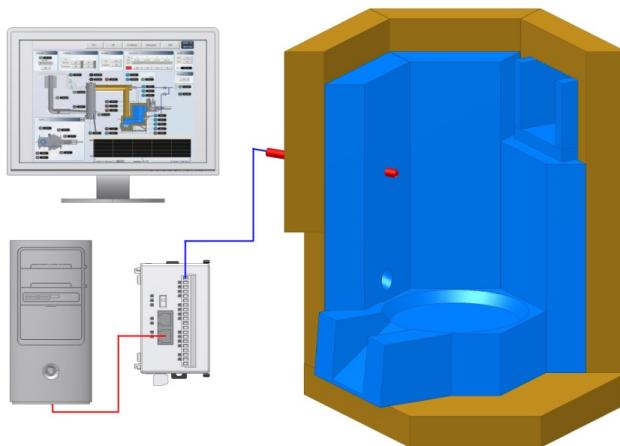


Fig. 6. Schematic of connection of measurement chain for the type B thermoelement

Voltage signal from type B thermocouple is transmitted with 10 m long copper cable to control cabinet. The copper wires from thermoelement are connected to thermocouple inputs of PLC controller by means of additional connector. The input module has its own temperature measurement inside the control cabinet for compensation of influence of cold-junction temperature. The module of thermocouple inputs

converts analog voltage signal obtained from type B thermoelement into digital signal and transmits it together with other values to SCADA, where it is visualized and recorded. Schematic of connection of example measurement chain for temperature measurement with type B thermoelement is shown in Fig. 6.

Standard uncertainty of temperature measurement in the system is evaluated using type B method [6, 10].

In the evaluation of standard uncertainty of temperature measurement were included:

- Uncertainty of measurement with thermocouple, given by the manufacturer: $e_T = 4 \text{ }^\circ\text{C}$, with assumption of rectangular distribution, measurement standard uncertainty equals $\frac{e_T}{\sqrt{3}} = 2.3 \text{ }^\circ\text{C}$.
- Measurement uncertainty associated with system of processing of measurement signal given by the manufacturer: $e_p = 3 \text{ }^\circ\text{C}$ with assumption of rectangular distribution, measurement standard uncertainty is equal to $\frac{e_p}{\sqrt{3}} = 1.7 \text{ }^\circ\text{C}$.
- Estimated measurement uncertainty associated with temperature gradients in the chamber: $e_g = 10 \text{ }^\circ\text{C}$ with a more optimistic assumption of triangular distribution, measurement standard uncertainty equals $\frac{e_g}{\sqrt{6}} = 4.1 \text{ }^\circ\text{C}$.

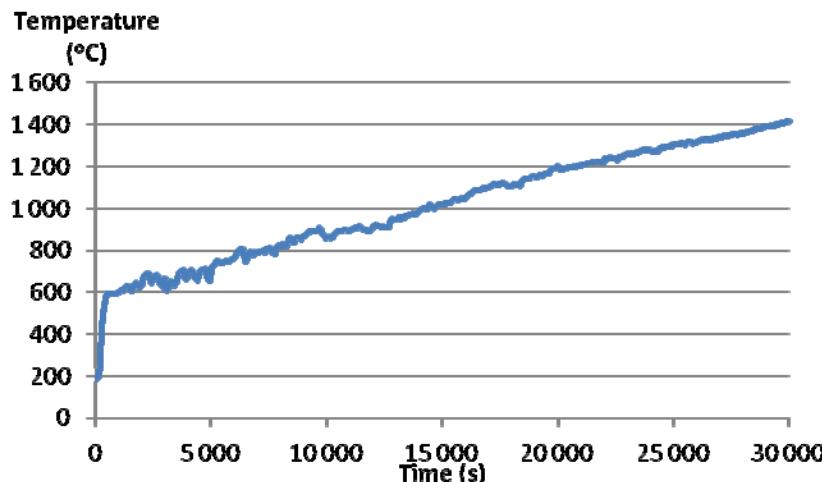


Fig. 7. Result of measurement of temperature in reactor using type B thermocouple during its heating up procedure

For the given assumptions combined standard uncertainty of result of temperature measurement e_s is equal to:

$$e_s = \sqrt{\left(\frac{e_T}{\sqrt{3}}\right)^2 + \left(\frac{e_p}{\sqrt{3}}\right)^2 + \left(\frac{e_g}{\sqrt{6}}\right)^2} = 5 \text{ }^\circ\text{C} \quad (1)$$

In order to simplify the analysis, coverage factor was assumed $k = 2$ [8]. Finally, expanded uncertainty of temperature measurement e reads:

$$e = k \cdot e_s = 10 \text{ } ^\circ\text{C} \quad (2)$$

A recorded time history of measurement of real temperature inside reactor is presented on Fig. 7.

4 Conclusion

The analysis of example measurement chain of measurement of temperature inside plasma reactor chamber, presented in the paper, demonstrated validity of reactor design, and the estimated expanded measurement uncertainty $e=10 \text{ } ^\circ\text{C}$ is within allowable range for the assumed applications. Temperature measurement realized using type B thermocouple is in most cases within the range $1200\text{--}1700 \text{ } ^\circ\text{C}$, in which measurement uncertainty of $10 \text{ } ^\circ\text{C}$ does not affect the process in any significant way.

On the graph of time history of temperature in the reactor (Fig. 7), measured using the analyzed measurement chain, noticeable changes of temperature are visible. The origin of those changes is not generated by the measurement chain, but is consequence of changes of temperature in the reactor during its heating up and stabilization of operating parameters. Reactor is heated with three plasmatrons that produce plasma streams, which are not as stable heat source as e.g. resistance heaters. Additional disturbances of temperature in reactor are also associated with other factors like opening of an inspection hole. All those factors were reproduced and related disturbances recorded, which confirms effectiveness and accuracy of the applied solution.

The developed measurement and control system of plasma reactor for processing of electronic printed circuit boards thanks to its functions like parallel acquisition and recording of multiple measurement quantities opens possibility of more comprehensive analysis of the obtained data and also increases safety and convenience of using the test stand. Acquisition of such wide range data allows application of multi-parametric optimization of the process for optimization of electrical energy use, or throughput etc. The implemented solutions based on a PLC controller and dedicated measurement modules (e.g. for thermocouples) create system of large reconfiguration possibilities and open structure that allows expansion with additional equipment for further enhancements of its functionality.

Acknowledgements. The work was financed from the means of the research and development project no. N R03 0083 10, founded by the Ministry of Science and Higher Education of Poland and realized through The National Centre for Research and Development (NCBiR).

References

1. Huisman, J.: 2008 Review of Directive 2002/96 on WEEE, Final Report, United Nations University (2007)
2. Cobbing, M.: Toxic Tech: Not In Our Backyard (2008), <http://Greenpeace.org>

3. Directive 2002/96/EC of the European Parliament and of the Council of 27 January 2003 on waste electrical and electronic equipment (WEEE) (2003),
http://ec.europa.eu/environment/waste/weee/index_en.htm,
<http://www.mg.gov.pl/English/ECONOMY/Industrial+Policy/ROHS+Directive>, <http://www.zipsee.pl/aktualnosci/112/zmiana-dyrektyny-weee.html>
4. Szałatkiewicz, J., Szewczyk, R., Budny, E., Missala, T., Winiarski, W.: Construction aspects of plasma based technology for waste of electrical and electronic equipment (WEEE) management in urban areas. Procedia Engineering 57, 1100–1108 (2013)
5. Szałatkiewicz, J.: Energy recovery from waste of printed circuit boards in plasmatron plasma reactor. Polish Journal of Environmental Studies 23(1) (2014)
6. Fotowicz, P.: Obliczanie niepewności rozszerzonej metodą analityczną opartą na splocie rozkładów wielkości wejściowych. Pomiary Automatyka Robotyka 1, 5 (2005) (in Polish)
7. Joint Committee for Guides in Metrology, Evaluation of measurement data — Guide to the expression of uncertainty in measurement. JCGM 100 (2008)
8. Arendarski, J.: Niepewność pomiarów, Oficyna Wydawnicza Politechniki Warszawskiej (2006) (in Polish)
9. Dorozhovets, M., Propozycje, Z.L.: Propozycje rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (1) Uwzględnianie wpływu auto-korelacji i nieadekwatności rozkładu wyników obserwacji w niepewności typu A. Pomiary Automatyka Robotyka 11(1), 6–15 (2007) (in Polish)
10. Dorozhovets, M., Propozycje, Z.L.: rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (2) Uściślenie metod obliczeń niepewności typu B. Pomiary Automatyka Robotyka 11(2), 45–52 (2007) (in Polish)

Influence of Operating Conditions on Functional Properties of High Resolution Analog to Digital Converter

Krzysztof Trzcinka¹, Roman Szewczyk¹, and Oskar Gińko²

¹ Industrial Research Institute of Automation and Measurements PIAP, Warsaw, Poland
{ktrzcinka, rszczek}@piap.pl

² Institute of Metrology and Biomedical Engineering, Warsaw University of Technology
oskarginko@gmail.com

Abstract. This paper presents the results disturbing of impact of the factors on the properties of the high resolution analog-to-digital converter. The researches were held for the following operating conditions ADC converter: with battery supply or with switching power supply, for different ambient temperatures, with the optocouplers through the signal lines, with using two different voltage reference source, and with using shield. Of the results of measurements were calculated standard deviation in the number of points. Standard deviation were used to quantify the impact on the accuracy of the measurements AD converter.

Keywords: analog to digital converter, high resolution, weighting, operating conditions.

1 Introduction

Analog-to-digital converters are very important components of electronic measurement systems, since they account for link between digital and analog information [1]. In view of a possible great resolutions of those devices it is possible to obtain increasingly high accuracy of measurement systems. One of the branches where super precision is requisite is weighing, which require usage of analog-to-digital converters with higher resolution. Scales with high resolution transducer facilitate very accurate assessment of a mass and ensures repeatability of a mass measurement results (substantial mainly in a pharmaceutical industry during a preparation of drugs according to precise formulas). Due to the realization of research project executed in association with a polish scales producer, the Radwag company, research work upon developing a possibly highest resolution weighting machine was undertaken. Elimination of all possible source of error and distraction is essential to achieve that goal. Every part of the design, from mechanical part to every electronic circuit, should be analyzed. One of the important elements of the scales is a system of processing strength signal to digital signal with AD converter. Hence, it was decided to investigate how operating conditions affect properties of the high resolution ADC converters. The one that was chosen to this research was an analog-to-digital converter by Analog Devices, model AD7191 with 24 bits resolution that ensures $2^{24} = 16\,777\,216$ points when fully utilized. This device is dedicated to

weighting machines and offers low noise and sigma-delta processing. Still, due to occurrence of noise, the resolution of the converters is narrowed and in this case, for the AD converter's configuration with gain equal to one, a typical resolution can shrink to 21.5 bits, which gives 2 965 820 points. The example given shows that operations leading to reduction of noise level to minimum have to be undertaken, so that the resolution of the AD converter could be fully utilized. Many different factors can influence worsening of the AD converter's resolution [2-4]. Some of them concern conditions in which the device work and they can be singled out [5]:

- occurrence of interferences arisen in a power-supply source transferred by conduction throughout power-supply lines,
- occurrence of conducted interference through the signal lines,
- changes in ambient temperature,
- electromagnetic radiation disturbances.

To enable researches on the influence of the factors listed above on the analog-to-digital converter's system, an electronic system with AD7191 converter was built, based on the analysis of those problems.

2 Electronic System with Analog-to-Digital Converter AD7191

In the Fig. 1 block process flow sheet of electronic system with 24 bits analog-to-digital converter AD7191 was presented, electric scheme was presented in the Fig. 2.

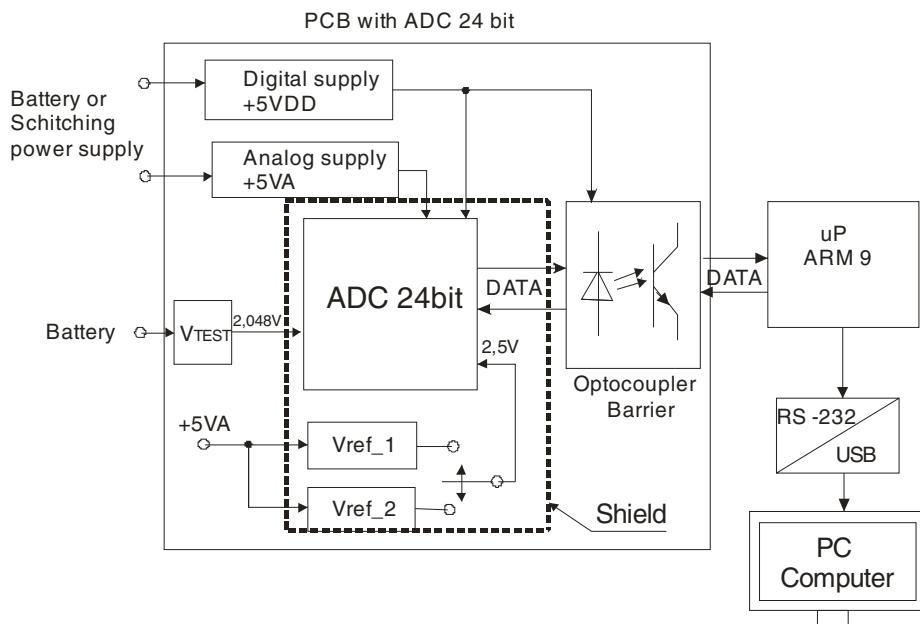


Fig. 1. Block process flow sheet of electronic system with 24 bits analog-to-digital converter AD7191

On analog input of converter AD7191 was provided voltage V_{TEST} (2.048 V) which was measured. The reference voltage source of V_{TEST} was powered by battery.

To supply the ADC converter were used two other voltage sources 5 V: separate from power supply of digital element (VDD) and from analog element (VA). On the ADC system converter was used two variable jumpers reference voltage source.

Integrated circuits AD converter and a voltage reference source were placed in cooper shield on PCB. Data lines from the converter connected to the microprocessor through the optocoupler barrier assuming galvanic insulation. The output data rate was setting to 10 Hz. Data from the AD converter was recorded (about 6 thousands samples) on computer's disc and then it was submitted to analysis and filtration when using MALTAB software.

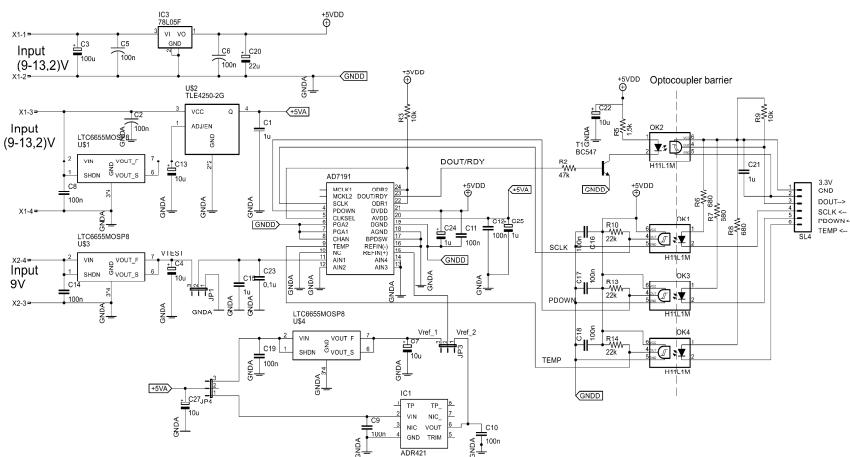


Fig. 2. Electronic scheme with analog-to-digital converter AD7191

Parameter of analog-to-digital converter AD 7191, 24 bits sigma-delta [8]

Gain: 1,

Output data rate: 10 Hz,

Typical resolution (bits): 24 (21.5),

Analog input: 0–2.5 V,

Voltage noise: 15 nV (RMS noise @ 10 Hz, gain 128),

Offset drift: 5 nV/ $^{\circ}\text{C}$,

Gain drift: 1 ppm/ $^{\circ}\text{C}$

Temperature range: -40°C to $+105^{\circ}\text{C}$,

24-lead TSSOP package.

Parameter of reference voltage sources LTC6655 – Vref_1 [6]

Input Voltage: 3–13.2 V

Output Voltage, V_{OUT} : 2.5 V

Voltage noise (0.1 Hz – 10 Hz): $625\text{nV}_{\text{p-p}}$

Temperature coefficient: 2 ppm/ $^{\circ}\text{C}$

Temperature range: -40°C to $+125^{\circ}\text{C}$

Precision: 0.025%

Parameter of reference voltage sources ADR421 – Vref_2 [7]

Input Voltage: 4.5 – 18 V

Output Voltage, V_{out}: 2.5 VVoltage noise (0.1 Hz – 10 Hz): 1.75 μ Vp-p

Temperature coefficient: 3 ppm/°C

Temperature range: -40°C to +125°C

Precision: 0.04%

Parameter of researched voltage sources LTC6655 - VTEST [6]

Input Voltage: 3–13.2 V

Output Voltage, V_{out}: 2.048 VVoltage noise (0.1 Hz – 10 Hz): 512nV_{p-p}

Temperature coefficient: 2 ppm/°C

Temperature range: -40°C to +125°C

Precision: 0.025%

3 Research of Efficiency High Resolution Analog-to-Digital Converter

The recorded of measurements was converted to its value in the decimal system so that the result was in code which means the number of plots. The results were subjected to filtration in MATLAB software in usage third order digital low pass filter Butterworth with cut-off frequency 25 Hz. From the obtained (filtered) signal values was determined by the standard deviation showing the scatter of the results from mean value. The calculated standard deviation is used qualify the impact of measurements made by AD converter.

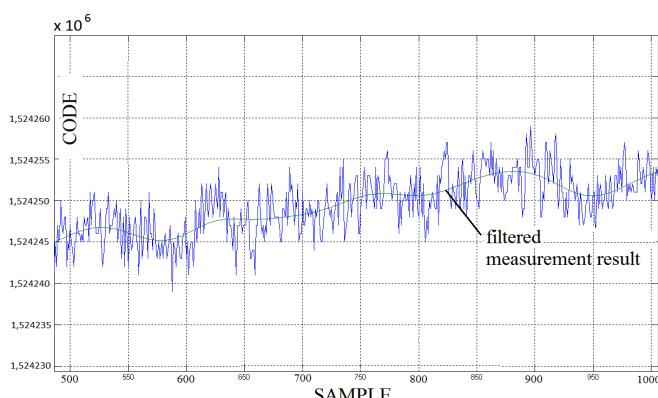


Fig. 3. Filtration of measurements results

Histogram of measurement scatter from filtrated men value was used to graphical analysis (see figure 4.).

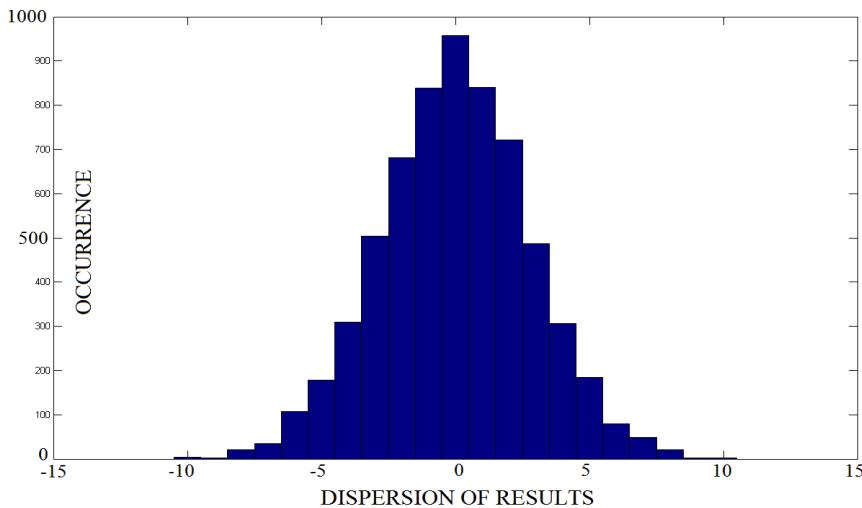


Fig. 4. Histogram of value dispersion measurements from mean value (standard deviation = 4.25; 5500 Samples)

3.1 Impact of Disturbances from the Power Supply

At the time of the measurements, circuits, as in Fig. 2 was fed in the first case of the 9 V battery, in the second case the same circuit was fed from switching power supply connected to mains. Switching power supply was applied to feed the system with AD converter and it was typical consumer with CE signature with output voltage of 13.2 V and ripples 40 mV_{p-p} with frequency about 50 kHz (used by the manufacturer to feed the weight) and maximum output current 2.1 A. To obtain repeatability results, every time the switching power supply was connected to the mains by power line filter (EMI Filter 1-phase type 2010). Every time during the measurements environmental condition were recorded in laboratory.

3.2 Impact of Ambient Temperature

The system of AD converter was subjected to the influence of ambient temperature in the range from 0°C to +50°C, it is the temperature range for measurement devices used indoors. Converter AD with reference voltage source was subjected thermal effects (just as it is in reality), source VTEST was thermally isolated. On

the table 1 shows comparisons between the battery and the switching power supply for two different reference voltage sources using galvanic isolation of the signal line.

Table 1. The standard deviation of the AD converter measurement results depending on the temperature

Temp [°C]	Battery powered		Switching power supply powered	
	σ of Vref_1	σ of Vref_2	σ of Vref_1	σ of Vref_2
0	2.98	3.38	3.13	3.28
10	2.78	3.33	2.82	3.52
20	2.71	3.95	2.93	5.53
30	2.76	5.42	2.78	5.86
40	2.93	7.01	2.97	8.8
50	2.98	11.53	3.06	12.1

where:

σ of Vref_1 – standard deviation for analog-to-digital converter with LTC6655 reference voltage source,

σ of Vref_2 – standard deviation for analog-to-digital converter with ADR421 reference voltage source.

These measurements (see Table 1) showed a significant effect of temperature on the growth of measurement errors of the AD converter. The largest measurements errors demonstrated at the temperature +50 °C during operation of AD converter with reference voltage source ADR 421. In the entire temperature range investigated results are better (they have lower standard deviation) for operation of AD converter with source LTC6655. For the worst case, the operation at +50 °C errors for LCT6655 reference voltage source are 4 times lower than with operation with ADR421 source.

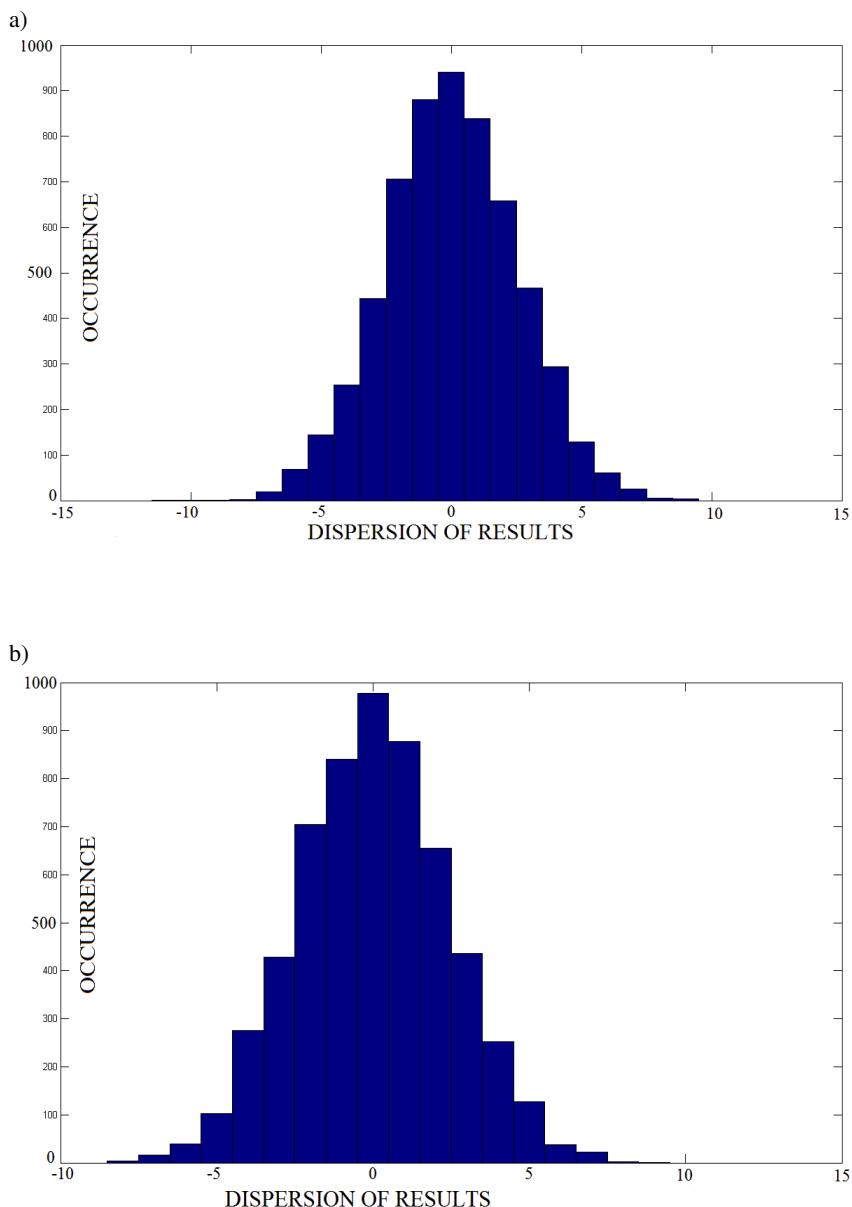


Fig. 5. Comparison of histograms of value dispersion measurements for mean value at AD converter work with source of LT6655 reference voltage (ADC battery powered): a) at 0°C temperature (standard deviation = 2.98; 5900 Samples), b) at +50°C temperature (standard deviation = 2.98; 5800 Samples)

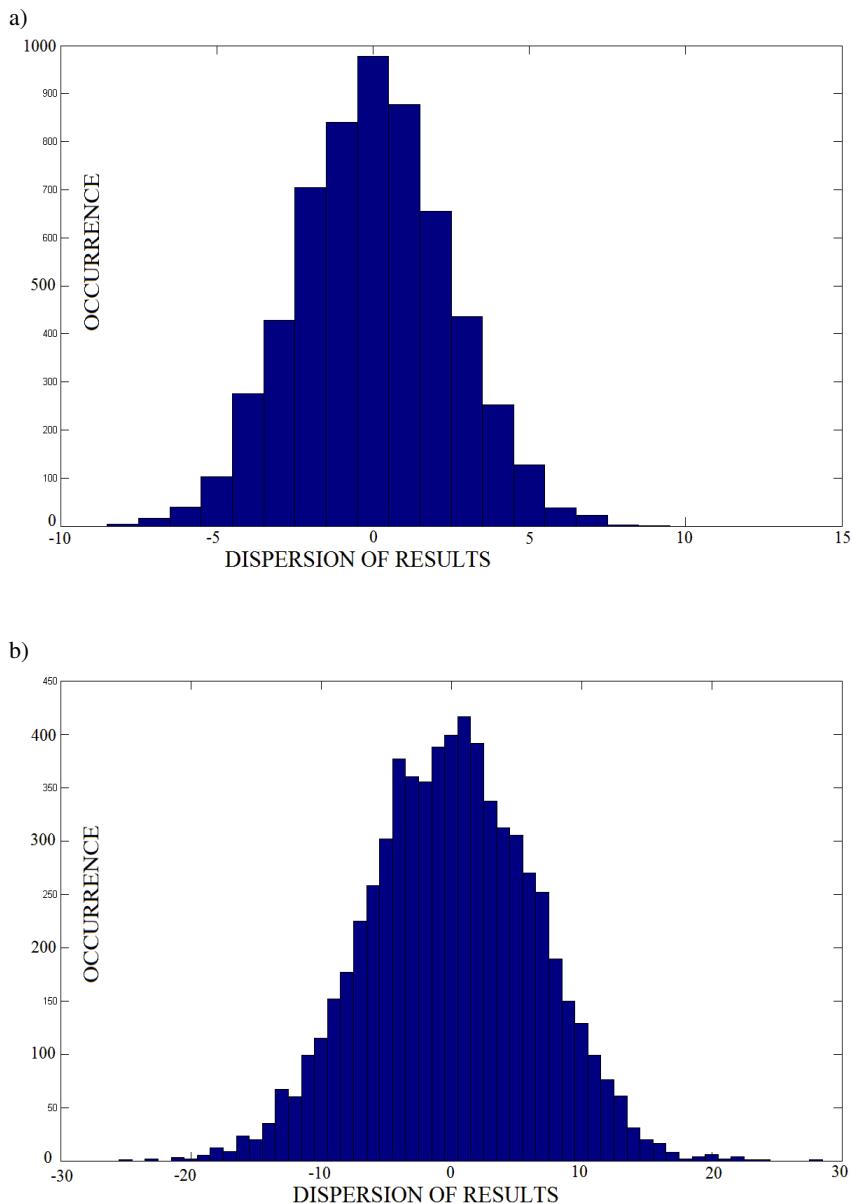


Fig. 6. Comparison of histograms of value dispersion measurements for mean value at AD converter work at +50°C temperature (ADC battery powered): a) with source of LT6655 reference voltage (standard deviation = 2.98; 5800 Samples), b) with source of ADR421 reference voltage (standard deviation = 11.53; 6300 Samples)

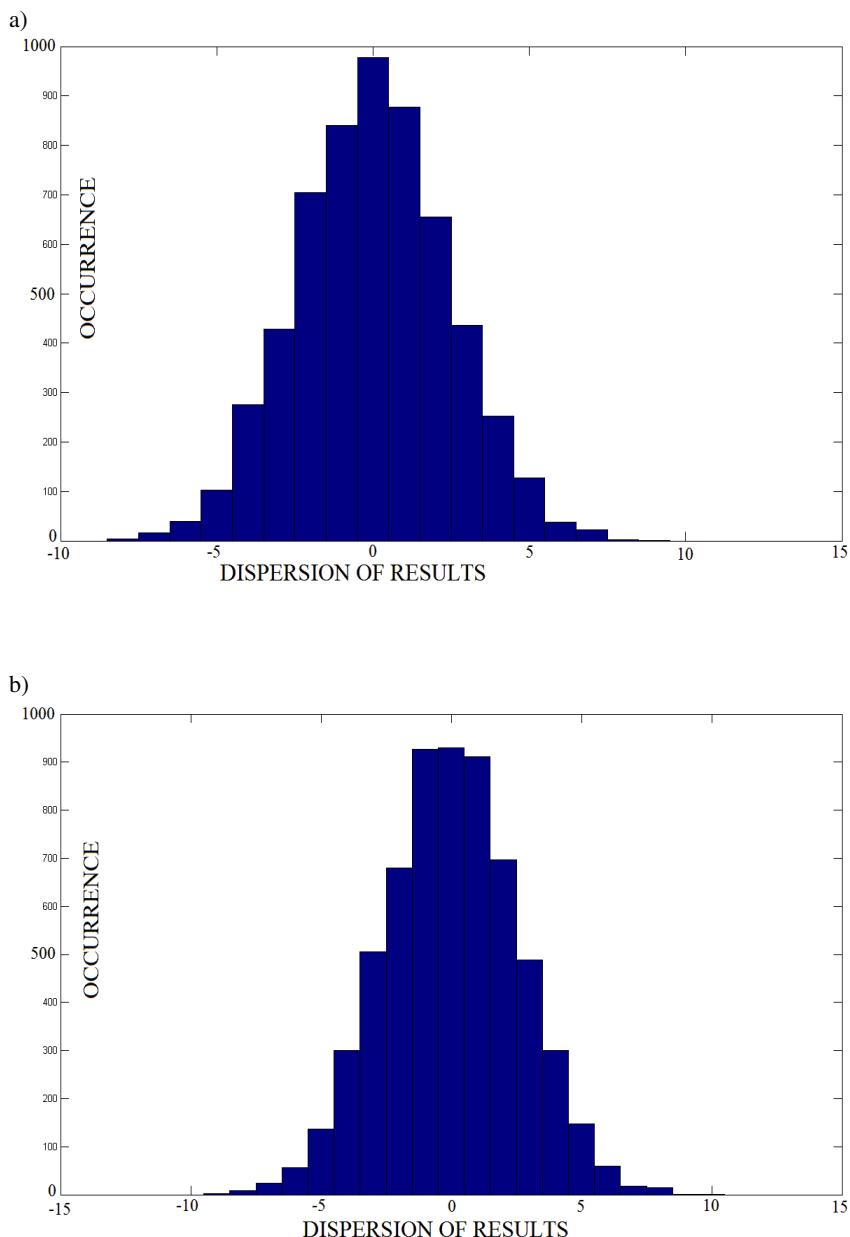


Fig. 7. Comparison of histograms of value dispersion measurements for mean value at AD converter work with source of LT6655 reference voltage at +50°C temperature: a) battery powered (standard deviation = 2.98; 5800 Samples), b) switched-mode power supply powered (standard deviation = 3.06; 6000 Samples)

3.3 Impact of Interference of the Signal Lines

Connection AD7191 converter with microprocessor system were performed with usage of optocouplers barriers make galvanic isolation for data signals (limiting conducted interference of the signal lines). During the research made a series of measurements without galvanic isolation and a series of measurements with galvanic isolation (see table 2), with comparisons between the battery and the switched-mode power supply for two different reference voltage sources.

Table 2. Standard deviation of the AD converter measurement results depending on the application of galvanic isolation of the signal line

Configuration	Battery powered		Switching power supply powered	
	σ of Vref_1	σ of Vref_2	σ of Vref_1	σ of Vref_2
Galvanic isolation	2.74	4.14	2.82	4.42
Absence of galvanic isolation	2.77	4.34	2.87	4.51

where: σ of Vref_1 – standard deviation for analog-to-digital converter with LTC6655 reference voltage source,

σ of Vref_2 – standard deviation for analog-to-digital converter with ADR421 reference voltage source.

According to the measurements presented in the table 2, an application of the galvanic isolation on the signal lines allows lower standard deviation from mean value of the measurements value powered both (with battery and switching power supply). When AD converter powered from battery, difference in standard deviation for galvanic isolation and its absence are inconsiderable, e.g. at work of the AD converter with LTC6655 source, standard deviation with galvanic isolation amounts to 2.74, and 2.77 without the isolation. When the AD converter powered from switching power supply, standard deviation with galvanic isolation amounts to 2.82 in case without galvanic isolation 2.87.

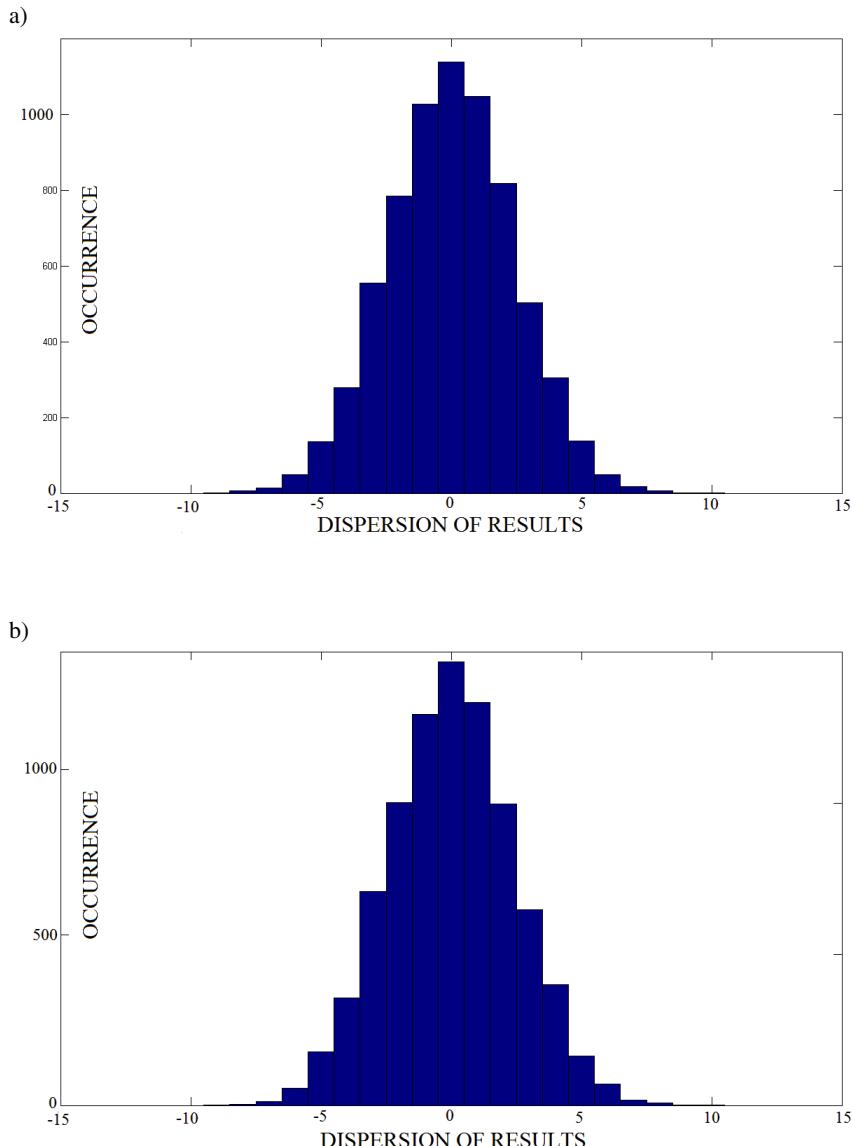


Fig. 8. Comparison of histograms of value dispersion measurements for mean value at AD converter work with reference voltage source of LT6655 supplied with battery: a) with the usage of signals line's galvanic isolation (standard deviation = 2.74; 7000 Samples), b) without signals line's galvanic isolation (standard deviation = 2.77; 7200 Samples)

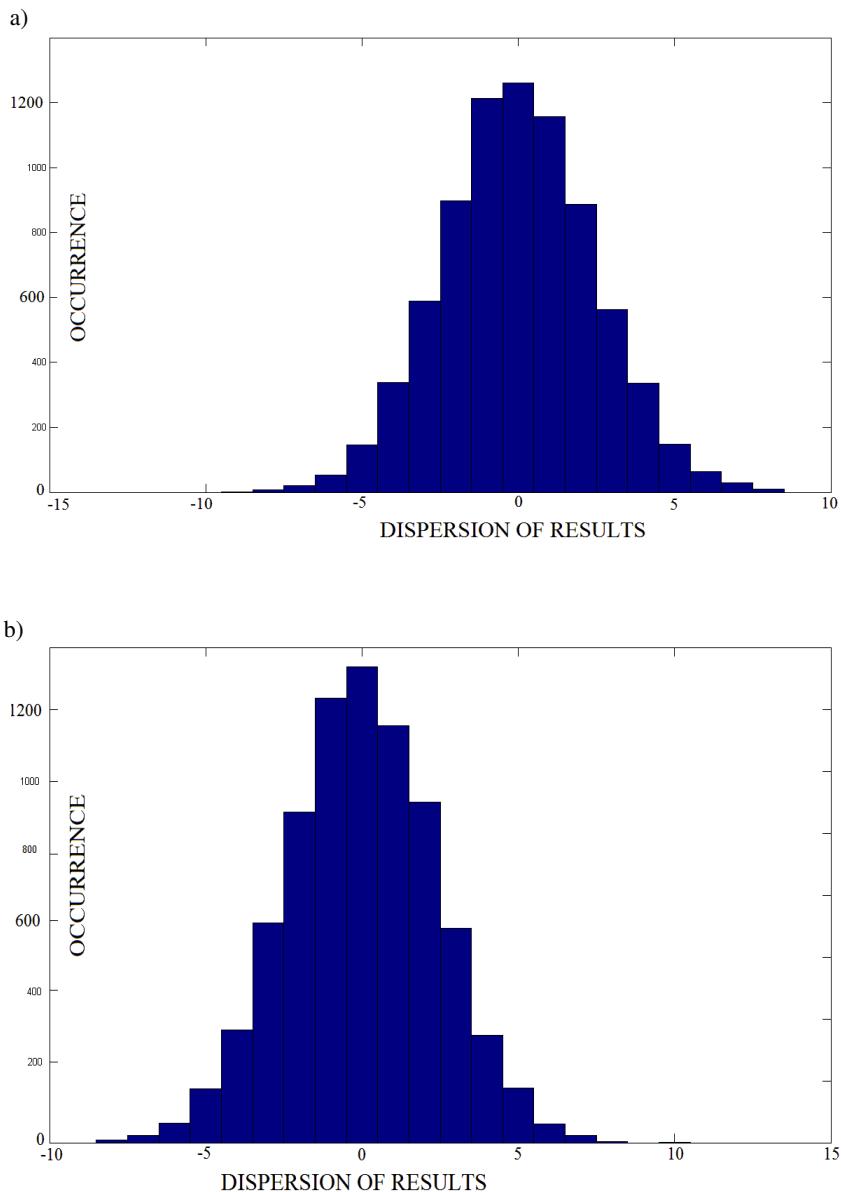


Fig. 9. Comparison of histograms of value dispersion measurements for mean value at AD converter work with reference voltage source of LT6655 supplied with switching power supply: a) with the usage of signals line's galvanic isolation (standard deviation = 2.82; 7000 Samples), b) without signals line's galvanic isolation (standard deviation = 2.87; 7500 Samples)

3.4 Effect of the Quality of the Reference Voltage Source

The use of two reference voltage sources switchable by jumper to the AD7191 allowed to research the impact of the reference voltage source quality on measurements. The source of reference voltage used was one recommended by a manufacturer of the AD converter, i.e. ADR421 (noise level of $1,75\mu\text{V}$, output voltage with an accuracy of 0,04%, temperature coefficient: 3 ppm/ $^{\circ}\text{C}$) and another reference source LTC6655 with better parameters (noise level of 625 nV, the output voltage with an accuracy of 0.025%, temperature coefficient: 2 ppm/ $^{\circ}\text{C}$).

Based on the measurements (table 1, table 2 and table 3), a significant effect of the reference voltage source quality on measurement results made with high resolution AD converter was concluded a much larger scatter of measurements with ADR421 source than with LTC6655 source. For instance, working AD converter at $+50^{\circ}\text{C}$ temperature (battery powered – table 1) has standard deviation on the level of 2.98 for LTC6655 source, and 11.53 for ADR421, which is almost four times greater. For the measurement results shown in table 2, in each case of work with ADR421 source, standard deviation is in the range 4.14–4.51, and is greater than the work with the LTC 6655 where the standard deviation is in the range 2.74–2.87. Parameters of the reference voltage source such as the temperature coefficient, noise level and the accuracy of the reference voltage source should have the lowest possible values in order to reduce measurement errors of the AD converter.

3.5 Effect of Shielding

AD converter and the reference voltage sources were placed in a copper shield on PCB. To assess the impact of shielding, a series of measurements without the shield on the circuit board were performed. Due to the low efficiency of copper shields, a series of measurements were made when the entire PCB with the AD converter were placed inside the housing made of a soft magnetic material. Shielding efficiency results are shown in table 3 for configuration with a battery and a switching power supply for two different reference voltage sources with using galvanic isolation of the data line.

According to the test results shown in table 3, it is possible to obtain better performance of measurements with the shielding, however the differences are very small, e.g. while using the battery power (when working with LTC6655) without the shielding, the standard deviation equals 2.76 while with the use of a copper shield – 2.73. The result is about 0.03 lower, and in the additional shield of soft magnetic steel deviation was at the level of 2.72. It should be noted that significant improvement of the shielding was not achieved (standard deviation value lowering), so that the system with an AD worked in the configuration in which low standard deviation values has been achieved already. The second reason was the lack of full control over the surrounding electromagnetic environment. To learn more about the impact of shielding on measurements results, the process should be performed in a controlled electromagnetic environment with a known electromagnetic radiation disturbances.

Table 3. The standard deviation of the AD converter measurement results depending on the applied shielding

Configuration	Battery powered		Switching power supply powered	
	σ of Vref_1	σ of Vref_2	σ of Vref_1	σ of Vref_2
Without shield	2.76	4.48	2.89	4.68
Copper shield	2.73	4.42	2.82	4.57
Additional shield of soft magnetic steel	2.72	–	2.82	–

where:

- σ of Vref_1 – standard deviation for AD converter with LTC6655 reference voltage source,
 σ of Vref_2 – standard deviation for analog-to-digital converter with ADR421 reference voltage source.

4 Conclusions

In the present study only a portion of the factors affecting the work of the high resolution AD converter was examined. Nevertheless, on the basis of the performed researches interesting practical conclusions were obtained. It has been shown that the most significant effect among studied factors on the standard deviation of the AD converter measurement results is ambient temperature (that influences the transducer and the reference voltage source). The main source of errors resulting from the influence of ambient temperature has been the lack sufficient thermal stability of ADR421 reference voltage (recommended by the converter manufacturer) at ambient temperature of +50°C, where the standard deviation of the measurements was 11.53 while at work at the same temperature with LTC6655 source standard deviation of 2.98 (with battery-powered AD converter).

Subsequently the most significant factor affecting the accuracy of measurements in systems with the AD converter is a noise level of reference voltage sources. Using ADR421 reference voltage source of 1.75 μ V noise level, it introduced a standard deviation of 4.14–4.51 (table 2), while using the LTC6655 source of 625 nV noise level standard deviation of 2.74–2.87 was received (table 2). This is another important parameter for which special attention should be paid in the selection of a reference voltage source.

Theoretically a little effect of the power source on the quality of the AD converter measurements was shown, and the case of powering the ADC from the battery and the switching-mode power supply was compared. Similar conclusions were observed while using a galvanic separation of signals lines as well as using a shield. These

three factors, although no significant impact on the quality of their measurements was proved, are very important in the real working conditions of equipment utilizing high resolution ADC. With the usage of the switching power supply as the powering source, remember that this is a direct way of coupling with the interference existing in the supply network, which in the unfavorable cases can not only significantly degrade the accuracy of measurements, but even interfere with proper operation of the measuring device. Similarly, based on the comparison of results when working with electrical separation and without separation of line data signals, influence of disturbances from the microprocessor is merely noticeable (see table 2), but the existence of opto-coupler barriers is an effective blockade of disturbances coming from the powering lines that can couple with the data lines and interfere with the work of the conductor. When comparing the results obtained from the shielded AD converter with the results obtained without shielding the differences were slight (see table 3). Yet, shielding of the AD is important and it is recommended to be used specifically for measuring devices with high resolution AD converters, because of the existence of electromagnetic radiation disturbances of unknown levels in the surrounding environment. In the worst case electromagnetic disturbances may even interfere with proper operation of the measuring equipment.

To sum up, the system with high resolution AD converter should include selected AD converter and the selected reference voltage source. The elements of the smallest possible temperature coefficients and the least noise levels should be selected. By following the converters manufacturer's instructions, pay attention to the recommended source reference voltage in an application note and carry out market research to find the equivalent of better performance. Additional treatments like the use of battery power, galvanic separation of signals of data lines and shielding PCB with the AD converter provides a stable and reliable operation of the AD converter with the possibility to achieve the highest accuracy measurements.

Acknowledgment. This work was partially supported by The National Centre of Research and Development (Poland).

References

1. Kulka, Z., Libura, A., Nadachowski, M.: Przetworniki analogowo-cyfrowe. WKŁ, Warszawa (1987)
2. Korytkowski, J.: The present monolithic sigma-delta converters for root-mean-square voltage measurement and their comparison to classic monolithic integrated RMS converters. Pomiary Automatyka Robotyka 16(7-8), 84–89 (2012)
3. Korytkowski, J.: Digital-analog converter electronic circuits and performance monolithic CA converters. PIAP Press, Warsaw (2012)
4. Jun, S.-W., Yasutomi, K., et al.: Linearized settling error calibration for a pipeline A/D converter using non-slewing amplifiers. JAMRIS (4), 204–206 (2009)
5. Horowitz, P., Hill, W.: Sztuka elektroniki. WKŁ, Wydawnictwo (2001)
6. <http://www.linear.com/LTC6655>
7. http://www.analog.com/static/imported-files/data_sheets/ADR420_421_423_425.pdf
8. http://www.analog.com/static/imported-files/data_sheets/AD7191.pdf

FPGA Based Processing Unit for a Checkweigher

Robert Ugodziński¹, Łukasz Gosiewski¹, and Roman Szewczyk²

¹ Industrial Research Institute for Automation and Measurements PIAP,
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
rugodzinski@piap.pl

² Warsaw University of Technology, Faculty of Mechatronics,
sw. A. Boboli 8, 02-525 Warsaw, Poland
szewczyk@mchtr.edu.pl

Abstract. This paper presents FPGA based module for a checkweigher utilizing Zynq type advanced signal processor. Developed module creates new possibility of implementation of advanced data processing algorithms for real-time signal filtering. Moreover, developed module makes possible data fusion from both position sensor, as well as driving coil of checkweigher. As a result the dynamic characteristic of weighting module can be effectively corrected, leading to significant improvement of functional characteristic of checkweigher. As a result the higher accuracy of measurements may be achieved.

Keywords: FPGA, signal, signal filtering, checkweigher.

1 Introduction

Precision mass measurements are critical to different advanced production systems. Especially high-tech production lines for pharmacy or other chemical production require high accuracy, robustness, and high speed measurements. Such measurements are realized by dynamic weighting systems called checkweigher.

Checkweighers have been intensively developed for last 20 years [1, 2]. As a result, the mechanical system of high quality industrial checkweigher guarantee low level of vibrations generated by belt-based weighted element transport system, high dynamic range of mass measurements, as well as effective vibration suppression. An example of such advanced dynamic weighing instruments is HPW series of checkweighers [3], produced by RADWAG Company.

In spite of important advances in the development electric and mechanical systems of checkweighers, further improvement of its accuracy, reliability and speed of operation is required by the industry. Analyses of possibilities of such improvements lead to the conclusion that these improvements are possible mainly by application of new signal processing methods available by market introduction of more effective, FPGA-based signal processors. An example of such signal processors is ZYNQ family introduced in 2011 by Xilinx Corporation [4]. It should be highlighted that ZYNQ family of signal processors integrates in their silicon structure (developed in 28 nm technology) the complete ARM Cortex-A9 processors. As a result, such solution offers new

possibilities of enhancement of efficiency and quality of signal processing in high-end industrial checkweighers. It seems that new possibilities offered by ZYNQ family of signal processors are still not fully utilized in advanced checkweighers. The presented paper fulfills this gap.

2 Hardware Architecture

2.1 Unit Architecture Overview

A schematic block diagram of proposed signal processing unit for advanced checkweigher is presented in figure 1.

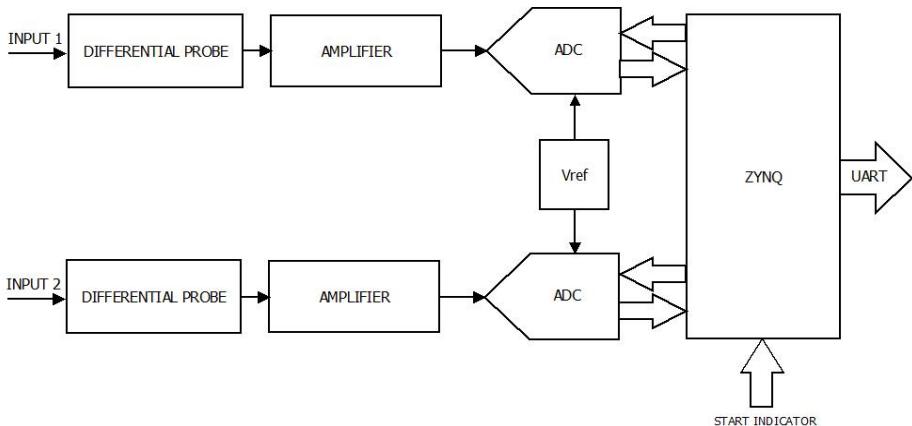


Fig. 1. Schematic block diagram of ZYNQ-based signal processing unit for checkweigher

To enable advanced signal processing, the processing unit consists of two channels, separated into modules. Signals are obtained from high precision resistor and scale position indicator via differential probes that assure enough input impedance to eliminate interference between channels and minimize impact on measured signals. Then, they are transferred to amplifiers that magnify and shift signals to fit full range of ADC and maximize measurement resolution. Results of conversion are sent to ZYNQ, in three wire ADC-custom serial communication protocol, where are digitally filtered and processed into outcome weight, which is transferred to main computer via UART or if necessary – by Ethernet.

2.2 Analog to Digital Converters

For presented solution Analog Devices AD7767-1 ADC was chosen [5]. It provides linearly scalable sampling rate up to 64 kS/s in proportion to external master clock, rendered in this case by ZYNQ and resolution of 24 bits. Interface consists of three lines – data output, serial clock driving it, and data readiness indicator. ADC built on-chip low-pass digital FIR filter with cut-off frequency set on half of sampling rate.

Reference voltage is rendered by ADR433 source at level of 3 V. ADC is driven by differential unipolar signal.

This model of ADC was chosen because of its 24-bit resolution needed to achieve required weigh accuracy and unusually high for precise ADC sampling rate, necessary for accurate imaging of characteristics. Research in industrial environment show that highest frequency of vibrations, which have impact on checkweigher is 800 Hz, and it has to be considered in filtering, which means that to get good image, it's needed to have at least ten times greater sampling rate, and if possible even more. That leaves a little choice of models. Another, less significant restriction was for ADC to operate on 3,3 V logic, which excluded a few considered ones.

2.3 ZYNQ

ZYNQ is so called system-on-chip (SoC) – an integrated circuit that joins all components electronic system into a single chip. ZYNQ consists of programmable logic – FPGA, processing system – two processor cores, programmable I/O blocks and two 12-bit ADCs. Integrating FPGA and processors gives us possibilities, that wouldn't be achievable by using only one of them. FPGA provides us with fast digital processing and real time operations thanks to logic gates, while processors enable implementation of complex algorithms or usage of embedded operating systems [6-8]. Integration makes possible fast data transfer between them and easy fitting whole system to our needs.

Flexibility and computing power is what made us to choose ZYNQ. Digital signal processing especially need the latter one. ZYNQ due to having two cores, enabled us to use double threading if application needed it, and each of them is clocked with frequency up to 600 MHz, which is few times more than typical high-end microcontroller. As our application has to process tens of thousands of samples with complicated algorithm in less than second, we had to use something more than a mere microprocessor. Programmable logic was another feature that could take some computing off from processor cores and simplify interfacing.

3 Software

3.1 Software Overview

Software is scattered between programmable logic and processing systems. FPGA part of ZYNQ is responsible for obtaining data from ADCs and initial filtering. Interfacing peripherals in VHDL provides us the possibility of easy multiplications of supported channels, also we can configure I/Os as we see fit at the moment. Digital filters are configured as FIR with cut-off frequency set on 0.1 of sampling rate. Filtering is done in real time, as data from ADC is constantly transferred regardless of whether measurement is in process, or not. It was decided to put initial digital filters in programmable logic, because implementation of them here gives us speed of filtering – one sample per clock cycle, which is unavailable in processor cores.

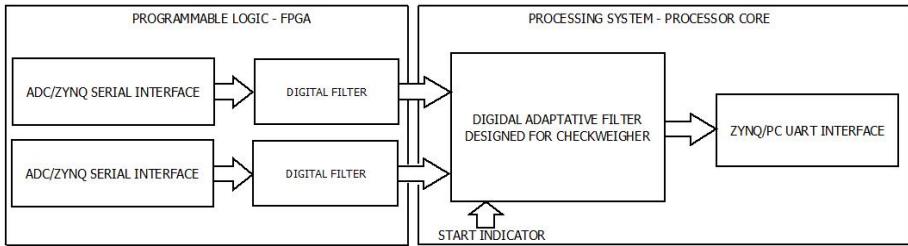


Fig. 2. General block diagram of embedded software for checkweigher's signal processing unit

Connection between FPGA and processor is realized by internal high-speed bus. In every conversion cycle the data from digital filters is put on two 32 bit parallel channels, then interrupted signal is being sent to processing system, where rest of operations are being held. It's 32 bit instead of 24 to ease further processing – there are no 24 bit signed variable in C, and conversion U2 24 bit into 32 in VHDL is just a matter of replicating most significant bit in bit vector. Standalone application implemented in processing system is awaiting on signal from checkweigher, that object to weigh is entering the scale. When it happens, samples are being gathered into buffer and processed by filter designed for checkweigher.

When measurement result is available it is send to computer via UART. It's worth to mention that UART peripheral in ZYNQ is treated as a console, therefore there is no need of configuration – data are sent with “printf” procedure, and received with “getch” function. If connected to PC, we can use serial terminal, to operate like on desktop console application. If there's need ZYNQ gives us also possibility of booting Linux distribution, which provides us with TCP/IP protocol if unit has to communicate with other computers in network.

4 Digital Filter for Checkweigher

Prior to the research of the designed hardware, it was possible to develop innovative and adapting digital filter. It uses brand new method to compensate external noise of surrounding machines thanks to additional measurement of signal from the position indicator. This allows us to simulate PID controller and obtain data needed to calculate the mass on the conveyor.

Below it is described step by step how the filter works.

Step 1: Obtain data from the hardware

Measurements are carried out on both channels, as it is presented in figure 3. Resampling of measurement can be done to average samples and reduce computing.

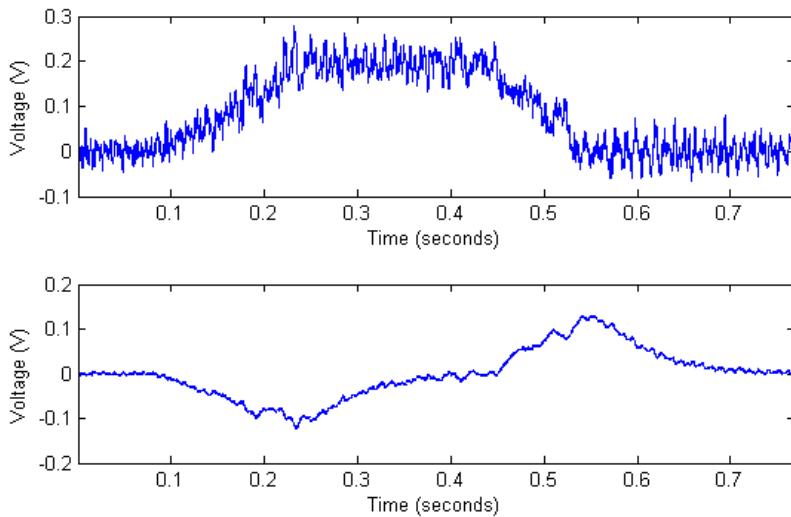


Fig. 3. Time series obtained from the hardware: a) from high precision resistor b) from position indicator

Step 2: Obtain updated PID controller model

As mentioned before, measurement from the input of PID controller (position indicator) and its output (high precision resistor) is obtained, so it enables to compute model of this controller. Identification occurs in every single measurement to compensate drifts of analog components of PID controller. To speed up computing, an algorithm uses the model from previous measurement.

Step 3: Recognize and delete an integrator component from model.

In this step a filter based on PID model used on the next step is created.

Step 4: Simulate the output of designed filter and subtract it from the time series from precision resistor.

It creates time series that only have the response of the integrator of PID controller.

Step 5: Obtain specific time series section from filter output and fit step response of first order inertial object function.

Figure 4 shows a line that describes the fitting section. It starts from the sample indicated by lowest value from position indicator and ends on sample indicated by maximum value of filtered time series.

These values indicate that measured object is completely on the conveyor and we can proceed to obtain the measuring.

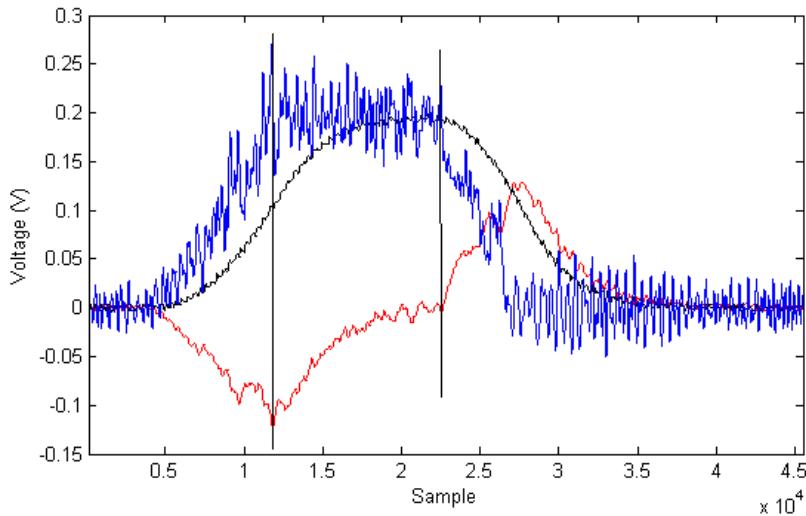


Fig. 4. Time series from high precision (blue), obtained integrator response from subtraction (black) and time series from position indicator (red)

Then we can fit the curve – parameterized step response function of first order inertial object.

$$h(t) = -a \cdot e^{-\frac{t}{b}} + c \quad (1)$$

Step 6: Obtain measurement

The most important parameter is ‘c’. It described the value of function at the infinite amount of time as it is a value of integral component of PID controller in full stability. Thanks to it, this parameter, after scaling, can be treated as a final measurement of unknown mass.

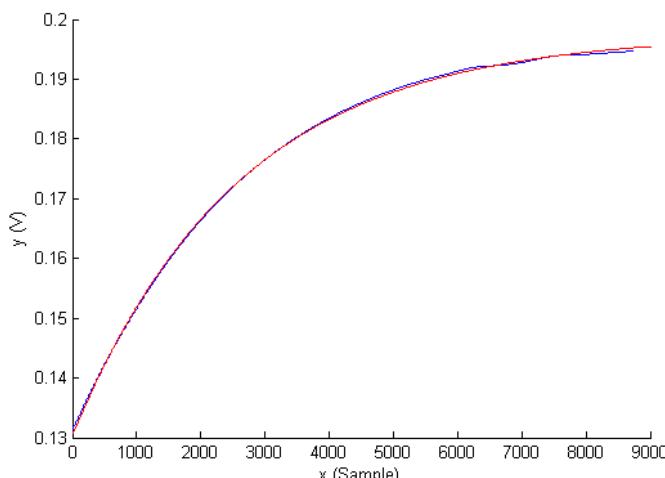


Fig. 5. Fitted curve (red) and fitting section (blue)

Analysis of measurement series shown that described implementation of filter gives significant improvement (from 20% to 40%, according to measuring condition) compared with old IRR and FIR filtering system. It will not be achievable without additional measurement that gives a view of input noise of PID controller.

5 Conclusion

Presented solution enables technological break-thru in digital signal processing in checkweighers. Due to application of advanced FPGA signal processor it is possible to perform data fusion of signals from checkweigher's position sensor and current measurements resistor. As a result, the significant improvement of checkweigher's accuracy can be achieved. Moreover, presented algorithm of data fusion generates redundant information about the weight measurement process. On the base of this information, it is possible to perform validation of the measurement process and detect possible malfunction of the weighting system. This function is especially important in critical life-dependent systems used in pharmacy or other kinds of chemical production.

Acknowledgements. Authors would like to express their gratitude to the RADWAG Company for its help and providing checkweighers for experiments.

This work was partially supported by The National Centre of Research and Development (Poland).

References

1. Yamazaki, T., Sakurai, Y., Ohnishi, H., Kobayashi, M.: Continuous mass measurement in checkweighers and conveyor belt scales. In: Proceedings of the 41st SICE Annual Conference 2002, vol. 1, pp. 470–474 (2002)
2. Pietrzak, P.: Dynamic mass measurement using a discrete time-variant filter. In: IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI), pp. 151–155
3. Dynamic Checkweighers DWT/HL/HPW, RADWAG Information Note (2013)
4. Fleming, S.T., Thomas, D.B.: FPGA based control for real time systems. In: 23rd International Conference on Field Programmable Logic and Applications (FPL), pp. 1–2 (2013)
5. AD7767: 24-Bit, 8.5 mW, 109 dB, 128 kSPS/64 kSPS/32 kSPS ADCs, Analog Devices Data Sheet, Rev. C (May 2010)
6. Maldonado, Y., Castillo, O., Melin, P.: Novel genetic optimization of membership functions of fuzzy logic for speed control of a direct current motor for hardware applications in FPGA. JAMRIS 4(4), 53–63 (2010)
7. Brzozowski, B., Henzel, M., Mazurek, P.: The FPGA implementation of the PID controller for the active magnetic bearing with permanent magnets. Pomiary Automatyka Robotyka 16(2), 437–442 (2012) (in Polish)
8. Kozien, K., Mysinski, W.: Implementacja sprzętowa protokołu PS/2 w układzie FPGA XILINX Spartan III XC3S1000. Pomiary Automatyka Robotyka (3), 81–83 (2010) (in Polish)

Improving of the Type A Uncertainty Evaluation by Refining the Measurement Data from a Priori Unknown Systematic Influences

Zygmunt L. Warsza¹ and Jerzy M. Korczyński²

¹ Industrial Research Institute of Automation and Measurement PIAP, Warsaw, Poland
zlw@op.pl

² Lodz University of Technology, Lodz, Poland
jerzykor@p.lodz.pl

Abstract. A new approach to improving the type A uncertainty evaluation by cleaning of the collected data from unwanted influences which appears as non-periodical and periodical systematic components identified in the data is presented in the paper. The approach refers to regularly in time sampled data. The cleaning process comply with the main stream of ISO GUM recommendation and can be recognized as good practice in the proper estimation of the type A uncertainty. The proposed approach is discussed in the paper and the numerical example is presented as well.

Keywords: uncertainties, data cleaning, type A uncertainties.

1 Introduction

A graphical model of the measuring process and proposed type A measurement uncertainty evaluation upgrading, of the collected data are presented in Fig. 1. Let assume, that the measurand is characterized by only one value, which is subjected to disturbances while it is passing through the tested object and measurement channel. Disturbances of the measurement signal depend on different physical phenomena and come from outside, effecting jointly with imperfection and aging of elements of the measurement channel. Upgraded evaluation of the measurement data processing includes two stages. The collected data are passing first through the stage 1, in which corrections of known influences are applied as well as “outliers” are identified and treated for example by the procedure described in [6]. Then data are passing through the stage 2, in which data sampled regularly or collected in known time spacing, are processed by cleaning procedure from non-periodical component – linear or nonlinear trend and also from periodical components – oscillations identified in collected data. This cleaning data procedure is proposed for application before any further data handling and uncertainty evaluation. The data after cleaning are more appropriate for estimation of the true value of measurand and for calculation of uncertainties by applying GUM recommendation. Further evaluation of the measurement uncertainty can be carried out on new set of data according to GUM recommendation and is as

follows: when the sample of n observations i_q was corrected from identified components then the new mean value $\bar{x} = \bar{q}$ is calculated, as the best estimate of the whole population of trials (collected data – observations) of the measured value x .

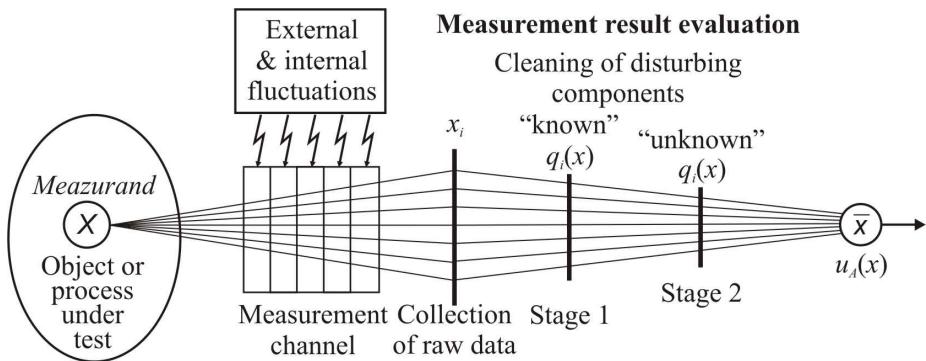


Fig. 1. Model of the measurement process and upgraded evaluation of the uncertainty type A and measurement result estimator x of the regularly sampled measurand x

The standard deviation $s(\bar{q})$ of that mean is the standard uncertainty $u_A(x)$. The evaluation is based on the statistical method and is one of two basic component of combined standard uncertainty expressed by formula: $u_c(x) = \sqrt{u_A^2(x) + u_B^2(x)}$. The coverage interval, according to GUM, is calculated as expanded uncertainty expressed as a product of coverage factor and combined uncertainty: $U_p(x) = k_p u_c(x)$.

The strictly theoretical background for application of the type A method for uncertainty calculation of cleaned data requires:

- observations q_i do not carry any recognizable systematic component,
- results q_i after cleaning are uncorrelated (statistically independent) and are of the equal weight,
- results q_i are randomly distributed and the statistical parameters of collected data are estimate of the population with needed level of confidence,
- mean value $\bar{x} \equiv \bar{q}$ of the sample could be accepted as the proper result of measurements, its standard deviation $s(\bar{q}) = u_A$ is calculable as for Normal distribution and it is the pure statistical component of overall accuracy.

Procedure of the uncertainty $u_A(x)$ evaluation according to actual ISO GUM [1] for type A method recommended for corrected values of observations: $q_1, q_2, q_3, \dots q_n$ (n – number of observations) comprise the equations (1) to (4), as follows:

Mean value:

$$\bar{x} \equiv \bar{q} = \frac{1}{n} \sum_{i=1}^n q_i \quad (1)$$

Variance of the sample:

$$s^2(q_i) = \frac{1}{n-1} \sum_{i=1}^n (q_i - \bar{q})^2 \quad (2)$$

Experimental standard deviation:

$$s(\bar{q}) = \sqrt{s^2(q_i)} \quad (3)$$

Standard uncertainty of type A:

$$u_A(x) = s(\bar{q}) = \frac{s(q_i)}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (q_i - \bar{q})^2} \quad (4)$$

The literature [2–7], dealing with uncertainty evaluation, relates mainly to the above GUM [1] approach. The Monte Carlo based method, MCM, presented for example in [1] and in GUM Supplement 1 do not refers to data cleaning as proposed here, so any data handling before MCM is applied are not forbidden. So it is allowed if appropriate or even welcome if improves the final result of uncertainty evaluation. Considerations in GUM [1] and in related literature refers for the sample of limited number of only independent trials, then correlation of data is not required.

There are still many areas in measurement science, research and technology to which today's GUM recommendation cannot be applied due its limitations. GUM do not treats uncertainty calculation properly if:

- the set of raw observations is not the sample of the pure random population.
- the set of raw observations is of not the normal distribution
and also
- the method does not take into consideration the order and relations between samples, do not refers to:
— time variable samples, or
— sample of which elements are influenced by ambient conditions, which require stochastic stationary and non stationary processes modelling as more adequate to the real world
and
- the GUM method does not refers to the evaluation of uncertainties of dynamic parameters and parameters obtained as a result of digital signal processing (DSP) for different algorithms.

Measurement data handling by removing all a priori "known" systematic components from raw observations still do not guarantee obtaining the sample free of unknown disturbing the data regular components. These components should be also eliminated, but identification of all of them might not be possible. The identification and removing such elements from the set of data can be recognized as "cleaning of data" or signal filtering as the data are similar to set of digital values of recorded signals. If an additional information is known, e.g. procedure: how observations x_i of the constant

value x of the measured quantity are collected, i.e. regularly sampled as series in time or space or by other known way, then some of undesirable components as outliers, trend or harmonics in relation to the length of the sample, could be cleaned up. It can be done only partly by the input filtration, more - by algorithms including identification of the components. For regularly sampled observations the stage 2 of data cleaning as presented in Fig. 1 is proposed to be added to good practice of GUM recommendation. The Least Square Method for components identification was applied for examples presented in [11]. It is worth to say clearly, that enough properly "cleaned" observations additionally may be also statistically dependent, i.e. they could be auto-correlated, especially if they have been collected in relatively too short periods between trials. Also Normal distribution may not be the best distribution for the real measurement data. The mean value of the sample of such observations is not always the most likelihood parameter of their distribution and other unbiased estimators should be used, as the midrange of rectangular distributions and median (MED) of Laplace double-exponential ones [3, 10]. The last two problems were explained and discussed in [12]. The methods of achieving the best possible values of the uncertainties: $u_B(x)$ and $U_p(x)$ are presented in [10, 13].

These problems are very important as new generation of instruments and measurement systems can be built in which calculation of uncertainty of measurement can be incorporated as one of their function [8].

2 Detection and Discrimination of Systematic Components from Regularly Sampled Measurements

As it is pointed out in the previous chapter after correction of sample observations from known systematic disturbances some unknown components of the regular, systematic nature may still remain. If uncertainty, u_A , is calculated for such data, still contaminated by systematic nature of effects, the obtained value is too high, as not only random dispersion of measurement results is present. Such raw measurement data shouldn't be treated as random stationary process, i.e. should not be characterized by statistical parameters only. The "cleaning" task is to investigate, in collected data, a priori unknown such effects as no periodical trend and as periodical interference components. This is possible only if the collected data are regularly sampled or time interval between sampled data is known.

The proposed cleaning process should be proceeded by elimination of outliers, the data which are inconsistent to other collected data, if any are present. That problem was deeply analyzed by Pavese and Ichim [6] and the specific procedure of elimination of outliers proposed.

As time dependent trend seems to be identified applying mathematical rules including LSM method for optimisation, the constant value which may be incorporated in measurement data will remain unknown until the instrument calibration process will take place. However the constant may be time dependent due to instrument aging, but such constant value wouldn't influence uncertainty u_A . That constant value practically does not vary if observations are performed in relatively short time, but

may be time dependent during the life of instrument or full time of testing of the particular object due to variation of measuring conditions and aging of its and instrumentation parameters [2]. The measuring circuit might require also the extraordinary internal calibration. The electrical drift of the instrument may be corrected by self calibration procedure, by manual or automatic internal calibration, but it not covers the correction of the trend of entrance signal to the instrument. For sensors special stand is needed. If the instrument self-calibration is not available in situ, the constant value of its non-recognizable trend must be treated as a estimated additional component of the uncertainty $u_B(x)$.

There are few methods to identify and to remove from raw measurement data the trend and other interfering components. One of the simplest – the Least Square Method (LSM) was applied as it is shown in the Example described below.

Example: The digital voltmeter of 4 1/2 digit was used to collect measurement data of the parameter of some process. The data was sampled uniformly in time and $n=121$ results in (V) are as follows:

Table 1. Raw values of observations of the Example

1.2200	1.2080	1.2186	1.2263	1.2497	1.2725	1.2981	1.2731	1.2500
1.2286	1.2181	1.2183	1.2162	1.2247	1.2253	1.2108	1.2409	1.2529
1.2696	1.2577	1.2397	1.2300	1.2341	1.2562	1.2449	1.2378	1.2203
1.1920	1.2056	1.2092	1.2198	1.2227	1.2210	1.2134	1.2064	1.2138
1.2154	1.2220	1.2352	1.2479	1.2385	1.2277	1.2206	1.2320	1.2466
1.2679	1.2412	1.2279	1.1897	1.2123	1.2291	1.2498	1.2450	1.2343
1.2356	1.2420	1.2239	1.2101	1.2057	1.2044	1.2011	1.1940	1.1941
1.1836	1.1956	1.2002	1.2159	1.2142	1.1963	1.1840	1.1726	1.1657
1.1553	1.1726	1.1932	1.2146	1.1983	1.1904	1.1736	1.1874	1.2003
1.1950	1.1911	1.1754	1.1594	1.1748	1.1799	1.1817	1.1816	1.1907
1.1937	1.1982	1.1956	1.1977	1.1868	1.1684	1.1455	1.1648	1.2019
1.2126	1.2086	1.1885	1.1760	1.1729	1.1706	1.1692	1.1921	1.2036
1.2229	1.1996	1.1810	1.1609	1.1314	1.0975	1.0704	1.0845	1.0954
1.1146	1.1172	1.1148	1.1263					

The task was to calculate the average value from the 121 trials of collected readings of the voltage and to estimate the uncertainty as its standard deviation after the identification and elimination of earlier unknown systematic effects (like linear or other no periodical trend and periodic components). There is no available information regarding corrections which could be applied to measurement results in the beginning of calculations. Uncertainty $u_B(x)$ is not in the scope of interest in this numerical example.

Solution:

The raw data values v_i (in order of collection) are presented in Fig. 2a. The declining trend of that data is observed. The trend and oscillation were discovered in the collected data as it is presented in Fig 2a. Physical analysis of the measurement

process should confirm that such systematic components could exist in the raw sample. For comparison on Fig. 2b are given raw observations and their values after each of two steps of elimination: firstly trend and secondly oscillation. Additionally in Fig 3 it is shown how sets of observations ordered by values change by these two steps of data cleaning. Ideal shape of such curve for Normal distribution is very near to the integral probability curve (Laplace) turned by 900. Mean values and uncertainties, u_A of all three sets of measurement data are given in Table 2.

Deviations Δv_i of raw data from the mean value $\bar{x} \equiv \bar{q}$ are seen on the Fig 2a and curve 1 in Fig 4 is connecting points of their histogram intervals (j – from $j=1$ to $j=8$ represents number of intervals). On axis “y” are given the empirical occurrences $w_j = n_j/n$, where n_j – number of data in the interval j , n – number of all collected data. The width of intervals are calculated using as follows

$$h = \frac{\max(\Delta v_i) - \min(\Delta v_i)}{m} = \frac{0,228}{8} = 0,0285 \quad (5)$$

where: $m=8$ - number of intervals.

Even from the first glimpsed to the curve 2 of Fig 4, it obvious that histogram is non symmetric and do not seems to have a Normal distribution shape. Despite of that the mean value \bar{V} and standard deviation $s(v_i)$ were calculated using formulas: (1) and (3), which are suggested by GUM [1]. These results will be compared to the similar results after each cleaning process of the data as it is quoted in Tab. 2.

Assuming “zero” value of the trend is on mean value and applying the LSM method to trend line we obtain:

$$y(i) = (-7.82E-06)i^2 + (7.82E-05)i + 0,033 \quad (6)$$

The data $\Delta q'_i$ deviations of the sample corrected by elimination increments of the nonlinear trend in relation to its mean value are presented in Fig. 2b. Their mean value is the same as before but range is now smaller, it is changed from 0.228 to 0.143. The new histogram for corrected data $\Delta q'_i$ is presented by curve 2 in Fig. 4. The Normal distribution can be expected, but it is appropriate to test the thesis of compliances using criterion χ^2 and in which χ^2 is expressed by:

$$\chi^2 = n \sum_{j=1}^{n_p} \frac{(w_j - p_j)^2}{p_j} \quad (7)$$

where: p_i – is a probability of interval j according to Normal distribution

Periodical component – see Fig. 2a, is detected using also the LSM method. It has quite small amplitude but is influencing on the calculated value of uncertainty u_A . Curve 3 of histogram obtained for data after two steps of cleaning is given also on Fig. 4. Range of that data is practically the same about 0,143 and the mean value is also practically not changed as the whole number of sinusoid period is in the length of data.

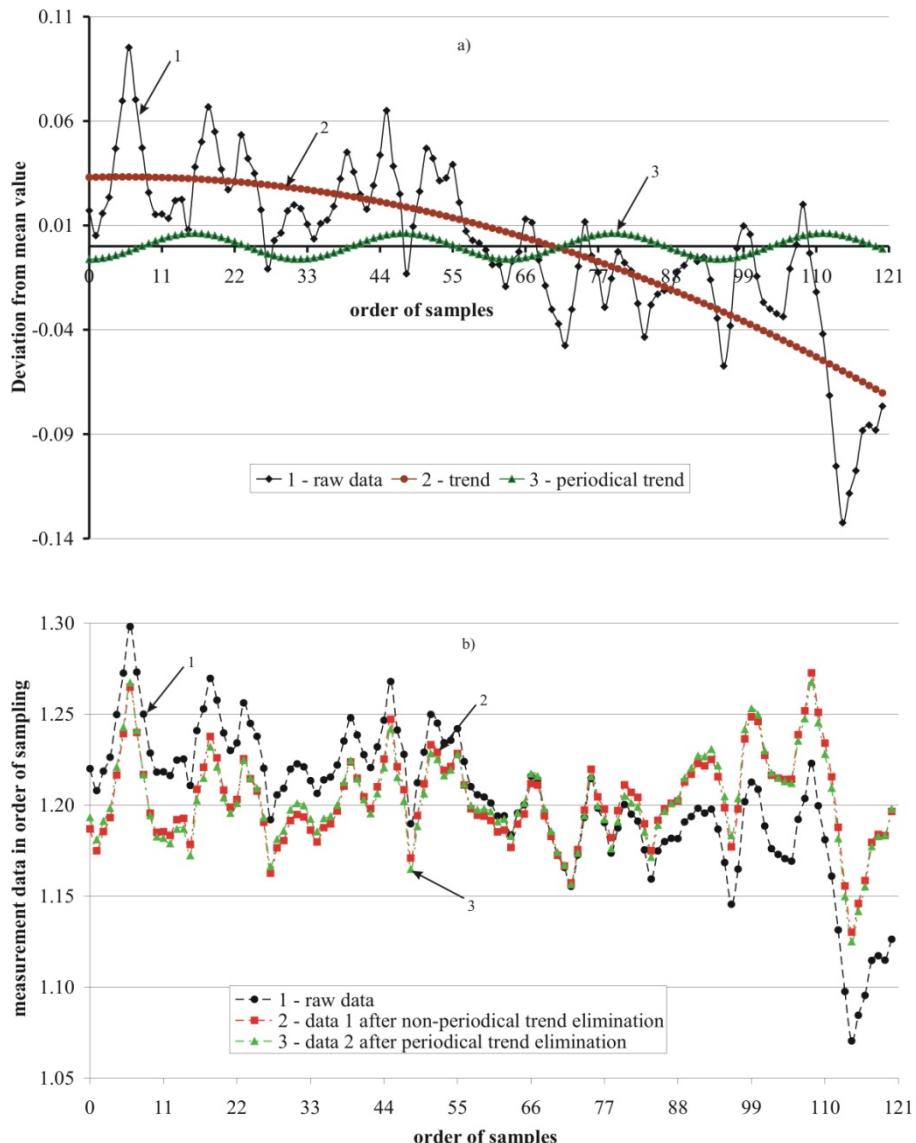


Fig. 2. Sets of measurement data: a) deviations of raw measurement data from mean value in order of the regular sampling and identified systematic components incorporated in data, b) Sets of measurement data after cleaning from trend and oscillation

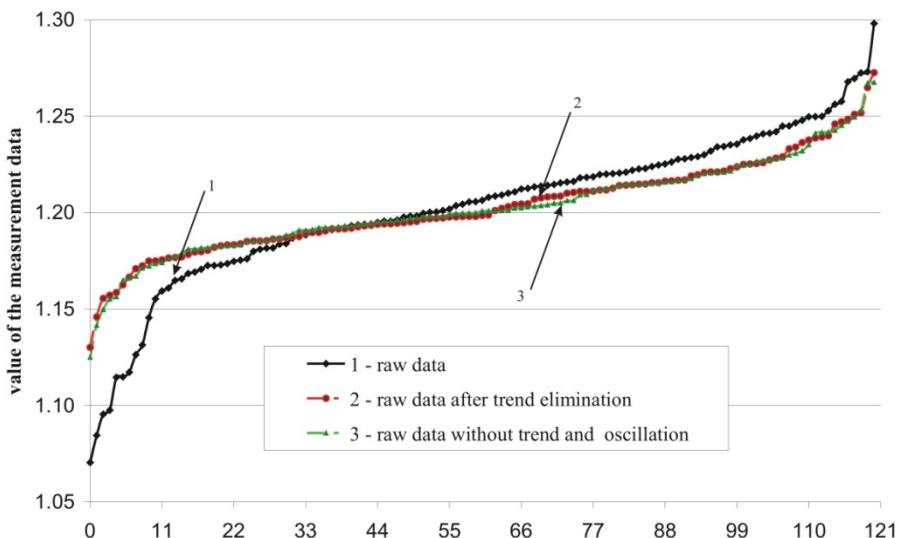


Fig. 3. Raw measurement data and after each step of cleaning ordered by values

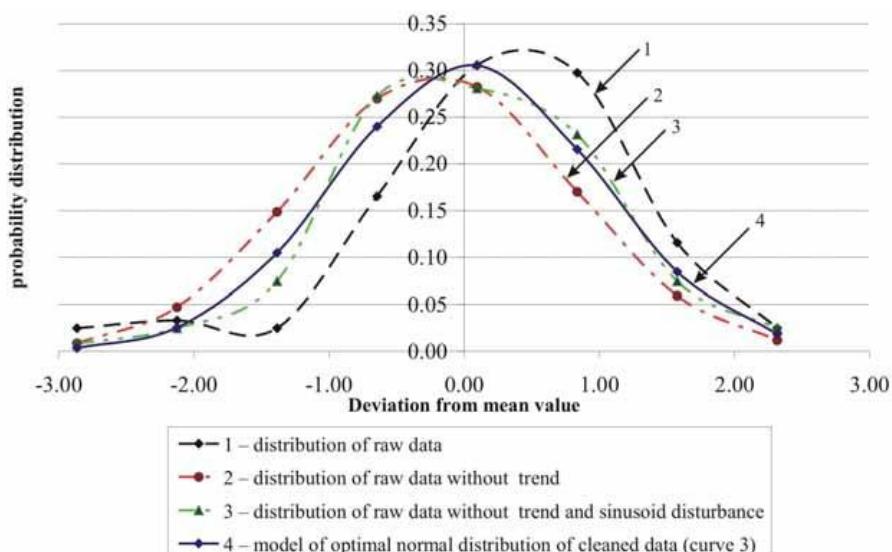


Fig. 4. Distributions of the measurement data deviations from the mean value

Steps of calculations of parameters for raw data and corrected data were performed according to formulas as stated before and are presented in Tab. 2. In practice, as the two significant digits of the u_A uncertainty are enough in the most cases to describe result of measurements, so the proposed here data cleaning by identification of one or two components seems to be quite sufficient in that numerical example. Some numerical intermediate results presented in table have more digits, but it is done deliberately for better presentation how cleaning can be effective.

Table 2. The parameters of collected measurement data sample of raw observations

Measurement data	Raw data v_i	Data cleaned from:	
		trend q'_i	trend and single oscillation q_i
χ^2 criterion for normal distribution	$\chi^2 = 52.28 > \chi^2_{5, 0.05} = 11.1$	$\chi^2 = 3.83 < \chi^2_{5, 0.05} = 11.1$	$\chi^2 = 3.25 < \chi^2_{5, 0.05} = 11.1$
	negative result	positive results	
Mean value	$\bar{v} = 1.2029$	$\bar{q}' = 1.2028$	$\bar{q} = 1.20262$
Standard deviation	$s(v_i) = 0.03953$	$s(q'_i) = 0.02409$	$s(q_i) = 0.02407$
Uncertainty u_A	0.003593	0.002190	0.002188
Ratio of standard deviations	$\frac{s(v_i)}{s(q'_i)} = \frac{0.03953}{0.02409} \approx 1.6407$		$\frac{s(v_i)}{s(q)} \approx 1.6420$
Result of measurements	$x = 1.2029 \pm 0.0036$		$x = 1.2026 \pm 0.0022$

Conclusion about numerical example data: The measurement data cleaning is resulting in lowering experimental standard deviation by 64.00 % after trend elimination and by 64.08 % for trend and periodical component elimination. So, the cleaning process is resulting in lowering the standard u_A uncertainty on the same ratio. No further periodical component is observed in corrected data. But to check the correctness of that statement the harmonic analysis for all collected data was applied. The amplitudes of harmonics are calculated as fraction of standard uncertainty of $s(q_i)$ – see Fig. 5. The graph represents typical character for process of the low frequency noise. It proves the earlier expectations about no predominant harmonic observation.

Based on cleaned measurement data, it is possible to calculate others parameters like expanded uncertainty based on t-Student distribution for small number of observations n. This approach is well developed in GUM recommendations [1] and was treated widely in literature.

3 Verification of Data Cleaning Method

To verify the proposed procedures of measurement data cleaning the authors use the simulation method. It is based on adding known systematic components to the sample of collected random measurement data and next application of the proposed cleaning procedure. Such verification is presented in Example 2 given in [11]. The original collected data were classified as sample from the population of probability distribution function uniformly distributed, and the linear and periodical component were added to that data. The parameters of the both added components were not known a priori for the person, who applied the cleaning procedure. This procedure was just the same as presented in Example of this paper.

If the set of measurement data is contaminated by periodical character, then the mean value is slightly changed due to it, unless not complete number of periods is in the collected set. Such situation is very likely in practice, and then it is worth to remove such periodical component, what was elaborated by authors and results of some

numerical example are presented above. The basic parameters of such periodical signal were identified, i.e.: amplitude, frequency and phase shift. The LSM was applied as the criterion for their optimization.

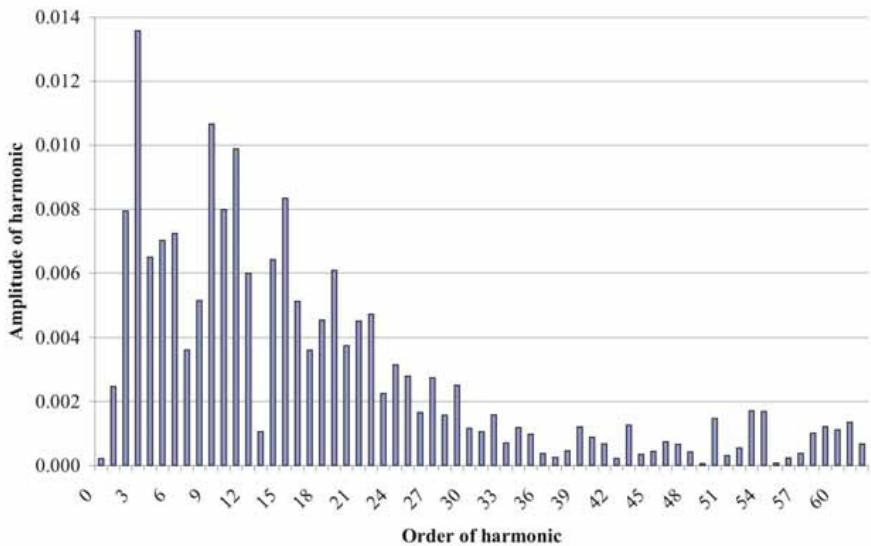


Fig. 5. Frequency spectrum of measurement data deviations after cleaning. (Frequencies of harmonics are related to of the reciprocal of the period of measurement data collecting time).

4 Summary and Final Conclusions

This paper propose upgrading of the procedure recommended by guide ISO GUM for evaluation of the uncertainty by type A method in the case of regularly sampled data measurements. This is done by "cleaning" the raw measurement data from unknown systematic components, which is based on identification and removing from set of data the trends of linear and periodical characters. The rule after data cleaning is such that the standard uncertainty lowers. It was proved by many elaborated examples of which one for normal distribution is presented as Example.

Example for uniform distribution is quoted in [11]. In that Example the precision of the periodic component identifications by LSM method was also tested and its result was enough satisfied for uncertainty u_A estimations. Calculations of the uncertainty u_A in the case when nearer measurement observations are correlated, and standard uncertainty as accuracy measure of other then the mean value estimators of random population (e.g. midrange – for uniform distribution, median – for double exponential one) are also described for comparison in [10, 12, 13]. In the most cases in practice it is enough to remove from raw data the linear trend component and the main harmonic function only. That should satisfied accuracy commonly applied in uncertainty u_A evaluation and could be added to GUM recommendations in the way as and summarized on Fig. 6 (data cleaning operations described in this paper are

shown over the dotted line). Without application of the cleaning process of unknown systematic components they could influencing twice on final measure of result accuracy, i.e.: on $A u$ value and second time in the estimation of u_B uncertainty components.

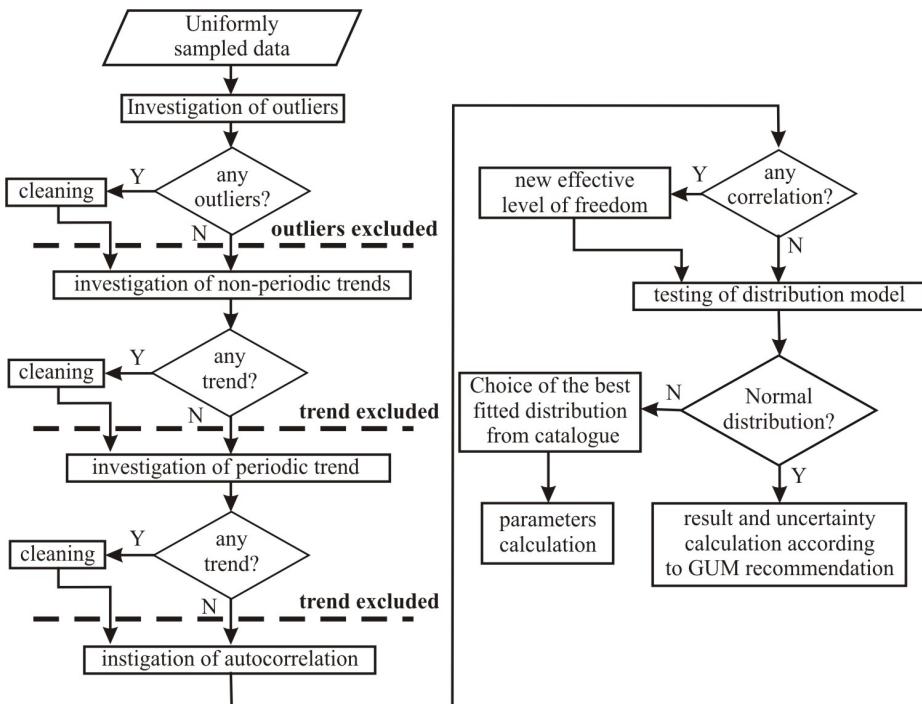


Fig. 6. Scheme of upgraded procedure of uncertainty u_A calculations

It is very important in the measurement practice to be sure that identified regular components in the raw sample are really become from undesirable interference signals and they are not just the result of random dispersion of observations in this particular sample. It is recommended to check them in few samples collected in nearly similar circumstances and compare them by the properly choose criterion. Or on other way by recording the sample as long as possible and then to test if it is stationary, e.g. by dividing it to parts and compare if regular components identified in these parts are similar.

This paper do not draw out all aspects of investigation of trends and other disturbances in raw measurement results and “data cleaning” by application signal filtration method [2, 3, 9] – but is contributing to the such wide problem. It is obvious that one publication do not refers all.

This cleaning approach is an early stage of data handling before investigation the influence of their autocorrelation and choosing the adequate probability distribution as next stages proposed the procedure of the uncertainty of type A evaluation are presented in [12, 13], while the improvement procedures the uncertainty of type B

[14] and overall uncertainty U in [10, 13]. All these proposals may be included in activity of developing the GUM recommendations on expression of the uncertainty in measurement [7], but before that they also need some task of standardization.

References

1. Guide to the Expression of Uncertainty in Measurement, revised and corrected with supplements. BIPM (2008)
2. Piotrowski, J.: Theory of Physical and Technical Measurements. Elsevier Amsterdam, PWN Warszawa (1992)
3. Novitski, P.V., Zograf, I.A.: Ocena pogreshnostej rezul'tatov izmereni, p. 248. Energoatomizdat, Leningrad (1985) (in Russian)
4. Rabinovich, S.G.: Measurement Errors and Uncertainties Theory and Practice, 3rd edn. Springer, Heidelberg (2005)
5. Kirkup, L., Frenkel, B.: An Introduction to the Uncertainty in Measurement Using the GUM. Cambridge Univ. Press, Physics (2006)
6. Pavese, F., Ichim, D.: SAODR: Sequence analysis for outlier data rejection. Measurement Science and Technology 15, 2047–2052 (2004)
7. Bich, W., Cox, M.G., Harris, P.M.: Evolution of the ‘Guide to the Expression of Uncertainty in Measurement’. Metrologia 43, 161–166 (2006)
8. Warsza, Z., Korczynski, J.: Elimination of the influence of a priori unknown systematic components on the uncertainty type A of regularly sampled measurements. Pomiary Automatyka Robotyka 12(7-8), 45–53 (in Polish)
9. Korczynski, M.J., Hetman, A.: Advanced Mathematical & Computational Tools In Metrology. In: New Approach to Presentation of Measurement Results in Virtual Instruments, pp. 309–314. Word Scientific Publishing Co. Pte. Ltd., River Edge (2004) ISBN 981-238-904-0
10. Dorozhovets, M., Warsza, Z.L.: Upgrading calculating methods of the uncertainty of measurements results in practice. Przegląd Elektrotechniki (1), 1–13 (2007) (in Polish)
11. Warsza, Z., Dorozhovets, M., Korczynski, M.: Methods of upgrading the uncertainty of type A evaluation (1). Elimination the influence of unknown drift and harmonic components. In: Proceedings of 15th IMEKO TC4 Symposium, TU Iasi Romania, pp. 193–198 (September 2007)
12. Dorozhovetz, M., Warsza, Z.: Methods of upgrading the uncertainty of type A evaluation Elimination of the influence of autocorrelation of observations and choosing the adequate distribution. In: Proceedings of 15th IMEKO TC4 Symposium, TU Iasi Romania, pp. 199–204 (September 2007)
13. Dorozhovets, M., Warsza, Z.L.: Propozycje rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (1) Uwzględnianie wpływu autokorelacji i nieadekwatności rozkładu wyników obserwacji w niepewności typu A. Pomiary Automatyka Robotyka 11(1), 6–15 (2007) (in Polish)
14. Dorozhovets, M., Warsza, Z.L.: Propozycje rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (2) Uściślenie metod obliczeń niepewności typu B. Pomiary Automatyka Robotyka 11(2), 45–52 (2007) (in Polish)

Evaluation of the Standard Deviation of the Random Component of the Measured Signal from Its Autocorrelated Observations

Zygmunt L. Warsza

Industrial Research Institute of Automation and Measurement PIAP, Warsaw, Poland
zlw@op.pl

Abstract. Described is the proposal of evaluation the standard deviation of the stationary random component of measured signal from its regularly sampled observations when they are auto-correlated. As, the first step is the identification and removing the regularly variable components from the raw sample data. Then formulas for standard deviation of the sample and of the mean value are expressed with use the correction coefficients or the so-called "effective number" of observations. These quantities depend on number of observations and on the autocorrelation function of the sample cleaned from regular components. How to estimate the autocorrelation function for the sample data is also described. Few numerical examples to illustrate problems are included.

Keywords: probability distribution, standard deviation, autocorrelation, uncertainty type A.

1 Introduction

Most of measurement signals have regular and random components. Sources of randomness may be existing as well in the measured object and in the measurement system. Mainly it is because of influence of environments. If one or more regular components of the signal have to be measured then its random component can be considered as non needed the random error. This error has the form of random process and can be treated as the stationary ones in some period of time of collecting the measurement observations. If random component has also the limited frequency range then observations which are very near each other are autocorrelated. Randomly distributed deviations from constant value or other function established in particular measurements as measurand are described by their distribution. Standard deviation of this distribution is named uncertainty type A. The international guide, known under acronym GUM, gives recommendations how to evaluate in measurement the standard and expanded uncertainties of required probability. GUM covers only measurement of the variable which data is randomly distributed but no related statistically (e.g. without autocorrelation of their data). Then use of actual version of GUM [8] is limited in many types of measurements. In particular, for measurements of signals which are the

processes variable in time or in space it was not established yet how to estimate standard deviation (SD) and uncertainties type A.

In this paper is synthetically discussed how to determinate standard deviation and measurement uncertainties of regularly sampled measurand when observations are autocorrelated. It was covered in detail in publications [1–6, 9] and their bibliographies. This includes preparing of the raw data sample to further statistical calculations by removal regularly variable components from it. Then the estimation of sample proper standard deviation for a priori known autocorrelation function ρ is given. For the cases when this function is unknown, presented is method, how to find the estimator r of the autocorrelation function from the sample data.

2 Removing the Regular Components from Raw Data

The values of the measurement signal or the output readings are discrete as a result of sampling of the input analogue signal with properly chosen frequency and A/D conversion. The process of collecting measurement observations is now usually automated.

The dispersion of the "raw" values of measurement observations is caused by reason of both random and determined type. That is result of changes of measurand itself, and changes of the internal parameters of the measuring circuit and of environmental conditions. It is different for each of the samples taken at different times, measured by different instruments, and even by the same device over its lifetime. Random changes of the observations can be stationary or no stationary. There are short-term noise (called outliers), which prior to the assignment of values and the uncertainty of the measurements should be identified and eliminated from the raw data. Changes in the form of a regular (non periodic) component, i.e. the trend and periodic components affect the shape of the histogram of the sample, the mean value and uncertainty of type A. And there are also undesirable.

Calculations of the Standard Deviation and result and accuracy of measurement are made usually with the following assumptions:

- time T_n to collect the sample of n observations (width of the measurement window) was chosen properly, i.e. the random component of the sample can be assumed stationary,
- the sampling is uniform with the constant frequency n/T_n , and
- the random component is additive only, i.e. independent from the measurand value, and is described by the normal probability distribution.

Then values $y(i)$ of the discrete output signal obtained by sampling can be considered as:

$$y(i) = f(i) + f_R(i) + N(0, \sigma) \quad (1)$$

where: $i \in (1, n)$ – the current number of the sample observation $y(i)$ obtained in time t_i , $f(i)$ – values (estimated from measurements) of the known a priori measurand function at times t_i , $f_R(i)$ – the values of other a priori unknown and unwanted regular components in time t_i , and $N(0, \sigma)$ – normally distributed random component with standard deviation σ .

The standard deviation SD as statistical uncertainty u_A of measurements should be estimated only from values of the last random component $N(0,\sigma)$. The values of function $f(i)$ estimating the measurand and of the residual signal contains the unwanted non-known a priori regular component $f_R(i)$ should be subtracted from the sample data before SD calculations.

Example 1

Figure 1a shows a series of successive $n = 121$ "raw" results of measurement observations (black squares), obtained by regular sampling of the tested process [3]. You can see that the sample contains non-linear decreasing trend, there are not noticeable periodic components, and rather a small laceration of the data graph indicates that this data can be autocorrelated. After removing one outlier, using LSM (least squares) the polynomial was determined:

$$y(i) = (-7,65 \cdot 10^{-7}) i^2 + (6,6 \cdot 10^{-5}) i + 0,034$$

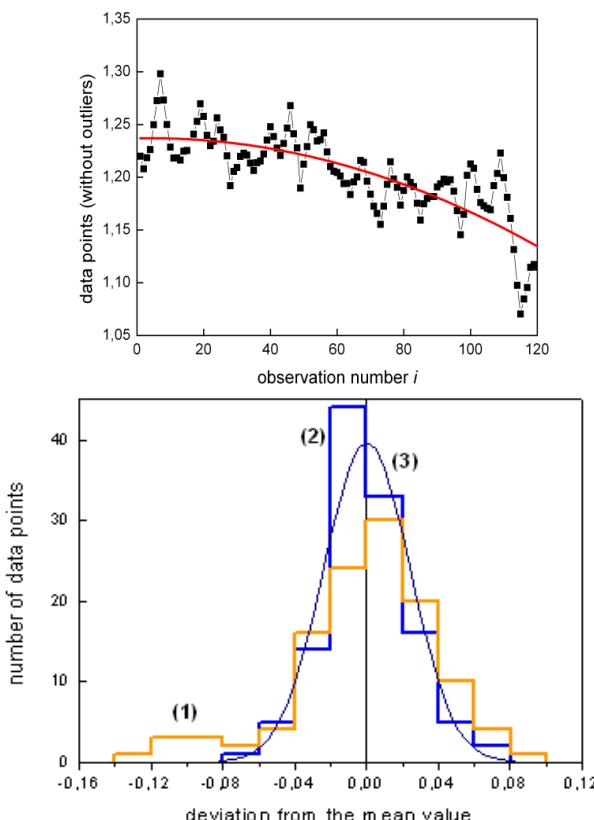


Fig. 1. An example of the measurement data sample collected sequentially by regular sampling: a) deviations from the mean value of the raw data and their systematic regularly variable component - drift identified by the least squares method; b) distributions: of the raw data (1), after subtraction of the drift (2) and Gaussian pdf curve fitted to that data (3)

Trend is modelled and removed from the sample data, assuming that it passes through the average value. In Figure 1b a histogram of the raw sample is given – (1) and histogram after removal of the trend – (2) and a matching criterion χ^2 its normal (Gaussian) probability density function (pdf) – (3).

Trend causing asymmetry and the existence of a large "tail" on the left of histogram (1), which does not meet the criterion of χ^2 , and with trend uncertainty u_A is about 64% higher than $u_A=0.00219$ without it. Purification of the raw data significantly reduced u_A . For $n=120$ the relative standard deviation $s(u_A)/s \approx 8.5\%$.

3 The Uncertainty of Mean Value for Correlated Observations

Presented is a brief description of the problem. The sequence of measurement data obtained from the sampling process and purified from the deterministic component can be described by a stationary time series. Statistical correlations between realizations X_i , X_{i+k} of such series is characterized by the autocorrelation function

$$\rho_k = \frac{\text{cov}(X_i, X_{i+k})}{\sigma^2} \quad (2)$$

Function ρ_k depends on the frequency spectrum of the test process and is known or its estimate r_k should be found from the measurement data. In measurement of physical quantities the correlation function is positive.

The relationship between standard deviation $\sigma(\bar{x})$ of the mean value and σ of the individual correlated observation x_i results from the variance of the sum of random variables [1, 2, 4, 9]

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n_{eff}}} \quad (3)$$

where

$$n_{eff} = \frac{n}{1 + 2 \sum_{k=1}^{n-1} (1 - k/n) \rho_k} \equiv \frac{n}{1 + D_\rho} \quad (3a)$$

For the statistically independent observations $\rho_k \rightarrow 0$ (for $k \geq 0$), consequently $D_\rho = 0$ and formula (2) passes to the commonly known relation

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

In opposite, when the observations are fully correlated (closely linked), i.e. $\rho_k \rightarrow 1$, with (3a) results

$$D_\rho \rightarrow \frac{2}{n} \sum_{k=1}^{n-1} (n-k) \cdot 1 = n - 1 \quad (4)$$

Then the standard deviation of the mean is the same as for a single observation of the sample because in the limit $\rho_k \rightarrow 1$ all subsequently repeated observations will be the same.

The value of n_{eff} is needed to proper estimation of the standard deviations $s_a(x_i)$, $s_a(\bar{x})$ for sample of autocorrelated observations [4–6]. Their relations to GUM parameters $s(x_i)$, $s(\bar{x})$ are given in table 1. Autocorrelation data have the effective number degrees of freedom v_{eff} . Defined approximately value of v_{eff} [4–6] is given in Table 1 by eq. (8). In this case $v_{\text{eff}} \neq n_{\text{eff}} - 1$ [4, 5].

The relative dispersion of the standard deviation of autocorrelated observation depends on v_{eff} . It is also given in table 1 – eq. (9).

Formulas for calculation expanded uncertainties of non-correlated and correlated data are given in the last row of table 1.

Table 1. Formulas for evaluation the standard deviation and uncertainty of the measurement sample of non correlated and autocorrelated data

Parameter	Non autocorrelated data (GUM)	Autocorrelated data
Effective number of observations	n	$n_{\text{eff}} = \frac{n}{1 + 2 \sum_{k=1}^{n-1} (1 - k/n) r_k} \equiv \frac{n}{1 + D\rho} \quad \dots\dots(5)$ <p>where: ρ_k - autocorrelation function in point k</p>
Standard deviation of single observation	$s(x_i) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ <p>where: \bar{x} - mean value</p>	$s_a(x_i) = k_a s(x_i) \quad \dots\dots(6)$ <p>where: $k_a = \sqrt{\frac{n_{\text{eff}}(n-1)}{n(n_{\text{eff}}-1)}} \approx 1 \quad \dots\dots(6a)$</p>
Standard deviation of Mean value \bar{x}	$s(\bar{x}) = \frac{s(x_i)}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}$ <p>$\equiv u_A$</p>	$s_a(\bar{x}) = \frac{s_a(x_i)}{\sqrt{n_{\text{eff}}}} = k_b s(\bar{x}) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n_{\text{eff}}-1)}} \quad \dots\dots(7)$ <p>where: $k_b = \sqrt{\frac{n-1}{n_{\text{eff}}-1}} \quad \dots\dots(7a)$</p>
Number of freedom	$v=n-1$	$v_{\text{eff}} \equiv \frac{n}{1 + 2 \sum_{k=1}^{n-1} \rho_k^2} - 1 \quad \dots\dots(8)$
Related standard dispersion of standard deviation	$\frac{u(s)}{s} = \frac{u(s(\bar{x}))}{s(\bar{x})} \equiv \frac{1}{\sqrt{2v}}$	$\frac{u(s_a)}{s_a} = \frac{u(s_a(\bar{x}))}{s_a(\bar{x})} \equiv \frac{1}{\sqrt{2v_{\text{eff}}}} \quad \dots\dots(9)$
Expanded uncertainty for probability p (coefficient k_p)	$U = k_p u = k_p \sqrt{u_A^2 + u_B^2}$	$U = k_p u = k_p \sqrt{s_a^2 + u_B^2} \quad \dots\dots(10)$

4 Estimator of the Autocorrelation Function of Measurement Data Sample

The autocorrelation function is usually not known and needs to be estimated from the measurement data. The most commonly used and implemented in computer programs has form

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{s^2(q_i)} \quad (11)$$

Estimate r_k (Fig. 2a) has two qualitatively different parts. For small distances k is the falling edge, in which contains real information about the autocorrelation function. The remainder tail is the image of a rather large fluctuations of the correlated noise.

According Zieba [4, 5], the replacement of the function ρ_k in formula (3) by its estimate r_k gives an estimate of the effective number of observations n_{eff} not of the satisfied properties. The reason is the influence of the tail of autocorrelation function. Zięba proposed to reduce summation in (11) to only a few initial estimates of r_k elements, i.e.

$$\hat{n}_{\text{eff}} = \frac{n}{1 + 2 \sum_{k=1}^{n_c} \left(1 - \frac{k}{n}\right) r_k} \quad (12)$$

The border value n_c is determined by the last non-zero element of the r_k estimate before its first passage through zero (FTZ method – so called from the first transit through zero). For example, this value for the curve on Figure 2a is $n_c=3$. FTZ method is valid only for positive correlations.

Figure 2b [6, 4] shows two examples of probability distributions of the estimators obtained by Monte Carlo method. The distribution marked as teor on figure 2.b is calculated from the formula for uncorrelated observations of the standard deviation $z=s/\sigma$ resulting from the distribution of χ^2 , where v was substituted by the effective number of degrees of freedom v_{eff} .

The simulation studies using MC [5] show that in this case estimator \hat{n}_{eff} reduce the negative bias of the mean value \bar{x} . Obtained value of \hat{n}_{eff} is used for calculations of $s_a(x_i)$ and $s_a(\bar{x})$ by (6) and (7).

Example 2

Let us calculate the standard deviation of single measurement $s_a(q_i)$ and of the mean value $s_a(\bar{q})$ of 120 data of observations q_i from Example 1 after withdrawal trend.

The estimator r_k of their autocorrelation function is taken for $n_c=3$ according to Figure 2a. The large obtained value $r_k=0.81$ confirms autocorrelation of these data. Standard deviation without considering the autocorrelation is $s(q_i)=0.0241$. At this value of r_k the formula (12) implies the estimate $n_{\text{eff}} = 32.1$ and from (6a) and (7a) the coefficient

$k_a=1.012$ and $k_b=1.96$. Correlation do not significantly affect the $s_a(q)=0.0244$, and much more the u_A , which increases about 2 times from $s(\bar{q})=0.00219$ to $s_a(\bar{q})=0.00472$. From (9) and (10) is $u(s_a)/s \approx 11\%$.

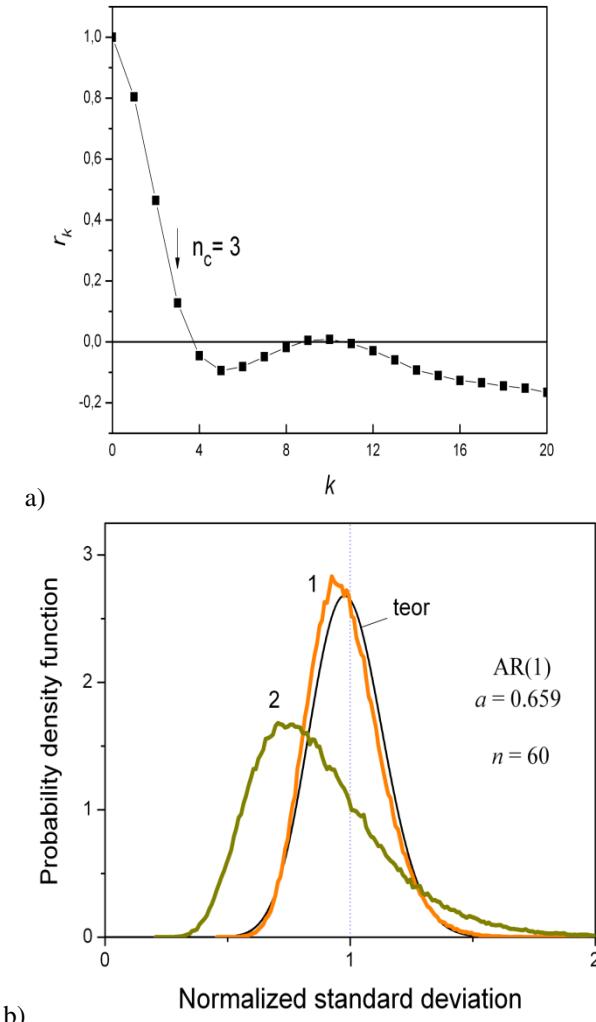


Fig. 2. a). The initial part of the estimate $\{r_k\}$ of autocorrelation function ρ_k computed from data of Figure 1a (after removing the trend from the raw data); b). Probability density functions for normalized estimators of the standard deviation by the first order autoregressive model AR(1)¹ and the random sample size $n=60$ [5, 6]. Curves 1 and 2 are derived from MC simulations and relate, respectively, s_a / σ and $s_a(\bar{x}) / \sigma(\bar{x})$. Theoretical curve 'teor' has been calculated theoretically using the model (9) for $v_{\text{eff}} = 22.7$.

¹ Model AR (1) - autoregressive time series of the first order, for which $r_k = a^k$ [4].

5 Summary and Conclusions

Discussed issues aroused from the purpose of determining the standard deviation and uncertainty for the measurement data obtained by sampling. This paper summarizes the research results of cleaning such raw data from their regular components [3], and gives the proper method of calculation the standard deviation for the sample of auto-correlated data. The developed method is easy to use and allows to extend the range of application of GUM method type A of estimation uncertainty to such data [1, 2, 4-7, 9, 10].

Conclusions:

- Before calculating the standard deviation of the sample of measurement observations as the uncertainty u_A appropriate computational methods must first be used to identify and remove from the raw results the systematic (non periodic and periodic) components.
- For such a randomly distributed values of the observations one needs to know or estimate their autocorrelation function. This function causes a significant increase in uncertainty u_A compared to the calculated according to GUM. It corresponds to the lower effective number of independent measurements to be taken into account in estimating the standard uncertainty. The content specified adjusted formulas.
- For a limited time to collect measurement observations, a reduction of their sample SD as measurement uncertainty by increasing the sample size (by increasing the sampling rate) is unreliable, because it leads to the necessity to reflect the impact of autocorrelation function of observations.
- Evaluation of the uncertainty u_A for autocorrelated data discussed here, is valid, as in GUM, for the model of normal distribution. For other distributions, this method requires further investigation.
- Programs for the calculation of uncertainty should be supplemented by algorithms for the identification and elimination from the "raw" data the regular components and to obtain estimators of the autocorrelation function from such cleaned measured data.
- Presented method allows to upgrade the GUM recommendations by include the proper calculation of the standard and expanded uncertainty for measurand with the stationary random component of their signal.

References

1. Dorozhovets, M., Warsza, Z.: Methods of upgrading the uncertainty..., Part 2 Elimination of the influence of autocorrelation of observations and choosing the adequate distribution. In: Proceedings of 15th IMEKO TC4 Symposium, Part 1, Iasi Romania, pp. 199–204 (September 2007)
2. Warsza, Z.L., Dorozhovets, M.: Uncertainty type A evaluation of autocorrelated measurement observations. Biuletyn WAT (Military Technical Academy) LVII(2), 2143–2152 (2008)

3. Warsza, Z.L., Dorozhovets, M., Korczynski, M.J.: Methods of upgrading the uncertainty of type A evaluation (1). Elimination the influence of unknown drift and harmonic components. In: Proceedings of 15th IMEKO TC4 Symposium, Part 1, Iasi Romania, pp. 1193–1198 (September 2007)
4. Zięba, A.: Effective number of observations and unbiased estimators of variance for auto-correlated data – an overview. *Metrology & Measurement Systems* 17, 3–16 (2010)
5. Zięba, A., Ramza, P.: Standard deviation of the mean of autocorrelated observations estimated with the use of the autocorrelation function estimated from the data. *Metrology & Measurement Systems* 18, 529–542 (2011)
6. Warsza, Z.L., Zięba, A.: Niepewność typu A pomiaru o obserwacjach samoskorelowanych (Uncertainty type A in measurements of autocorrelated observations). *Pomiary Automatyka Kontrola* 58(2), 157–161 (2012)
7. Warsza, Z.L.: Evaluation of the type A uncertainty in measurements with autocorrelated observations. In: Proceedings of Jointed TC 1 - TC 7 Imeko Symposium, Genua (2013) (in print)
8. Guide to the Expression of Uncertainty in Measurement, revised and corrected, BIPM 2008 (2008)
9. Dorozhovets, M., Warsza, Z.L.: Propozycje rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (1) Uwzględnianie wpływu autokorelacji i nieadekwatności rozkładu wyników obserwacji w niepewności typu A. *Pomiary Automatyka Robotyka* 11(1), 6–15 (2007) (in Polish)
10. Dorozhovets, M., Warsza, Z.L.: Propozycje rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (2) Uściślenie metod obliczeń niepewności typu B. *Pomiary Automatyka Robotyka* 11(2), 45–52 (2007) (in Polish)

Author Index

- Adamczyk, Mirosław 3
Arent, Krzysztof 353, 497
- Bocewicz, Grzegorz 11
Banaszak, Zbigniew 11
Bazydło, Piotr 567
Będkowski, Janusz 441
Belter, Dominik 343
Bieńkowski, Adam 607, 665
Broel-Plater, Bogdan 21
Budny, Eugeniusz 265, 687
Busłowicz, Mikołaj 33, 45
- Cholewiński, Mateusz 353
Ciuk, Tymoteusz 631
Cupriak, Małgorzata 55
Czajka, Piotr 579
Czarnowski, Jan 365
Czyżewski, Adam 579
- Dąbek, Przemysław 507
Dąbrowski, Szymon 649
Domański, Kamil 649
- Fraś, Jan 365
Frydrych, Piotr 591
- Gąsior, Przemysław 377
Galas, Jacek 579
Gardecki, Stanisław 377, 387
Giernacki, Wojciech 377, 387
Gińko, Oskar 697
Główka, Jakub 365, 475
Gosiewski, Łukasz 713
Gośliński, Jarosław 377, 387, 623
- Goszczyński, Tadeusz 597
Grabowski, Dariusz 409
Granosik, Grzegorz 553
Gruszecka, Magdalena 665
- Hendzel, Zenon 507
Henzel, Maciej 155
Hetmańczyk, Mariusz Piotr 63, 73
- Jackiewicz, Dorota 607, 665
Jamro, Marcin 81, 91
Janiak, Mariusz 497
Jasiński, Sławomir 55
Juszkiewicz, Łukasz 497
- Kachniarz, Maciej 607
Kacprzyk, Janusz 507
Kaczmarek, Piotr 441
Kaczorek, Tadeusz 103
Kaliczyńska, Małgorzata 55
Kamiński, Marcin 567
Kasprzak, Włodzimierz 399, 421
Klimasara, Wojciech J. 409
Kłos, Sławomir 115
Kochanowski, Maciej 579
Kociszewski, Rafał 127
Konowrocki, Robert 205
Korczyński, Jerzy M. 721
Kornuta, Tomasz 399, 421, 543
Kowalski, Grzegorz 431
Kozłowski, Artur 677
Kryca, Marek 677
Kuna-Broniowski, Marek 615
- Ławryńczuk, Maciej 329
Litwin, Dariusz 579

- Maciąś, Mateusz 365, 431, 475
 Majek, Karol 441
 Makarewicz, Adam 33
 Makarski, Piotr 615
 Malec, Marcin 453
 Marusak, Piotr M. 137
 Mateusiak, Piotr 147
 Mazur, Alicja 353
 Mazurek, Paulina 155
 Milecki, Andrzej 3, 167
 Missala, Tadeusz 175, 265, 687
 Mizak, Wojciech 579
 Morawski, Marcin 453
 Mrozek, Bogumiła 187
 Musialik, Paweł 441
 Nowicki, Michał 631, 641, 665
 Owczarek, Piotr 167, 623
 Owczarkowski, Adam 623
 Pasternak, Iwona 631
 Paszkiel, Szczepan 197
 Pawlewski, Paweł 11
 Petruk, Oleg 631, 641, 649
 Pochanke, Andrzej 205
 Pręgowska, Agnieszka 205
 Rokosz, Tomasz 475
 Rostkowska, Marta 463
 Ruszewski, Andrzej 217
 Rybarczyk, Dominik 167, 623
 Rzońca, Dariusz 81, 229
 Sadolewski, Jan 81, 229
 Safinowski, Marcin 649
 Sajewski, Łukasz 239
 Salach, Jacek 607, 631, 641, 665
 Sitek, Paweł 251
 Skrzypczyński, Piotr 463
 Sobierajski, Waldemar 677
 Socjusz, Maciej 579
 Sprońska, Agnieszka 475
 Stec, Andrzej 81, 295
 Stefańczyk, Maciej 421, 487
 Strupiński, Włodzimierz 631
 Świdler, Jerzy 63, 73
 Świdler, Zbigniew 81, 295
 Szałatkiewicz, Jakub 265, 687
 Szewczyk, Roman 147, 265, 275, 567, 591,
 607, 631, 641, 649, 687, 697, 713
 Szolc, Tomasz 205
 Szpakowska-Peas, Ewelina 285
 Tchoń, Krzysztof 497
 Topolski, Michał 463
 Trojnacki, Maciej 507
 Trybus, Bartosz 81, 229
 Trybus, Leszek 81, 295
 Trzasko, Wojciech 307
 Trzcinka, Krzysztof 631, 649, 697
 Typiak, Rafał 523
 Ugodziński, Robert 713
 Urbański, Michał 567
 Walas, Krzysztof 343
 Wałęcki, Michał 487, 533
 Warsza, Zygmunt L. 721, 733
 Wikarek, Jarosław 317
 Winiarski, Tomasz 533
 Winiarski, Wojciech 265, 631, 649, 687
 Wołoszczuk, Adam 431
 Wszołek, Grzegorz 63, 73
 Wysocki, Antoni 329
 Zieliński, Cezary 399, 543
 Zubrycki, Igor 553