



FACULTY OF ELECTRICAL ENGINEERING
UNIVERSITY OF BANJA LUKA

ELECTRONICS

VOLUME 15, NUMBER 1, JUNE 2011

FACULTY OF ELECTRICAL ENGINEERING UNIVERSITY OF BANJA LUKA

Address: Patre 5, 78000 Banja Luka, Bosnia and Herzegovina

Phone: +387 51 211824

Fax: +387 51 211408

ELECTRONICS

Web: www.electronics.etfbl.net

E-mail: electronics@etfbl.net

Editor-in-Chief:

Branko L. Dokić, Ph. D.
Faculty of Electrical Engineering
University of Banja Luka, Bosnia and Herzegovina
E-mail: bdokic@etfbl.net

International Editorial Board:

- Goce Arsov, St. Cyril and Methodius University, Skopje, Republic of Macedonia
- Petar Biljanović, University of Zagreb, Croatia
- Milorad Božić, University of Banja Luka, Bosnia and Herzegovina
- Đemal Kolonić, University of Banja Luka, Bosnia and Herzegovina
- Vladimir Katić, University of Novi Sad, Serbia
- Vančo Litovski, University of Niš, Serbia
- Danilo Mandić, Imperial College, London, United Kingdom
- Vojin Oklobdžija, University of Texas at Austin, USA
- Zorica Pantić, Wentworth Institute of Technology, Boston, USA
- Aleksandra Smiljanić, University of Belgrade, Serbia
- Slobodan Vukosavić, University of Belgrade, Serbia
- Volker Zerbe, Technical University of Ilmenau, Germany
- Mark Zwoliński, University of Southampton, United Kingdom
- Deška Markova, Technical University of Gabrovo
- Chakresh Kumar, Galgotias College of Engineering and Technology, Greater Noida, India

Secretary:

Petar Matić, M.Sc.
E-mail: pero@etfbl.net

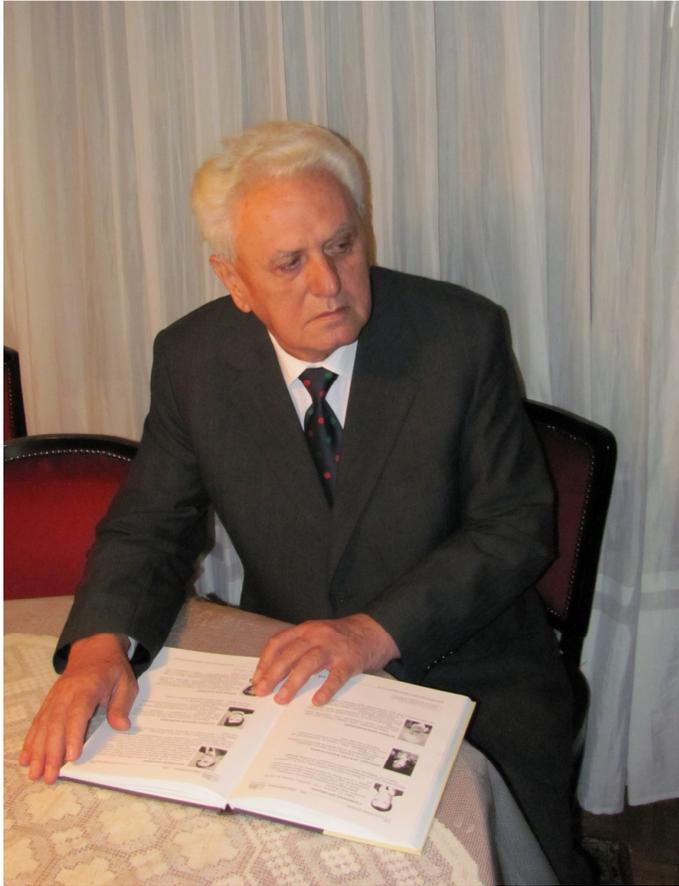
Mladen Knežić, M.Sc.
E-mail: mladen_knezic@etfbl.net

Željko Ivanović, M.Sc.
E-mail: zeljko@etfbl.net

Publisher:

Faculty of Electrical Engineering
University of Banja Luka, Bosnia and Herzegovina
Address: Patre 5, 78000 Banja Luka, Bosnia and Herzegovina
Phone: + 387 51 211824
Fax: + 387 51 211408
Web: www.etfbl.net

Prof. dr Milić Stojić Biography



Milić Stojić, Full Professor in Electrical Engineering, University of Belgrade, was born in February 27, 1940 in Užice, by father Radoje and mother Kosara, born Cerović. He completed his elementary and grammar school in Užice. In the academic 1958/59 year he enrolled the Faculty of Electrical Engineering, University of Belgrade, where he graduated in January 1963. In 1962 he spent two months in London at *Telephone and Cables Company* under the IASTE (International Association of the Exchange of Students in Technical Experience). In October 1963, he enrolled the postgraduate studies at the Electrical Engineering Faculty, where he obtained his M.Sc. degree by defending his M.Sc. thesis entitled *An Approach to Solving the Sensitivity Problem in Control Systems*.

From the University of Belgrade, he obtained his Ph.D. degree in 1967 with the thesis entitled *Analysis and Design of Nonlinear Control Systems Utilizing the Algebraic Criteria of Asymptotic Stability* defended also at the Faculty of Electrical Engineering. The main results of the Ph.D. thesis were published in 1969 in two long papers in the renowned scientific journal *Avtomatica and Telemekhanika*, edited by the Soviet Academy of Science and later published in U.S.A. as a cover-to-cover English translation in *Automation and Remote Control*. An abridged version of the thesis was also

published in Russian as a monograph, edited by the *Publications of Electrical Engineering Faculty* in Belgrade.

After graduation, from February to October 1963, he was with the Institute for Nuclear Science in Vinča, Belgrade. Since October 1963, he is with the Faculty of Electrical Engineering, University of Belgrade, where he has been promoted to the academic positions Teaching Assistant, Assistant Professor, Associate Professor, and Full Professor in 1963, 1969, 1975, and 1980, respectively. He taught courses in automatic control at the undergraduate and postgraduate studies at the electrical engineering faculties of universities in Belgrade, Niš, Banja Luka, and East Sarajevo (Republic of Srpska), at the Faculty of Technical Science in Novi Sad and at the faculties of Mechanical Engineering in Belgrade and Kragujevac. Thirty eight academic years he taught exceptional courses at the High Military Academy in Žarkovo (Belgrade). In 1980, 1984, and 1988, he spent several months at the Physics Department, University of Birmingham (UK), in the frame of joint research program. He delivered numerous invited lectures at the conferences, universities, and scientific institutes in former Yugoslavia and abroad: at the Manchester University (UK), Moscow Aviation Institute, and Institute of Control Science (Moscow). He also delivered lectures at the *Autumn School in Computational Intelligence and Information Technologies* held at the Faculty of Electronics in Niš.

At the Faculty of Electrical Engineering, in three two-year election periods he was the head of the Department of Automation and in two four-year election period he was the head of the Division of Electronics, Telecommunication, and Automation. In one four-year election period, he was the head of the Council of Donor Services at the Republic Assembly of Science. Over ten years he was the head of the Council of Experts in Electrical Engineering at the University of Belgrade. In one four-year election period, he was the member of Administrative Committee of the Institute for Nuclear Science "Vinča".

As a head of program committee, he organized over 20 conferences of the Yugoslav Society of ETAN (Electronics, Telecommunication, Automation and Nuclear Science), at different centers all over former Yugoslavia. At the beginning of civil warfare in Yugoslavia in 1991, the ETAN ended his existence. Then Professor Stojić founded the new Society of ETRAN (Electronics, Telecommunication, Computers, Automation and Nuclear Science) as a successor of previous Society of ETAN. In two four-year election periods, from 1992 to 2001, he was the president of ETRAN. Now he is honorary member of the ETRAN Presidency. He is full member of the Serbian Scientific Society and up to 2002 he was the secretary of its Technical Department. Professor Stojić is the full member of the Engineering Academy of Serbia and Montenegro.

Professor Stojić is married, his wife Radmila, sons Radoje and Đorđe, and daughter Borjana are all graduated electrical

engineers. He has two granddaughters Milena (13) and Milica (5).

Professor Stojić was the Associate Editor of *Automatica*, the Journal of IFAC (International Federation of Automatic Control) and the Editor-in-Chief of *Publications of Electrical Engineering Faculty*, Series: *Electronics, Telecommunication, and Automation*, the journal of Electrical Engineering Faculty in Belgrade. At the same time as he was the secretary of the Serbian Scientific Society, he was the Editor-in-Chief of its journal *Scientific Review: Science and Engineering*. At the present, he is the member of the editorial boards of *Facta Universitatis*, Series: *Electronics and Energetics*, journal of University of Niš, *Tehnika, Vojnotehnički Glasnik*, and guest editor of *Electronics*, journal of the Electrical Engineering Faculty in Banja Luka (Republic of Srpska).

For his outstanding merits on the fields of science, education, and results achieved in education of professionals and scientists, he was awarded with the Gold Wreath Labor Medal, Charter of Medallion of University of Niš, and Great Charter and Medallion of ETRAN.

There are many awards that Professor Stojić received for his outstanding scientific and professional contributions in his papers published in scientific journals or presented at international and domestic conferences. For these activities he also received a number of prizes: October Reward of the City of Belgrade for mathematics and technical science, Award of the International Conference of IFIP (International Federation of Information Processing), together with Professor Stanoje Bingulac, for the best paper presented at the conference, three awards "Professor Dušan Mitrović" for the best papers presented at the ETRAN conferences. Together with Professor Slobodan Vukosavić, he received the Award "Professor Branko Raković" for the best papers published in international scientific journals.

For outstanding engineering achievements, he received the Decoration of Military Merits from Marshal Tito, president of Yugoslavia, and Tesla's Award for Scientific Merit (together with Professor Slobodan Vukosavić).

Professor Stojić was very active in cooperation with other faculties, institutes, and professional organizations (Electronic Faculty of Niš, JUREMA in Zagreb, High Military School in Belgrade, Institute of Nuclear Science in Belgrade, ETRAN, TELSIS, INDEL, INFOTEX, ...) from which he obtained a number of charters, medallions, and letters of thanks.

The main scientific contributions of Professor Stojić lie in the stability theory, sensitivity analysis of dynamic systems, system simulation, and microcomputer-based real-time control of electrical drives and industrial processes. He has published two textbooks; he is author/co-author of two monograph in

Serbian, co-author of one monograph in English, and author of one monograph in Russian. He has published 194 scientific papers, among them 26 in renowned international journals: *IEEE Transactions on Automatic Control*, *Avtomatika I Telemekhanika*, edited by the Academy of Science USSR, *International Journal on Circuit Theory and Applications*, *Automatica*, *International Journal of IFAC*, *International Journal on System Science*, *IEEE Transactions on Industrial Electronics*, *IEEE Transactions on Power Electronics*, *IEEE Transactions on Education*, and *Facta Universitatis*, Series: *Electronics and Energetics*. Other papers have been published in domestic journals (40), in the proceedings of conferences in Yugoslavia (94), and in the proceedings of international conferences (34). Several hundred times he has been cited in the papers, books, M.Sc and Ph.D thesis of Yugoslav authors. According to the Science Citation Index, he has been cited in the international journal and books 117 times.

He is the author of the book *Continuous Control Systems* (ten editions) and *Digital Control Systems* (five editions), which are used as standard textbooks of automatic control at universities of Yugoslavia. Although the textbooks are written in Serbian, they were reviewed in the Russian Reporting Journal *Новые книги за рубежом* (New Books in Abroad), issued by the Soviet Academy of Science, by Professor Yakov Zalmanovich Tsytkin. In the review, Professor Ya.Y. Tsytkin wrote: Эта книга отличается ясностью и лаконичностью выражения (This book is distinguished by clear and laconic expressivity).

Professor Stojić gave numerous technical solutions within projects carried out by the Faculty of Electrical Engineering for needs of the industry, scientific institutes, and Yugoslav Army. In three five-year fiscal periods he headed the strategic project *Electronic Systems of Measuring and Control* supported by the Ministry of Science and Technology of Serbia. He designed the Laboratory of Automatic Control at the Faculty of Electrical Engineering. Together with Professor Dejan Živković, he designed the Laboratory of Control Systems in the Air Force at the High Military School in Žarkovo (Belgrade). For the project of electronic simulator of guided missile, he was awarded the Yugoslav Decoration for Military Merits, and for the project of microprocessor-based positioning servomechanism with induction motor he obtained, together with Professor Slobodan Vukosavić, Tesla's Award for the prominent scientific and professional merit.

He supervised over 20 M.Sc. and 16 Ph.D. thesis: 12 of them at the Faculty of Electrical Engineering in Belgrade and by one at electrical engineering faculties in Niš and Zagreb, Technical Faculty in Čačak, and Faculty of Mechanical Engineering in Kragujevac.

Editorial

IT is an honor and pleasure to introduce you to the special edition of the journal Electronics. This edition is dedicated to a great scientist and a university professor, PhD Milić Stojić, regarding his 70th birthday.

For this occasion I have invited eight prominent scientists and university professors, associates and professor Stojic's PhD students to write author papers. On the following pages there are papers of professor S. Vukosavic, professor M. Naumovic, professor P. Maric, PhD V. Arandjelovic and others. Among invited papers, there is a paper named "Extraction of External Torque Disturbance in Positioning Servomechanism" of professor Milic Stojic. Papers of authors S. S. Cvetkovic M. Srbinovska and Z. Kokolanski are regular papers.

Unfortunately, I was not a student of professor Stojić, nor did we have the same scientific interests. Nevertheless, I was fortunate enough to be his associate in the Presidency of the ETRAN and it was a great pleasure to have him in the Programme Committee of the INDEL and the Editorial Board of the journal Electronics.

In 1997, as president of the ETRAN, he encouraged me and gave me his generous support to start a symposium on the subject of industrial electronics, INDEL, and to start publishing a journal, Electronics, at the Faculty of Electrical Engineering in Banja Luka. Thanks to him the INDEL and Electronics have become internationally recognized and that is his invaluable contribution.

I owe gratitude to professor Stojić for learning that even the greatest scientific truths can be simplified enough, so that they become understandable to those who are not even remotely familiar with them. Since the science is in the service of truth, professor Stojić, using his approach to science and his interpretation of science, confirms that every truth is simple, only the path to it is difficult.

About the relationship of professor Milić Stojić to his students and associates speaks his former student, today a prominent scientist and a university professor, PhD Radomir S. Stanković:

"Whenever I need some advice in my learning or in professional work, the first to be asked is professor Stojic. Whenever I have asked, it was possible to meet and talk with him at any time and at any place, including his office, some conference venue, his home, or even his garden in Grocka. We

asked him not just about scientific or professional things, but also about everything that we wanted to understand and learn, or simply to interpret or anticipate, at the moment.

As a member of the Pen Club, and at the same time having a perfect command of Russian, professor Stojic has always been an excellent advisor in communicating with either Western or Eastern European research communities – a master to find proper wording, tailor sentences, or select the most appropriate way of addressing.

When we asked about some scientific topics, the answers were hardly ever given directly, but rather in terms of a series of simple counter questions. By answering them we were unnoticeably guided towards a discovery of the right answer on what we initially asked, leaving us a pleasant feeling that we found eventually the solution by ourselves.

When asking for advices about professional work considering teaching or similar topics, or simply when discussing various issues in the everyday life, the advices or answers were usually given in terms of some anecdotes from history, mostly the Serbian history, or quoting details from biographies of important scholars or statesman. The comments were given in such a manner that we had to think twice if we shall take them straightforwardly or in the completely opposite sense, was that what he said a straightforward strong comment, a bitter irony, or just an innocent joke. This way of talking and working with younger researchers is my main impression of working with and learning from professor Stojic.

If I was asked what professor Stojic was teaching us, his students, I would shortly say, everything – meaning science, professional work, such as writing, publishing, teaching, communicating in the research community, or shortly, professionalism, and simply, the life."

It is a custom to put the editorial of the chief editor at the beginning, because the editorial announces what follows. Nevertheless, this time I am making an exception by putting the biography of PhD Milić Stojić at the beginning, because it does not need any announcements. To professor Stojić I wish many years of scientific activity and to be a guest editor of the journal Electronics at least a few more times.

Ph.D. Branko Dokić, full professor
Editor-in-Chief

Extraction of External Torque Disturbance in Positioning Servomechanism

Milić R. Stojić and Đorđe M. Stojić

Abstract—This paper presents the design of the digitally-controlled positioning servomechanism whose steady state target position is unaffected by arbitrary class of load torque disturbances. For rejection of the torque disturbance effects on the steady state value of motor target angular position, the IMPACT (Internal Model Principle and Control Together) controlling structure is proposed. As an example, the servomechanism is considered in which the output drive torque is produced by using IFOS (Indirect Field Oriented Control) of an induction motor. Some improvements of the controlling structure are proposed in order to eliminate the ringing of control variable and thus to enable easier physical realization of the positioning servomechanism. The theoretical results are confirmed by simulation.

Index Terms—Positioning servomechanism, Load torque disturbance, IMPACT controlling structure.

I. INTRODUCTION

ONE of the key tasks in the design of feedback control systems is to eliminate, or to suppress as much as possible, the influence of external immeasurable disturbances on the steady state value of controlled variable (system output). To perform this task, a number of various controlling structures has been proposed, which can be classified into two main groups: (i) the control schemes that use the IMP (Internal Model Principle), and (ii) the schemes using IMC (Internal Model Control), which are often referred to as the pseudo inverse control.

The IMP based solutions have been proposed in a number of papers [1-5]. Similar task has been performed by using the disturbance observer [6] and a “phase-locked loop” structure [7], for extraction of sinusoidal disturbances. Starting from the idea of Ya.Z. Tsytkin [4], the authors of the survey paper [8] gave the review of previous results based on the application of IMP, focusing on the development of adaptive control for the case of immeasurable time-varying frequency narrow disturbances applied on an active suspension. As a special case, variable frequency sinusoidal disturbances were considered.

A new disturbance observer (DOB) is proposed in [9], is

based on the IMC with the improved performance by modifying the original DOB structure. Furthermore, the proposed observer is adapted to the extended task space formulation. In [10], the state and disturber observer algorithm for linear time-invariant systems is proposed, with the disturbance estimation treated as a plant inversion problem. Since the inverse of a physical system is usually noncausal, the derivatives of output signals, causing that the accurate values of output derivatives become difficult to obtain, which is a common problem for numerous output feedback-disturbance observers. In [10], the tuning parameter is introduced, which can be adjusted to reduce the effects of measuring noise.

Although the disturbance observer, IMP, and IMC based solutions effectively eliminate the influence of immeasurable loads, in general, to a certain degree they suffer from the increased sensitivity to the measuring noise in the feedback signal. In these cases, the problem of measuring noise cannot be completely solved by filtering the feedback signal, since the inclusion of a low-pass digital filter causes an error in the disturbance estimation and rejection of the immeasurable load influence from the feedback control.

In this paper, the IMPACT controlling structure is applied for design of a digitally-controlled positioning servomechanism. The proposed structure eliminates the effects of arbitrary load torque disturbances on the steady state value of the motor target angular position, while keeping the dynamics of the adopted position control loop almost intact. This task is completed by using suitable modification of the original IMPACT structure proposed in [4]. The paper is organized as follows. After the Introduction in Section 1, In Section 2, the IMPACT controlling structure, adopted for the position control of electrical drive with an induction motor, is described. It will be shown that the proposed structure enables that the desired continuous-time set-point transient response of closed-loop system and rejection of load torque disturbance are achieved independently. Section 3 gives a set of simulation results that illustrates the efficiency of the proposed IMPACT structure in rejecting of three typical disturbances from the steady state value of the target angular position of the drive. Section IV considers the problem of ringing of control variable.

Milić R. Stojić is with the Faculty of Electrical Engineering, Belgrade, Serbia (e-mail: estojic@etf.rs).

Đorđe M. Stojić is with the Institute of Electrical Engineering, Belgrade, Serbia (e-mail: djolestojic@yahoo.com).

II. IMPACT STRUCTURE OF POSITIONING SERVOMECHANISM

Fig. 1 shows the IMPACT structure primarily proposed by Ya. Z. Tsytkin [4]. Actually, Fig. 1 represents the original IMPACT structure modified for application to the design of position-controlled servo drive with induction motor. Likewise, the modification of general IMPACT structure suitable for control of disturbance invariant processes having large dead-times has been proposed in [11]. In [12], similar control structure has been used in the design and realization of speed-controlled electrical drives.

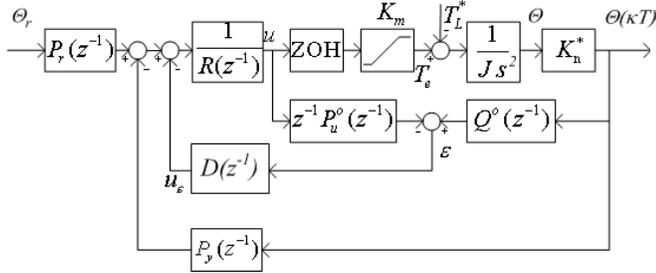


Fig. 1. IMPACT controlling structure of digitally-controlled positioning servomechanism.

The control portion of the structure comprises: two-input nominal plant model, control polynomials $P_r(z^{-1})$, $P_y(z^{-1})$, $R(z^{-1})$, and prediction polynomial $D(z^{-1})$ included into the minor local feedback loop of the structure. The nominal model of the plant consists of the vector-controlled torque driven induction motor and counting-type shaft encoder measuring the motor speed [13]. The encoder generates feedback variable

$$\theta^*(kT) = \frac{K_n}{2\pi} \theta(t) = K_n^* \theta(t) \quad (1)$$

where $\theta(t)$ denotes shaft angular position in radians, T is the sampling period, and K_n is the total number of quantum marks on the encoder disc. Notice that the control portion of the structure in Fig. 1 includes both the two-input nominal plant model explicitly and the model of external disturbance embedded implicitly into the prediction polynomial $D(z^{-1})$.

The zero-hold equivalent nominal model of the plant comprising the vector-controlled induction motor and shaft encoder may be approximated, in the linear regime, by

$$W^0(z^{-1}) = \frac{\theta^*(z^{-1})}{u(z^{-1})} = Z \left[\frac{K_m K_n^* (1 - e^{-Ts})}{J s^3} \right] \quad (2)$$

wherefrom one obtains

$$W^0(z^{-1}) = \frac{z^{-1} P_u^0(z^{-1})}{Q_0(z^{-1})} = C_m \frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^2} \quad (3)$$

where $C_m = K_m K_n^* T^2 / 2J$ is the synthetic plant parameter and K_m , K_n^* , and J denote respectively the electromagnetic torque coefficient, number of quantum marks per radian, and motor inertia. Hence, in the IMPACT structure of Fig. 1, $z^{-1} P_u^0(z^{-1}) = C_m z^{-1}(1+z^{-1})$ and $Q^0(z^{-1}) = (1-z^{-1})^2$.

For minimal phase plants, the proper choice of polynomial $R(z^{-1})$ is $R(z^{-1}) = P_u^0(z^{-1}) = C_m(1+z^{-1})$ [11, 12]. In the nominal case, $P_u(z^{-1}) = P_u^0(z^{-1})$, $Q(z^{-1}) = Q^0(z^{-1})$ and for $R(z^{-1}) = P_u^0(z^{-1})$, the closed-loop transfer function $\theta^*(z^{-1}) / \theta_r(z^{-1})$ is easily derived from Fig. 1 as

$$\frac{\theta^*(z^{-1})}{\theta_r(z^{-1})} = \frac{z^{-1} P_r(z^{-1})}{Q^0(z^{-1}) + z^{-1} P_y(z^{-1})}. \quad (4)$$

After approximation of torque disturbance $T_L(t)$ by $T_L^*(t) = T_L(t)$ for $kT \leq t < (k+1)T$, $k = 0, 1, 2, \dots$, the system closed-loop transfer function $\theta^*(z^{-1}) / T_L^*(z^{-1})$ becomes

$$\frac{\theta^*(z^{-1})}{T_L^*(z^{-1})} = \frac{Q^0(z^{-1}) [1 - z^{-1} D(z^{-1})]}{Q^0(z^{-1}) + z^{-1} P_y(z^{-1})} W_L(z^{-1}) \quad (5)$$

where

$$W_L(z^{-1}) = Z \left[\frac{K_n^*}{J} \frac{1 - e^{-Ts}}{s^3} \right] = \frac{K_n^* T^2}{2J} \frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^2} \quad (6)$$

A. Rejection of disturbance

From Eqs. (5) and (6) and $Q^0(z^{-1}) = (1-z^{-1})^2$, the steady-state error in the presence of a known class of external disturbances $T_L^*(t)$ will become zero, after a certain finite number of sampling periods, i.e.,

$$\lim_{z \rightarrow 1} (1-z^{-1}) \frac{[1 - z^{-1} D(z^{-1})]}{Q^0(z^{-1}) + z^{-1} P_y(z^{-1})} \frac{K_n^* T^2 z^{-1}(1+z^{-1})}{2J} T_L^*(z^{-1}) = 0 \quad (7)$$

Since

$$\lim_{z \rightarrow 1} \frac{1}{Q^0(z^{-1}) + z^{-1} P_y(z^{-1})} \frac{K_n^* T^2 z^{-1}(1+z^{-1})}{2J} \neq 0 \quad (8)$$

equation (7) is satisfied if

$$\lim_{z \rightarrow 1} (1-z^{-1}) [1 - z^{-1} D(z^{-1})] T_L^*(z^{-1}) = 0. \quad (9)$$

Suppose that the class of load torque disturbances is known

and given by its z -transform $T_L^*(z^{-1}) = A(z^{-1})/B(z^{-1})$. Then Eq. (9) is satisfied and consequently the disturbance is completely rejected in the steady-state if

$$1 - z^{-1}D(z^{-1}) = B(z^{-1}) \quad (10)$$

wherefrom one obtains the prediction polynomial

$$D(z^{-1}) = \frac{1 - B(z^{-1})}{z^{-1}}. \quad (11)$$

For example, for the constant, ramp, parabolic, and sinusoidal ($T_L(t) = \sin \omega t$) torque disturbances, the denominator of disturbance model $B(z^{-1})$ equals $1 - z^{-1}$, $(1 - z^{-1})^2$, $(1 - z^{-1})^3$, and $1 - 2z^{-1} \cos \omega T + z^{-2}$, respectively. Furthermore, the disturbance polynomial $B(z^{-1})$ can also be determined for the cases of more complicated disturbances. For example, for the composed disturbance of superposed ramp and sinusoidal signals, the disturbance model $B(z^{-1}) = (1 - z^{-1})^2(1 - 2z^{-1} \cos \omega T + z^{-2})$ should be used for the IMPACT structure design. In such way, for any kind of single or more complicated class of disturbances, the corresponding prediction polynomial $D(z^{-1})$ can be immediately determined by using Eq. (11). Nevertheless, torque disturbance $T_L(t)$ is usually slow varying and therefore the implementation of disturbance polynomial $B(z^{-1}) = (1 - z^{-1})^2$ or the corresponding prediction polynomial $D(z^{-1}) = 2 - z^{-1}$, which corresponds to extraction of ramp disturbances, will effectively reject the influence of torque disturbance on the steady state value of the angular target position of the motor shaft. Moreover, the use of $B(z^{-1}) = (1 - z^{-1})^2$ for calculation of prediction polynomial $D(z^{-1})$ by Eq. (11) will strongly suppress low frequency stochastic disturbances that can be generated by double integration of white noise.

B. Parameter setting

Control polynomials $P_r(z^{-1})$ and $P_y(z^{-1})$ in the main control loop of the structure in Fig. 1 may be determined by the specified closed-loop system pole spectrum or by the desired closed-loop system transfer function. Since the closed-loop system is of the second order, the set-point closed-loop continuous system response may be specified by the closed-loop system transfer function

$$\frac{\theta(s)}{\theta_r(s)} = \frac{\sigma^2}{(s + \sigma)^2} \quad (12)$$

which guarantees a strictly aperiodical and sufficiently fast set-point response.

Let us suppose the desired bandwidth of the closed-loop system $f_c \approx 6$ Hz. Then, with sampling period $T = 0.01$ s, one calculates $e^{-\sigma T} = e^{-2\pi f_c T} = 0.6859$ and $\sigma = 37.7$.

The zero-hold equivalent pulse transfer function of (12) is

$$\frac{\theta(z^{-1})}{\theta_r(z^{-1})} = Z \left[\frac{1 - e^{-\sigma T}}{s} \frac{37.7^2}{(s + 37.7)^2} \right] = \frac{0.05549z^{-1} + 0.04316z^{-2}}{1 - 1.3720z^{-1} + 0.4705z^{-2}}. \quad (13)$$

After substituting $Q^0(z^{-1}) = (1 - z^{-1})^2$ into (4) and then equating identically Eqs. (4) and (13), one determines control polynomials $P_r(z^{-1})$ and $P_y(z^{-1})$ as

$$P_r(z^{-1}) = 0.05549 + 0.04316z^{-1} \quad (14a)$$

$$P_y(z^{-1}) = 0.6280 - 0.5295z^{-1}. \quad (14b)$$

III. SIMULATION

In order to examine the efficiency of the IMPACT structure in extraction of load torque disturbances, several simulation runs have been carried out. In the considered positioning servomechanism [14], the employed induction motor has inertia of $J = 0.0459$ kg·m². The incremental encoder giving 2500 pulses per revolution is used. The synthetic parameter $C_m = K_m K_n^* T^2 / 2J = 0.025$ is adopted. Note that this parameter can be easily measured on an experimental setup. Thus, other parameters required for the simulation are: $K_n^* = 2500 / \pi = 397.887$, $K_m = 2C_m J / K_n^* T^2 = 0.05768$, and $K_m K_n^* / J = 500$. In all simulation runs that follow, the disturbance model polynomial $B(z^{-1}) = (1 - z^{-1})^2$ and the corresponding prediction polynomial $D(z^{-1}) = 2 - z^{-1}$ calculated from (11) are used. Recall that these polynomials correspond to extraction of the constant, ramp, and slow varying disturbances.

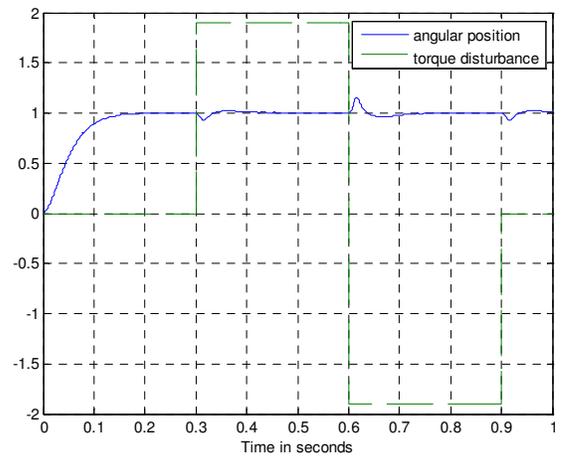


Fig. 2. Extraction of constant step disturbance.

In the first simulation run, two successive step constant torque disturbances are applied. Fig. 2 shows that the influence of the disturbance on the steady state value of target position is completely eliminated after relatively brief transient responses.

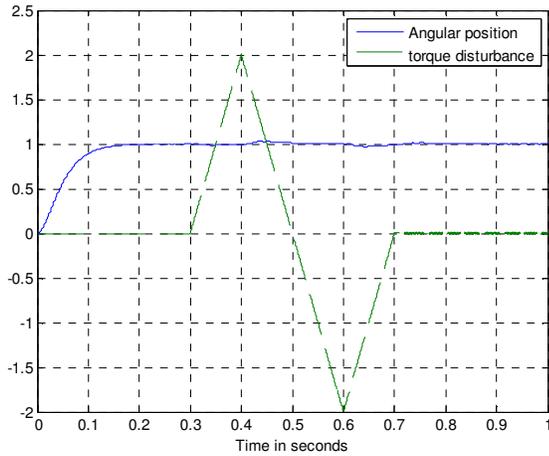


Fig. 3. Extraction of ramp disturbances.

In the second simulation run, the combined ramp torque disturbance shown in Fig. 3 is applied. Fig. 3 shows that the controlling structure completely rejects the disturbance from the steady state value of the servomechanism angular position.

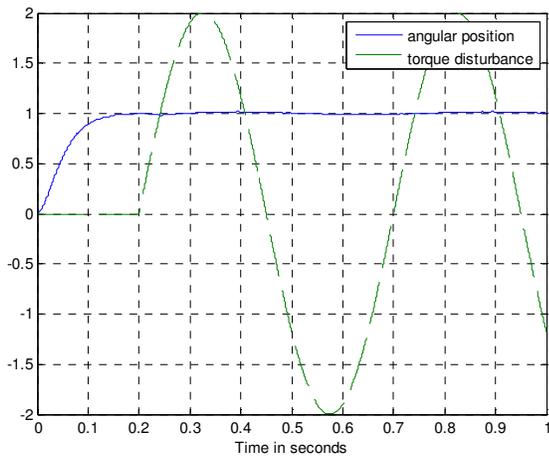


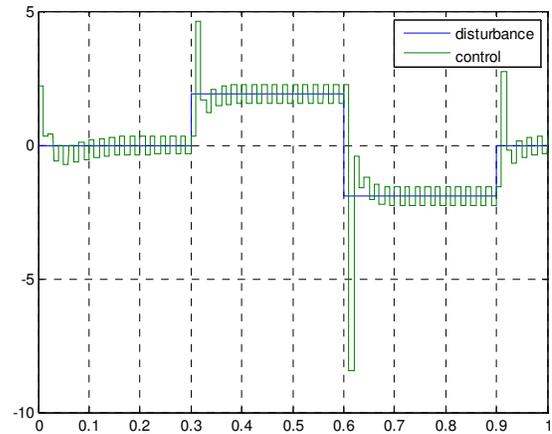
Fig. 4. Extraction of sinusoidal disturbance.

Finally, the sinusoidal torque disturbance shown in Fig. 4 is applied. Fig. 4 illustrates the efficiency of the proposed IMPACT structure in extraction of this kind of external disturbances. Namely, Fig. 4 shows that the disturbance practically does not affect the steady state target position of the positioning servomechanism.

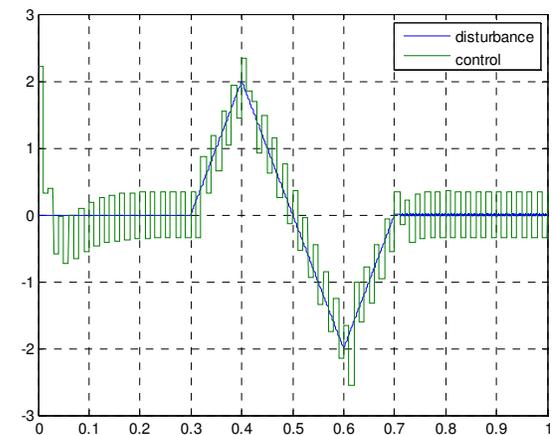
IV. RINGING OF CONTROL

Figs. 2, 3 and 4 show the efficiency of the IMPACT controlling structure in rejection of different external torque disturbances from the steady state target position of the servomechanism. Nevertheless, the physical realization of the servomechanism requires the additional analysis of the shape

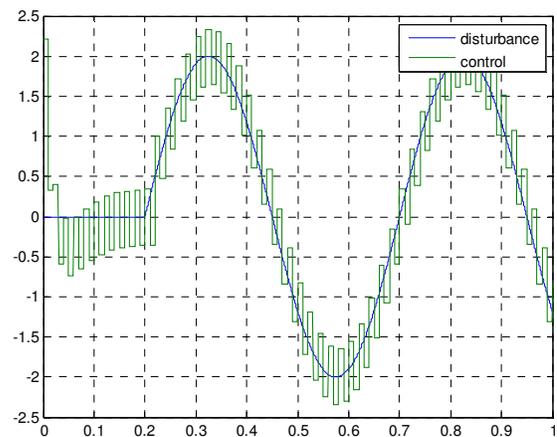
and amplitude of control variable. To this end, in Fig. 6 the disturbances and their controls are shown. From the figure, it is seen that in all cases the controls exhibit fluctuations that are often called “ringing of control”, which could produce serious difficulties in practical realization of driving inverter.



(a)



(b)



(c)

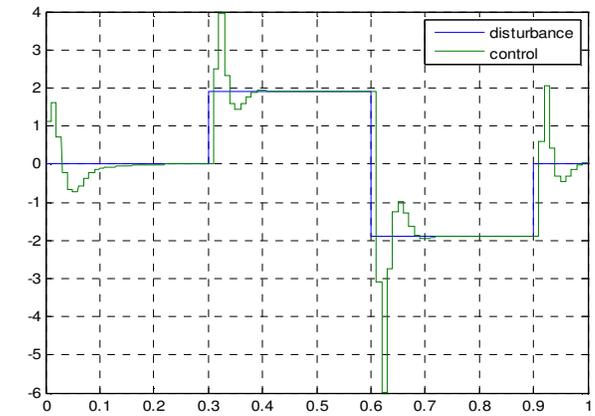
Fig. 5. Ringing of control variables for different types of torque disturbances: (a) constant, (b) ramp, and (c) sinusoidal.

The fluctuations arise due to the present of pole $z = -1$ in the cascade filter $R(z) = z/0.025(z+1)$ within the control portion of the IMPACT structure. This pole generates periodical fluctuations of control variable, having the period equal to $T/2$. To prevail over these difficulties, one can use the idea of Dahlin, which consists in eliminating the critical pole, and thus the ringing of control, by setting $z=1$ in $R(z)$ to obtain $R(1) = 1/0.05$. In doing so, the new structure similar to the IMPACT is obtained. The structure was simulated and the results of simulation runs are shown in Fig. 5. By comparing Figs. 4 and 5, it is noticed that the ringing of control does not exist anymore.

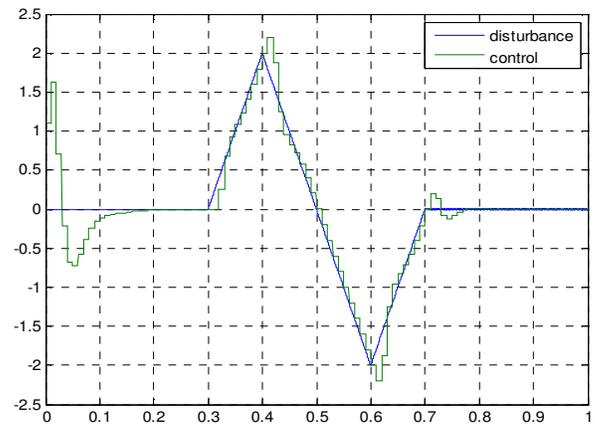
After substitution of $R(z) = z/0.025(z+1)$ by $R(1) = 1/0.05$, the obtained controlling structure is not the proper IMPACT structure any longer and therefore its quality of disturbance transient response and ability of extraction of external disturbances are reduced. This is illustrated by simulation of the new structure and simulation results are given in Figs. 7, 8, and 9. Comparing Figs. 2, 3, and 4 with the corresponding Figs. 7, 8, and 9, one can conclude that, in the new control structure, the setting time and overshoot of disturbance transient response prolongs and increases, respectively. Furthermore, the ability of the new structure in rejection of torque disturbances is slightly reduced. Namely, in the steady state, a small amount of disturbances is noticed. However, this disadvantage of the modified IMPACT structure does not disqualify its practical application in designing of positioning mechanisms.

V. CONCLUSION

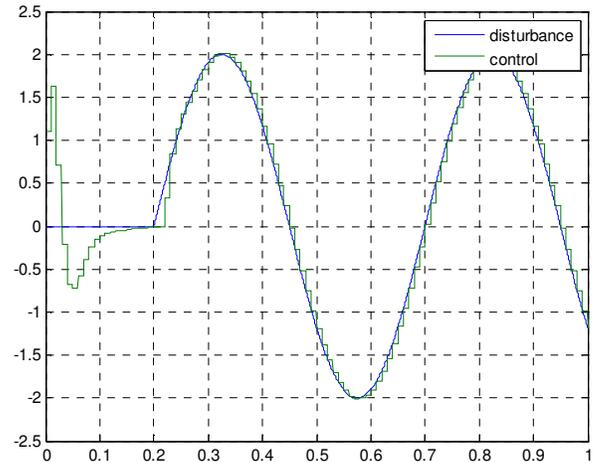
One of the tasks of a control system is that in the presence of external disturbances it tracks the reference signal without steady-state error. In most practical applications, some a priori information about the class of disturbances is available. In such cases, extraction of immeasurable external disturbance is possible by using the IMPACT controlling structure proposed in this paper. The structure has been applied for design of positioning servomechanism in which the steady state target angular position of motor shaft is invariant in respect of external load torque disturbances. For any kind of known class of disturbance it is possible to design the controlling structure that completely rejects or suppresses the influence of disturbance on the steady state value of motor angular position. The simulation results confirm the ability of the structure for disturbance extraction. For the sake of clarity, the procedure outlined in this paper is illustrated by the design of disturbance invariant positioning servomechanism in which the control plant includes a vector controlled induction motor. It is of particular interest to note that the procedure stays unchanged if some other type of driving motor is used. The only difference is the inclusion of the corresponding nominal plant model.



(a)



(b)



(c)

Fig. 6. Control variables without ringing for different types of torque disturbances: (a) constant, (b) ramp, and (c) sinusoidal.

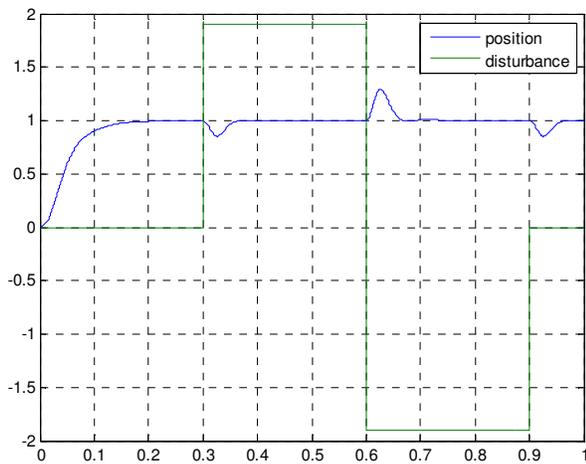


Fig. 7. Extraction of constant step disturbances.

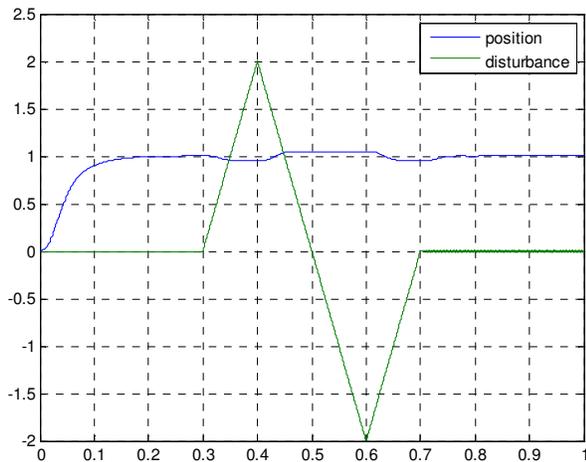


Fig. 8. Extraction of ramp disturbance.

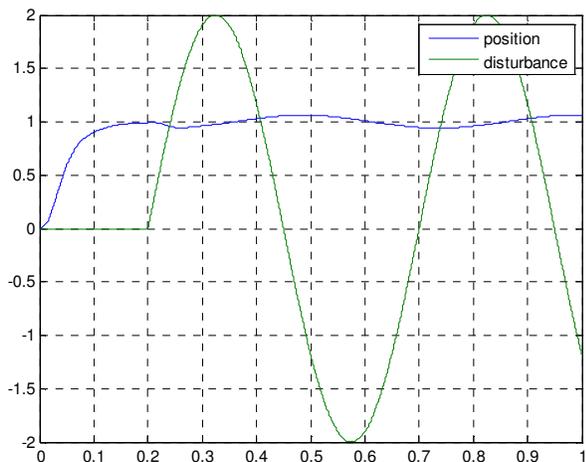


Fig. 9. Extraction of sinusoidal disturbances.

REFERENCES

- [1] Francis B.A, Wonham W.M. (1976) The internal model principle of control theory. *Automatica* vol. 12, pp. 457-465
- [2] Johnson C.D. (1976) Theory of disturbance-accommodating controllers. In Leonides C.T. (Ed.), *Control of dynamic systems* vol. 12, pp. 387-489
- [3] Bengtsson G. (1977) Output regulation and internal models – a frequency domain approach. *Automatica* vol. 13: 333-345
- [4] Tsytkin Ya.Z. and Holmberg U. (1995) Robust stochastic control using the internal model principle and internal model control. *International Journal of Control* vol. 61, no. 4, pp. 809-822
- [5] Amara F.B, Kabamba P.T. and Ulsoy A.G. (1999) Adaptive sinusoidal disturbance rejection in linear discrete-time systems. Parts 1 and 2, *Journal of Dynamic Systems Measurement and Control* vol. 121, pp. 648-659
- [6] Marino M, Santosuosso G.I. and Tomei P. (2000) Robust adaptive compensation of biased sinusoidal disturbances with unknown frequency. *Automatica* vol. 39, pp. 1755-1761
- [7] Bodson M, Douglas S.C. (1997) Adaptive algorithms for the rejection of sinusoidal disturbances with unknown frequency. *Automatica* vol. 33, pp. 2213-2221
- [8] Landau I.D. Constantinescu A, Rey D (2005) Adaptive narrow band disturbance rejection applied to an active – suspension – an internal model principle approach. *Automatica* vol. 41, pp. 563-574
- [9] Oh Y. and Chung W.K. (1999) Disturbance-observer-based motion control of redundant manipulators using inertially decoupled dynamics. *IEEE/ASME Trans on Mechatronics* vol. 4, no 2, pp. 133-145
- [10] Liu C.S and Peng H. (2002) Inverse-dynamics based state and disturbance observers for linear time-invariant systems. *Journal of Dynamic Systems, Meas. and Control* vol. 124, pp. 375-381
- [11] Stojić M.R., Matijević M.S. and Draganović Lj.S. (2001) A robust Smith predictor modified by internal models for integrating process with dead time. *IEEE Trans. on Automatic Control* vol. 46, no 8, pp. 1293-1298
- [12] Stojić M.R. and Stojić Dj.M. (2006) Design of the disturbance-invariant speed-controlled drive with Tesla's induction motor. *Proc. of the Sixth International Symposium Nikola Tesla*, 65-69, Academic Mind, Belgrade
- [13] Stojić M.R. (1984) Design of microprocessor-based system for DC motor speed control. *IEEE Trans. on Industrial Electronics* vol. EI-31, no 3, pp. 243-249
- [14] Stojić M.R. Vukosavić SN (1991) Design of microprocessor-based system for positioning servo-mechanism with induction motor, *IEEE Trans. on Industrial Electronics*, vol. IA-38, no 5, pp 369-378

High Frequency Power Supply for Electrostatic Precipitators in Thermal Power Plants

Slobodan Vukosavić

Abstract—Electrostatic precipitators (ESP) or *electrofilters* remove flying ashes and fine particles from the flue gas in thermal power plants, before passing the gas into the chimney. Maximum allowable value of dust is 50 mg/m^3 and it requires that the efficiency of the ESPs better than 99%, which calls for an increase of active surface of the electrodes, hence increasing the filter volume and the weight of steel used for the filter. In previous decades, electrostatic precipitators in thermal power plants were fed by thyristor controlled, single-phase fed devices having a high degree of reliability, but with a relatively low collection efficiency, hence requiring large effective surface of the collection plates and a large weight of steel construction in order to achieve the prescribed emission limits. Collection efficiency and energy efficiency of the electrostatic precipitator can be increased by applying high frequency high voltage power supply (HF HV). Electrical engineering faculty of the University of Belgrade (ETF) has developed technology and HF HV equipment for the ESP power supply. This solution was subjected to extensive experimental investigation at TE Morava from 2008. to 2010. High frequency power supply is proven to reduce emission two times in controlled conditions while increasing energy efficiency of the precipitator, compared to the conventional thyristor controlled 50 Hz supply. Two high frequency high voltage unit AR70/1000 with parameters 70 kV and 1000 mA are installed at TE Morava and thoroughly testes. It was found that the HF HV power supply of the ESP at TE Morava increases collection efficiency so that emission of fine particles and flying ashes are halved, brought down to only 50% of the emissions encountered with conventional 50 Hz thyristor driven power supplies. On the basis of this study, conclusion is drawn that the equipment comprising HF HV supplies are the best solution for new ESP installations, as well as for the reconstruction of existing facilities. The paper describes the topology of the HF HV power supply, power management and controls, and brings the most important details of the implementation. It is found that the HF HV solution achieves several significant improvements over the conventional thyristor system. It is possible to provide more precise control of the ESP parameters such as the output voltages and currents. It is also possible to make a rapid increase or decrease in voltage and to effectuate a very fast response to load changes. Due to this advantages it is possible to suppress the supply quickly in the case of sparking, reducing the spark energy and the quantity of ionized gasses produced by the electric arc. Reduction in the spark energy is up to 10 times compared to conventional thyristors solution. This means that the erosion of the electrode system is significantly reduced, and that the quality of the collection plates is preserved for much longer periods. At the same time, lower quantity of ionized gasses produced by the spark contribute to much shorter de-ionization intervals, required to

quit sparking and evacuate charged particles in order to reinstate the voltage and proceed with the operation. In addition, HF HV power supply provides a significant reduction in size and weight of the complete ESP installation, hence reducing the tons of steel that has to be built in. Therefore, the HF HV power supply may be the key instrument to reducing the cost of the de-dusting ecological equipment. Besides, size and weight reduction leads to cost savings of installation and maintenance. According to estimates, savings in steel may reach 30%, contributing to the overall cost savings of roughly 20%. Within this paper, in addition to describing the AR70/1000 unit topology and principles of operation, the paper presents the results and measurements obtained during extensive experimental investigations wherein performances of 50 Hz based thyristor units with T/R sets are compared to HF HV power supply.

Index Terms—Electrostatic precipitator, High-voltage high-frequency supply.

I. INTRODUCTION

ALTHOUGH many readers are already acquainted with the functioning of the ESP, this introduction provides a brief reminder for those who have not come across this topic. ESP are used for removal of ultrafine dust particles from flue gas. The most frequent use is the removal of tiny particles of flying ashes in thermal power plants. In order to achieve this, it is necessary for the flue gas to move in the horizontal direction through a strong electric field established between sets of large, parallel electrode plates. The distance between the plates ranges from 300 mm to 500 mm. The plates extend vertically and along the gas flow. The positively charged electrodes are grounded, while the negative ones are connected to the DC voltage source ranging from 50 kV to 100 kV. The negatively charged electrodes have a number of spikes or small diameter wire, which contributes to corona and creates large amount of ions. The current density due to the corona discharge reaches 1 mA/m^2 and it affects a great deal the precipitation and dust collection. Dust particles are being charged by means of diffusion and the field effect and they move towards positively charged electrode, where they are collected. Therefore, the negative electrode is also called the emission plate, while the positive, grounded electrode is also called the collecting plate. Particle migration speed towards the collecting plate is one of most significant parameters of electrofilters. The collecting electrodes are periodically shaken off, i.e. rapped. The rapping process is performed by means of the spinning hammers which

periodically knock on the electrode supports and cause mechanical vibration which helps detaching the dust layer deposits from the plates. The ash being removed during rapping falls into the V-shaped hoppers, located below the ESP chamber. From there, the dust and ashes are transported further by means of water or pressurized air. For the sake of an efficient dust removal, each ESP has at least three series connected sections. The flue gas enters the ESP through the entry section, and then passes through one or more middle sections. Upon leaving these sections, the flue gas passes through the exit section. By passing through the sequence of the serially connected sections the precipitation efficiency (i.e. filtering efficiency) reaches 99.9%. With this efficiency and with every cubic meter of gas leaving the boiler and entering the ESP with 50 g of flying ashes, the gas leaving the electrofilter and entering into the chimney may have the total of 50 mg of dust per every cubic meter of gas. In order to create the electric field and the corona discharge current, the electrodes should be connected with a controllable DC voltage source which provides voltages from 50 kV to 100 kV. Considering the fact that sparks often occur between the electrodes and the electric arc occasionally appears, the source is exposed to periodic short circuit conditions. In the event of arcing, it is necessary to switch off the power supply for an interval of time so as to allow for recombination or removal of charged particles created by the arc. Premature voltage rise will find residues of conductive, ionized gas between the plates, re-entering hence the short circuit condition. After re-establishing of the dielectric strength of the gas, the voltage between the plates may be increased and the precipitation process may continue. The required DC current depends on the surface of the plates. It is necessary to provide the current of 1 mA for every square meter of the electrode surface. Hence, an electrofilter with 10000 m² requires the current of 10 A. The ESP power supply and controls are traditionally based on one pair of antiparallel thyristors which alter the amplitude of the primary AC voltage. The primary supply is usually single-phase, 400 V 50 Hz. By changing the firing angle, the voltage is being changed in the range from 0 to 400 V and brought to the primary of the line frequency power transformer whose

secondary voltage reaches 45 kV to 90 kV (Fig. 1). The high voltage rectifier made out of the great number of series connected diodes is placed in the oil-filled container, along with the transformer itself. The transformer and the diode rectifier are denoted as T/R, Transformer – Rectifier set. One T/R set provides the DC output voltages from 50 kV to 100 kV and currents from 1 A to 2 A. The supply of larger filter section is achieved by putting several T/R units in parallel.

II. T/R UNITS CHARACTERISTICS

Considering the occurrence of sparks and short circuits, it is necessary to foresee the short circuit current limitation measures. The switched on thyristor cannot be switched off before the expiration of the current half-period of the mains. Hence, the short circuit current is limited by the series reactance only. The short circuit reactance of single phase transformers rated several tenths of kVA is relatively small and insufficient for the proper limitation of the short circuit current. Eventual design of dedicated transformer with increased leakage flux would increase the losses as well as the size and weight of the transformer. One of acceptable ways for reactance increase is the HV and LV coil fabrication with increased axial “apertures” between the coil sections in order to increase the leakage. This leads to a fairly complicated insulating structure which should withstand relatively high voltage spikes. Instead, practical T/R sets use the standard, concentrated HV/LV construction and employ a separate external inductance. Transformers are usually put into mineral oil in the common tank with the rectifier and are frequently located on the top of the filter, thus minimizing HV connection length between the transformer and electrodes.

The dielectric tests of T/R units as well as the limitations concerning the upper oil temperature are different from the tests foreseen for conventional HV transformers. The induced overvoltage test is commonly performed at 1,5 rated voltage during one minute. Dielectric strength for short pulses is not performed, since there is no risk of lightning strikes. Considering that the transformer and rectifier share the same

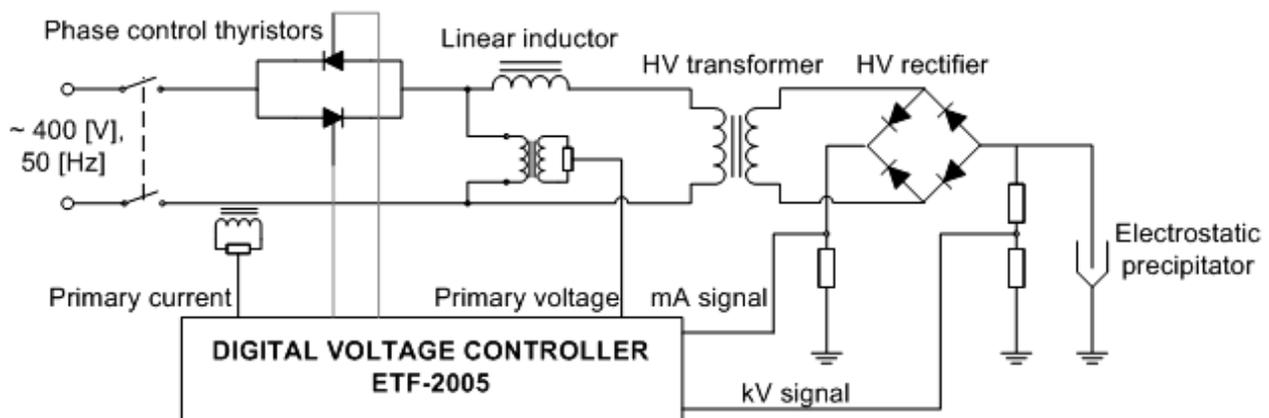


Fig. 1. The electric scheme of thyristor T/R unit.

container (housing) and the same oil, it is necessary to ensure that the increase in the upper oil temperature does not exceed the values that can damage the rectifier. Hence, the ESP transformer losses should be measured with a great care. Among other things, a salient feature of the transformer is the capability to resist frequent short circuits which occur due to sparking between the electrodes. Hence, one part of these transformers testing is the repetitive occurrence of short circuits with the analysis of thermal and mechanical robustness.

Although T/R units are reliable and guarantee a long operating life, they have a very poor power factor and unfavorable waveform of electric currents and voltages (Fig. 2). The thyristor commutation creates harmonic distortion, while the presence of relatively high series reactance along with phase delays due to thyristor firing angle create substantial reactive power. In addition, electrode voltage has a large pulsating component at the fundamental frequency of 100 Hz, which causes a decrease in the average voltage. Namely, the peak voltage must not be higher than breakdown voltage U_{bd} . For that reason, the average voltage must be significantly lower. If the voltage waveform is such that the pulsing component at 100 Hz has an amplitude of ΔU , the average voltage across the electrodes must be less than or equal to $U_{bd} - \Delta U$. As a consequence, the electric current density is lowered as well as the precipitation efficiency. Hence, for the given gas flow and specified precipitation efficiency, it is necessary to foresee the electrode system with increased surface and weight. Such an increase raises the weight, size and cost of the ESP.

The short summary of T/R supplied ESP is summarized below:

- Due to a large voltage ripple, the average voltage and the corona current density are lowered;
- It is necessary to have a larger surface of the plates and spend a larger quantity of steel;
- The reactive power, distortion power and losses are relatively high;
- T/R unit represents a single-phase load;
- Spark energy amounts 130 J – 200 J due to a slow thyristor reaction;
- High spark energy causes a rapid erosion of collecting plates;
- The eroded collecting plates are more difficult to de-dust, so the inner ash layers are permanently retained and fused

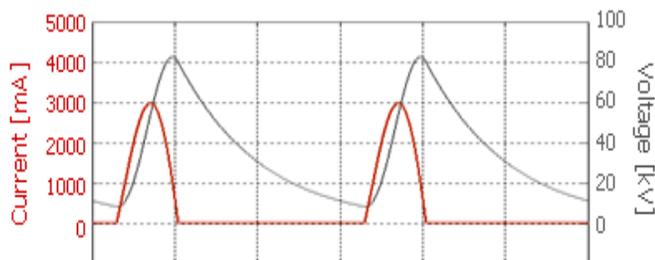


Fig. 2. The characteristic voltage and electric current waveforma obtained with the T/R power supply.

- Very long de-ionization intervals (from 40 ms to 100 ms) are necessary due to higher spark energy.

III. BASIC CHARACTERISTICS OF THE HF POWER SUPPLY

The high frequency (HF) power supply of the electrofilter implies the utilization of a power transformer with a high frequency rectifier on the high voltage secondary side, similar to standard thyristor supply. The difference is that instead antiparallel thyristors the HF supply uses an IGBT converter which supplies the primary of transformer with 5 kHz – 20 kHz AC voltage (Fig. 3). The transformer operates at higher frequencies, but it handles the primary and secondary voltage roughly the same as those encountered in a T/R set. Its dimensions and weight are significantly lower than the line frequency transformer, while its core is made of ferrite materials. The HV diode rectifier receives HF high voltage and provides a DC voltage across the plates. Due to a high operating frequency, the voltage ripple is significantly reduced. Therefore, it is possible to achieve much higher average voltages and hence improve the ESP efficiency.

The low voltage IGBT power converter receives the three-phase mains supply and presents a balanced load.. Since the HV supply as a whole is a symmetrical three-phase load, conditions are met for the distortions, power losses, asymmetries, as well as the reactive power to be improved with respect to single-phase thyristor T/R unit. The HF transformer has a significantly lower reactance and a smaller number of turns, as well as a considerably smaller volume. Therefore, the overall system losses are lower, while the energy efficiency is significantly improved. With regard to voltage control, the HF unit provides much faster reaction time, measured in the tens of μs , while a typical T/R unit produces control delays defined by one period of the mains, hence, either 16,7 ms or 20 ms. A quick control reaction in case of arcing reduces the arc energy, hence allowing for a very brief de-ionization intervals and quickly re-applied voltage. The most important characteristic of the HF supply is a higher average voltage and increased current density. For the frequencies ranging between 10 kHz and 25 kHz, the voltage ripple is lower than 3%, while with thyristor T/R units the ripple amounts 35% to 45%. As a consequence, the HV supply yields up to 20% higher average voltage, which significantly increases the corona current and the precipitation efficiency.

In his book entitled “Industrial Electrostatic Precipitation”, Harry White shows the results of the experiment aimed to establish the link between the peak voltage and number of sparks per minute. The equation

$$S = \exp(b(U_{\max} - U_1)) \quad (1)$$

gives the number of sparks per minute in the function of peak voltage U_{\max} , wherein the voltage U_1 is the one leading to precisely one spark per minute. The exponential relation signifies that the probability of spark appearance increases quickly with relatively small rise in the peak operating voltage. Therefore, the fact is that the voltage ripple made by thyristor

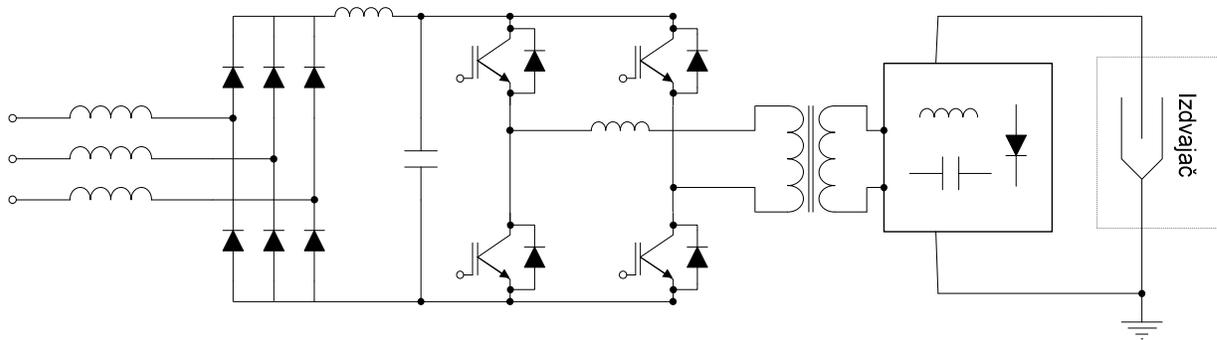


Fig. 3. The electric scheme of HF HV ESP power supply.

T/R units is very high, making it a bare necessity to decrease the average voltage in order to minimize the number of sparks to an acceptable range. Considering that the T/R unit voltage ripple is greater than 35%, while the acceptable number of sparks is only 10 per minute, it may be concluded that the HF supply provides a voltage increase of roughly 20%. It is necessary to mention that the further increase of sparks number is unacceptable, due to every spark event requiring prolonged de-ionization power-down intervals contributing to an increased emission.

Norbert Grass [1,3,4], a Siemens engineer, investigated the effects of the HF power supply implementation, and determined that the corona current doubles with HF power supply, compared to the one obtained with 50 Hz T/R units. Therefore, prediction is that emissions can be halved at least with the HF power supply. Moreover, an increased migration velocity of dust particle towards collecting plates contribute that the layers of collected dust are more compact, and this facilitates the rapping process and decreases the probability of the particle retrieval into the gas flow during rapping.

According to measurements performed in the "Morava" thermal power plant, equipped with four HF units, the high frequency ESP power supply provides for a significant increase in both the energy efficiency and the collection efficiency. Considering that the HF supply has a three-phase rectifier, while the T/R unit has a single phase mains supply, the power factor and the line current waveforms of the HF supply more favorable, while the reactive and apparent powers are reduced. According to measurements, the comparison of the two gives the following results:

	50 Hz	HF
$\text{Cos}(\varphi)$	< 0.65	> 0.95
$\lambda=P/S$	< 0.5	> 0.75

By measuring the voltage and current waveforms at TE Morava, as well as by logging the gas opacity data from the ESP monitoring system, it is concluded that the HF supply gives the following improvement in relation to 50 Hz T/R supply:

- Spark energy is decreased four to five times. While in the case of a 50 Hz T/R supply the spark energy includes the amount supplied through the thyristors until they cease to conduct, the HF power supply reduces the spark energy down

to $W_c = \frac{1}{2} CU^2$. Namely, the residual spark energy is the one accumulated in the electric field between the electrodes, as the HF source itself has a negligible contribution due to a quick turn-off of IGBT power switches;

- In case of occurrence of an arc or spark, the de-ionization power-down interval is very short, ranging only 2 ms to 10 ms, so that the effects of the power interruption on the dust emission becomes very small;
- According to the evaluation, the necessary surface of electrodes is diminished by 30%;
- It is possible to enable quick voltage changes in coordination with the rapping, further enhancing the ESP efficiency.

The advantages of the HF supply have made a great number of renowned manufacturers to project and plan the production of ESP HF power supply devices. Siemens has made an effort with the device PIC410F, aimed to operate at frequency of 10 kHz and to provide high output DC voltage. Due to problems [1, 3, 4], they moved on to the experiments with the high voltage transformer which operates at the lower frequencies, that is, at 500 Hz. The British company Genvolt made a transformer operating at frequencies of 20 kHz [2] and with a relatively low power of $P < 50$ kW. The efforts to manufacture a reliable HF power source for the ESP with sufficient power, voltage, and an acceptable life time were made by Alstom, who designed SIR device. Similarly, FL Schmidt designed the HF power supplies operating at frequencies next to 10 kHz. The American company NWL manufactures Power Plus HF power supply for the ESP, widely used in cement mills in the USA, as well as in the factories which include the process of sintering. The NWL devices are somewhat limited in the output voltage, and they cannot reach the voltages above 100 kV, required evermore for the operation of new electrofilters with interelectrode distance in excess to 500 mm, recently quite common in thermal power plants. For majority of manufacturers, a wider application of the HF devices is impeded by reliability issues and limited output voltages. In some cases, frequent insulation breakdowns are observed in the field, caused by the simultaneous presence of the high voltage, high frequency, high temperature, high dV/dt and high di/dt values within the device.

IV. THE PROBLEMS OF THE EXISTING HF SOLUTIONS

The problems encountered in the HF electrofilter supply exploitation are frequently reflected in (i) the breakdown of semiconductor power switches in the primary circuit, (ii) secondary circuit insulation breakdown or (iii) the problems of EM noise and control. The first problem emerges due to the circumstance that the IGBT power switches deliver next to 100 kW while commutating at the frequencies from 10 kHz to 25 kHz. Hence, the IGBT power switches operate at the limit of their performance. Therefore, unless zero current switching (ZCS) or the zero voltage switching (ZVS) is ensured, they may suffer serious overheating. The second problem is caused by the rapid insulation degradation and aging in the presence of high dV/dt stress, while the third one is most frequently attributed to an inadequate hardware and software solutions. In particular, low voltage and highly sensitive DSP technology is implemented in a noisy environment of high power converters.

The primary voltages and currents within HF power supply assume values next to 600 V and 300 A. Therefore, the semiconductor power switches to be used are the IGBT devices. In addition to their favorable characteristics, IGBTs also have their commutation losses which limit the maximum switching frequency. The recommended switching frequency for IGBT power switches depends on the rated current, and it is lower as the current rating increases. For transistors of 300 A, commutation and conduction losses are equal at the frequencies between 2 kHz and 5 kHz, which represents the technical optimum for their operation. Therefore, it is necessary to utilize resonant topologies in high frequency range, enabling the semiconductor power switches to operate without commutation losses. This enables the elimination or at least a significant decrease in commutation losses.

Up to date literature includes the analysis of a number of power converter topologies [6-18] providing reduction in commutation losses. Attention has been focused mostly to ZVS – zero voltage switching, ZCS – zero current switching as well as to decreasing the dV/dt in the primary circuit of the converter. The HF power supply for the ESP may have high dV/dt values in the secondary circuit, across high voltage diodes, at the ends of secondary inductors and other elements of the secondary circuit. At instances where the secondary voltage switches from -100 kV to +100 kV within commutation intervals shorter than 1 μ s, extremely high voltage slopes and extremely quick changes in the electric field may subject the insulating materials under severe stress. As a consequence, the dielectrophoresis occurs as well as the chemical reactions within the dielectric material, caused by the catalytic effect of the high speed electric field changes. The ultimate consequence of the described processes is a rapid degradation of the insulating material. According to the experience acquired by many electric equipment manufacturers, the insulation breakdown in the secondary circuit takes place after 6-12 months, which is substantially unfavorable for the end-user. This outcome has caused a great number of withdrawals of HF devices from the field. Solution

described in this paper eliminates the insulation aging and reliability issues and enables the longevity of the HF power source.

A number of reliability issues is caused by an inadequate application of DSP technology in environment of high power, high voltage, high frequency conversion. Due to high speed voltage changes, as well as high currents at high frequencies, control circuits are submersed by a significant electromagnetic noise which jeopardizes the integrity of analogue and digital signals. Therefore, it is necessary to implement hardware measures in order to protect the integrity of the input signals, output signals, and the elements of electronic circuits. Common measures include the installation of the SMD transformer and the miniature resistors and inductance in places where the electromagnetic disturbance should be prevented and the reinforcement of relevant signals. Besides, during the acquisition of analog signals, the oversampling technique should be used in order to remove the undesirable noise, thus preserving the response speed and the control integrity.

V. THE BASIC CHARACTERISTICS OF THE AR70/1000

At the School of Electrical Engineering, the University of Belgrade, a group of researchers worked on the development of the HF power supply for the ESP over the past four years. The project was funded by the Ministry of science and technology of The Republic of Serbia. As the result, the first specimens of the HF power supply have been manufactured in Serbia. Two years ago, “Morava” power plant was equipped with HF power supplies AR70 which are constantly in operation ever since. During these two years, a series of tests was performed in order to investigate the impact of HF power supplies on the ESP operation. The basic characteristics of HF power source AR70-1000:

- Distributed multiresonant topology in the secondary circuit enables decreasing in commutation losses and reduces the insulation stress, hence suppressing the catalytic effects of the electric field high speed changes and preventing chemical reactions leading to accelerated dielectric aging;
- Precipitator control and diagnostics are based on the spectral analysis of corona current and plate voltages, providing a cheap and reliable way for early detection of sparking, an efficient detection of the back corona, and a reliable estimation of the thickness of the ash layer accumulated on the collecting plates.
- Converter topology enables ZCS commutation of IGBT power switches, significantly lowering the overall converter losses;
- Voltage and current control includes the possibility controlling the number of sparks per minute;
- From the accelerated aging, the MTBF of the device can be estimated to 20 years.

The total mass of the active material in AR70/1000 transformer with rated power of 100 kW and operating frequency of 10 kHz amounts only 50 kg. Distributed

multiresonant secondary circuit is made of conventional reactive components having a relatively low weight, and it uses standard HV diodes (Figs. 4–6). Within the secondary HV rectifier, high speed diodes are not required as the converter topology enables the rectifier operation in ZVS mode. Besides, it is the ZVS mode that reduces the dV/dt in the secondary circuit and ensures the longevity of the insulation materials.

In comparison to standard 50 Hz sources including the thyristor cabinet and the T/R group, the HF power supply AR70/1000 is five times smaller in volume, about 5 times lighter and up to 30% cheaper. It should be noticed that the HF source features explained above provide a significant increase

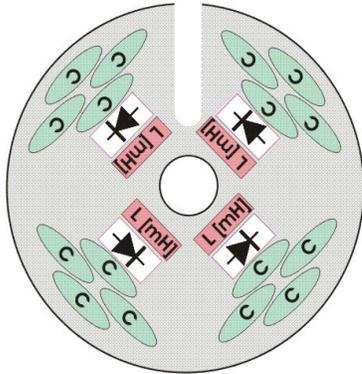


Fig. 4. One element of the multiresonant circuit within the secondary of the AR70-1000.

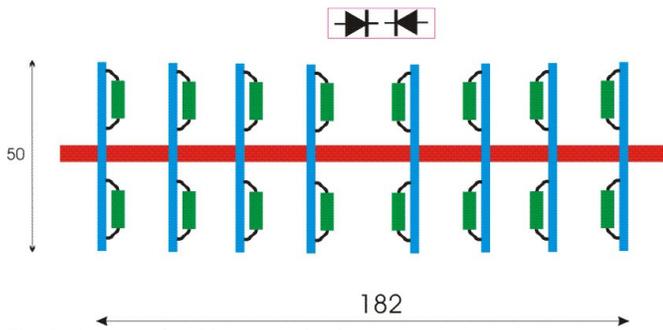


Fig. 5. A group of multiresonant circuits.

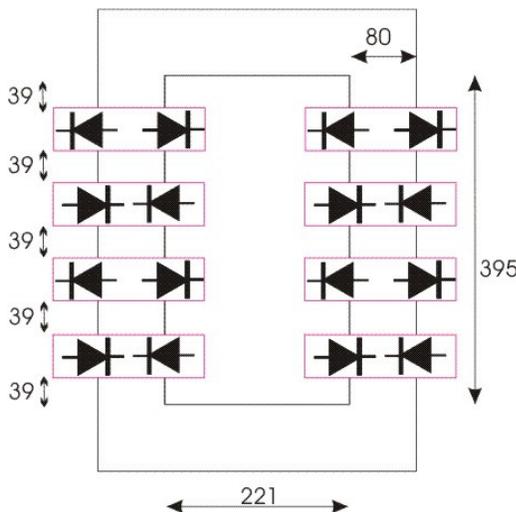


Fig. 6. Connecting multiresonant circuits in groups.

in precipitation efficiency, contribute to a more favorable shape of the line current, ensure much lower reactive power and reduces significantly the erosion of the plates due to a significant decrease in the spark energy. Due to the original multiresonant topology, DSP unit for control and supervision as well as the original solution of thermal handling, and mechanical protection, the AR70/1000 meets the following standards and regulations,

- Low Voltage Directive (73/23/EEC);
- EMC directive (89/336/EEC);
- CEI EN 60204-1, par. 6.2.3, 20.3, 20.4;
- IP Code, EN60529;
- CEI EN60800-3;
- EN60800-3/A11.

The appearance of the AR70/1000 device is given in Fig. 7. One of the problems encountered by other manufacturers is the ferrite core protection. Namely, in cases of overheating, larger ferrite elements used for assembling the magnetic core of the HF transformer may crack, hence disabling further operation due to an increased equivalent air gap and magnetic resistance. This happens due to positive power loss coefficient of the ferrite, and due to difficulties in cooling the innermost ferrite blocks. Such an outcome is prevented in AR70 unit by using the original cooling design and a custom shape of the core. Fig. 8 illustrates the temperature distribution of the ferrite core obtained by using the thermal imaging. This figure shows that the heat distribution is uniform. The external dimensions of the AR70 device are given in Fig. 9.

VI. THE FIELD EXPERIENCE

Since June 13th 2008 until July 21st 2010, the testing took place at “Morava” thermal power plant in Svilajnac, equipped with AR70/1000 units. The measurements were organized in order to determine the effects of the HF power supply on the precipitation efficiency and to establish the expected reduction in dust emission. The following equipment was installed at the Morava PP:

- Four HF power sources for the ESP, two of them being AR70/1000 and the other two NWL Power Plus devices.
- HV splitter with selector switches which enables the ESP

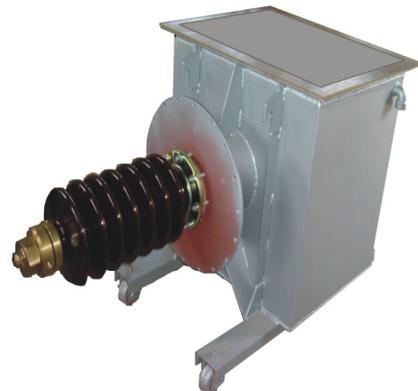


Fig. 7. The appearance of AR70/1000 device.

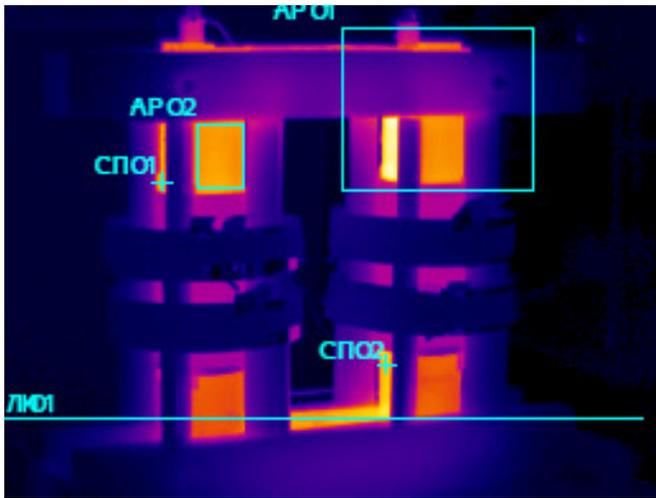


Fig. 8. The temperature distribution in ferrite core.

supply either from conventional 50 Hz T/R units or from the HF power sources;

- The P/Q/S/Energy measurement group for logging of the the active, reactive, apparent, and distortion power, as well as active and reactive energy;
- One triboelectric dust sensor, mounted in the back feeding channel of the right ESP branch.

Fig. 10 shows the block diagram with the specified equipment mounted in “Morava” power plant. The HV splitter system enables hot-swap selecting either the conventional T/R units or the HF power supplies in the course of the ESP operation. Hence, comparison of the two was possible in almost identical conditions. In HF mode, one of the ESP branches is supplied from the NWL units, while the other

feeds from AR70/1000 sources. All the devices may be selected in the course of operation, so it is possible to compare T/R units with AR70/1000 units, T/R units to NWL units, or NWL units to AR70/1000 units. In this way it is possible to determine the best solution with regard to precipitation efficiency. The locations where the HF units were mounted are shown in Fig. 11.

The triboelectric sensor is shown in Fig. 12 and mounted in the back feeding channel of the right ESP branch. The mounting location is shown in Fig. 13. It has been selected so that the gas flow is laminar so as to insure stability and accuracy. Alternative solutions, i.e. i.e. eventual mounting within the post-confusor channel has been abandoned, since it was not possible to identify location that would be far enough from the channel curves or the fan, so as to ensure a laminar flow. As it is well known, mounting of the opacity meters in zones with pronounced turbulence contributes to significant errors.

After mounting the first two HF units in the Morava PP, the series of measuring was carried out in order to perform the comparison of the precipitation efficiency of ESP which are supplied from T/R units and ESP which is supplied from the HF source. Two configurations were tested on that occasion:

- The supply of all sections from T/R units (the total of 4 T/R units);
- The supply of 2 sections from the HF source and two sections from T/R units.

The amount of ash which is separated in the exit section funnel was measure indirectly, on the basis of the measured temperature of the funnel wall. On the basis of the known ash

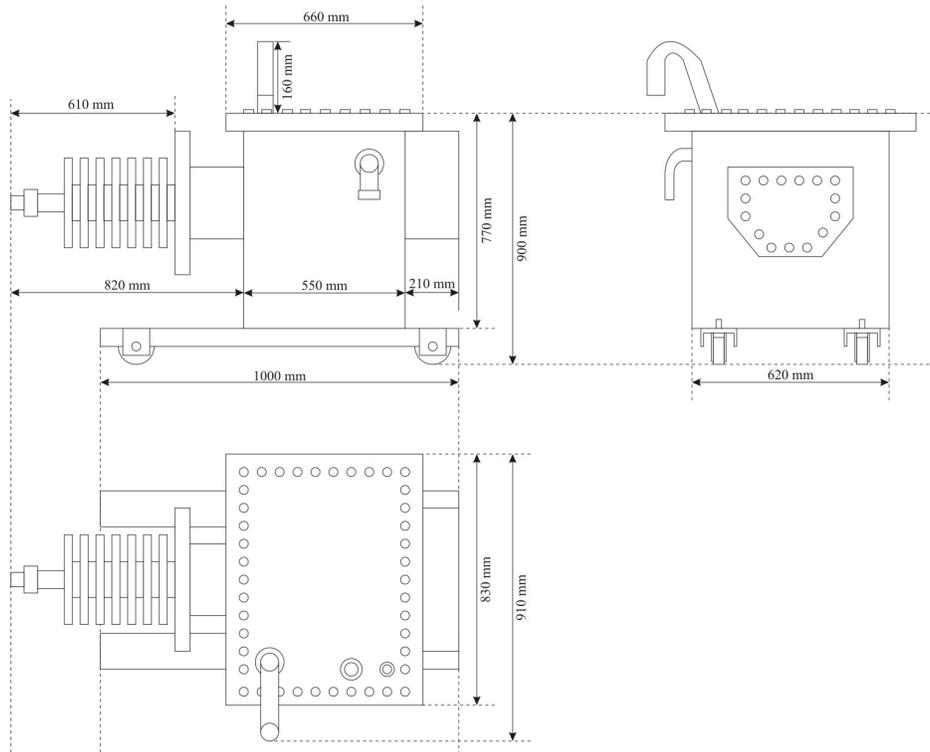


Fig. 9. External dimension of AR70/1000 device.

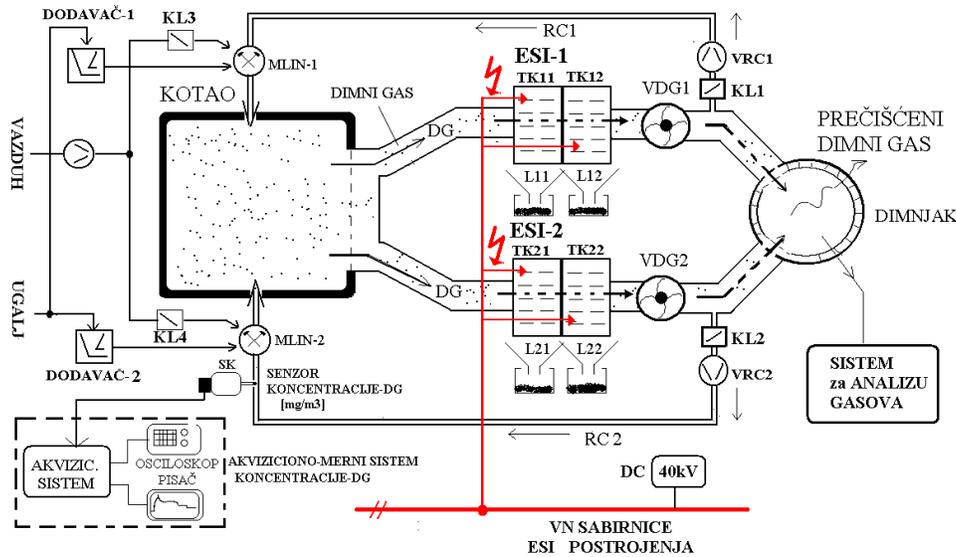


Fig. 10. The equipment installed at Morava PP.

temperature which is collected in the funnel, the known specific ash temperature, as well as the known conditions of heating and cooling of funnel, the relative quantity ash change was noticed, which is separated after the HF supply implementation. It was noticed that the ash quantity in the exit section funnel was increased by 80% by application of HF supply. This finding was confirmed by the fact that after the HF supply implementation, it was necessary to empty the

funnel twice often (The exit section funnel is emptied manually in the Morava PP).

During measuring it has not been disposed of transducer for tracking changes of the flue gas dusting. The HF supply effects for the total emission can be evaluated on the basis of the smoke appearance from the chimney, which is illustrated in Fig. 14. The left part of the figure shows the smoke appearance in the configuration of 4 x 50 Hz devices. The right part of the figure shows the smoke appearance in the configuration of 2 x 50 Hz + 2 x HF devices.



Fig. 11. HF power supplies.



Fig. 12. Triboelectric sensor.

VII. MEASUREMENTS FROM THE OPACITY TRANSDUCER

Fig. 15 shows the voltage waveform between the electrofilter plates. The CH1 trace shows the voltage waveform in case of supply from T/R units of 50 Hz system. The CH2 trace shows the case when the ESP is supplied from HF units. It can be noticed that HF supply contributes to the significant decreasing of voltage ripple. As a consequence, it is possible to achieve the significantly higher mean voltage value between the electrodes. Fig. 15 shows also system reaction time for voltage control. It is possible to notice that the HF system enables substantially faster voltage changes.

After mounting the triboelectric transducer, the possibility is opened to measure relative changes in the emission of dust during the filter operation. Fig. 16 shows the obtained results. The left part of the Figure shows the emission difference in case of using the HF supply, that is, if the right filter branch is supplied by two HF units. The transition from HF system to 50 Hz system was made in the middle. Considering the fact that the HV dividers are commuted in the non-voltage state, electrodes are left without supply for a short time interval, so that the reading from transducer becomes saturated due to very high particle emission. After switching the HV divider into the location for 50 Hz supply, the voltage is re-established between the electrodes. The right part of the Figure shows the



Fig. 13. Mounting location.



Fig. 14. The left part of the figure illustrates the smoke appearance in configuration of 4 x 50 Hz. The right part of the figure illustrates the smoke appearance in configuration of 2 x 50 Hz + 2 x HF.

emission change from 50 Hz T/R units. In the observed figure it is noticed that the HF supply diminishes the emission from 650 mg/m^3 , which is obtained for the supply from 50 Hz T/R units, to only 150 mg/m^3 .

It is necessary to mention that the particle emission is significantly changed depending on the coal content, kettle regime, as well as the other process parameters. So, in order to validly compare two systems of supply, it is necessary to perform measuring in somewhat longer time interval. Fig. 17 shows the emission comparison of the HF supply and 50 Hz supply in time interval of three hours. The measuring was performed during 2010. The blue color denotes the emission in case of supply 4 x HF, that is, in case when the ESP is completely supplied by HF units. The red color denotes the

emission in configuration 4x50 Hz, that is, when the total ESP is supplied by 50 Hz T/R units. Mean emission value determination in the interval of three hours shows that for conventional 50 Hz supply with T/R units, mean emission value amounts to 700 mg/m^3 , while for the HF supply the mean emission value totals 350 mg/m^3 . So, by implementation of the HF supply, the particle emission is halved on the same electrode system.

VIII. PRICE COMPARISON OF HF EQUIPMENT AND 50 Hz EQUIPMENT

The wider implementation of the HF supply in thermal power plants depends on the price of equipment also. Thus, this section compares prices of equipment needed for supply of one filter section which has parameters of 70 kV and 1000 mA.

A. T/R Unit and Thyristor Converter for 50 Hz Supply

T/R units of European manufacturers have prices in the range of 19 k€ (FLSmidth Airtech). Alongside the T/R unit, it is necessary to have also a pair of antiparallely connected thyristors with the controller and the appropriate equipment. Cabinet including controller, thyristors, and the appropriate equipment costs 11 k€ (Castelet, Kraft).

So, in order to provide the supply of 70 kV and current of 1000 mA, it is necessary to purchase equipment which costs around 30 k€.

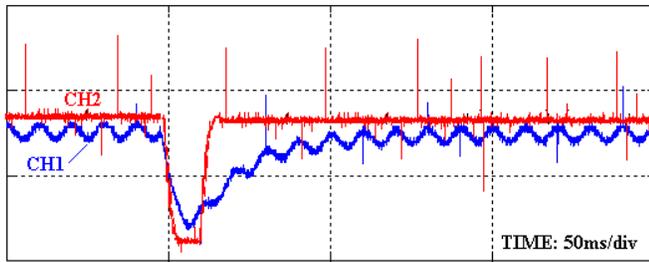
B. HF Supply

The HF supply unit is very frequently applied so that the high frequency power transformer and low voltage converter are located in the same housing and have a single heat removal system. The fabrication cost of the HF voltage supply unit of 70 kV and current of 1000 mA will total 24 k€.

It may be concluded from the above that, besides other significant advantages, the HF power supply also has a lower price, compared to conventional T/R units.

IX. CONCLUSION

- The particle emission is diminished twice by implementation of the HF ESP power supply.
- The HF equipment is five times lighter than conventional T/R 50 Hz units;
- The HF supply equipment has lower price than the equivalent thyristor-controlled T/R units;
- The necessary surface of the electrodes and the quantity of steel built into a new ESP can be lowered by using HF units;
- There is a consolidated Serbian-manufactured solution for the HF supply based on the original multiresonant HF HV technology and intermittent control. The solution has the potential of involving local industries in ESP building and reconstruction.



CH1: napon 50Hz ESI -20kV/div , CH2: napon VF ESI-20kV/div

Fig. 15. The comparison of the waveform voltage between the plates in case of 50 Hz supply and HF supply.

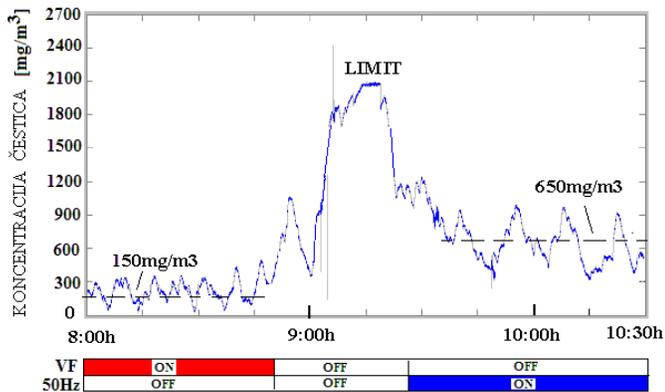


Fig. 16. The emission of flue ash depending on the way of electrode system supply. The left part of the figure shows the emission in case of using HF supply. The transition from HF to 50 Hz system was made in the middle. The right part of the figure shows the emission in case of 50 Hz T/R units.

REFERENCES

- [1] Norbert Grass, "150kV/300kW High Voltage Supply with IGBT Inverter for Large Industrial Electrostatic Precipitator", in *Industry Applications Conference, 2007, 42nd IAS Annual Meeting of the 2007*, pp. 808-811.
- [2] John C. Fothergill, Philip W. Devine, and Paul W. Lefley, "A Novel Prototype Design for a Transformer for High Voltage, High Frequency, High Power Use", in *IEEE Trans. on Power Delivery*, vol. 16, no. 1, January 2001, pp. 89-98.
- [3] Norbert Grass, Werner Hartmann, Michael Klöckner, "Application Of Different Types Of High-Voltage Supplies On Industrial Electrostatic Precipitators", *IEEE Trans. on Industry Applications*, vol. 40, no. 6, pp. 1513-1520, Nov/Dec 2004.
- [4] N. Grass, "Fuzzy logic-optimising IGBT inverter for electrostatic precipitators," *IEEE-IAS Annu. Meeting*, vol. 4, Phoenix, AZ, Oct. 4-7, 1999, pp. 2457-2462.
- [5] Laurentiu M. Dumitran, Pierre Atten, Didier Blanchard, and Petru Notingher, "Drift Velocity of Fine Particles Estimated From Fractional Efficiency Measurements in a Laboratory-Scaled Electrostatic Precipitator" *IEEE Trans. on Industry Applications*, vol. 38, no. 3, pp. 852-857, May/June 2002.
- [6] Won-Ho Kim, Jong-Soo Kim, Iouri Kang, Geun-Hie Rim C. U. Kim "A High Voltage Pulsed Power System for Electrostatic Precipitators" *IEEE IAS Annual Meeting 1999*.
- [7] Junming Sun, Mutsuo Nakaoka, Hiroshi Takano, "High Voltage Transformer Parasitic Resonant PWM DC-DC High Power Converters and their Performance Evaluations, in ISIE 1997 Conference, pp. 572-577.
- [8] R. L. Steigerwald, "A Comparison of Half-Bridge Resonant Converter Topologies" in *IEEE Trans. on PE*, vol 3, no. 2, April 1988, pp. 174-182.
- [9] R. L. Steigerwald, R. W. De Doncker, M. H. Kheraluwala, "A comparison of High Power DC-DC Soft Switched Converter Topologies", *IEEE Trans. Industry Applications*, vol. 32, no. 5, pp. 1139-1145, Sept/Oct 1996.
- [10] O. D. Patterson, D. M. Divan "Pseudo-Resonant Full Bridge DC/DC Converter", *IEEE Trans. Power Electron.*, vol. 6, no. 4, pp. 671-678, October 1991.
- [11] J. Zhang, X. Xie, X. Wu, G. Wu, and Z. Qian, "A novel zero-current transition full bridge DC/DC converter", *IEEE Trans. Power Electron.*, vol. 21, no. 2, pp. 354-360, Mar. 2006.
- [12] T. T. Song, N. Huang, "A novel zero-voltage and zero-current switching full bridge PWM converter", *IEEE Trans. Power Electron.*, vol. 20, no. 2, pp. 286-291, Mar. 2005.
- [13] J. Dudrik, P. Spanik, and N. D. Trip, "Zero voltage and zero current switching full bridge DC/DC converter with auxiliary transformer", *IEEE Trans. Power Electron.*, vol. 21, no. 5, pp. 1328-1335, Sep. 2006.
- [14] Sanbao Zheng and Darisuz Czarkowski, "High-Voltage High-Power Resonant Converter For Electrostatic Precipitator", in *APEC Conf. Record 2003*, vol. 2, pp. 1100-1104.
- [15] J. F. Chen, R. Y. Chen, T. J. Liang, "Study and Implementation of a Single Stage Current Fed Boost PFC Converter With ZCS for High Voltage Applications, *IEEE Trans on PE*, Vol. 23, No. 1, January 2008, pp. 379-385.
- [16] E. H. Kim, B.H. Kwon, "Zero voltage and zero current switching full bridge converter with secondary resonance", *IEEE Trans. Ind. Electron.*, vol. 57, no. 3, pp. 1017-1025, Mar. 2010.
- [17] J. A. Martin-Ramos, A.M. Pernia, J.Diaz, F. Nuno, J.A. Martinez, "Power supply for high voltage application", *IEEE Trans. Power Electron.*, vol. 23, no. 4, pp. 1608-1619, July. 2008.
- [18] J. Liu, Sheng, J. Shi, Z. Zhang, X. He, "LCC Resonant Converter Operating under Discontinuous Resonant Current Mode in High Voltage, High Power and High Frequency Applications", in *Applied Power Electronics Conference APEC2009*, pp. 1482-1486.

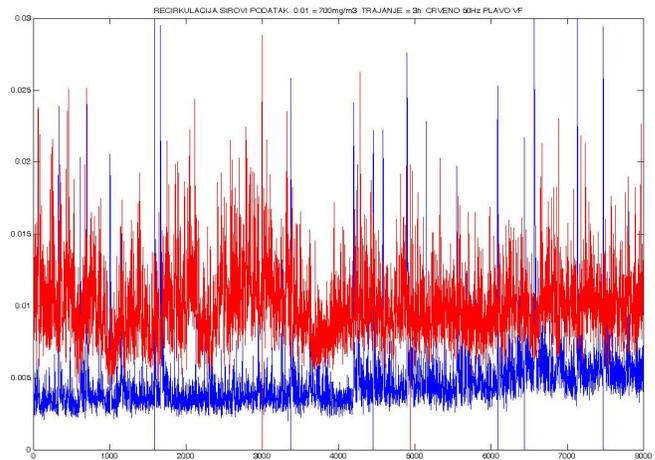


Fig. 17. The comparison of emission for HF supply and 50 Hz supply in the time interval of three hours. The measuring was performed during 2010. The blue color denotes the emission for 4 x HF supply, that is, in case when the ESP is completely supplied by the HF unit. The red color denotes the emission for configuration 4 x 50 Hz, that is, when the total ESP is supplied by 50 Hz T/R units.

Realization of an Electronic Load for Testing Low Power PEM Fuel Cells

Djordje Šaponjić, Vojislav Arandjelović, Milica Marčeta Kaninski and Aleksandar Maksić

Abstract—A realized electronic load system intended for testing and characterization of hydrogen fuel cells is described. The system is based on microcontroller PIC16F877 by applying the concept of virtual instrumentation. The accomplished accuracy of the developed electronic system allows performing efficiently investigations of the electro-chemical phenomena involved in the process of designing hydrogen fuel cells.

Index Terms—Programmable current source, microcontroller, fuel cell.

I. INTRODUCTION

Hydrogen is an almost ideal medium for energy conservation and potentially it could be a part of the basis for further development of the present day civilization. The ever increasing demand for energy is an incentive for investing additional efforts in developing the production methods, storage systems, as well as the methods of releasing the stored energy of hydrogen and its conversion to electricity or heat. Fuel cells are used for obtaining electrical energy from hydrogen [1]. At present, depending on particular application, several types of fuel cells are in use: Proton Exchange Membrane Fuel Cells (PEMFC), Solid Oxide Fuel Cells (SOFC), Alkaline Fuel Cells (AFC), Direct Methanol Fuel Cells (DMFC), Metal hydride Fuel Cells (MHFC), etc.

For the purpose of characterization of fuel cells, investigation, and understanding of the phenomena involved, different methods and specific electro-chemical instrumentation are used. A most frequent instrument used for characterization of current-voltage characteristics of fuel cells is electronic load [2]. In essence, electronic load is a programmable current source fed by a chemical power source, a fuel cell in this case. The main requirement concerns measurement of the cell voltage while its current is modulated by the programmable current source. In this way, one obtains the current-voltage characteristic which contains numerous data as regards electro-chemical processes involved in a fuel cell.

II. BASIC ELEMENTS OF AN ELECTRONIC LOAD

An electronic load consists of two main parts: master application of a virtual instrument [3] on a PC and microcontroller system accompanied by a programmable

current source. Block diagram of an electronic load is shown in Fig. 1.

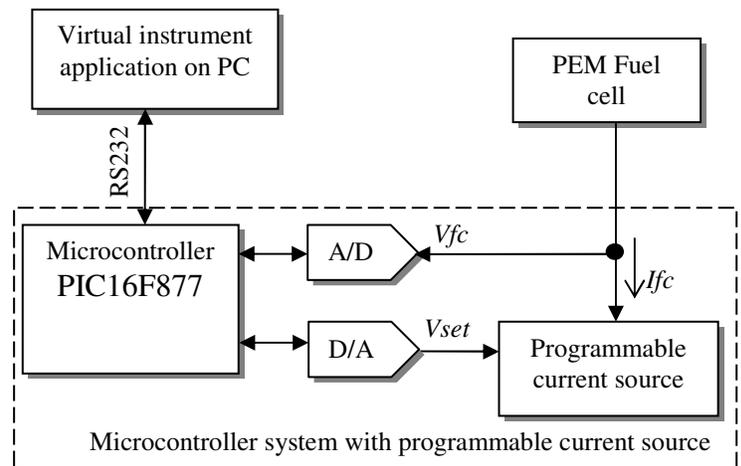


Fig. 1 Block diagram of electronic load.

The microcontroller system consists of an 8-bit PIC16F877 microcontroller [4], a 12-bit A/D converter MCP3204 together with an input circuit, a 12-bit D/A converter MCP4921 together with an adaptation circuit, and a programmable current source. Essentially, the microcontroller system is designed as a slave instrument fully controlled by the master virtual instrument application. The task of the microcontroller system is to load the examined fuel cell by the current which is defined by the master PC application and then to measure voltage across the fuel cell. The master-slave data exchange is realized a series RS232 channel. The communication protocol is very simple. The master application sends two bytes carrying information on the 12-bit value of the required current, and the slave microcontroller system returns four bytes, two repeating the required current value (confirmation of the received message) and two carrying 12-bit value of the voltage measured across the cell.

The programmable current source is realized according to the block diagram shown in Fig. 2. Basically, the source consists of the high power MOSFET BUZ11A and operational amplifier LM4250. The programmable current source is controlled by the control voltage V_{set} . Resistor R_{set} defines the maximum current of the source. In order to accomplish an improved measurement of IU characteristic of fuel cells, it is required that the value of R_{set} is as low as possible which would reduce its influence on measurement of the cell voltage at short circuit (diffusion region).

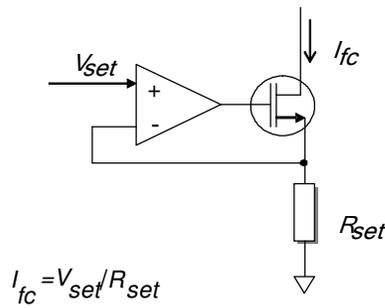


Fig. 2 Block diagram of the programmable current source.

The developed electronic load is intended for testing fuel cells, either mono-blocks or stacks up to 5 cells, thus for the purpose of input circuit protection, the voltage range is limited to 10V.

III. MASTER APPLICATION OF A VIRTUAL INSTRUMENT

The main task of the master application of a virtual instrument is control of the microcontroller system of the electronic load. In the process of the IU characterization of a fuel cell, several different measurements have to be carried out: measurement of the IU characteristic by varying a range of currents from I_{min} to I_{max} within a predefined number of measuring points, measurement of the IU characteristic during a long term load of the fuel cell by a load current I_{const} , and measurement of the IU characteristic during a long term load of the fuel cell by a constant voltage V_{const} .

Measurements of the IU characteristic by varying the value of current over a range from I_{min} to I_{max} , within a predefined number of measurement points, is realized by an open loop control algorithm. The master application of a virtual instrument determines the value of current of the microcontroller system by the RS232 serial interface. Variation of the current is updated each 0.5s within the range from I_{min} to I_{max} . The voltage measured across the fuel cell, together with the value of the preassigned current, are stored in a separate file for further analysis. On the basis of the obtained IU characteristic, the current-power (IP) characteristic is formed. This type of measurement serves for determination of the three regions of fuel cell operation: activation, ohmic, and diffusion regions which is significant for defining operating regime of a fuel cell.

For measurement of IU characteristic of a fuel cell during a long-term loading by a constant current I_{const} , the virtual instrument application is also realized by an open loop control algorithm. The master application defines the value of current I_{const} and passes it to the microcontroller system. The master application also determines the length of time of the fuel cell voltage measurement. This measurement serves for testing stability of the fuel cell and of its constituents: electrodes, membrane, catalyst, and activator.

For measurement of IU characteristic of a fuel cell during a long-term loading by a constant voltage V_{const} , the virtual instrument application is realized by a proportional control algorithm. The feedback is closed in the master application in order to avoid change of the microcontroller system and software of the slave. Each 0.1s the master application takes measurement of the voltage across the fuel cell, makes a

comparison with the pre-assigned voltage V_{const} , and modifies current I_{fc} in order to ensure the pre-assigned voltage across the cell. Since the speed of control is not specified, the algorithm of proportional control is applied.

IV. THE REALIZED ELECTRONIC LOAD

The microcontroller system of the electronic load is based on PIC16F877 microcontroller made by Microchip, while the corresponding software has been developed by using the development tools EASYPIC5 and MicroC Pro made by Mikroelektronika [5,6]. The realized hardware has been optimized for a series production with the aim of offering this product as a part of the experimental set intended for education in the area of application of hydrogen for energy conservation.

The master application of virtual instrument has been realized by using software package made by National Instruments CVI. The realized application consists of three parts: IU characterization of fuel cells within current range from I_{min} to I_{max} called EASY eLOAD, IU characterization of fuel cells long-term loaded by a constant current I_{const} called EASY eGalvanostat, and IU characterization of fuel cells long-term loaded by a constant voltage V_{const} called EASY eVoltstat.

The obtained characteristics of the electronic load have been examined by measuring a set current I_{fc} and by making a comparison of the voltage V_{fc} measured across the fuel cell by the microcontroller system and by a reference voltmeter. Fig. 3 shows block diagram of the system for measurement of the electronic load characteristics.

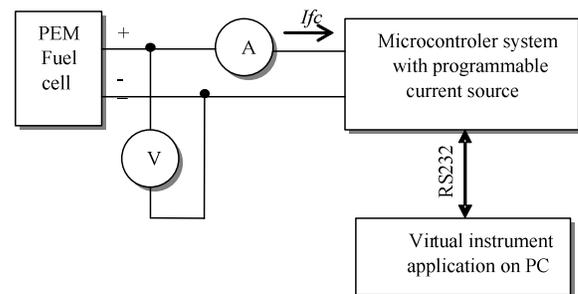


Fig. 3 Block diagram of the system for measurement of the electronic load characteristics.

The reference measurements of the current I_{fc} and voltage V_{fc} , which control operation of the electronic load, have been made by a Hp3457A multi-meter.

A PEM fuel cell has been tested. The Membrane Electrode Assembly (MEA) consists of two electrodes made of carbon paper having a thin layer of the catalyst deposited on the paper surface (0.2mg/cm² of Pt), Nafion 117 membrane, and all pressed to make a compact whole.

Fig. 4 shows a photograph of the measurement system associated with the electronic load developed for testing the characteristics of PEM fuel cells during the process of verification of the properties of the electronic load by reference measurements.

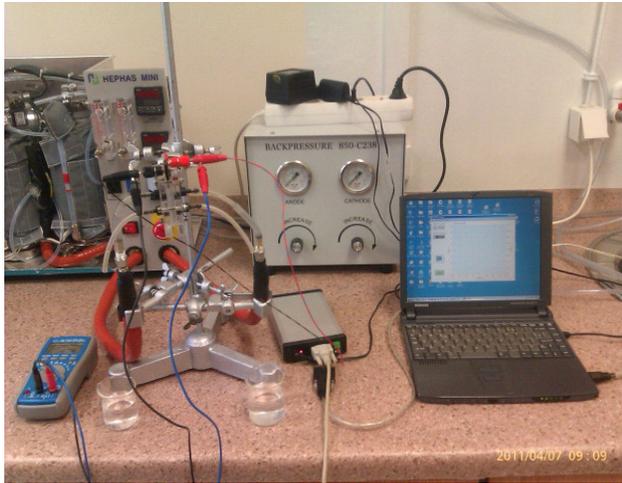


Fig. 4 Measurement system of the electronic load in the process of testing PEM fuel cells.

Fig. 5 shows the application of the virtual instrument during measurements of a PEM fuel cell intended for verification of the electronic load properties

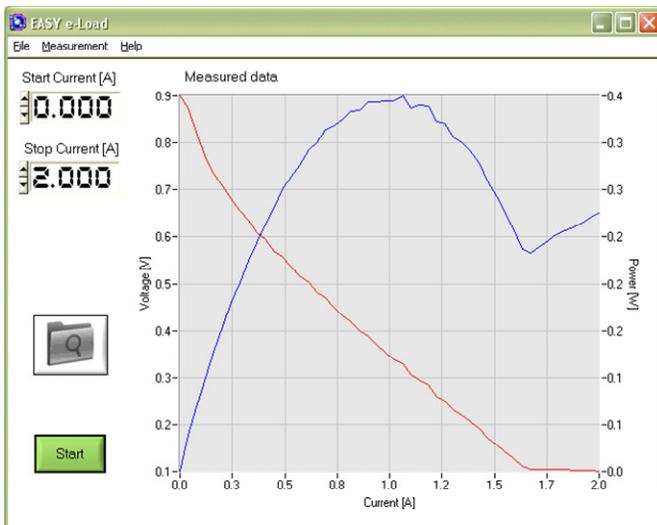


Fig. 5 Application of the virtual instrument of the electronic load.

The obtained measurement results are presented in Table 1. The table shows values of the pre-assigned current I_{set} , the set current I_{fc} measured by the electronic load, the voltage across fuel cell V_{fc} measured by the electronic load, the power P_{fc} measured by the electronic load, the set current I_{ref} measured by the reference meter, the voltage across fuel cell V_{ref}

measured by the reference meter, and the power P_{ref} measured by the reference meter.

TABLE I
RESULTS OF THE MEASUREMENTS OF THE CURRENT, VOLTAGE, AND POWER OF THE ELECTRONIC LOAD OF A PEM FUEL CELL

I_{set}	I_{fc}	V_{fc}	P_{fc}	I_{ref}	V_{ref}	P_{ref}
0,2	0,2343	0,688	0,1612	0,2338	0,6912	0,1616
0,4	0,4296	0,585	0,2513	0,4288	0,5906	0,2532
0,6	0,625	0,504	0,315	0,618	0,5055	0,3124
0,8	0,8203	0,426	0,3494	0,818	0,4289	0,3508
1	1,0156	0,353	0,3585	1,0087	0,3548	0,3579
1,2	1,2109	0,281	0,3402	1,2058	0,2841	0,3426
1,4	1,4062	0,216	0,3037	1,402	0,2202	0,3087
1,6	1,6015	0,143	0,229	1,5989	0,1436	0,2296
1,8	1,7968	0,136	0,2443	1,792	0,1388	0,2487
2	1,9983	0,133	0,2657	1,9922	0,1345	0,2680

On the basis of the presented results one may conclude that the electronic load shows satisfactory characteristics of the set current I_{fc} and measured voltage V_{fc} , therefore it can be used for testing of fuel cells and their constituent parts.

V. CONCLUSION

The realized system of electronic load intended for testing and characterization of fuel cells is based on a microcontroller system and the concept virtual instrument and represents a modern engineering design meeting the requirements for accuracy of the set current and measured voltage across a fuel cell. The electronic load system allows a better insight in electro-chemical processes in a fuel cell and facilitates investigations in this area.

REFERENCES

- [1] A.B. Stambouli, E. Traversa, Fuel cells, an alternative to standard sources of energy, *Renewable and Sustainable Energy Reviews*, 6 (2002) 295–304.
- [2] M. Cifrain, K.V. Kordesch, *Handbook of Fuel Cells*, John Wiley, 2003.
- [3] GARY W. JOHNSON, RICHARD JENNINGS, "LABVIEW GRAPHICAL PROGRAMMING", MCGRAW-HILL PROFESSIONAL, 2006.
- [4] PIC16F877 Data Sheet, 40-Pin 8-Bit CMOS FLASH Microcontroller, Microchip Technology Inc, 2001.
- [5] EASY PIC 5, Mikroelektronika Development Tool for Microchip PIC microcontrollers, Version 5.0, Mikroelektronika, Belgrade, 2009.
- [6] MicroC Pro Advanced C Compiler for PIC MCU, Mikroelektronika C Compiler for Microchip PIC microcontrollers, Version 3.2, Mikroelektronika, Belgrade, 2009

Robust RST Controller Design by Convex Optimization

Milan S. Matijević, Ranko Sredojević and Vladimir M. Stojanović

Abstract—A convex optimization-based robust RST controller design approach is proposed in this paper. It is shown that the system specifications given as the Nyquist robust stability criteria, absorption of effects for different types of external disturbances and the pole placement problem for LTI systems can be transformed into convex constraints. The controller design problem is then reduced to solving the standard quadratically-constrained convex program. The analysis also illustrates the potential limitations of the method through inherent tradeoff between robustness stability and performance specifications. The design procedure is illustrated on a practical industrial problem, showing that the proposed method can provide very robust solutions with good performance.

Index Terms—Control design, Digital control, Optimization

I. INTRODUCTION

General procedure to design and tune a good controller is [1]: I) To specify the desired control-loop performances; II) To obtain a dynamic model of the plant to be controlled (e.g. from real data by identification); III) To develop a suitable controller design methodology, compatible with the desired performances and the corresponding plant model; IV) To have a procedure for controller validation and onsite re-tuning; V) To develop and implement software packages with real-time capabilities for data acquisition, system identification, control design and on-site commissioning. Energy and material savings as well as improvement in the quality of the products should be a result of a well designed control system.

This paper presents a new approach for the design of the RST controller, which can be shown to cover all the linear control laws for linear SISO system [1]. In previous work, the pole placement methodology for synthesis of linear SISO systems produced the RST controller in the final step [2] as a

This work was supported in part by the Fulbright Grant and the Ministry of Education and Science of Republic Serbia, under projects TR 33022 and TR 33047.

M.S.Matijević is with the University of Kragujevac, Faculty of Mechanical Engineering, Kragujevac, Serbia (e-mail: matijevic@kg.ac.rs). During 2010/11 school year he is Fulbright Scholar at MIT, Cambridge, MA 02139, USA.

R. Sredojević is with MIT, Cambridge, MA 02139, USA. He is now with the Integrated Systems Group, RLE, Department of EECS at MIT, Cambridge, MA02139 (e-mail: rasha@mit.edu).

V.M Stojanović is with MIT, Cambridge, MA 02139, USA. He is head of the Integrated Systems Group, RLE, Department of EECS at MIT, Cambridge, MA02139 (e-mail: vlada@mit.edu).

solution of one or more of the Diophantine equations [2]-[5]. However, Diophantine equations do not have a unique solution, and different possible solutions of the RST controller parameters have different implications related to the control objectives [2],[3]. To choose the RST polynomials that better fit the control system requirements can be a very difficult numerical problem, especially in auto- and self-tuning control systems. Because of these difficulties general RST controller design for industrial applications is still a challenge [7],[5].

In this paper, we develop a procedure for design of robust RST controllers based on the use of convex optimization. We develop methods to turn the Diophantine equations and robust stability specifications into convex constraints, and formulate the RST controller problem as a quadratically-constrained convex feasibility problem that can be solved very efficiently on regular compute hardware. We also utilize the absorption principle to specify the control objectives for the steady-state tracking trajectory in the presence of disturbances. The methodology is illustrated through an example of the controller design for the flexible coupled motor servo drive with load [12].

II. PROBLEM FORMULATION AND DESIGN STEPS

In the RST control structure shown in Fig.1, the plant is described by its pulse transfer function or by polynomials $B(z^{-1})$ and $A(z^{-1})$, the control structure is given by polynomials $R(z^{-1})$, $S(z^{-1})$ and $T(z^{-1})$, r is the reference signal, and signals d and v model the influence of external disturbances and noise on the system output y .

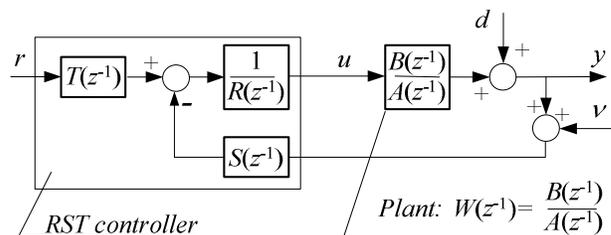


Fig.1 RST control structure.

The nominal model of the plant in Fig.1 is given by

$$W^0(z^{-1}) = \frac{B^0(z^{-1})}{A^0(z^{-1})}. \quad (1)$$

The uncertainties of the plant modeling may be adequately described by a multiplicative bound $\alpha(\omega)$ [2],[6]

$$W(z^{-1}) = W^0(z^{-1})(1 + \delta W(z^{-1})) \quad (2)$$

$$|\delta W(e^{-j\omega T})| \leq \alpha(\omega), \quad \omega \in [0, \pi/T], \quad (3)$$

where $W(z^{-1})$ represents the actual plant behavior.

Under the nominal conditions ($B(z^{-1}) \equiv B^0(z^{-1})$ and $A(z^{-1}) \equiv A^0(z^{-1})$), the closed-loop transfer functions $y(z^{-1})/r(z^{-1})$, $y(z^{-1})/d(z^{-1})$ and $y(z^{-1})/v(z^{-1})$ are easily derived from Fig. 1 as

$$G_{ry}(z^{-1}) = \frac{y(z^{-1})}{r(z^{-1})} = \frac{T(z^{-1})B^0(z^{-1})}{A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1})} \quad (4)$$

$$G_{dy}(z^{-1}) = \frac{y(z^{-1})}{d(z^{-1})} = \frac{A^0(z^{-1})R(z^{-1})}{A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1})} \quad (5)$$

$$G_{vy}(z^{-1}) = \frac{y(z^{-1})}{v(z^{-1})} = -\frac{B^0(z^{-1})S(z^{-1})}{A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1})}. \quad (6)$$

The system set-point response can be adjusted according to desired system closed-loop transfer function (pole placement problem [2])

$$G_{de}(z^{-1}) = \frac{y(z^{-1})}{r(z^{-1})} = \frac{B_{de}(z^{-1})}{A_{de}(z^{-1})}. \quad (7)$$

Namely,

$$\frac{T(z^{-1})B^0(z^{-1})}{A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1})} = \frac{B_{de}(z^{-1})A_0(z^{-1})}{A_{de}(z^{-1})A_0(z^{-1})} = \frac{N_{pid}(z^{-1})}{K_{pid}(z^{-1})}, \quad (8)$$

where $A_0(z^{-1})$ is the so-called observer polynomial [2]. The condition of the desired dynamical behavior of the nominal closed-loop system is then given by Diophantine equation

$$A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1}) \equiv K_{pol}(z^{-1}). \quad (9)$$

In order to specify the desired steady-state behavior of the system, according to Fig. 1, we use the absorption principle [6] and derive the error signal as

$$e = r - y = r - \frac{TB^0}{A^0R + B^0S}r - \frac{B^0S}{A^0R + B^0S}v - \frac{A^0R}{A^0R + B^0S}d \quad (10)$$

where, for the sake of brevity, variable arguments are omitted from notations of variables.

The absorption principle has the intention to embed the disturbance model into the control algorithm in order to suppress or reject the influence of the disturbance on the steady-state value of the process output. In the case of regular disturbances belonging to a certain class [6]:

$$f \in \{\Phi(z^{-1})f(k) = 0, \text{ for } k \geq (\deg \Phi)\}, \quad (11)$$

we will call the polynomial $\Phi(z^{-1})$ the absorbing polynomial.

The absorption filter $\Phi(z^{-1})$ is designed for a known class of disturbances and can be simply resolved as [3]

$$\Phi(z^{-1}) = D_w(z^{-1}), \text{ from } f(z^{-1}) = \frac{N_w(z^{-1})}{D_w(z^{-1})}. \quad (12)$$

In the case of a stochastic disturbance $d(t)$, the absorption filter should suppress as much as possible the effects of the disturbance on the system output. Thus, for a low frequency disturbance $d(t)$, which can be generated by double integration of the white noise, an appropriate choice of the absorption filter is $\Phi_d(z^{-1}) = (1 - z^{-1})^2$ that corresponds to absorption of a linear (ramp) disturbance [6].

The absorption conditions of the steady-state influence of the external disturbance $d(t)$ and the reference input $r(t)$ on the error tracking system signal (10) are derived from (11) and (10) as

$$A^0(z^{-1})R(z^{-1}) - \Phi_d(z^{-1})M_d(z^{-1}) \equiv 0 \quad (13)$$

$$B^0(z^{-1})T(z^{-1}) + \Phi_r(z^{-1})M_r(z^{-1}) \equiv K_{pol}(z^{-1}) \quad (14)$$

Then, the system of Fig. 1 satisfies the condition of robust stability if the nominal plant is stable and the following inequality holds [2]6

$$\alpha(\omega) \leq \left| \frac{A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1})}{B^0(z^{-1})S(z^{-1})} \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \cdot \quad (15)$$

By taking into account (9) the condition of robust stability (15) can be transformed to

$$\left| S(z^{-1}) \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \leq \frac{1}{\alpha(\omega)} \left| \frac{K_{pol}(z^{-1})}{B^0(z^{-1})} \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \quad (16)$$

Suppression of disturbance and noise effects on the system output can also be expressed by relations

$$\left| \frac{A^0(z^{-1})R(z^{-1})}{A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1})} \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \leq \left| G_{dy}^{desired}(z^{-1}) \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \quad (17)$$

$$\left| \frac{B^0(z^{-1})S(z^{-1})}{A^0(z^{-1})R(z^{-1}) + B^0(z^{-1})S(z^{-1})} \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \leq \left| G_{vy}^{desired}(z^{-1}) \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \quad (18)$$

where

$$\left| G_{dy}^{desired}(z^{-1}) \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]}, \left| G_{vy}^{desired}(z^{-1}) \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]}$$

are the desired or required transfer functions whose magnitudes at all frequencies should be as small as possible. From (18), the condition of noise suppression can be expressed as

$$\left|S(z^{-1})\right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \leq \left|G_{vy}^{desired}\right| \left|\frac{K_{pol}(z^{-1})}{B^0(z^{-1})}\right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \quad (19)$$

Given the above formulation, the controller design procedure can be formulated in the following steps: 1) Define the desired characteristic polynomial $K_{pol}(z^{-1})$, frequency function $G_{vy}^{desired}(\omega)$ $\omega \in [0, \pi/T]$, and the desired absorption filters $\Phi_r(z^{-1})$ and $\Phi_d(z^{-1})$ based on *a priori* information about signals $d(t)$ and $r(t)$; 2) Identify the plant model yielding $A^0(z^{-1})$, $B^0(z^{-1})$ and $\alpha(\omega)$; 3) Check whether the control specifications are realistic. If not - redesign the specifications under 1) and/or do a more accurate identification procedure under 2); 4) Solve the system of three Diophantine equations (9), (13), (14) with inequalities (16) and (19) to obtain the controller polynomials R , S and T .

Challenges and tradeoffs involved in choosing the right set of control specifications are discussed in [1]-[3], and in general represent nontrivial design decisions. A possible choice of characteristic polynomial $K_{pol}(z^{-1})$ [6] is given as

$$K_{pol}(z^{-1}) = \prod_{i=1}^n (1 - b_i z^{-1}), \quad 0 \leq b_i \leq 0.9, \quad (20)$$

which corresponds to a strictly aperiodic closed-loop system step response. Smaller values of n and b_i correspond to higher speed of the system response and lower degree of system robustness. Thus in tuning of n and b_i , it is necessary to start with a certain value of n and smaller values of b_i and then to increase b_i gradually. If for the allowable values of b_i the desired criterions are not satisfied, the value of n should be increased to the next integer and so on.

Another desired criterion is a condition for disturbance suppression (17), which can be reformulated as

$$\left|R(z^{-1})\right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \leq \left|G_{dy}^{desired}(z^{-1})\right| \left|\frac{K_{pol}(z^{-1})}{A^0(z^{-1})}\right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]} \quad (21)$$

While directly specifying $G_{dy}^{desired}$ is desirable from user perspective, it can be overly constraining in practice due to the inherent tradeoff with the robustness criterion (15), leading often to infeasible optimization problem formulations. Instead, to allow additional degrees of freedom for the optimization problem, while retaining some performance guarantees, we specify the steady-state error behavior utilizing the absorption principle as shown in (13) and (14).

To further illustrate this inherent tradeoff we can write starting from (9), (13), (17),

$$\begin{aligned} |K_{pol}| &= |A^0 R + B^0 S| \leq |R| |A^0| + |S| |B^0| \\ &\leq |\Phi_d M_d| + \frac{|K_{pol}|}{\alpha} = |G_{dy}| |K_{pol}| + \frac{|K_{pol}|}{\alpha} \leq |G_{dy}^{desired}| |K_{pol}| + \frac{|K_{pol}|}{\alpha}, \end{aligned} \quad (22)$$

where the first inequality is the triangle inequality and the second and third follow from imposing constraints in (16) and (21). Simplifying, by dividing with non-zero polynomials, we see that (wherever K_{pol} is non-zero, which is almost everywhere as the set of zeros is finite for any polynomial) we must have

$$1 \leq \frac{|\Phi_d M_d|}{|K_{pol}|} + \frac{1}{\alpha} \leq |G_{dy}^{desired}| + \frac{1}{\alpha} \quad (23)$$

at every discrete frequency. This means that frequencies where our model's uncertainty is high are also critical for disturbance rejection. Also, from (23)

$$\alpha(\omega) \leq \frac{1}{1 - \left| \frac{\Phi_d(z^{-1}) M_d(z^{-1})}{K_{pol}(z^{-1})} \right|_{z^{-1}=e^{-j\omega T}, \omega T \in [0, \pi]}} \quad (24)$$

Relations (23) and (24) represent a useful check if the given performance and robustness specifications are not simultaneously achievable.

The control design problem given by (9), (13), (14), (16) and (19) can be described as a convex optimization problem. Unlike a pole placement problem [2], this problem formulation enables us to look at a broader range of solutions to Diophantine equations, while still constrained by the robustness stability and noise suppression criterions, leading to potentially better controller performance.

III. CONVEX OPTIMIZATION AS A DESIGN TOOL

In this section we show how the previously described design specifications of the robust RST controller can be transformed into a standard convex program. This step enables us to use readily available optimization software for efficient resolution of design constraints and obtaining the desired RST parameters. In our case the resulting convex programming formulation is a quadratically constrained (QC) feasibility problem.

In order to use the available convex optimization packages, a design problem must be convex and formulated as one of the standard optimization programs [10],[11]. In general, convex optimizations are problems of the following form

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, i = 1, \dots, k, \\ & Ax = b \end{aligned} \quad (25)$$

where the optimization objective (cost), $f_0(x)$, must be a convex function. All the constraints must be convex sets

Finally, we have not included any performance metric in our robust RST model, thus we can trivially assign $f_0(x) = 0$, to complete our optimization program. In general, however, we could define a (convex) function to measure the quality of the controller to be optimized.

From the previous we see that, once the desired performance is specified in a convenient form, synthesis of the RST controller that ensures robust stability to the plant uncertainty, certain level of performance in terms of disturbance and noise suppression and internal consistency of the model can be formulated as a QC optimization problem.

In view of relations (22), (23) and (24) we should note that it might happen that the optimization has no valid solutions for a particular set of design specifications. Unfortunately, due to necessity but not sufficiency of these constraints there is no simple procedure to ensure that the desired controller will exist so that the optimization has a feasible solution. Although a number of obviously infeasible formulations can be filtered-out by using the relations (22)-(24), a significant number of possibly infeasible specifications still exist, leaving this an open research question.

Finally we should note that the described optimization model of the RST controller is by no means a complete one. For example we might want to impose additional constraints to ensure that the solver only allows for stable R , S and T polynomials. This is known to be a non-convex constraint if the sets of all the stable polynomials are considered. However, certain convexifications (that basically reject some solutions in order to construct a convex, inner approximation of the set of stable polynomials) and parameterizations can be used to include such constraints in optimization design flow. Currently the stability of the solution is checked *ex post facto* and improving this aspect of the model is the next logical step in this line of work.

IV. DESIGN EXAMPLE

The proposed controller synthesis method is shown here for the class of motion control systems with flexible coupling. Our aim is to control the load shaft speed in the presence of torsion vibrations, system parameter variations, disturbance torque, and in the absence of a dedicated load side speed sensor.

Note that many controllers already exist in the field of motion control, but most of them are designed by assuming an ideal, rigid transmission train, adopting $W^0(s) = 1/(J_m + J_l)s$ (J_m, J_l - motor and load inertia) as a simple plant model. As an actual plant model we utilize a model of flexible coupling of motor axle and load [12]. We distinguish the following important data (see Fig. 2) $J_m=0.000620 \text{ kgm}^2$, $J_l=0.000220 \text{ kgm}^2$, $c_s=350 \text{ Nm/rad}$, $b_v=0.004 \text{ Nms/rad}$.

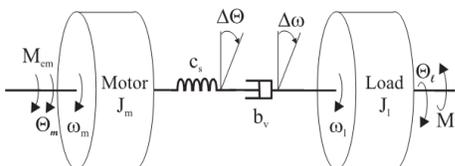


Fig. 2. Flexible coupling of motor axle and load.

The desired close-loop system transfer function is specified by undamped natural frequency $\omega_n=400 \text{ rad/s}$ and relative damping coefficient $\zeta=0.7$. The sample rate is $T = 0.5 \text{ ms}$.

Adopted absorption polynomials are $\Phi_r = 1 - z^{-1}$, $\Phi_d = (1 - z^{-1})^2$, which fit the step and ramp disturbances.

This formulation of the RST controller design can be solved within minutes, for couple of hundreds sampling points on the unity circle, on standard 2GHz personal computer with 2GB of RAM and running Ubuntu Linux. Our particular implementation of this optimization is done through YALMIP [10] parser and solved with SeDuMi [11] SDP solver.

With the above input specifications the solver's calculation of polynomials R , S and T are: $R(z^{-1})=0.64-0.77z^{-1}+0.395z^{-2}-0.264z^{-3}$, $S(z^{-1})=0.5-1.1z^{-1}+1.034z^{-2}-0.41z^{-3}$ and $T(z^{-1})=0.039z^{-1}$. The designed system has met control specifications as illustrated by simulation results in Figs. 3 and 4, for a system reference $\omega_r(t)=3h(t-0.05) \text{ rad/s}$, and a disturbance via load torque $M_l = 1h(t-0.1) \text{ Nm}$ as in [12]. The disturbance effect d (Fig. 1) for a nominal plant ($W^0, d^0 = W^0 M_l$) is a ramp, while for the real plant in Fig. 2, the torque disturbance manifests itself as the output disturbance $d \equiv W(s)M_l$, which is a ramp with superposed quasi-oscillation at plant resonance frequency (1468 rad/s). The results encompass the time response of the nominal system, and the motor and load speed time responses of realistic plant model structure in Fig. 2. Figure 3 shows that the RST controller has satisfied the robust stability.

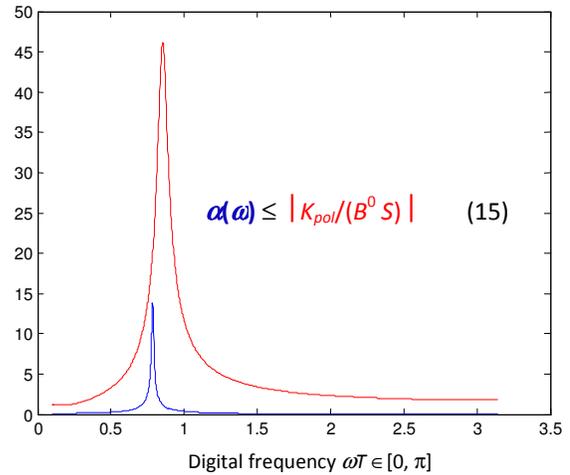


Fig.3. Illustration of the robust stability test.

Multiplicative bound of uncertainties $\alpha(\omega)$ in Fig. 3 is calculated from the realistic and nominal plant models. The successful design of the robust stability and the width of the robustness region are obvious advantages of the proposed design method. The control of the plants with uncertainties at resonant frequencies is a very difficult task in general [6] and control methods for flexible drives mostly do not address these robust stability issues [3], [12].

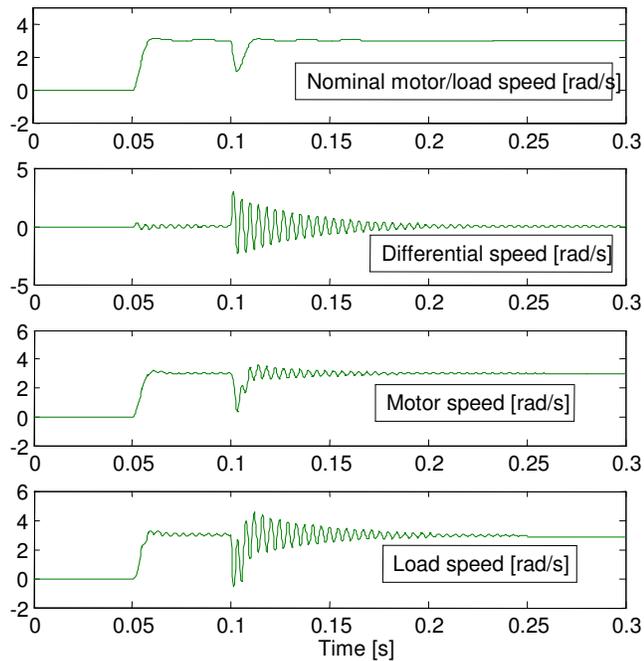


Fig.4. Simulation of the RST controller design efficiency.

V. CONCLUSION

Developing design methodology for an RST controller is a very attractive task since this topology covers the whole space of linear discrete SISO controllers. While some well-known subsets, like PID controller, have been heavily used in industrial applications, other, potentially more-efficient designs have not transitioned yet into industrial settings, mainly due to the lack of a computationally efficient and robust design methodology. The main challenges lie in controller tuning and plant identification procedures and their realization in industrial environment.

In this paper we presented a methodology for design of robust RST controllers, which utilizes the computational efficiency of convex optimization to find the controller designs

from a larger set of possible solutions than considered previously in the literature. The convex relaxations allow the search space to be still efficiently constrained by the robust stability and steady-state disturbance and noise rejection specifications, yielding good robustness performance at critical resonant frequencies.

REFERENCES

- [1] I. Landau, "The R-S-T digital controller design and applications", *Control Eng. Practice*, Vol.6, 155-165, 1998.
- [2] K. Astrom and B. Wittenmark, *Computer Controlled Systems. Theory and Design*, Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [3] M.S. Matijevic, *Development of new structures of digitally controls electromotor drives and industrial processes*, PhD thesis, Faculty of Mechanical Engineering, Kragujevac, 2001. (in Serbian)
- [4] E. Ostertag, E. Godoy, RST-Controller Design for Sinewave References by Means of an Auxiliary Diophantine Equation, *44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05*, 12-15 Dec. 2005
- [5] H. Gharsallaoui, M. Ayadi, M. Benrejeb, P. Borne, Flatness-based control and conventional RST polynomial control of a thermal process, *Int. J. of Computers, Communications & Control*, ISSN 1841-9836, E-ISSN 1841-9844, Vol. IV (2009), No. 1, pp. 41-56
- [6] Ya.Z. Tsytkin and U. Holmberg, "Robust stochastic control using the internal model principle and internal model control" *Int. J. Control*, vol. 61, N^o4, P 809-822, 1995.
- [7] J. Langer, I.D. Landau, "Combined pole placement/sensitivity function shaping method using convex optimization criteria", *Automatica*, 35, P 1111-1120, 1999.
- [8] S.P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [9] S.-P. Wu, S. Boyd, and L. Vandenberghe, "FIR filter design via semidefinite programming and spectral factorization," *Proc. IEEE Conf. on Decision and Control*, 1996, pp. 271-276.
- [10] J. Löfberg, "YALMIP : A Toolbox for Modeling and Optimization in {MATLAB}," *In Proceedings of the CACSD Conference (2004)* , Taipei, Taiwan: 2004.
- [11] J.F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," 1998
- [12] S.N. Vukosavic and M.R. Stojic, „Suppression of torsional oscillations in a high-performance speed servo drive“, *IEEE Trans. on Industrial Electronics*, vol. 45, N^o1, 108-117, 1998.

PI-like Observer Structures in Digitally Controlled DC Servo Drives: Theory and Experiments

Milica B. Naumović

Abstract—This paper deals with the problem of the shaft speed estimation in a digitally controlled DC servo drive. Some different observer structures are presented and compared. The developed extended observers enable proper estimation of the plant state variables, even under the action of a constant or slow varying load torque disturbance. Two useful procedures for the adjustment of observer gains are proposed and verified by both numerical simulations and real-time experimental results.

Index Terms—Shaft speed estimation, Extended observer, Digitally controlled DC servo drive.

I. INTRODUCTION

High-performance servo drives are required in many applications of digitally controlled machines. Two types of position sensors are most frequently applied: optical encoders (absolute or incremental), and electromagnetic resolvers (inherently absolute) [1], [2]. The transducer output signal is used as the position feedback signal in a position-controlled system; consequently, the signal in the inner velocity loop must be estimated. In speed-controlled high-performance servo drives, the feedback velocity signal is to be estimated from the torque command and measured angular position of the motor shaft, in the presence of the quantization noise and a constant or slow varying load torque disturbance. In order to obtain smooth and sufficiently accurate position and speed signals, the observer structure is often implemented.

This paper deals with the analysis and design of a dynamic system that is able to estimate state variables (position and speed signals) in an environment where the shaft position information is incomplete due to a limited resolution of the position transducer, even in the presence of a constant load torque disturbance. After comparing several different methods of velocity estimation, in this paper a novel approach to the extending of the discrete-time observer is proposed.

This paper is organized as follows. The problem formulation is given in Section II. Section III presents the design procedures of the observers extended by using the additionally introduced integral terms in a digitally controlled

servo drive. Procedures for calculating the observer gains are given in Section IV. Section V gives a concrete example to demonstrate the effectiveness of the proposed observers. Finally, Section VI presents the concluding remarks.

II. PROBLEM FORMULATION

In all cases, no matter what type of sensor is utilized in the digitally controlled servo system, the shaft position is read as a digital signal. Hence, the resolution of the shaft position measurement is limited. Due to the finite resolution, the actual shaft position differs from the digital word representing the position (lower resolution - the larger difference).

To estimate velocity signal, the least complicated algorithm yields

$$\hat{\omega}(k) = \frac{\theta(k) - \theta(k-n)}{nT}, \quad (1)$$

where T is sampling period, θ is the angular position of drive shaft, k is the sample number index, and integer $n \geq 1$. By setting $n = 1$ the average velocity over the preceding sampling interval is estimated as the well-known Euler's approximation of the derivative, that is a simple first difference. Note that the velocity resolution is limited directly by the transducer resolution and the time interval nT .

Due to the finite resolution of the angle measurement, the shaft speed signals estimated by (1) would be highly contaminated by the quantization noise. In order to improve the quality of the shaft velocity estimation, an observer structure is often implemented. Besides enabling an accurate state estimation of the control object, the applied observer can be used also for filtering the measurement noise.

Consider the discrete-time model of the plant

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{E}(T)\mathbf{x}(k) + \mathbf{F}(T)\mathbf{u}(k) \\ \mathbf{c}(k) &= \mathbf{D}\mathbf{x}(k) + \mathbf{H}\mathbf{u}(k) \end{aligned} \quad (2)$$

where $\mathbf{x}(k) = \mathbf{x}(kT) \in \mathbb{R}^n$ is the state vector to be observed, $\mathbf{u}(k) = \mathbf{u}(kT) \in \mathbb{R}^r$ and $\mathbf{c}(k) = \mathbf{c}(kT) \in \mathbb{R}^m$ are known control input vector and output vector, respectively. The sampling interval is T . Constant matrices \mathbf{E} , \mathbf{F} , \mathbf{D} and \mathbf{H} have appropriate dimensions; the pair (\mathbf{E}, \mathbf{F}) is controllable and the pair (\mathbf{E}, \mathbf{D}) is observable.

The observer or asymptotic state estimator is a dynamic

This paper is supported in part by project Grant III44004 (2011-2014) financed by Ministry of Education and Science, Republic of Serbia.

M. B. Naumović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (phone: +381 18 529-441; fax: +381 18 588-399; e-mail: milica.naumovic@elfak.ni.ac.rs).

system, with inputs $\mathbf{u}(k)$, $\mathbf{c}(k)$ and output $\hat{\mathbf{x}}(k)$, having the property $\lim_{k \rightarrow \infty} \tilde{\mathbf{x}}(k) = \mathbf{0}$, where

$$\tilde{\mathbf{x}}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k) \quad (3)$$

is the estimation error.

The system plant and the associate observer are shown in Fig. 1. We adopt the notation

$$\mathbf{G}(z) := \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{D} & \mathbf{H} \end{bmatrix}. \quad (4)$$

where

$$G(z) = \mathbf{D}(z\mathbf{I} - \mathbf{E})^{-1} \mathbf{F} + \mathbf{H}, \quad z - \text{complex variable}, \quad (5)$$

is the transfer function matrix derived from (2).

The state variables reconstructed by the observer can be used by the control law

$$\mathbf{u}(k) = -\mathbf{K} \hat{\mathbf{x}}(k). \quad (6)$$

Notice that the observer has a special structure

$$\hat{\mathbf{x}}(k+1) = \underbrace{\mathbf{E} \hat{\mathbf{x}}(k) + \mathbf{F} \mathbf{u}(k)}_{\text{a priori estimate}} + \mathbf{G} \underbrace{\{\mathbf{c}(k) - [\mathbf{D} \hat{\mathbf{x}}(k) + \mathbf{H} \mathbf{u}(k)]\}}_{\text{output estimation error}} \quad (7)$$

$$\hat{\mathbf{x}}(k) = \mathbf{x}_0(k).$$

In order to obtain a state-space description for the system in Fig. 1 supplemented with the state feedback (6), we define the state vector to be $[\mathbf{x}(k) \mid \tilde{\mathbf{x}}(k)]^T$, giving the equation for the closed-loop regulator system as follows

$$\begin{bmatrix} \mathbf{x}(k+1) \\ \tilde{\mathbf{x}}(k+1) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{E} - \mathbf{F}\mathbf{K} & \mathbf{F}\mathbf{K} \\ \mathbf{0} & \mathbf{E} - \mathbf{G}\mathbf{D} \end{bmatrix}}_{\mathbf{E}_c} \begin{bmatrix} \mathbf{x}(k) \\ \tilde{\mathbf{x}}(k) \end{bmatrix}. \quad (8)$$

Because the matrix \mathbf{E}_c is block upper-triangular, its eigenvalues are just the eigenvalues of the diagonal blocks. Thus we have that

$$\text{eig}(\mathbf{E}_c) = \text{eig}(\mathbf{E} - \mathbf{F}\mathbf{K}) \cup \text{eig}(\mathbf{E} - \mathbf{G}\mathbf{D}), \quad (9)$$

where the symbol \cup denotes the union. Equation (9) is known as the separation principle, which refers to the fact that the state feedback vector and the observer gains vector can be calculated separately [3]-[5].

In the case $\mathbf{H} = \mathbf{0}$ we can rewrite (7) as follows

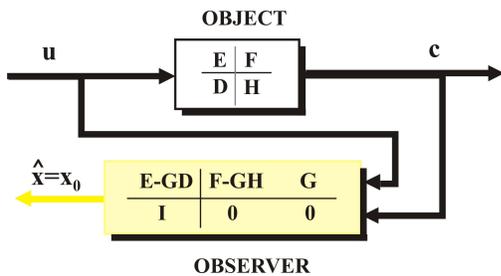


Fig. 1. Control Object and Observer.

$$\hat{\mathbf{x}}(k+1) = (\mathbf{E} - \mathbf{G}\mathbf{D}) \hat{\mathbf{x}}(k) + \mathbf{F}\mathbf{u}(k) + \mathbf{G}\mathbf{c}(k), \quad (10)$$

where all state values are estimated with the resolution limited only by the word length of the digital controller. The observer gain matrix \mathbf{G} in (7) is to be determined according to requirements for the desired speed of estimation.

Recall, for the sake of simplicity, instead of identity observer, the reduced-order one is proposed to estimate only the unmeasured states. In the case of reduced-order observer design, the object model, like (2), becomes

$$\begin{bmatrix} \mathbf{x}_a(k+1) \\ \mathbf{x}_b(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{aa} & \mathbf{E}_{ab} \\ \mathbf{E}_{ba} & \mathbf{E}_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{x}_a(k) \\ \mathbf{x}_b(k) \end{bmatrix} + \begin{bmatrix} \mathbf{F}_a \\ \mathbf{F}_b \end{bmatrix} \mathbf{u}(k) \quad (11)$$

$$\mathbf{c}(k) = [\mathbf{I} \quad \mathbf{0}] \begin{bmatrix} \mathbf{x}_a(k) \\ \mathbf{x}_b(k) \end{bmatrix},$$

where $\mathbf{x}_a(k)$ is the state vector portion directly measured, and $\mathbf{x}_b(k)$ is the remaining portion to be estimated. Then the observer model takes the form as follows [3], [4]

$$\hat{\mathbf{x}}_b(k+1) = (\mathbf{E}_{bb} - \mathbf{G}\mathbf{E}_{ab}) \hat{\mathbf{x}}_b(k) + (\mathbf{E}_{ba} - \mathbf{G}\mathbf{E}_{aa}) \mathbf{c}(k) + \mathbf{G}\mathbf{c}(k+1) + (\mathbf{F}_b - \mathbf{G}\mathbf{F}_a) \mathbf{u}(k). \quad (12)$$

As it is well known, in the presence of constant or slow varying disturbances that may not be considered as state disturbances, both the identity observer (10) and the reduced-order observer (12) are not able to estimate the proper values of the state variables. Some modifications, that are based on the special extending the standard observer structures with the integral action [6]-[11], provide the correct estimation even under the disturbance action.

III. POSITION AND VELOCITY OBSERVATION IN A DIGITALLY CONTROLLED SERVO DRIVE

By assuming the state vector as $\mathbf{x}(k) = [\theta(k) \quad \omega(k)]^T$, where $\theta(k)$ and $\omega(k)$ are the shaft angular position and speed respectively, the state-space model of the object in Fig. 2 becomes

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{E}(T)\mathbf{x}(k) + \mathbf{f}(T)\mathbf{u}(k) \\ \mathbf{c}(k) &= \mathbf{d}\mathbf{x}(k) \end{aligned} \quad (13)$$

with

$$\mathbf{E}(T) = \begin{bmatrix} 1 & T_m(1 - e^{-T/T_m}) \\ 0 & e^{-T/T_m} \end{bmatrix} = \begin{bmatrix} 1 & e_1 \\ 0 & e_2 \end{bmatrix}, \quad (14)$$

$$\mathbf{f}(T) = \begin{bmatrix} K_m(T + T_m e^{-T/T_m} - T_m) \\ K_m(1 - e^{-T/T_m}) \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \quad \text{and} \quad \mathbf{d} = [1 \quad 0].$$

In above equations K_m and T_m are the gain factor and the mechanical time constant of the considered drive, respectively.

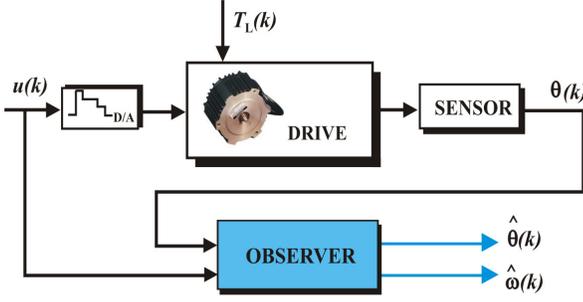


Fig. 2. Block diagram of a drive with observer.

The observer that joins the drive in Fig. 2 estimates the states $x_1(k)$ and $x_2(k)$ by using the control variable $u(k)$ and measured output $\theta(k)$ as inputs. Equations (10) and (12) of the ordinary identity observer and the reduced-order observer can be rewritten as

$$\begin{bmatrix} \hat{x}_1(k+1) \\ \hat{x}_2(k+1) \end{bmatrix} = \begin{bmatrix} 1-g_1 & e_1 \\ -g_2 & e_2 \end{bmatrix} \begin{bmatrix} \hat{x}_1(k) \\ \hat{x}_2(k) \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} u(k) + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} c(k) \quad (15)$$

and

$$\begin{aligned} \hat{x}_2(k+1) = & (e_2 - g_2 e_1) \hat{x}_2(k) + g_2 c(k+1) - g_2 c(k) \\ & + (f_2 - g_2 f_1) u(k). \end{aligned} \quad (16)$$

In numerous applications, as in robotics for example, the employed observer must enable the estimation of the plant state variables even under the action of the constant (gravitation) or a slow varying load torque disturbance $T_L(k)$. As it is well known [3], in the presence of such disturbances, that may not be considered as initial state disturbances, the ordinary observers (full or reduced-order) are not able to estimate the state variables.

For proper state estimation, the possibility of extending the ordinary discrete-time observer with additional integrators is discussed. The solution consists in the following: the observer gains multiply the generated errors of the state variables $\tilde{x}_i = x_i(k) - \hat{x}_i(k)$, $i = 1, 2$; the errors of position and velocity estimation are simultaneously processed through two discrete integrators assuring the zero steady-state estimation errors in the presence of a constant load torque disturbance T_L .

It is possible to choose the estimates of new state variables $\hat{x}_3(k)$ and $\hat{x}_4(k)$ to be the outputs of the inserted integrators resulting in the following discrete-time new state equations [7]

$$\hat{x}_3(k+1) = \hat{x}_3(k) + g_3 [c(k) - \hat{x}_1(k)] = -g_3 \hat{x}_1(k) + \hat{x}_3(k) + g_3 c(k) \quad (17)$$

and

$$\begin{aligned} \hat{x}_4(k+1) &= \hat{x}_4(k) + g_4 \left[\frac{c(k) - c(k-1)}{T} - \hat{x}_2(k) \right] \\ &= -g_4 \hat{x}_2(k) + \hat{x}_4(k) + \frac{g_4}{T} c(k) - \frac{g_4}{T} c(k-1). \end{aligned} \quad (18)$$

After adding new state variables $\hat{x}_3(k)$ and $\hat{x}_4(k)$ into state vector to be observed $\hat{\mathbf{x}}(k) = [\hat{\theta}(k) \ \hat{\omega}(k)]^T$, the vector difference equation of the identity observer (15) extended in the previously described manner becomes

$$\hat{\mathbf{x}}_e(k+1) = (\mathbf{E}_e - \mathbf{G}_{e1} \mathbf{D}_e) \hat{\mathbf{x}}_e(k) + \mathbf{f}_e u(k) + \mathbf{G}_{e2} \begin{bmatrix} c(k) \\ c(k-1) \end{bmatrix}, \quad (19)$$

where

$$\begin{aligned} \hat{\mathbf{x}}_e(k+1) &= \begin{bmatrix} \hat{\mathbf{x}}(k+1) \\ \hat{x}_3(k+1) \\ \hat{x}_4(k+1) \end{bmatrix}, \quad \mathbf{E}_e = \begin{bmatrix} 1 & e_1 & | & 1 & 0 \\ 0 & e_2 & | & 0 & 1 \\ \hline 0 & 0 & | & 1 & 0 \\ 0 & 0 & | & 0 & 1 \end{bmatrix}, \\ \mathbf{f}_e &= \begin{bmatrix} f_1 \\ f_2 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{D}_e = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \\ \mathbf{G}_{e1} &= \begin{bmatrix} g_1 & 0 \\ g_2 & 0 \\ g_3 & 0 \\ 0 & g_4 \end{bmatrix}, \quad \text{and} \quad \mathbf{G}_{e2} = \mathbf{G}_{e1} \begin{bmatrix} 0 & 0 \\ 1/T & -1/T \end{bmatrix}. \end{aligned} \quad (20)$$

Certainly, the described proportional double-integral observer (PI²O) to be devised by using two additionally introduced integral terms of the output estimation errors can offer some degrees of freedom in the observer design.

In the case of the reduced-order observer we must return to (16) and introduce a new state variable $\hat{x}_4(k)$ as the output of the discrete integrator by which the observer is extended. In such a way obtained the reduced-order proportional integral observer (PIO) is described by the following equations:

$$\begin{aligned} \hat{x}_2(k+1) &= (e_2 - g_2 e_1) \hat{x}_2(k) + \hat{x}_4(k) \\ &+ g_2 [c(k+1) - c(k) - f_1 u(k)] + f_2 u(k) \end{aligned} \quad (21)$$

$$\hat{x}_4(k+1) = -g_4 \hat{x}_2(k) + \hat{x}_4(k) + \frac{g_4}{T} [c(k) - c(k-1)].$$

IV. PROCEDURES FOR CALCULATING THE OBSERVER GAINS

A. A Procedure for Parameter Adjustment of Reduced-Order PI Observer

Recall, the unmeasured velocity variable can be properly estimated by the reduced-order PI observer even in the presence of constant or slow varying load torque disturbance

T_L acting on the drive of the system given in Fig. 2. The observer equations (21) can be rewritten in the form as

$$\begin{bmatrix} \hat{x}_2(k+1) \\ \hat{x}_4(k+1) \end{bmatrix} = \begin{bmatrix} e_2 - g_2 e_1 & 1 \\ -g_4 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_2(k) \\ \hat{x}_4(k) \end{bmatrix} + \begin{bmatrix} g_2 & 0 \\ 0 & g_4 \end{bmatrix} \begin{bmatrix} c(k+1) - c(k) - f_1 u(k) \\ c(k)/T - c(k-1)/T \end{bmatrix} + \begin{bmatrix} f_2 \\ 0 \end{bmatrix} u(k). \quad (22)$$

The observer gains g_2 and g_4 are calculated according to the desired quality and speed of state estimation. The parameter setting may be conveniently performed assuming both observer poles to be the same and equal to

$$\sigma_z = \exp(-2\pi f_0 T), \quad (23)$$

where f_0 denotes the observer bandwidth. Thus the characteristic equation of the extended observer given by (22) is

$$\det \begin{bmatrix} z - e_2 + g_2 e_1 & -1 \\ g_4 & z - 1 \end{bmatrix} = 0. \quad (24)$$

The desired characteristic equation, founded by multiplying out the observer poles, becomes

$$(z - \sigma_z)^2 = z^2 - 2\sigma_z z + \sigma_z^2 = 0. \quad (25)$$

Hence, the simple relations, obtained by matching coefficients of like powers of z in (24) and (25), yield two unknown gains of PI observer as follows

$$\begin{aligned} g_2 &= (1 + e_2 - 2\sigma_z)/e_1 \\ g_4 &= (1 - \sigma_z)^2. \end{aligned} \quad (26)$$

B. A Procedure for Adjustment of PI² Observer Parameters

Let us denote by

$$\Delta_0 = \det [z\mathbf{I} - (\mathbf{E}_e - \mathbf{G}_{e1}\mathbf{D}_e)] = z^4 + a_3 z^3 + a_2 z^2 + a_1 z + a_0 = 0 \quad (27)$$

the characteristic equation of the extended identity observer, whose model is given by (19)-(20), and may be written in the form

$$\begin{bmatrix} \hat{x}_1(k+1) \\ \hat{x}_2(k+1) \\ \hat{x}_3(k+1) \\ \hat{x}_4(k+1) \end{bmatrix} = \begin{bmatrix} 1 - g_1 & e_1 & 1 & 0 \\ -g_2 & e_2 & 0 & 1 \\ -g_3 & 0 & 1 & 0 \\ 0 & -g_4 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_1(k) \\ \hat{x}_2(k) \\ \hat{x}_3(k) \\ \hat{x}_4(k) \end{bmatrix} + \begin{bmatrix} f_1 & g_1 & 0 \\ f_2 & g_2 & 0 \\ 0 & g_3 & 0 \\ 0 & g_4/T & -g_4/T \end{bmatrix} \begin{bmatrix} u(k) \\ c(k) \\ c(k-1) \end{bmatrix}. \quad (28)$$

The observer gain matrix \mathbf{G}_{e1} in (20) is calculated according to the desired observer pole spectrum determining

the speed of convergence between the state of the system and the state estimated by the observer. The setting of gains $g_i, i=1,2,3,4$ may be conveniently performed assuming all observer poles to be the same and equal to $\sigma_z = \exp(-2\pi f_0 T)$, where f_0 denotes the observer bandwidth, as in (23). Therefore, the observer gains are calculated from

$$\det \begin{bmatrix} z - 1 + g_1 & -e_1 & -1 & 0 \\ g_2 & z - e_2 & 0 & -1 \\ g_3 & 0 & z - 1 & 0 \\ 0 & g_4 & 0 & z - 1 \end{bmatrix} = (z - \sigma_z)^4. \quad (29)$$

Finally, after equating coefficients of like powers of z , the following relationships may be written

$$\begin{aligned} g_1 &= 3 - 4\sigma_z + e_2 \\ e_1 g_2 + g_3 + g_4 &= 3 + 6\sigma_z^2 + (e_2 - 4\sigma_z)(e_2 + 2) \\ (1 - e_2)g_3 + (3 + e_2 - 4\sigma_z)g_4 &= 4(1 - \sigma_z)^3 \\ g_3 g_4 &= (1 - \sigma_z)^4. \end{aligned} \quad (30)$$

V. ILLUSTRATIVE EXAMPLE

In this section an example of shaft velocity estimation in a low power DC motor servomechanism is presented. The desired quality of the transient response of the considered system is matched by the conventional controller whose parameters are calculated by using the standard pole placement method. In view of the fact that it is well-known from classical control theory, the design procedure of the controller is not explained in detail. The experimental setup described in this section has been built to demonstrate proper estimation of the plant state variables, even in the case of the actions of constant or slow varying load torque disturbances. The simulation results of the proposed observers studied in the previous section are compared with the experimental ones. The goal is to illustrate that in all cases the estimated and the experimentally obtained results are good matched.

A. The Experimental Setup

Fig. 3 visualizes the structure of the experimental environment for rapid control prototyping that was realized during the PhD thesis research in Control Engineering Laboratory at the University of Niš, Faculty of Electronic Engineering [12]. The experimental setup of the servo system consists of several functional elements as follows: **1.** DC motor with incremental encoder, **2.** Personal computer upgraded to a powerful development system for rapid control prototyping, **3.** Led panel for signal connection, **4.** PWM power amplifier, and **5.** Power supply.

As control object, a DC motor Type Bautz E586MGB is used with the following rating [12]: $M_{\max} = 0.22 \text{ Nm}$, $I_{\max} = 3.7 \text{ A}$, $n_{\max} = 6000 \text{ min}^{-1}$, the voltage constant $k_e = 5.85 \text{ V/1000 min}^{-1}$, the torque constant $k_t = 0.056 \text{ Nm/A}$.

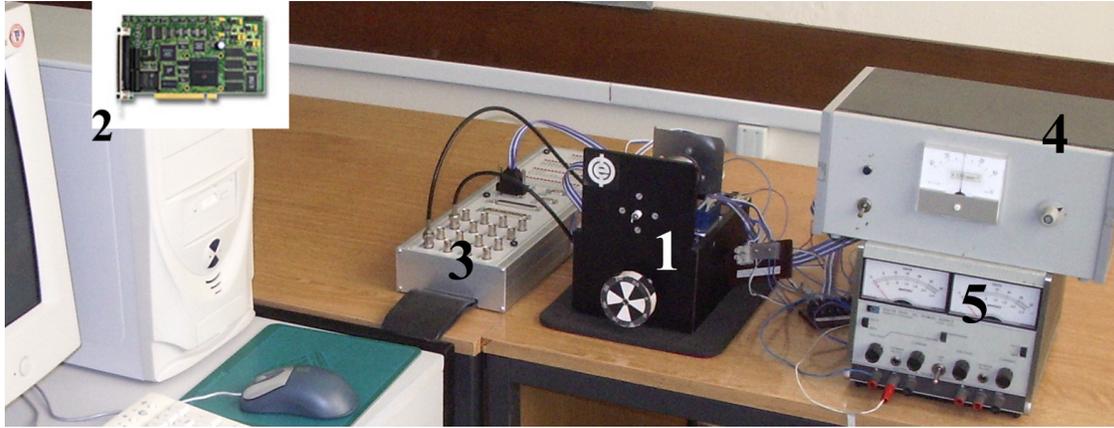


Fig. 3. Experimental setup [12].

For the position measurement $\theta(t)$ the quadrature incremental encoder with 1000 lines is used. The observer-based control algorithms are implemented by using the dSPACE system [13]. Namely, as a standard PC card, dSPACE DS1104 R&D Controller Board is slotted into PC using 5 V PCI bus as a backplane, upgrading in that way the personal computer to a powerful development system for rapid control prototyping.

The controller board is based on the Motorola 32-bit floating-point MPC 8240 processor with 250 MHz CPU, and contains all the necessary peripherals for AC and DC motor control in various fields. Moreover, because of demanding I/O operation, a slave DSP subsystem based on the Texas Instruments TMS320F240 16-bit fixed point digital signal processor with 20 MHz clock frequency is provided. Led panel CLP1104 indicates the status of the board's digital signals. Interfacing the computer DS1104 controller board with the control object input and output, as well as indication the status of the board's digital signals is possible through the LED panel CLP1104.

Using the information related to the measured angular position of the motor shaft $\theta(t)$ and the reference signal $\theta_r(t)$, the dSPACE system generates, based on the implemented control algorithm, a control signal which after PWM power amplifier with the carrier frequency of 15 kHz produces a voltage to the motor.

dSPACE Prototyper is a flexible development system that enables rapid control design of the real controlled system without manual programming. Namely, the dSPACE Real-Time Interface allows to implement the considered MATLAB[®]/Simulink model onto dSPACE hardware via code generated by Real-Time Workshop automatically. This software offers an application that makes observation of the processed variables in real time possible.

B. Simulation and Experimental Results

In this section the simulation and experimental results for the observers studied in the previous section are presented. To verify the usefulness of suggested procedures for setting of observer gains, ensuring the proper speed estimation of the drive given in Fig. 3, the system simulation has been carried out in all details, taking into account the limited resolution (the increment of $2\pi/4000$ rad) of position sensor.

The electrical subsystem dynamics of the motor and the inertial dynamics of the power amplifier can be neglected. The plant in this example is a type-1 servo with transfer function from input current to output angular position as follows

$$G_p(s) = \frac{K_m}{s(T_m s + 1)} \quad (31)$$

The motor's gain factor $K_m = 24.8$ and the mechanical time constant $T_m = 0.0379$ s are computed on the basis of the experimentally recorded open-loop step response given in Fig. 4. Note, that due to the finite resolution of angle measurement, the shaft speed signals estimated by (1), are contaminated by the quantization noise, especially in the case $n = 1$.

The sampling period $T = 0.001$ s was adopted. The speed of continuous-time closed-loop system response and stability margin are specified by the dominant pole pair (the damping ratio $\zeta = 0.707$, and the natural frequency $\omega_n = 10$ rad/s) located in Nyquist frequency region. The desired quality of transient response is matched by the gains of the position PI regulator $K_p = 0.52024$ and $K_I = 0.0012885$.

According to relations (15) and (16), the gains of the ordinary identity and the reduced-order digital observers were adjusted to values given in Table I, insuring the bandwidth of 4.5 Hz and proper speed estimation. Also, the gains for both digital reduced-order PI observer and full-order PI² observer

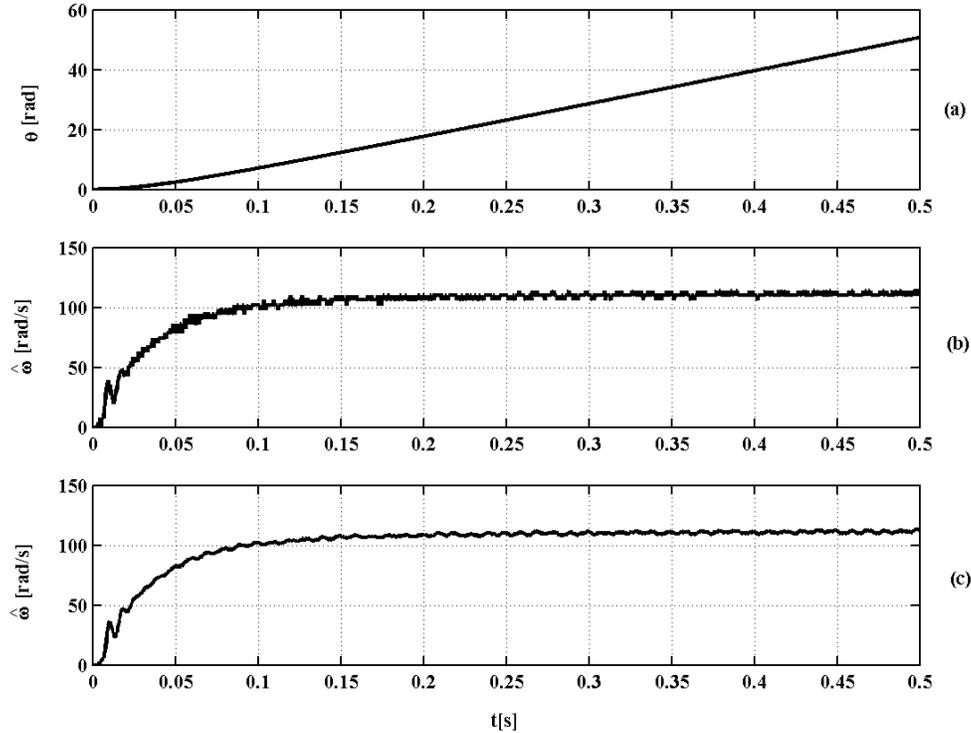


Fig. 4. (a) Open-loop step response $\theta(t)$; (b) Estimate of shaft speed $\hat{\omega}(t)$ derived by Euler's approximation of the derivative; (c) Estimate of shaft speed $\hat{\omega}(t)$ derived by simple algorithm (1) and $n = 5$.

were set according to relations (26) and (30) to values given in Table I, providing the same transient behavior.

In the simulations, as well as in the experiment the system was excited by the step reference signal $\theta_{\text{ref}}(t) = 10h(t-2)$ rad, and by the external disturbance over the period 6 to 10 seconds. The disturbance was a constant load torque $T_L = 0.1$ Nm, which is 53% of the rated torque. During the control object modeling the electrical time constant was neglected, and the effect of disturbance can be mapped onto the object input, and presented by the appropriate voltage signal $M_o^* = 3.82$ V which acts inside the control channel.

Figs. 5-7 and Figs. 8-10 compare experimental versus full-model simulation results for both the reduced-order observer and the identity observer, respectively. These results show a remarkable agreement between the simulated and measured

quantities of the considered system.

Note that the control object is low power DC motor with some dry friction problems which are especially expressive in the case of different positioning tasks.

The results presented in Figs. 5-7 indicate that under relatively unfavorable real conditions the proper shaft speed estimation can be provided by using the reduced-order PI observer whose setting is proposed in the previous section. The same conclusion can be drawn about the full-order PI² observer from the results given in Figs. 8-10.

Some differences between simulation and experimental results, and the presence of the observation errors, shown in Fig. 11, are caused by the final resolution of the applied encoder, the quantization noise of the digital hardware, as well as by the unmodeled dynamics.

TABLE I
OBSERVERS SETTING

Type of Observer Structure	Observer Gains			
	g_1	g_2	g_3	g_4
Identity Observer	0.0301626	0.00659052	-	-
PI ² Observer	0.0853859	0.92137800	0.000762402	0.000762402
Reduced-Order Observer	-	2.58350000	-	-
PI Observer	-	30.5470000	-	0.000762402

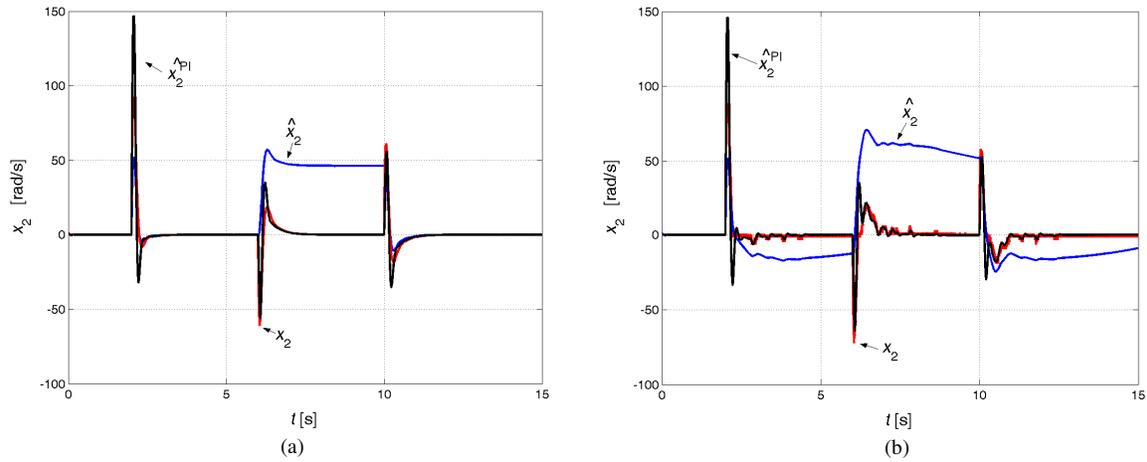


Fig. 5. True and estimation values of the shaft speed using reduced-order observer (16) and reduced-order PI observer (22) (a) simulation, (b) experiment.

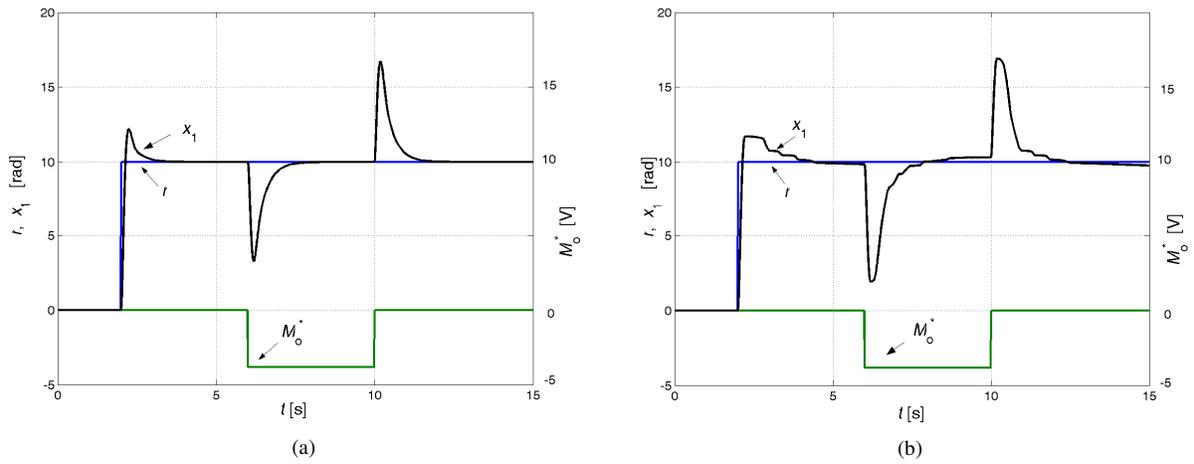


Fig. 6. Reference position (r), step response of shaft position (θ) and on the plant input mapped load torque (M_o^*) (a) simulation, (b) experiment.

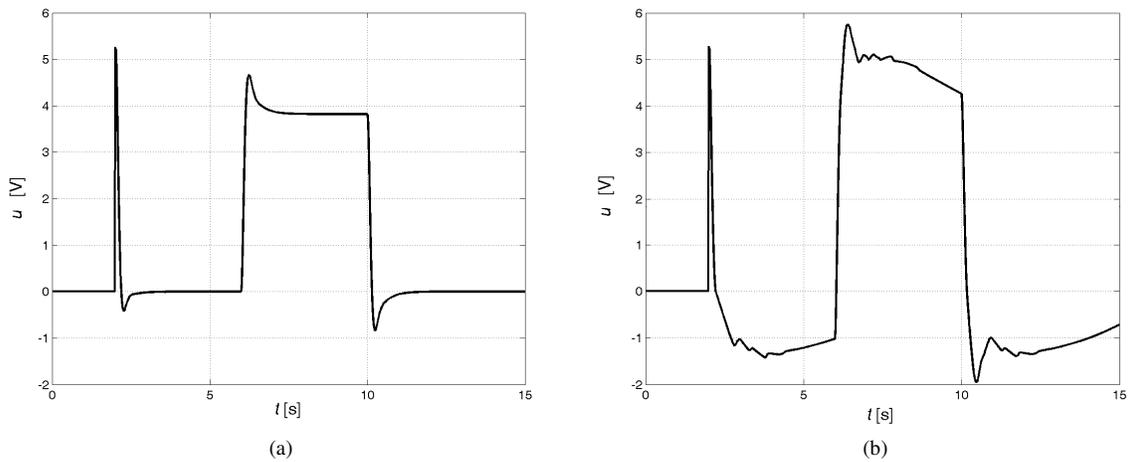


Fig. 7. Control signal in system with reduced-order observers (a) simulation, (b) experiment.

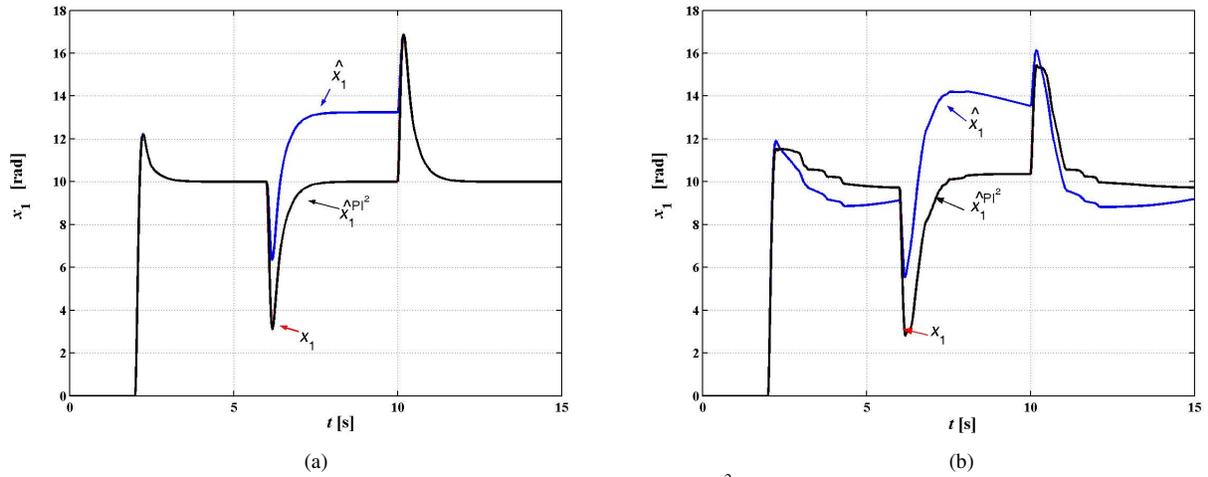


Fig. 8. True and estimation values of the shaft position using identity observer (15) and PI^2 observer (28) (a) simulation, (b) experiment.

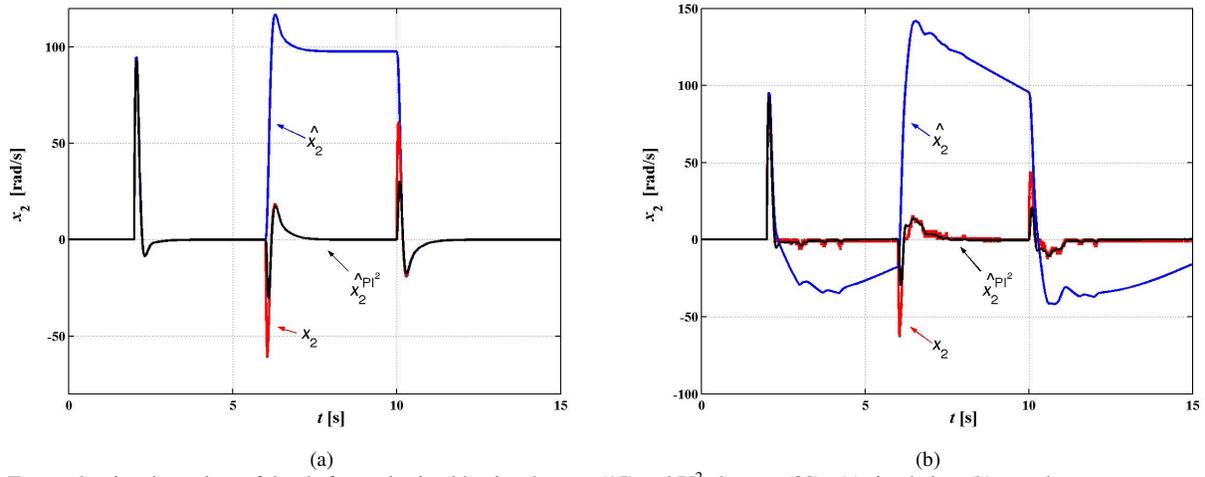


Fig. 9. True and estimation values of the shaft speed using identity observer (15) and PI^2 observer (28) (a) simulation, (b) experiment.

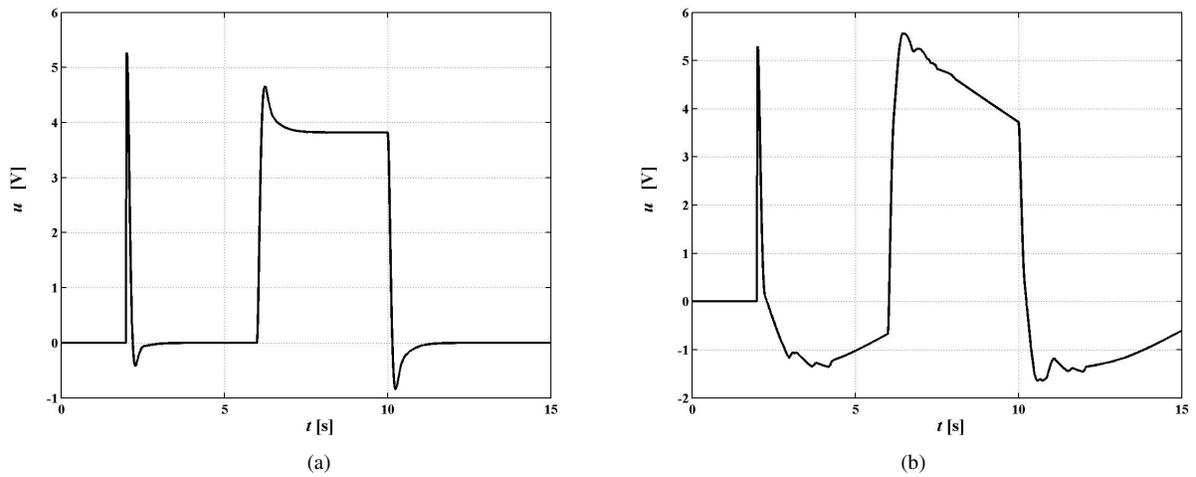


Fig. 10. Control signal in system with full-order observers (a) simulation, (b) experiment.

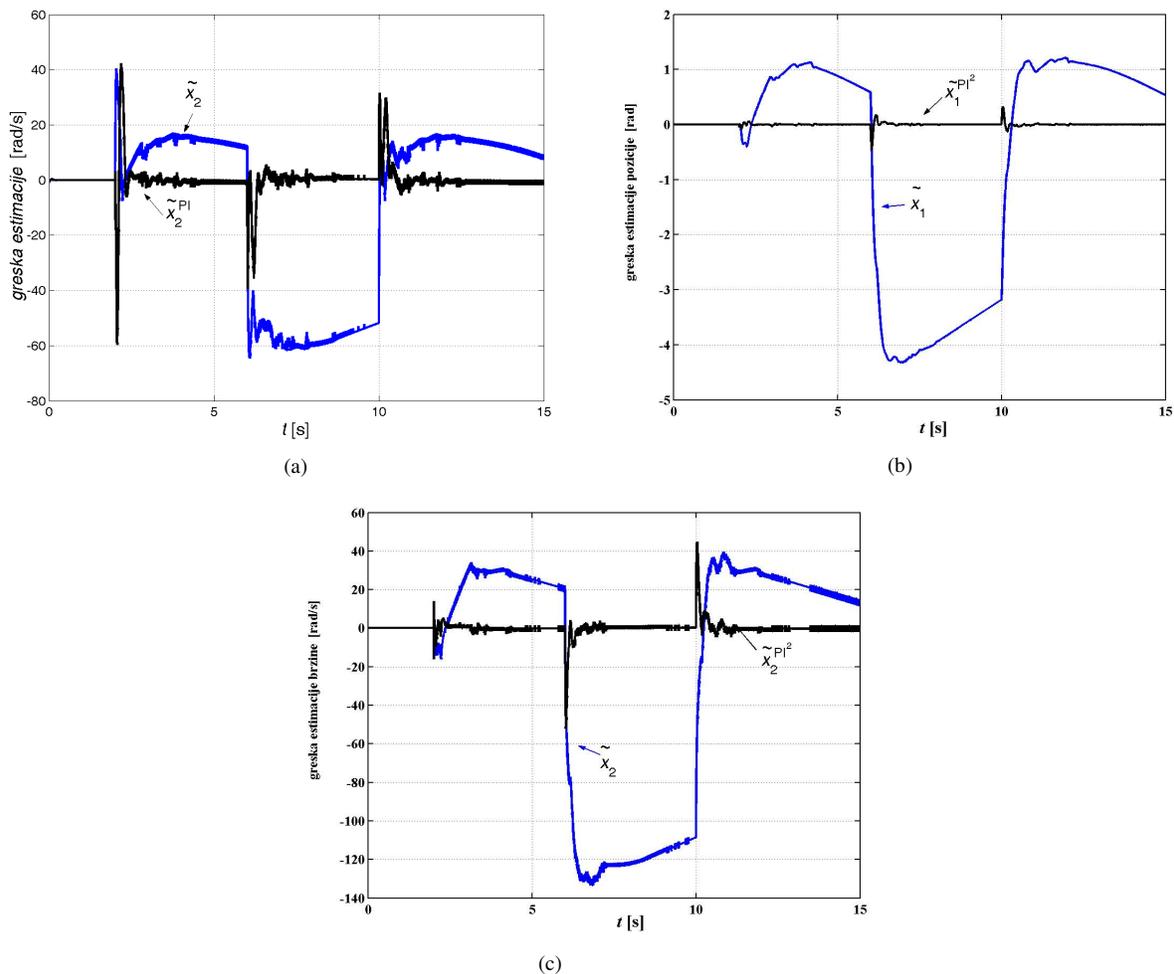


Fig. 11. Estimation errors: (a) Estimation error of the shaft speed using reduced-order observer (16) and reduced-order PI observer (22), (b) Estimation error of the shaft position using identity observer (15) and PI^2 observer (28), (c) Estimation error of the shaft speed using identity observer (15) and PI^2 observer (28).

Although a perfect observation paradigm cannot be obtained, the proposed algorithms can effectively control the estimation errors of system states even in the presence of external disturbances.

VI. CONCLUSION

The aim of this paper is to consider the possibility of using the ordinary discrete-time observers full and reduced-order, and their modifications called PI^2 and PI observer for proper speed estimation in the case of the constant or slow varying load torque disturbances. For gain adjustment of observers extended with integral actions the suitable simple procedures are proposed. Simulation results, as well as the real-time experimental results validate the superior performances of the proposed new state observer structures.

REFERENCES

- [1] R.D. Lorenz, K.W. Van Patten, "High-Resolution Velocity Estimation for All-Digital, ac Servo Drives", *IEEE Trans. Ind. App.*, vol. 27, no. 4, pp. 701-705, Aug. 1991.
- [2] D.C. Hanselman, "Techniques for Improving Resolver-to-Digital Conversion Accuracy", *IEEE Trans. Ind. Electron.*, vol. 38, no. 6, pp. 501-504, Dec. 1991.
- [3] R. Isermann, *Digital Control Systems*, Springer-Verlag, Berlin, 1989.
- [4] R.J. Vaccaro, *Digital control, A state-space approach*, McGraw-Hill, Inc, New York, 1995.
- [5] G.F. Franklin, J.D. Powell, and M.L. Workman, *Digital control of dynamic systems*, Addison-Wesley, New York, 1990.
- [6] M.B. Naumović, M.R. Stojić, "Suitable Introductions of Integral Terms in the Design of Control Systems", *Scientific review*, Belgrade, No. 21-22, pp. 81-92, 1996.
- [7] M.B. Naumović, and M.R. Stojić, "Velocity Estimation in Digital Controlled DC Servo Drives," in *Proc. 24th Annu. Conf. IEEE Industrial Electronics Society (IECON'98)*, Aachen, September 1998, vol. 3/4, pp. 1505-1508.
- [8] M.B. Naumović, "Some Approaches to Velocity Estimation in Digital Controlled DC Servo Drives," in *Proc. VI International Conference on*

- Systems, Automatic Control and Measurements (SAUM'98)*, Niš, September 1998, pp. 140-144.
- [9] M. B. Naumović, and J. Popović, "On Preserving the Robustness of the Velocity Estimation in Digital Controlled DC Servo Drives," in *Proc. L ETRAN Conference*, Belgrade, June 2006, Vol. I, pp. 239-242 (in Serbian).
- [10] J. Popović, M. B. Naumović, and B. R. Veselić, "Digital Reduced Order PI Observer for Velocity Estimation in DC Servo Drives," in *Proc. Conf. with International Participation MECHATRONIC SYSTEMS – Development, Application and Perspective*, Niš, September 2006, pp. 81-86 (in Serbian).
- [11] M. B. Naumović, and B. R. Veselić, "Application of Digital PI² Observer in a DC Servo Drive," in *Proc. 5th Symposium of Industrial Electronics, INDEL 2006*, Banja Luka, November 2006, pp. 264-268, (in Serbian).
- [12] B.R. Veselić, "Synthesis of robust systems for coordinated tracking of complex trajectories by using discrete-time sliding modes," Ph.D. dissertation, Dept. Automatic Control, University of Niš, Faculty of Electronic Engineering, 2006 (in Serbian).
- [13] dSPACE International. Available: <http://www.dspaceinc.com/en/inc/home.cfm>

Frequency Domain Design of a Complex Controller under Constraints on Robustness and Sensitivity to Measurement Noise

Tomislav B. Šekara, Miloš B. Trifunović and Vidan Govedarica

Abstract—New general rules have been developed for designing complex controllers under constraints on robustness and sensitivity to measurement noise. The design is based on a compromise between robustness and performance. This solution makes possible obtaining practically realizable complex controllers. It is shown that the proposed method results into considerably better performance and robustness indices, compared to those obtained by the optimal PID controller.

Index Terms—Complex controller, PID controller, Robustness, Frequency domain.

I. INTRODUCTION

THE paper presents general rules for designing a complex controller $C(s)$, intended for the processes involving time delays and having an arbitrary order and multiple astaticism. The problem of control of complex processes (multiple instabilities, multiple astaticism, dominant time delay) cannot be solved adequately by applying PID controllers, which is the main reason for developing the methods for design of complex controllers.

It is well known that about 94% of feedbacks in industry are realized by PI/PID controllers [1], while in petrochemical industry this percentage is 97% [2,3].

Owing to a high significance of PI/PID, very efficient and simple procedures for tuning parameters of industrial controllers have been developed [4,5,6] as well as optimization procedures [7-22] for designing PI/PID controllers so the IAE (Integral of Absolute Error) is minimized under constraints on robustness, which satisfies the criterion defined in [23].

In addition to the mentioned methods, there are methods for design of PID controllers which are based on the IMC (Internal Model Control) controller [24-26]. The IMC method of controller design contains one adjustable parameter λ which, for a narrow class of processes, has direct influence on the time constant of the closed loop system.

Response of a process regulated by applying an IMC method to a Heaviside-type disturbance is dependent on the

dominant dynamics of the process. E.g., if a process is dominated by oscillatory dynamics, responses to any disturbance will be oscillatory.

For the purpose of accomplishing adequate indices of robustness and performance for a wider class of stable and unstable processes, new methods of designing complex controllers based on the modified IMC structure [27-29] have been developed. However, the design rules for complex controllers by applying these methods have not been designed for the general form of transfer function of the process $G_p(s)$, but only for certain classes of processes $G_p(s)$ [27-29].

Complex controller $C(s)$ defined in this work is designed for the general form of transfer function of the process, given in a rational form including delay $G_p(s)=H(s)\exp(-\tau s)/Q(s)$ under constraints on robustness and sensitivity to measurement noise.

The adjustable parameters of the complex controller $C(s)$ are the time constant λ and relative damping factor ζ of the dominant poles of the process in the closed loop with the complex controller $C(s)$ [7]. By adjusting parameter ζ one can accomplish a compromise between the robustness and performance indices, which is not possible for complex controllers designed by IMC [24-26] or modified IMC [27-289].

The proposed complex controller $C(s)$ is compared with the PID controller [7] through a series of simulations of a wide class of industrial processes. It is shown that the application of the proposed method results in considerably better indices of robustness and performance compared to those obtained with the method described in [7].

II. DESIGN OF THE COMPLEX CONTROLLER FOR A PROCESS HAVING TRANSFER FUNCTION $G_p(s)$

The control structure involving complex controller $C(s)$ is presented in Fig. 1

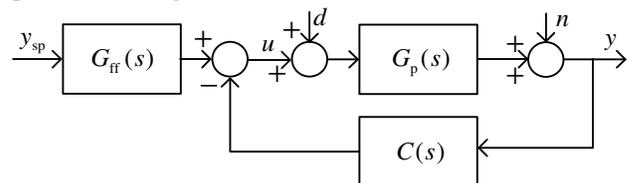


Fig. 1. The control structure.

T. B. Šekara, is with the Faculty of Electrical Engineering, University of Belgrade, Serbia (e-mail: tomi@etf.rs).

M. B. Trifunović, is with the Faculty of Electrical Engineering, University of Belgrade, Serbia (e-mail: milosht trifunovic@gmail.com).

V. Govedarica, is with the Faculty of Electrical Engineering, University of Eastern Sarajevo, Bosnia and Herzegovina (e-mail: vidangov@yahoo.com).

In general, $G_p(s) = H(s)e^{-\tau s} / Q(s)$, where $Q(s)$ and $H(s)$ are polynomials of the order $\deg Q(s) = n \geq \deg H(s) = m$ and $H(0) = h_0 \neq 0$. In order to facilitate the process of deriving, with no loss in generality, it is assumed that $\deg H(s) = 0$, i.e., $H(s) = h_0$. The complementary sensitivity function of the controlled process $G_p(s)$ of Fig.1 is given by relation $T_p(s) = L(s)/(1+L(s))$, with the feedback function of the form $L(s) = C(s)G_p(s)$. Let the desired complementary sensitivity function $T(s)$ be given by

$$T(s) = \frac{N(s)e^{-\tau s}}{P(s)}, \quad N(s) = 1 + \sum_{j=1}^n \eta_j s^j, \quad P(s) = 1 + \sum_{k=1}^p \lambda_k s^k \quad (1)$$

with $p \geq 2n$, $p, n \in N$ and adjustable parameters $\lambda_k > 0$, $k = \overline{1, p}$, $\eta_j \in R$, $j = \overline{1, n}$, which are determined on the basis of the desired performance of the closed loop system. From relations (1), the controller $C(s)$ of the process having transfer function $G_p(s)$ resulting in maximum suppression of disturbance d or n is

$$C(s) = \frac{1}{G_p(s)} \frac{T(s)}{1-T(s)} = \frac{1}{h_0} \frac{Q(s)N(s)}{F(s)}, \quad (2)$$

where $F(s) = P(s) - e^{-\tau s}N(s)$.

In general, parameters $\overline{\eta_1, \eta_n}$ are determined so that the poles of process $G_p(s)$ are cancelled by the zeros of function $F(s)$. Let the poles of process $G_p(s)$ are: $s_1 = s_2 = \dots = s_v = 0$, $s_{v+1} = s_{v+2} = \dots = s_{v+r} \neq 0$, i.e. zeros of polynomial $Q(s)$, of the order v and r . Let the remaining zeros of polynomial $Q(s)$ $\overline{s_{v+r+1}, s_n}$ be simple, then parameters $\overline{\eta_1, \eta_n}$ are determined according to the following rules.

Rule 1. If the zeros of polynomial $Q(s)$ are $s_1 = s_2 = \dots = s_v = 0$, parameters $\overline{\eta_1, \eta_v}$ are determined from condition

$$\left. \frac{d^j F(s)}{ds^j} \right|_{s=0} = 0, \quad j = \overline{1, v}. \quad (3)$$

Rule 2. If the zeros of polynomial $Q(s)$ are $s_{v+1} = s_{v+2} = \dots = s_{v+r} \neq 0$, parameters $\overline{\eta_{v+1}, \eta_r}$ are determined from condition

$$\left. \frac{d^{j-1} F(s)}{ds^{j-1}} \right|_{s=s_{v+j}} = 0, \quad j = \overline{1, r}. \quad (4)$$

Rule 3. If the zeros of polynomial $Q(s)$ are simple $\overline{s_{v+r+1}, s_n}$, parameters $\overline{\eta_{v+r+1}, \eta_n}$ are determined from condition

$$F(s) \Big|_{s=s_{v+j}} = 0 \quad \text{for } j = \overline{1, n-v-r}. \quad (5)$$

Polynomial $P(s)$ is usually taken in the form

$$P(s) = (\lambda s + 1)^p. \quad (6)$$

For the purpose of achieving better compromise performance/robustness in this work a new form of polynomial $P(s)$ is proposed

$$P_1(s) = (\lambda^2 s^2 + 2\zeta \lambda s + 1)^n, \quad \zeta \in O(1), \quad n \geq 1. \quad (7)$$

Rule 4. If in Rule 2 or Rule 3 some of the zeros of polynomial $Q(s)$ has a positive real part (unstable process), in controller (2) canceling of these zeros in the denominator and zeros in the numerator has to be carried out (elimination of dipoles).

Remark 1. If $\deg H(s) > 0$, the relations given by (1) - (7) remain the same and polynomial $N(s)$ in (1) becomes

$$N(s) = (\eta_n s^n + \eta_{n-1} s^{n-1} + \dots + \eta_1 s + 1)H(s) / h_0, \quad h_0 \neq 0, \quad N(0) = 1,$$

On the basis of (7), free parameters of the complex controller (2) are the time constant $\lambda > 0$ and relative damping factor $\zeta > 0$ of the closed loop system, like in [7]. The damping factor which is introduced in the design of complex controller plays a significant role in accomplishing a compromise between the performance and robustness indices. It is shown later that through the damping factor one can exert influence upon sensitivity to measurement noise at high frequencies M_n

$$M_n = \lim_{\omega \rightarrow \infty} \left| \frac{C(i\omega)}{1 + C(i\omega)G_p(i\omega)} \right|. \quad (8)$$

In order to strike a compromise between desired performance IAE and $M_s = \max_{\omega} |1/(1+L(i\omega))|$, time constant λ should satisfy condition

$$\max_{\omega, \lambda} |1/(1 + C(i\omega)G_p(i\omega))| = M_s. \quad (9)$$

For given ζ and M_s (9), time constant λ is determined by solving two nonlinear algebraic equations like in [7].

$$|1 + C(i\omega)G_p(i\omega)|^2 - 1/M_s^2 = 0, \quad (10)$$

$$\frac{\partial (|1 + C(i\omega)G_p(i\omega)|^2)}{\partial \omega} = 0, \quad (11)$$

Initially, parameter ζ should be taken as $\zeta=1$ and parameter λ close to the estimated transport delay. By determining time constant λ for different values of parameter ζ , one accomplishes a compromise between the values IAE, M_n and M_p . A comparison of the qualities of control for different values of parameter ζ is analyzed in detail in the next section.

III. COMPARATIVE ANALYSIS AND SIMULATIONS

A comparison of the proposed method for design of controller $C(s)$ (2) for different values of parameter ζ is given in Table 1 for sixteen representative typical dynamic characteristics:

$$G_{p1}(s) = \frac{2e^{-s}}{(10s+1)(5s+1)}, \quad G_{p2}(s) = \frac{1}{(s+1)^4},$$

$$G_{p3}(s) = \frac{1}{\prod_{k=0}^3 (0.7^k s + 1)}, G_{p4}(s) = \frac{e^{-5s}}{(s+1)^3}, G_{p5}(s) = \frac{1-s}{(s+1)^3},$$

$$G_{p6}(s) = \frac{1}{\prod_{k=0}^3 (0.2^k s + 1)}, G_{p7}(s) = \frac{(2s+1)e^{-4s}}{(10s+1)(7s+1)(3s+1)},$$

$$G_{p8}(s) = -\frac{(13.81s+1)(18.4s+1)}{(59s+1)^5}, G_{p9}(s) = \frac{e^{-s}}{(s^2+0.1s+1)},$$

$$G_{p10}(s) = \frac{e^{-0.5s}}{s}, G_{p11}(s) = \frac{-1.6(-0.5s+1)}{s(3s+1)}, G_{p12}(s) = \frac{1}{s(s+1)^3},$$

$$G_{p13}(s) = \frac{4e^{-2s}}{4s-1}, G_{p14}(s) = \frac{e^{-0.5s}}{(5s-1)(2s+1)(0.5s+1)},$$

$$G_{p15}(s) = \frac{e^{-0.1s}}{(s-1)^3}, G_{p16}(s) = \frac{e^{-0.5s}}{s^3}.$$

The controllers for all processes except the unstable ones ($G_{p13}(s)$, $G_{p14}(s)$ i $G_{p15}(s)$) are of the form (2) with the corresponding parameters λ and ζ from Table 1. For unstable processes, on the basis of rule 4, in order to eliminate unstable dipoles of the controller, time delay $e^{-\tau s}$ in polynomial $F(s)$ is approximated by Pade approximation of the order N/N , where N is chosen so that the robustness and performance indices are preserved. For processes $G_{p14}(s)$ and $G_{p15}(s)$ it is sufficient to take $N=2$

$$e^{-\tau s} \approx \frac{12 - 6\tau s + \tau^2 s^2}{12 + 6\tau s + \tau^2 s^2}, \quad (12)$$

i.e. $N=3$ for process $G_{p13}(s)$

$$e^{-\tau s} \approx \frac{120 - 60\tau s + 12\tau^2 s^2 - \tau^3 s^3}{120 + 60\tau s + 12\tau^2 s^2 + \tau^3 s^3}. \quad (13)$$

E.g., on the basis of (1) – (7), general form of the controller (2) for unstable process $G_{p13}(s)$ is given by

$$C(s) = \frac{(4s-1)(\eta_1 s + 1)}{4((\lambda^2 s^2 + 2\zeta\lambda s + 1) - e^{-2s}(\eta_1 s + 1))}.$$

For $\lambda=2.335$ and $\zeta=1$ obtained on the basis of relations (9)-(11) and replacing transport delay e^{-2s} by approximation (13) for $\tau=2$ one obtains the controller

$$C(s) = \frac{2.3(s+2.322)(s+0.0797)(s-0.25)(s^2+3.678s+6.459)}{s(s+9.186)(s-0.2499)(s^2+0.2203s+7.0334)}$$

and upon canceling the dipole $s \approx 0.25$, the final transfer function of controller from Table 1 for process $G_{p13}(s)$ is

$$C(s) = \frac{2.3(s+2.322)(s+0.0797)(s^2+3.678s+6.459)}{s(s+9.186)(s^2+0.2203s+7.0334)}.$$

By using the equivalent procedure, all other controllers of unstable processes have been determined, with approximation (12) applied for processes $G_{p14}(s)$ and $G_{p15}(s)$.

In order to reduce the order of the controllers of stable processes obtained by applying rules 1 to 3, it is desirable to apply the described reduction by cancellation of dipoles. For stable processes having dominant delays, this reduction of

TABLE I

THE RESULTS OBTAINED BY THE PROPOSED METHOD FOR DIFFERENT VALUES OF PARAMETER Z OF CONTROLLER C(S), $M_s = \text{CONST.}$

Process	λ	ζ	M_n	IAE	M_s	M_p
$G_{p1}(s)$	1.605	1.000	51.82	1.0767	2.00	1.56
$G_{p1}(s)$	1.186	1.551	207.28	0.9846	2.00	1.33
$G_{p1}(s)$	1.923	0.842	25.91	1.3320	2.00	1.65
$G_{p2}(s)$	0.395	1.000	665.39	0.2213	2.00	1.59
$G_{p2}(s)$	0.438	0.955	332.69	0.2963	2.00	1.56
$G_{p2}(s)$	0.553	0.870	66.54	0.5627	2.00	1.49
$G_{p3}(s)$	0.228	1.000	657.38	0.1152	2.00	1.60
$G_{p3}(s)$	0.253	0.962	328.69	0.1576	2.00	1.57
$G_{p3}(s)$	0.321	0.885	65.74	0.3052	2.00	1.48
$G_{p4}(s)$	0.680	1.000	10.1	6.0813	2.00	1.01
$G_{p4}(s)$	0.540	1.150	40.4	5.7267	2.00	1.01
$G_{p4}(s)$	0.760	0.950	5.05	6.3336	2.00	1.00
$G_{p5}(s)$	0.730	1.000	6.557	2.6549	2.00	1.11
$G_{p5}(s)$	0.580	1.250	26.228	2.6006	2.00	1.02
$G_{p5}(s)$	0.800	0.935	3.278	2.7663	2.00	1.13
$G_{p6}(s)$	0.061	1.000	176.95	0.0115	2.00	1.72
$G_{p6}(s)$	0.070	0.965	88.47	0.0171	2.00	1.70
$G_{p6}(s)$	0.091	0.932	17.69	0.0384	2.00	1.62
$G_{p7}(s)$	3.200	1.000	28.26	4.5657	2.00	1.29
$G_{p7}(s)$	2.576	1.250	113.04	4.0419	2.00	1.17
$G_{p7}(s)$	3.591	0.915	14.13	5.1513	2.00	1.32
$G_{p8}(s)$	19.00	1.000	7466	1.3675	2.00	1.96
$G_{p8}(s)$	26.76	0.910	746.6	8.4579	2.00	1.76
$G_{p8}(s)$	35.91	0.845	74.66	31.201	2.00	1.51
$G_{p9}(s)$	0.735	1.000	9.86	4.5654	2.00	1.14
$G_{p9}(s)$	0.599	1.800	39.44	6.9967	2.00	1.04
$G_{p9}(s)$	0.807	0.710	4.93	4.0345	2.00	1.18
$G_{p10}(s)$	0.628	1.000	4.45	1.1480	2.00	1.50
$G_{p10}(s)$	0.625	1.550	6.23	1.4792	2.00	1.31
$G_{p10}(s)$	0.674	0.700	2.225	1.1486	2.00	1.71
$G_{p11}(s)$	1.250	1.000	6.44	6.0566	2.00	1.68
$G_{p11}(s)$	1.130	1.780	19.32	9.8390	2.00	1.24
$G_{p11}(s)$	1.507	0.830	3.22	7.7678	2.00	1.80
$G_{p12}(s)$	0.670	1.000	58.64	2.4653	2.00	1.66
$G_{p12}(s)$	0.536	0.980	29.32	3.5307	2.00	1.67
$G_{p12}(s)$	0.748	1.060	234.56	1.1935	2.00	1.68
$G_{p13}(s)$	2.335	1.000	2.30	23.4905	3.00	2.73
$G_{p13}(s)$	2.510	1.200	2.40	28.3920	3.00	2.54
$G_{p13}(s)$	2.485	0.880	2.00	24.0484	3.00	2.87
$G_{p14}(s)$	1.200	1.00	12.77	2.8845	2.60	2.41
$G_{p14}(s)$	1.408	1.01	51.08	1.1734	2.60	2.29
$G_{p14}(s)$	1.407	1.04	6.385	4.9374	2.60	2.54
$G_{p15}(s)$	0.400	1.00	947.0	0.9270	5.90	6.60
$G_{p15}(s)$	0.459	1.05	735.4	1.7591	5.90	6.70
$G_{p15}(s)$	0.623	1.23	474.5	8.0659	5.90	6.88
$G_{p16}(s)$	1.820	1.00	4.04	231.30	3.00	2.99
$G_{p16}(s)$	1.260	1.12	16.16	74.089	3.00	2.85
$G_{p16}(s)$	1.006	1.23	40.4	39.553	3.00	2.75

a complex controller may lead to degradation of the robustness and performance indices, thus this reduction is not recommendable. Such an example is process $G_{p4}(s)$.

From Table 1 it is clear that for all processes $G_{pj}(s)$, $j=1, \dots, 16$, when applying $C(s)$ for the same M_s , the adjustable parameter ζ allows accomplishing a compromise between IAE, M_n , and M_p . This parameter is of key significance, since by its use one can decrease or increase value of M_n and improve the robustness and performance indices (Figs. 2 and 3).

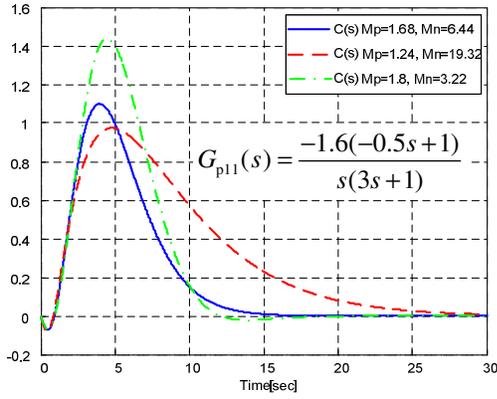


Fig. 2 Response to a Heaviside-type disturbance of process $G_{p11}(s)$ in closed loop with controller $C(s)$ for $M_s=2$.

The proposed method for design of the complex controller $C(s)$ (2) will be compared to the PID controller [7], which, as has been shown in [7], accomplished the robustness and performance indices the same as the optimal PID [21]. The comparison of these methods, assuming the same values of M_s and M_n , is presented in table 2 for the processes $G_{pj}(s), j=1,..,16$.

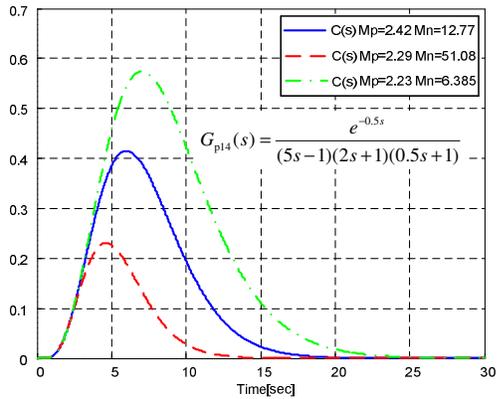


Fig. 3 Response to a Heaviside-type disturbance of process $G_{p11}(s)$ in closed loop with controller $C(s)$ for $M_s=2.6$.

In Table 2, parameter ζ for all complex controllers has been determined so that practically the same value of M_n as in the case of the PID controller is obtained. It can also be seen from Table 1 that the complex controller ensures a considerably better quality of control compared to that of the PID controller.

The following figures show the response to a Heaviside type of disturbance of the complex controller $C(s)$ and PID controller [7].

From Figs. 4-7 it is obvious that application of the complex controller results in a significantly lower IAE , with practically the same robustness as the one obtained with the PID controller. For unstable processes the complex controller gives considerably higher indices of robustness and performance compared to those of the PID controller. It should be mentioned that for the processes of higher order of instability and astaticism, complex controllers can be successfully designed as demonstrated with processes $G_{p15}(s)$ and $G_{p16}(s)$.

TABLE II

THE COMPARISON OF THE METHOD PROPOSED FOR COMPLEX CONTROLLER $C(s)$ WITH THE OPTIMALLY TUNED PID CONTROLLER [7]

Process-controller	λ	ζ	M_n	IAE	M_s	M_p
$G_{p1}(s)$ -C	1.5205	1.06	64.42	1.02	2.00	1.53
$G_{p1}(s)$ -PIDtun	1.8989	0.80	64.42	1.17	2.00	1.72
$G_{p2}(s)$ -C	0.6300	0.825	26.16	0.78	2.00	1.41
$G_{p2}(s)$ -PIDtun	1.0825	0.75	26.16	1.28	2.00	1.37
$G_{p3}(s)$ -C	0.3567	0.855	31.41	0.40	2.00	1.44
$G_{p3}(s)$ -PIDtun	0.6343	0.75	31.41	0.66	2.00	1.41
$G_{p4}(s)$ -C	0.9294	0.86	1.55	6.83	2.01	1.01
$G_{p4}(s)$ -PIDtun	2.5242	0.85	1.55	8.31	2.01	1.02
$G_{p5}(s)$ -C	0.7023	1.035	8.27	2.65	2.00	1.09
$G_{p5}(s)$ -PIDtun	1.0427	0.80	8.27	2.82	2.00	1.16
$G_{p6}(s)$ -C	0.0559	1.025	265.6	0.091	2.00	1.72
$G_{p6}(s)$ -PIDtun	0.0863	0.80	265.6	0.013	1.98	1.80
$G_{p7}(s)$ -C	3.2040	0.995	27.91	4.54	2.00	1.30
$G_{p7}(s)$ -PIDtun	5.0863	0.80	27.91	5.70	2.00	1.36
$G_{p8}(s)$ -C	40.3684	0.82	26.63	48.73	2.00	1.44
$G_{p8}(s)$ -PIDtun	72.3427	0.75	26.63	85.31	2.00	1.37
$G_{p9}(s)$ -C	0.8790	0.65	3.18	4.58	1.80	1.00
$G_{p9}(s)$ -PIDtun	1.1113	0.65	3.18	6.60	2.00	1.01
$G_{p10}(s)$ -C	0.6280	1.00	4.45	1.14	2.00	1.50
$G_{p10}(s)$ -PIDtun	0.3605	1.10	11.92	1.24	2.00	1.54
$G_{p11}(s)$ -C	1.1248	1.24	11.22	6.122	2.00	1.52
$G_{p11}(s)$ -PIDtun	1.1422	0.85	11.22	4.16	1.98	1.85
$G_{p12}(s)$ -C	0.9626	0.955	5.91	7.923	2.00	1.62
$G_{p12}(s)$ -PIDtun	1.9054	0.85	5.91	14.27	2.00	1.62
$G_{p13}(s)$ -C	1.8621	0.75	2.41	16.21	3.50	3.27
$G_{p13}(s)$ -PIDtun	1.6160	0.85	2.41	26.21	4.00	3.52
$G_{p14}(s)$ -C	1.0241	0.88	20.57	1.49	2.99	2.66
$G_{p14}(s)$ -PIDtun	1.8540	0.80	20.57	2.90	2.99	2.78
$G_{p15}(s)$ -C	0.6234	1.23	473.5	7.79	5.90	6.90
$G_{p15}(s)$ -PIDtun	-	-	-	-	-	-
$G_{p16}(s)$ -C	1.0057	1.23	40.15	39.55	3.00	2.75
$G_{p16}(s)$ -PIDtun	-	-	-	-	-	-

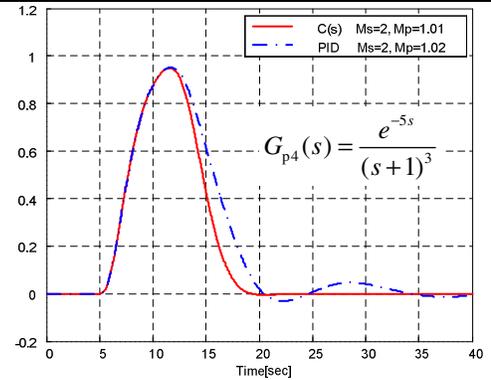


Fig. 4 Response to a Heaviside-type disturbance of process $G_{p4}(s)$ in closed loop with controllers from Table 2 for $M_s=2, M_n=8.27$.

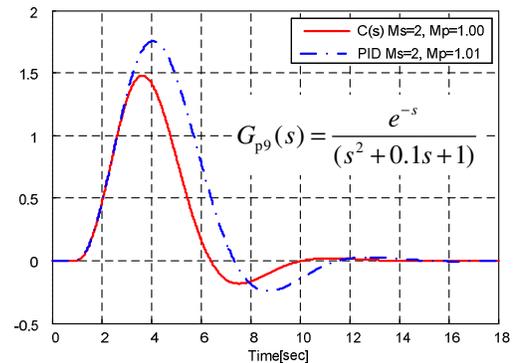


Fig. 5 Response to a Heaviside-type disturbance of process $G_{p9}(s)$ in closed loop with controllers from Table 2 for $M_s=2, M_n=3.18$.

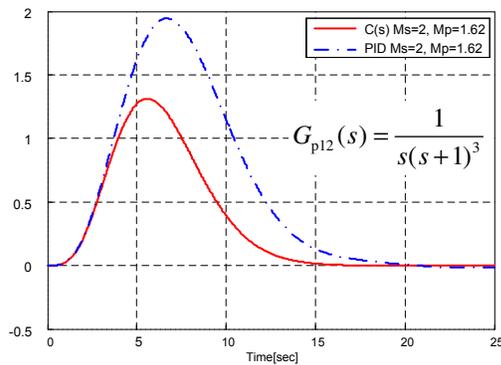


Fig. 6 Response to a Heaviside-type disturbance of process $G_{p12}(s)$ in closed loop with controllers from Table 2 for $M_s=2$, $M_n=5.91$.

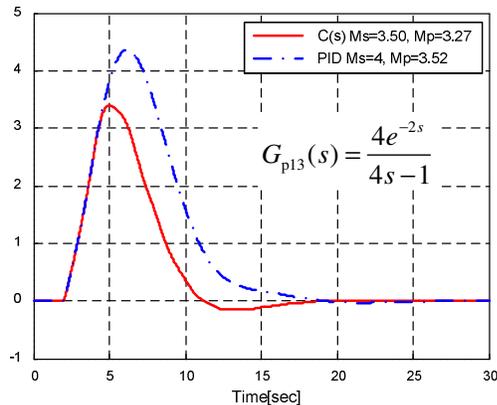


Fig. 7 Response to a Heaviside-type disturbance of process $G_{p13}(s)$ in closed loop with controllers from Table 2 for $M_n=2.41$.

IV. THE CONCLUSION

Design of complex controllers is aimed at increasing the robustness and performance indices compared to those obtainable with conventional controllers. For designing complex controllers an adequate knowledge of transfer function of the process is required. The paper presents general rules for designing complex controllers which have been tested on a wide class of processes. By applying suitable approximations of complex controllers, adequate conventional controllers are obtained for certain class of processes. The comparative analysis and simulations gave the expected results.

REFERENCES

- [1] S. Yamamoto and I. Hashimoto, "Present status and future needs: the view from Japanese industry", In Arkun and Ray, Eds., *Chemical Process Control-CPCIV. Proc. 4th Inter. Conf. on Chemical Process Control*, TX, 1991.
- [2] L. Desborough, R. Miller, "Increasing customer value of industrial control performance monitoring—Honeywell's experience", in: *Sixth International Conference on Chemical Process Control, AIChE Symposium Series Number 326*, Vol. 98, 2002.
- [3] K.J. Åström, T. Hägglund, "Revisiting the Ziegler-Nichols step response method for PID control", *Journal of Process Control*, vol. 14, pp. 635-650, September 2004.
- [4] M. R. Mataušek, G. S. Kvaščev, "A unified step response procedure for autotuning of PI controller and Smith predictor for stable processes", *Journal of Process Control*, Vol. 13, pp. 787-800, 2003.
- [5] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning", *Journal of Process Control*, Vol. 13, pp. 291-309, 2003.
- [6] M. Shamsuzzoha, M. Lee, "IMC-PID controller design for improved disturbance rejection of time-delayed processes", *Industrial & Engineering Chemistry Research*, vol. 46, no.7, pp. 2077-2091, 2007.
- [7] M.R. Mataušek, T.B. Šekara, "PID controller frequency-domain tuning for stable, integrating and unstable processes, including dead-time", *J. Process Control* vol. 21, Issue 1, pp. 17-27, 2011.
- [8] T.B. Šekara, M.R. Mataušek, "Classification of dynamic processes and PID controller tuning in a parameter plane", *J. Process Control* vol. 21, Issue 4, pp. 620-626, 2011.
- [9] H. Panagopoulos, K.J. Åström and T. Hägglund, "Design of PID controllers based on constrained optimization", *IEE Proceedings-Control Theory and Applications*, vol. 149, pp. 32-40 January 2002.
- [10] A. Wallén, K.J. Åström, and T. Hägglund, "Loop-shaping design of PID controllers with constant T_i/T_d ratio", *Asian Journal of Control*, vol. 4, pp. 403-409, December 2002.
- [11] C. Hwang and C-Y. Hsiao, "Solution of non-convex optimization arising in PI/PID control design", *Automatica* vol. 38, pp. 1895-1904, November 2002.
- [12] B. Kristiansson and B. Lennartson, "Evaluation and simple tuning of PID controllers with high-frequency robustness", *Journal of Process Control*, vol.16, pp. 91-102, February 2006.
- [13] B. Kristiansson and B. Lennartson, "Robust tuning of PI and PID controllers: using derivative action despite sensor noise" *IEEE Control Systems Magazine*, pp. 55-69, February 2006.
- [14] A.J. Isaksson and S.F. Graebe, "Derivative filter is an integral part of PID design", *IEE Proceedings-Control Theory and Applications*, vol. 149 pp. 41-45, January 2002.
- [15] A. Karimi, M. Kunze and R. Longchamp, "Robust controller design by linear programming with application to a double-axis positioning system", *Control Engineering Practice*, vol. 15, pp. 197-208, February 2007.
- [16] T. B. Šekara, M. R. Mataušek, "A simple effective method to obtain a well-tuned PID controller", Proceedings of the 51st, Conference on. ETRAN, AU3.1, 2007 (in Serbian).
- [17] T.B. Šekara, M.R. Mataušek, "Optimal tuning of a PI/PID controller for processes defined by a rational transfer function", *INFOTEH* Vol. 6, Paper A-2, p. 6-9, Jahorina, March 2007 (in Serbian).
- [18] T.B. Šekara and M.R. Mataušek, "Optimization of PID controller based on maximization of the proportional gain under constraints on robustness and sensitivity to measurement noise", *IEEE Trans. Automatic Control*, vol. 54, no.1, pp.184-189, Jan. 2009.
- [19] T.B. Šekara and M.R. Mataušek, "Revisiting the Ziegler-Nichols process dynamics characterization", *J. Process Control* Vol. 20, Issue 3, pp. 360-363, 2010.
- [20] T. B. Šekara, M. R. Mataušek, "A four-parameter optimization of a PID controller", Proceedings of 52. Conf. ETRAN, Vol. 1, Palić, Junne 2008 (in Serbian).
- [21] T.B. Šekara, M.R. Mataušek, "Optimal tuning of a PID controller in frequency domain," *INFOTEH*, Paper A-6, p. 24-27, Bosnia and Herzegovina, Jahorina, March 2009 (in Serbian).
- [22] T.B. Šekara, M.B. Trifunović, "Optimal tuning of a PID controller having a differential compensator connected in series in frequency domain", Proceedings of INDEL, pp. 258-261, Banja Luka, 4-6 November 2010 (in Serbian).
- [23] F.G. Shinskey, "How good are our controllers in absolute performance and robustness?", *Measurement and Control*, vol. 23, pp. 114-121, May 1990.
- [24] Chien, I. L. "IMC-PID controller design—an extension" *IFAC Proceedings Series*, vol. 6 (pp. 147-152), 1988.
- [25] Fruehauf, P. S., Chien, I. L., & Lauritsen, M. D. "Simplified IMC-PID tuning rules." *ISA Transactions*, 33(1), 43-59, 1994.
- [26] Lee, Y., Lee, M., Park, S., & Brosilow, C. "PID controller tuning for desired closed-loop responses for SI/SO system." *A.I.Ch.E. Journal*, 44(1), 106-115, 1998.
- [27] Liu T, Zhang W, Gu D. Analytical design of two-degree-of-freedom control scheme for open-loop unstable process with time delay. *J Process Control* 2005; 15:559-72.
- [28] M. Shamsuzzoha, M. Lee, "Analytical design of PID controller cascaded with a lead-lag filter for time-delay processes", *Korean J. Chem. Eng.*, vol. 26, pp. 622-630, 2009.
- [29] M. Shamsuzzoha, M. Lee, "Enhanced disturbance rejection for open-loop unstable process with time delay" *ISA Transactions*, vol. 48, pp. 237-244, 2009.

Choice of Window Size in Calibrating the Geometry of Manipulators Based on the Regions Correlation

Petar Maric and Velibor Djalic

Abstract—To achieve the full flexibility of industrial robots, in addition to the mechanical flexibility, it is necessary to achieve the flexibility in control. Precise automatic calibration of these robots manipulators is essential precondition for achieving this goal. Possibilities and limitations of stereo vision system implementation for automatic calibration of manipulators are presented in this paper. Furthermore, a practical solution for the problem of corresponding points using an area based algorithm is given. Analysis has been conducted of how the marker positioning on the manipulator end-effector influences the shape of the cost function. Choice of window size relative to marker size which provides the best reliability for corresponding area determination is proposed. Results of practical realization, which confirms conducted analysis and reliability of proposed calibration procedure, are presented.

Index Terms—Camera calibration, Computer vision, Modular reconfigurable robot, Robot calibration.

I. INTRODUCTION

A high level of positioning accuracy is an essential requirement in a wide range of industrial robots applications. This accuracy is affected by geometric factors (geometrical parameters accuracy) and non-geometric factors (gear backlashes, encoder resolution, flexibility of links, thermal effects, etc.).

The error due to geometric factors accounted for 90% of the total error. A common approach is to calibrate the current geometric parameters and treat the non-geometric factors as a randomly distributed error. The calibration procedure is very important for robot programming using CAD systems where the simulated robot must reflect accuracy the real robot. During a manipulator control system design, and periodically in the course of task performing, manipulator geometry calibration is required [1].

Vision systems have developed significantly over the last ten years and now have become standard automation components. They represent qualitative bounce in the area of metrology and sensing because they provide us with a remarkable amount of information about our surroundings, without direct physical contact [2]. At the same time, vision systems are the most complex sensors [3].

P. Maric and V. Djalic are with the University of Banja Luka, Faculty of Electrical Engineering, Banja Luka, Bosnia and Herzegovina (e-mail: {petar.maric, velibor.djalic}@etfbl.net).

Calibration of cameras is necessary first step in vision system using. Camera calibration is the process of determining the internal camera (geometric and optical) characteristics and the 3D position and orientation of the camera frame relative to a world coordinate system [2], [4]. There are a number of techniques which only requires the camera to observe a planar pattern(s) shown at a few different orientations. Precise set points are placed on the calibration plain. Calibration process is automatically performed based on the correspondence between the positions of calibration points and the positions of their images [5], [6].

If the camera calibration is performed then for every scene point in a world coordinate system it is possible to determine the position of its image point in image plain. This transformation is called perspective transformation. Inverse perspective transformation is very important for computer vision application in industrial automation. This transformation examines the problem of how to identify the point position in a world coordinate system based on the position of its image point for different camera positions or for several cameras at the same time.

Inverse perspective transformation and advances in processing and image analysis have a wide range of applications in industrial automation, and allow companies to achieve previously impossible levels of efficiency and productivity [7]. Effective co-operation in the computer aided manufacturing depends on the recognition and perception of typical production environments as well as on the understanding of tasks in their context. Vision systems are the basis for scene analysis and interpretation, both in time and 3D space. Measurement of the dimensions of objects and parts in many industrial fields is very important, as the quality of the product depends especially on the reliability and precision of each object part. The use of vision-based metrology allows calculating a set of 3D point coordinates and/or estimate the dimension and pose of a known object. They enable industrial robots to perform different and complex tasks reliably and accurately [8].

II. ROBOT CALIBRATION

A. Manipulator Geometry Modeling

The first step of manipulator calibration is concerned with a mathematical formulation that results in model which gives

relation between the geometric parameters, the joint variables and end-effector position. Many researchers have been looking for the suitable kinematic models for robot calibration, since Richard Paul's book [9]. The most popular among them is the Denavit-Hartenberg (D-H) method. For this reason we will use this notation.

Prior description kinematic model let us define the basic coordinate systems as follows (Fig. 1):

$O_B X_B Y_B Z_B$ – base coordinate system of the manipulator

$O_E X_E Y_E Z_E$ – end-effector (tool) coordinate system of the manipulator (we denote the origin O_E as the endpoint of the robot)

$O_i X_i Y_i Z_i$ ($i=1, n$) – coordinate system fixed to the i^{th} link ($O_n X_n Y_n Z_n$ – coordinate system fixed to the terminal link) of the manipulator.

The original D-H representation of a rigid link depends on geometric parameters. Four parameters a, d, α and θ denote manipulator link length, link offset, joint twist and joint angle, respectively. Composite 4×4 homogenous transformation matrix $A_{i-1,i}$ known as the D-H transformation matrix for adjacent coordinate system i and $i-1$, is:

$$A_{i-1,i} = \begin{bmatrix} \cos \theta_i & -\cos \alpha_i \sin \theta_i & \sin \alpha_i \sin \theta_i & a_i \cos \theta_i \\ \sin \theta_i & \cos \alpha_i \cos \theta_i & -\sin \alpha_i \cos \theta_i & a_i \sin \theta_i \\ 0 & \sin \alpha_i & \cos \alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

The homogenous matrix $A_{B,i}$ which specifies the location of the i^{th} coordinate system with respect to the base coordinate system is the chain product of successive coordinate transformation matrices $A_{i-1,i}$, and expressed as:

$$A_{B,i} = A_{B,1} A_{1,2} \dots A_{i-1,i}. \quad (2)$$

Particularly, for $i=n$ we have $A_{B,n}$ matrix which specifies the position and orientation of the end-effector of the manipulator with respect to the base coordinate system. Matrix $A_{B,n}$ is a function of the $4n$ geometrical parameters which are constant for constant robot geometry, and n joint coordinates that change their value when manipulator moves.

Moreover, a robot is not intended to perform a single operation at the workcell, it has interchangeable different tools. In order to facilitate the programming of the task, it is more practical to have transformation matrix defining the tool coordinate system with respect to the terminal link coordinate system $A_{n,E}$.

Thus, the transformation matrix $A_{w,E}$ can be written as:

$$A_{w,E} = A_{w,B} A_{B,n} A_{n,E}. \quad (3)$$

Since the world coordinate system can be chosen arbitrarily by the user, six parameters are needed to locate the robot base relative to the world coordinate system. From independence to some manipulator parameters it follows that consecutive coordinate systems are represented at most by four independent parameters.

Since the end-effector coordinate system can be defined arbitrarily with respect to the terminal link coordinate system ($O_n X_n Y_n Z_n$), six parameters are needed to define the matrix $A_{n,E}$. If we extend the robot notation to the definition of the end-effector coordinate system, it follows that the end-effector coordinate system introduces four independent parameters. For more details the reader can refer to [1].

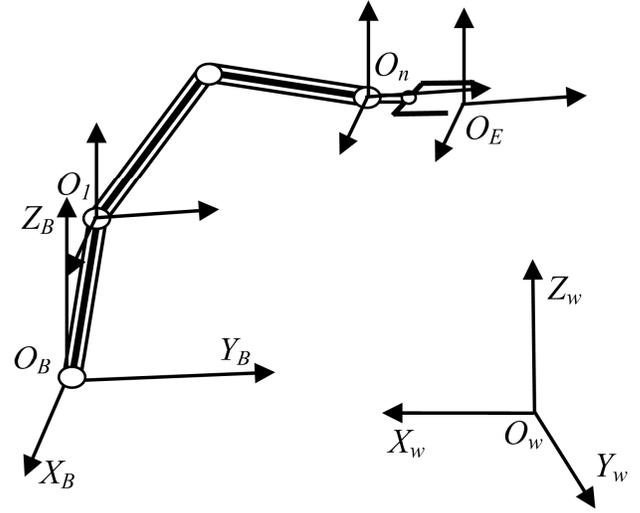


Fig. 1. Coordinate systems assignment for robot modeling.

With above mentioned equations (1), (2), (3) dependence between joint coordinates and geometrical parameters, and endpoint location of the tool can be written as:

$$x = f(q, g^0) \quad (4)$$

where x, q, g^0 denotes end-effector position vector expressed in the world coordinate system, vector of the joint variables, and vector of the geometric parameters, respectively. Dimension of the vector x is 6 if measurement can be made on the location and orientation of the end-effector. However, the most frequently only a location of the endpoint is measured, and therefore dimension of a vector x is 3. Dimension of the vector q is equivalent to the number of DOF for manipulator. Dimension of the vector g^0 is at most $4n+6$.

B. Geometric Parameters Estimation Based on the Differential Model

The calibration of the geometric parameters is based on estimating the parameters minimizing the difference between a function of the real robot variables and corresponding mathematical model. Many authors [10]–[13] presented open-loop methods that estimate the kinematic parameters of manipulators performing on the basis of joint coordinates and the Cartesian coordinates of the end-effector measurements. The joint encoders outputs readings are joint coordinates. It is assumed that there is a measuring device that can sense the position (sometime orientation) of an end-effector Cartesian coordinates.

A mobile closed kinematic chain method has been proposed that obviates the need for pose measurement by forming a manipulator into a mobile closed kinematic chain [14], [15]. Self motion of the mobile closed chain places manipulator in a number of configurations and the kinematic parameters are determined from the joint position readings alone.

The calibration using the end-effector coordinates (open-loop method) is the most popular one. The model represented by equation (4) is nonlinear in g^0 , and we must linearize it in order to apply linear estimators. The differential model provides the differential variation of the location of the end-effector as a function of the differential variation of the

geometric parameters. The function to be minimized is the difference between the measured (x) and calculated end-effector location (x^m). Let $\Delta x = x - x^m$, and $\Delta g = g^0 - g$ be the pose error vector of end-effector and geometric parameter error vector, respectively (g – vector of geometric parameters estimation). From equation (4), the calibration model can be represented by the linear differential equation

$$\Delta x = J_g \Delta g = x - x^m, \quad (5)$$

where:

g is the ($p \times 1$) vector of geometric parameters estimation

$\Delta x = x - x^m$ is the ($r \times 1$) pose error vector of end-effector

$\Delta g = g^0 - g$ is the geometric parameter error vector

J_g is the ($r \times p$) sensitivity matrix relating the variation of the endpoint position with respect to the geometric parameters variation (calibration Jacobian matrix) [10], [13].

To estimate Δg we apply equation (5) for a number of manipulator configurations. It gives the system of equations:

$$\Delta X = \Phi \Delta g + E \quad (6)$$

where is:

$$\Delta X = \begin{bmatrix} \Delta x^1 \\ \Delta x^2 \\ \vdots \\ \Delta x^k \end{bmatrix}, \quad \Phi = \begin{bmatrix} J_g^1(q^1, g) \\ J_g^2(q^2, g) \\ \vdots \\ J_g^k(q^k, g) \end{bmatrix}, \quad (7)$$

and E is the error vector which includes the effect of unmodeled non-geometric parameters:

$$E = \begin{bmatrix} e^1 \\ e^2 \\ \vdots \\ e^k \end{bmatrix}. \quad (8)$$

Equation (6) can be used to estimate iteratively the geometric parameters. This equation is solved to get the least-squares error solution to the current parameters estimate. The least-squares solution can be obtained from:

$$\Delta g = (\Phi^T \Phi)^{-1} \Phi^T \Delta X. \quad (9)$$

At the each iteration, geometric parameters are updated by adding Δg to the current value of g :

$$g = g + \Delta g. \quad (10)$$

By solving equations (9) and (10) alternately, the procedure is iterated until the Δg approaches zero.

Calibration a manipulator is an identification process, and hence, one should take a careful look at the identifiability of the model parameters [10], [14]. A general method to determine these parameters have been proposed in [14]. Determination of the identifiable (base) geometric parameters is based on the rank of the matrix Φ . Some parameters of manipulator related to the locked passive joints may become unidentifiable in the calibration algorithm due to the mobility constraints. It reduces number of identifiable parameters in general for the closed-loop kinematic chain approach, compared with open-loop case.

As the measurement process is generally time consuming, the goal is to use set of manipulator configurations that use limited number of optimum points on the parameters estimation. Furthermore, goal is to minimize the effect of noise on the parameters estimation. The condition number of the

matrix Φ gives a good estimate of the persistent excitation [1]. Therefore, much work was led on finding the so-called optimal excitation. The task of selecting the optimum manipulator configurations to be used during the calibration is discussed and solutions are proposed in [14]–[16]. It is worth noting that most of geometric calibration methods give an acceptable condition number using random configurations. The paper [17] presents an updating algorithm to reduce the complexity of computing and observability index for kinematic calibration of robots. An active calibration algorithm is developed to include an updating algorithm in the pose selection process.

III. COMPUTER VISION

A. Camera Model

This section describes the camera model. Fig. 2. illustrates the basic geometry of the camera model. The camera performs transformation from the 3D projective space to the 2D projective space. The projection is carried by an optical ray originating (or reflected) from a scene point P . The optical ray passes through the optical center O_c and hits the image plane at the point p .

Prior describing the perspective transformation, and camera model, let us define the basic coordinate systems. The coordinate frames are defined as follows:

$O_w X_w Y_w Z_w$ - world coordinate system (fixed reference system), where O_w represents the principal point. The world coordinate system is assigned in any convenient location.

$O_c X_c Y_c Z_c$ - camera centered coordinate system, where O_c represents the principal point on the optical center of the camera. The camera coordinate system is the reference system used for camera calibration, with the Z_c axis the same as the optical axis.

$O_i X_i Y_i Z_i$ - image coordinate system, where O_i represents the intersection of the image plane with the optical axis. $X_i Y_i$ plane is parallel to $X_c Y_c$ plane.

Let (x_w, y_w, z_w) are the 3D coordinates of the object point P in the 3D world coordinate system, and (u, v) position of the corresponding pixel in the digitized image. A projection of the point P to the image point p may be represented by a 3×4

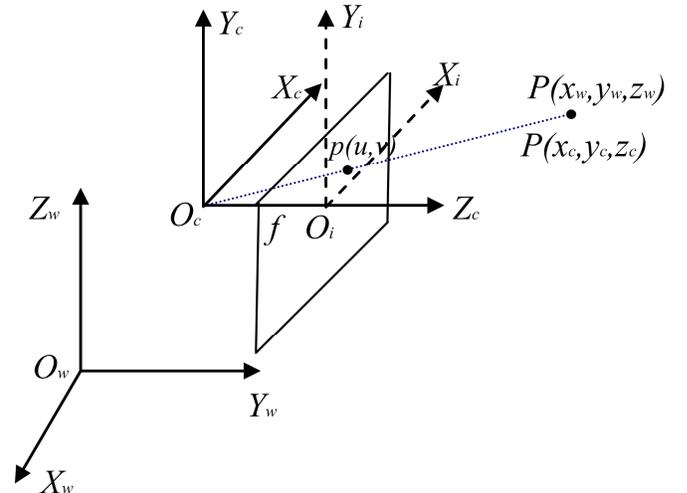


Fig. 2. The basic geometry of the camera model.

projection matrix (or camera matrix) M [2], [5]:

$$p = K[R \quad T]P = MP. \quad (11)$$

Matrix:

$$K = \begin{bmatrix} \alpha & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

is called the internal (intrinsic) camera transformation matrix. Parameters α , β , u_0 and v_0 are so called internal distortion-free camera parameters.

R and T , a 3x3 orthogonal matrix representing the camera's orientation and a translation vector representing its position, are given by:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \quad (13)$$

respectively. The parameters $r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}, r_{33}, t_x, t_y, t_z$ are external (extrinsic) parameters and represent the camera's position referred to the world coordinate system.

Projection in an ideal imaging system is governed by the pin-hole model. Real optical system suffers from a number types of distortion. The first one is caused by real lens spherical surfaces and manifests itself by radial position error. Radial distortion causes an inward or outward displacement of a given image point from its ideal (distortion free) location. This type of distortion is mainly caused by flawed radial curvature curve of the lens elements. A negative radial displacement (a point is imaged at a distance from the principle point that is smaller than predicted by the distortion free model) of the image point is referred to as barrel distortion. A positive radial displacement (a point is imaged at a distance from point that is larger than the predicted by the distortion free model) of the image point is referred to as pin-cushion distortion. The displacement is increasing with distance from the optical axis. This type of distortion is strictly symmetric about the optical axis. Fig. 3. illustrates the effect of radial distortion.

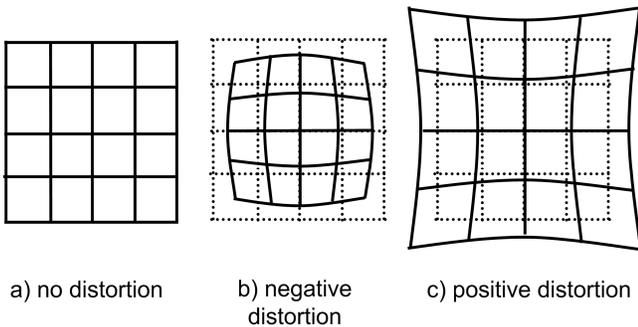


Fig. 3. Effect of radial distortion illustrated on a grid.

The radial distortion of a perfectly centered lens is usually modeled using the equations:

$$\Delta x_r = x_i(k_1 r^2 + k_2 r^4 + \dots), \quad (14)$$

$$\Delta y_r = y_i(k_1 r^2 + k_2 r^4 + \dots), \quad (15)$$

where r is the radial distance from the principal point of the image plane, and k_1, k_2, \dots are coefficients of radial distortion.

Only even powers of the distance r from the principal point occur, and typically only the first, or the first and the second terms in the power series are retained.

The real imagining systems also suffer from tangential distortion, which is at right angle to the vector from the center of the image. That type of distortion is generally caused by improper lens and camera assembly. Like radial distortion, tangential distortion grows with distance from the center of distortion and can be represented by equations:

$$\Delta x_t = -y_i(l_1 r^2 + l_2 r^4 + \dots), \quad (16)$$

$$\Delta y_t = x_i(l_1 r^2 + l_2 r^4 + \dots). \quad (17)$$

Fig.4. illustrates the effect of tangential distortion.

The reader is referred to [2]–[4] for more elaborated and more complicated lens models.

Note that one can express the distorted image coordinates as a power series using undistorted image coordinates as variables, or one can express undistorted image coordinates as a power series in the distorted image coordinates. The r in the above equations can be either based on actual image coordinates or distortion-free coordinates.

Bearing in mind the radial and tangential distortion, correspondence between distortion-free and distorted pixels image coordinates can be expressed by:

$$x_d = x_i + \Delta x_r + \Delta x_t, \quad (18)$$

$$y_d = y_i + \Delta y_r + \Delta y_t. \quad (19)$$

The parameters representing distortion of an image are: $k_1, k_2, \dots, l_1, l_2, \dots$. The distortion tends to be more noticeable with wide-angle lenses than telephoto lenses. Electro-optical systems typically have larger distortions than optical systems made of glass.

B. Camera Calibration

Camera calibration is considered as an important issue in computer vision applications. With the increasing need for higher accuracy measurement in computer vision, it has also attracted research effort in this subject. Task of camera calibration is to compute the camera projection matrix M from a set of image-scene point correspondences. By correspondences it means a set $\{(p_i, P_i)\}_{i=1}^m$ where p_i is a homogeneous vector representing image point and P_i is a homogeneous vector representing scene point, at the i^{th} step. Equation (11) gives an important result: the projection of a point P to an image point p by a camera is given by a linear mapping (in homogeneous coordinates):

$$p = MP. \quad (20)$$

The matrix M is non-square and thus the mapping is many-to-one. All scene points on a ray project to a single image point.

To compute M , it has to be solved the system of homogeneous linear equations

$$s_i p_i = M P_i, \quad (21)$$

where s_i are scale factors.

Camera calibration is performed by observing a calibration object whose geometry in 3D space is known with very good precision. The calibration object usually consists of two or three planes orthogonal to each other. These approaches

require an expensive calibration apparatus. Accurate planar targets are easier to make and maintain than three-dimensional targets. There is a number of techniques which only requires the camera to observe a planar pattern(s) shown at a few different orientation (Fig. 5). The calibration points are created by impressing a template of black squares (usually chess-board pattern) or dots on top of white planar surface (steel or even a hard book cover [5]). The corners of the squares are treated as a calibration points. Because the corners are always rounded, it is recommended to measure the coordinate of a number of points along the edges of the square away from the corners, and then extrapolate the edges to obtain position of the corners which lie on the intersection of adjacent edges.

Due to the high accuracy performance requirement for

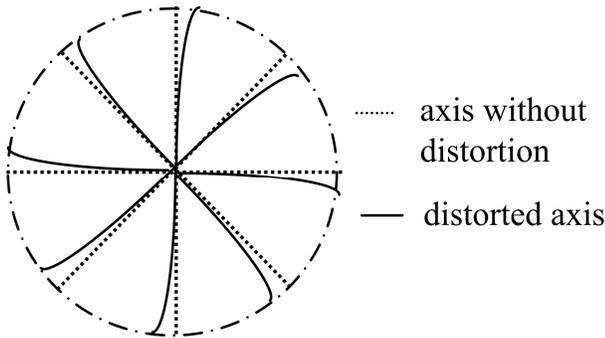


Fig. 4. Effect of tangential distortion.

camera calibration, a sub-pixel estimator is desirable. It is a procedure that attempts to estimate the value of an attribute in the image to greater precision than that normally considered attainable within restrictions of the discretization. Since the CCD camera has relatively low resolutions, interest in a sub-pixel method arises when one applies CCD-based image systems to the computer integrated manufacturing [6].

Camera calibration entails solving for a large number of calibration parameters, resulting in the large scale nonlinear search. The efficient way of avoiding this large scale nonlinear search is to use two-stage technique, described in [2]. The methods of this type in the first stage use a closed-form solution for most of the calibration parameters, and in the second stage iterative solution for the other parameters.

In [4] a two-stage approach was adopted with some

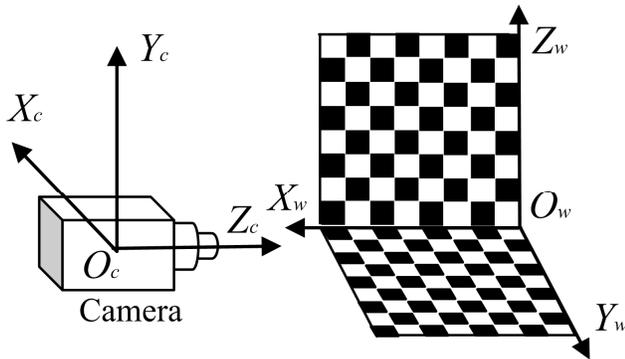


Fig. 5. Illustration of experimental setup for camera calibration using coplanar set of points.

modification. In the first step, the calibration parameters are estimated using a closed-form solution based on a distortion-free camera model. In the second step, the parameters estimated in the first step are improved iteratively through a nonlinear optimization, taking into account camera distortion. Since the algorithm that computes a closed-form solution is no iterative, it is fast, and solution is generally guaranteed. In the first step, only points near the optical axis are used. Consequently, the closed-form solution isn't affected very much by distortion and is good enough to be used as an initial guess for further optimization. If an approximate solution is given as an initial guess, the number of iterations can be significantly reduced, and the globally optimal solution can be reliably reached.

C. Stereo Vision

Calibration of one camera and knowledge of the coordinates of one image point allows us to determine a ray in space uniquely (back-projection of point). Given a homogeneous image point p , we want to find its original point P from the working space. This original point P is not given uniquely, but all points on a scene ray from image point p . Here, we will consider how to compute 3D scene point P from projections p_i in the several cameras, or projections p_i in one camera at different positions (different images are denoted by superscript i). Assume that m views are available, so that we have to solve linear system

$$s_i p_i = M_i P, \quad i=1, \dots, m. \quad (22)$$

This approach is known as triangulation (it can be interpreted in terms of similar triangles). Geometrically, it is a process of finding the common intersection of m rays given by back-projection of the image points by the cameras. In the reality, image points p_i are corrupted by noise, and the rays will not intersect and the system would have no solution. We might compute P as the scene point closest to all of the skew rays.

If two calibrated cameras observe the same scene point P , its 3D coordinates can be computed as the intersection of two of such rays. The epipolar geometry is a basis of a system with two cameras (principle of stereo vision). It is illustrated on Fig. 6.

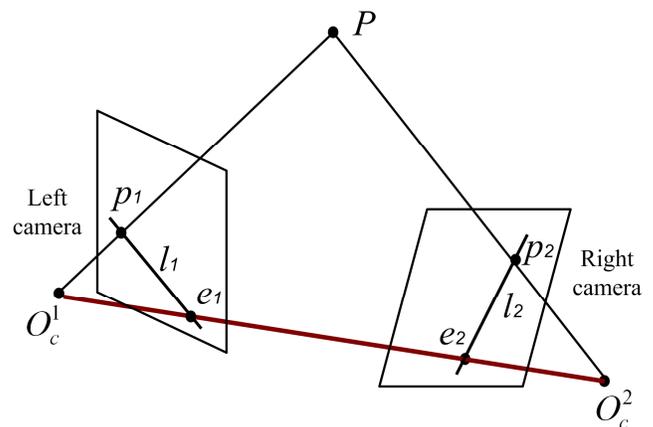


Fig. 6. The epipolar geometry.

Let O_c^1, O_c^2 represents the optical centers of the first and second camera, respectively. The same consideration holds if one camera takes two images from two different locations. In that case O_c^1 represents optical center of the camera when the first image is obtained, and O_c^2 represents the optical center for the second image. p_1 and p_2 denote the images of the 3D point P . The base line is the line joining the camera centers O_c^1 and O_c^2 . The baseline intersects the image planes in the epipoles e_1 and e_2 . Alternatively, an epipole is the image of the optical center of one camera in the other camera. Any scene point P and the two corresponding rays from optical centers O_c^1 and O_c^2 define an epipolar plane. This plane intersects the image plane in the epipolar line. It means, an epipolar line is the projection of the ray in one camera into the other camera. Obviously, the ray $O_c^1 P$ represents all possible positions of P for the first image and is seen as the epipolar line l_2 in the second image. The point p_2 in the second image that corresponds to p_1 must thus lie on the epipolar line in the second image l_2 , and reverse. The fact that the positions of two corresponding image points are not arbitrary is known as the epipolar constraint. This is a very important statement for the stereo vision. The epipolar constraint reduces the dimensionality of the search space for a correspondence between p_1 and p_2 in the second image from 2D to 1D.

A special relative position of the stereo cameras is called rectified configuration. In this case image planes coincide and

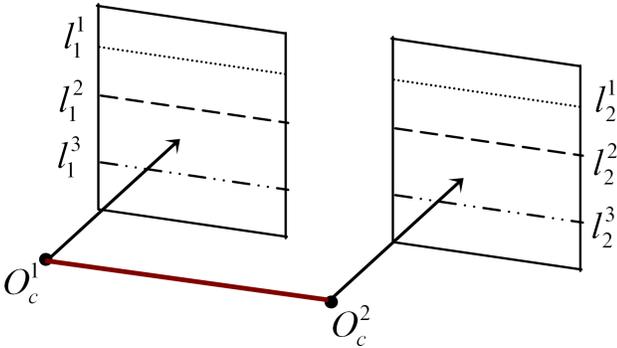


Fig. 7. The rectified configuration of two cameras.

line $O_c^1 O_c^2$ is parallel to them, as shown in Fig. 7.

The epipoles e_1 and e_2 go to infinity, and epipolar lines coincide with image rows, as a consequence. For the rectified configuration, if the internal calibration parameters of both cameras are equal, it implies that corresponding points can be sought in 1D space along image rows (epipolar lines).

The optical axes are parallel, which leads to the notion of disparity that is often used in stereo vision literature. Top view of two cameras stereo configuration with parallel optical axes is shown in Fig. 8. World coordinate system is parallel to cameras coordinate systems. The principal point O_w of the world coordinate system is assigned on the midway on the baseline. The coordinate z_w of point P represents its distance from the cameras ($z_w = 0$), and can be calculated from the

disparity $d = u_1 - u_2$. Values u_1, u_2 are measured at the same height (same rows of images). Noting that:

$$\frac{u_1}{f} = \frac{x_w + \frac{B}{2}}{z_w}, \quad \frac{u_2}{f} = \frac{x_w - \frac{B}{2}}{z_w}, \quad (23)$$

we have:

$$z_w = \frac{Bf}{d}. \quad (24)$$

The remaining two coordinates of the 3D point P can be calculated from equations:

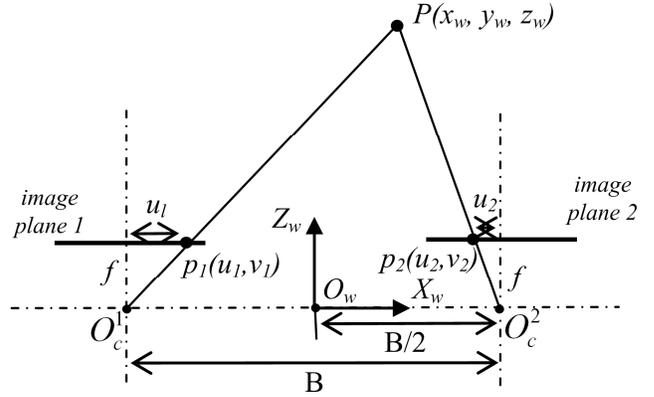


Fig. 8. Top view of two cameras with parallel optical axes rectified configuration.

$$x_w = \frac{-B(u_1 + u_2)}{2d}, \quad y = \frac{Bv_1}{d}. \quad (25)$$

The position of the point P in the 3D scene can be calculated from the disparity d . It is a question, how the same point can be found in two images if the same scene is observed from two different viewpoints. The solution of this correspondence problem is a key step in any stereo vision. Automatic solution of the correspondence problem is under extensive exploration. Until now there is not solution in general case. The inherent ambiguity of the correspondence problem can in practical cases be reduced using several constrains. A vast list of references about this task can be found in the [3].

The geometric transformation that changes a general cameras configuration with non-parallel epipolar lines to the parallel ones is called image rectification. More deep explanation about computing the image rectification can be found out in [3].

IV. ROBOT CALIBRATION USING COMPUTER VISION

Measurement of robot manipulator end-effector pose (i.e., position and orientation) in the reference coordinate system is unquestionably the most critical step towards a successful open-loop robot calibration. A variety of measurement techniques ranging from coordinate measuring machines, proximity measuring systems, theodolites, and laser tracking interferometer systems to inexpensive customized fixtures have been employed for calibration tasks. These systems are very expensive, tedious to use or with low working volume [12], [18], [19]. In general, the measurement system should be

accurate, inexpensive and should be operated automatically. The goal is to minimize the calibration time and the robot unavailability.

To overcome the above limitations, advances in robot calibration allow the start using a computer vision to calibrate a robot. Compared to those mechanical measuring devices, the camera system is low cost, fast, automated, user-friendly, non-invasive and can provide high accuracy [20].

There are two types of setups for vision-based robot pose measurement. The first one is to fix cameras in the robot environment so that the cameras can see a calibration fixture mounted on the robot end-effector while the robot changes its configuration. The second typical setup is to mount a camera or a pair of cameras on the end-effector of the robot manipulator.

The stationary camera configuration requires the use of stereo system placed at fixed location. It is not possible compute 3D scene point P position from only one projection p , on the camera plane. The stereo system has to be placed in location that maintains necessary field-of-view overlap. The proper camera position needs to be selected empirically. The stereo system must be calibrated before manipulator calibration. The manipulator is placed in a number of configurations. From pair of images the location (position and orientation) of the calibration board is computed for every configuration (Fig. 9.). At the each configuration, geometric parameters are updated by adding Δg (calculated in accordance with equation (9)) to the current value of g [21].

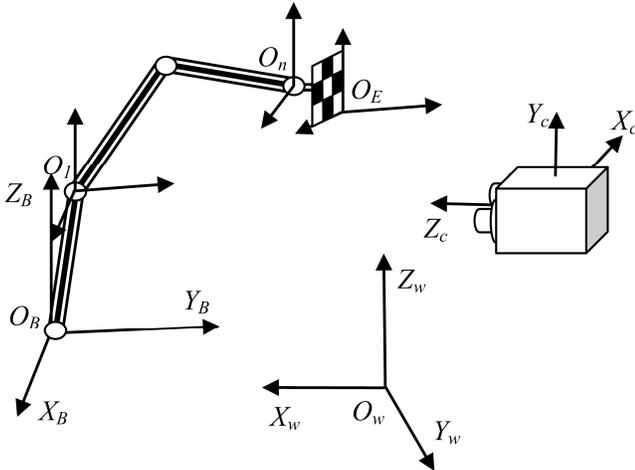


Fig. 9. A manipulator calibration using stationary camera configuration.

If it is enough to measure only the end-effector pose (usually tool's tip) for robot calibration, then it is not necessary to use a calibration plate. In that case it is enough to place a marker on the end-effector of the manipulator. Tests were conducted with square markers. In this way the calibration of the manipulators, from the viewpoint of practical implementation, is simplified. Also, this is an important prerequisite for increasing the flexibility of the manipulators, because in the case of automatically interchangeable tools, a marker can be placed on the end-effector of the manipulator. Then, for any change of tools, by tracking the position of marker it is possible to accomplish the recalibration of the manipulator.

The position of a marker (end-effector) is determined on the basis of pairs of images from stereo camera system [22]. Any point of the marker can be assumed as referent point of end-effector of the manipulator. For this reason, hereinafter a term „marker” is considered as a whole, rather than a one specified point. In this case the main problem is the automatic detection of corresponding points. The corresponding points are represented by a set of marker points on both images.

This paper uses algorithm based on the most similar intensity area correlation. The algorithm assumes that more pixels have similar intensity (color). Therefore, correlation of two pixels does not provide sufficient information because of the existence of more similar candidates. Thus, correlation of more adjacent pixels which are forming the window of $h \times w$ pixels is determined. When stereo system with parallel optical axes is used, the epipolar lines of both cameras lie on the same height on both images, as shown on Fig. 10.

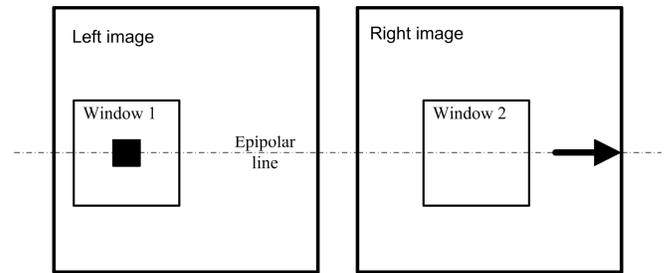


Fig. 10. Windows position of two corresponding points.

The algorithm principle is as follows. Window of $h \times w$ pixels is formed. The window central pixel represents the referent point on one of two images from stereo system (Ex. left image). This window is used as referent area to be searched on the second image (i.e. right image). On the second image the same size window is observed on the same height as on the first image. By changing window disparity d the second window is sliding along u axis. Measure of two windows intensity likelihood, i.e. cost function, is calculated as Sum of Squared Differences of all pixels intensities in both windows.

$$c(u, v, d) = \sum_{(h,w)} [Im_L(u+h, v+w) - Im_R(u+h+d, v+w)]^2 \quad (26)$$

The value of disparity d , for which is obtained minimal value of cost function, gives the position of window which is the best correlated with the reference window. Therefore, the corresponding windows are on the same height on both images, but shifted along u axis for:

$$disparity(u, v) = \min c(u, v, d) \quad (27)$$

For the purpose of calibration, stereo system with two standard web color cameras was used. The algorithm was tested with color marker placed on the end-effector of a modular Robix manipulator. Adopted window size is 5×5 pixels. Fig. 11. and Fig. 12. present images from left and right cameras, respectively. A detail of marker found on the second image is shown on Fig. 12. Fig. 13. shows graphical representation of cost function for disparity change along epipolar line, from minimum to maximum value. It is obvious, as it shown on Fig. 13, that a reliable method of determining the corresponding points is obtained by using marker and



Fig. 11. Image from left camera.

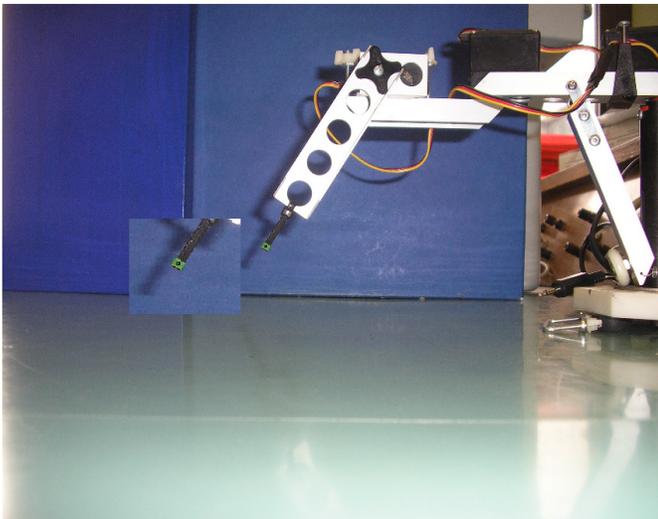


Fig. 12. Image from right camera.

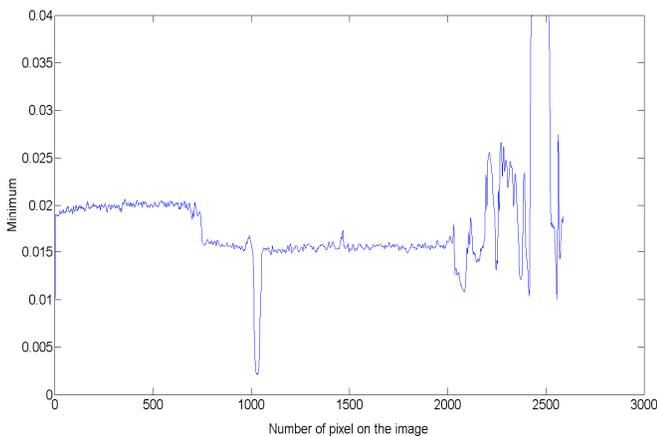


Fig. 13. Graphical representation of cost function for disparity change along epipolar line.

selected cost function. Selected cost function has a pronounced global minimum.

For successfully finding of corresponding points, choice of window size is crucial. In the classical problem of

correspondence, if the window size is too small, it increases probability of occurrence of a large number of candidates for correspondence.

This increases probability of wrong selection of corresponding points. On the other hand, if the window size is too large, there is a possibility for error because of a constant value of disparity within the window. Therefore, there is no single recommendation for the best window size. In special cases, even an adaptive window size is suggested, but such algorithms are generally very complex, compute demanding and not widely accepted in practice.

In accordance with previous demonstration, windows size will depend on the size of the marker when it is necessary to determine markers correspondence on two images. Marker is an area with nearly constant intensity (color). By adopting a window smaller than the marker, all the windows, which are contained in the marker, will be detected as candidates. Thus, there will be no single solution for the problem of correspondence. In another case, by adopting a window size larger than the size of marker, by increasing a size of window the value of cost function will not depend on the compliance of marker points. Value of the cost function will be influenced by compliance of region points out of marker. Therefore, it is reasonable assumption that the cost function will have less pronounced global minimum. Further augmentation of window size (relative to marker size) may produce error in correspondence estimation. Points surrounding marker will have significant influence on window correlation compared to marker points. Based on this analysis, it is concluded that the best results are achieved by adopting window size approximately equal to marker size.

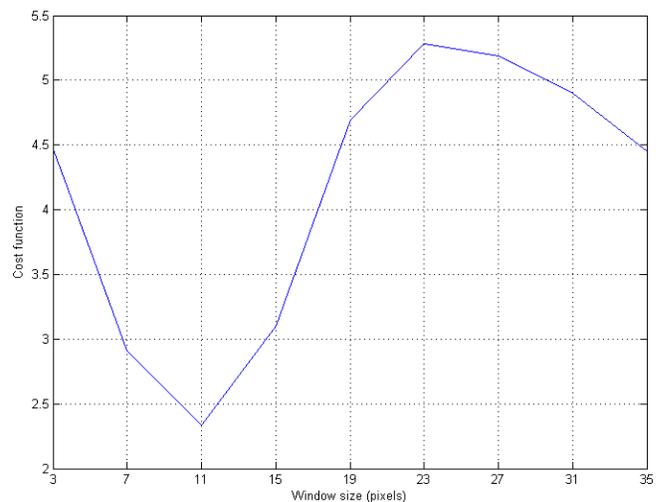


Fig. 14a. Graphical representation of cost function for disparity change along epipolar line.

Assumptions about the window size effects (relative to marker size) on the reliability of the correspondence procedure have been tested in several cases. Window size has been altered for different marker sizes. Diagrams of minimum values change of cost functions with change of window size, for three different sizes of markers, are illustrated on Fig. 14. In order to compare results, values of cost functions, shown on diagrams, are divided with number of pixels that belongs to

window. In this way, the cost functions represents the average inconsistency for every pixel of two windows. The first marker (Fig.14.a.) is the size of 11x11 pixels, the second marker (Fig.14.b) is the size of 21x21 pixels and a third marker (Fig.14.c) is the size of 37x37 pixels. The illustrations confirm that the best results are achieved by adopting that window size is equal to marker size.

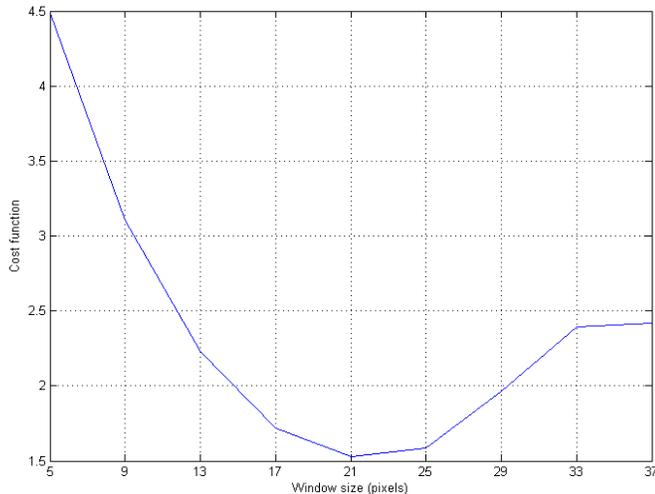


Fig. 14b. Graphical representation of cost function for disparity change along epipolar line.

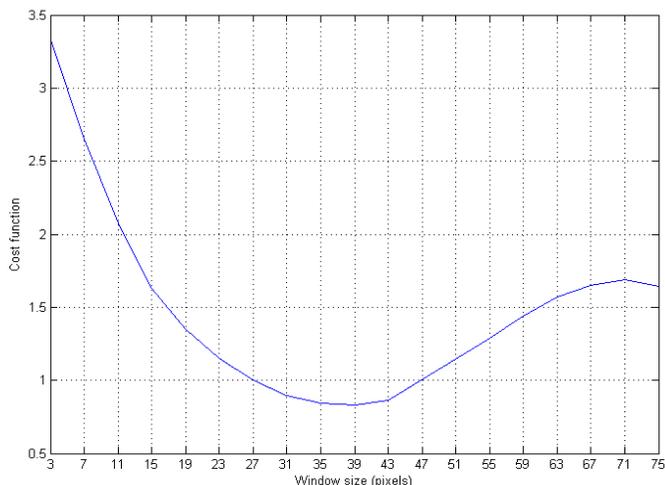


Fig. 14c. Graphical representation of cost function for disparity change along epipolar line.

V. CONCLUSION

An overview of new features for improving the calibration of industrial manipulators using visual systems is presented in this paper. The focus is on the stereo system with parallel optical axes. The practical aspects of using the algorithm based on area correlation are particularly analyzed. In order to increase the reliability of the corresponding areas determination the next is proposed: set the marker on end-effector of manipulator, choose the window size equal to marker size and scale cost function with total number of pixels for each window. Conducted experiments and shown illustrations confirm that presented calibration method allows

efficient and reliable manipulator calibration in standard conditions.

REFERENCES

- [1] W. Khalil and E. Dombre, *Modeling, Identification and Control of Robots*, Kogan Page Science, 2004.
- [2] R. Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses", *IEEE J. Robotics and Automation*, Vol. RA-3, No. 4, pp. 323-344, 1987.
- [3] M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis, and Machine Vision*, Thomson, 2008.
- [4] J. Weng, P. Cohen and M. Herniou, "Camera Calibration with Distortion Models and Accuracy Evaluation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 10, pp. 965-981, 1992.
- [5] Z. Zhuang, "A Flexible New Technique for Camera Calibration", Technical Report MSR-TR-98-71, 2008.
- [6] D. J. Kang, J. E. Ha, and M. H. Jeong, "Detection of Calibration Patterns for Camera Calibration with Irregular Lighting and Complicated Backgrounds", *Int. J. of Control, Automation, and Systems*, Vol. 6, No. 5, pp. 746-754, 2008.
- [7] P. Maric "Computer Vision Systems for the Enhancement of Industrial Robots Flexibility" (Invited Paper), Proc. X Triennial International SAUM Conference on Systems, Automatic Control and Measurements, pp. 80-88, 2010.
- [8] X. Tian, X. Zhang, K. Yamazaki and A. Hansel, "A study on three-dimensional vision system for machining setup verification", *Int. J. Robotics and Computer-Integrated Manufacturing*, No. 26, pp. 46-55, 2010.
- [9] B. Paul, *Robot manipulators: mathematics, programming, and control*, Cambridge, MIT Press, 1981.
- [10] W. Khalil, M. Gautier and Ch. Enguehard, "Identifiable parameters and optimum configurations for robots calibration", *Robotica*, Vol. 9, pp. 63-70, 1991.
- [11] E. Jackson, Z. Lin and D. Eddy, "A global formulation of robot manipulator kinematic calibration based on statistical considerations", *Proc. IEEE Conf. on Systems Man and Cybernetics*, pp. 3328-3333, 1995.
- [12] J. Renders, E. Rossignol, M. Besquetand and R. Hanus, "Kinematic calibration and geometrical parameter identification for robot", *IEEE Trans. on Robotics and Automation*, Vol. 7, No. 6, pp. 721-732, 1991.
- [13] P. Maric and V. Potkonjak, "Geometrical Parameters Estimation for Industrial Manipulators Using Two-step Estimation Schemes", *J. Of Intelligent and Robotic Systems*, vol. 24, pp. 89-97, 1999.
- [14] D. J. Bennett, J. M. Hollerbach, "Autonomous calibration of single-loop closed kinematic chains formed by manipulators with passive endpoint constraints", *IEEE Trans. on Robotics and Automation*, Vol. 7, No. 5, pp. 597-606, 1991.
- [15] W. Khalil, G. Garcia and J. Delagarde, "Calibration of geometrical parameters of robots without external sensors", *Proc. IEEE Int. Conf. On Robotics and Automation*, Vol. 3, pp.3039-3044, 1995.
- [16] S. Bay, "Autonomous parameter identification by optimal learning control", *IEEE Control Systems*, pp. 56-61, 1993.
- [17] Y. Sun and J. Hollerbach, "Active Robot Calibration Algorithm", *Proc. ICRA*, pp. 1276-1281, 2008.
- [18] M. Vincze, J. Prenninger and H. Gander, "A laser tracking system to measure position and orientation of robot end effectors under motion", *Int. J. Robotics Research*, Vol.13, No. 4, pp. 305-314, 1994.
- [19] M. Driels, "Automated partial pose measurement system for manipulator calibration experiments", *IEEE Trans. on Robotics and Automation*, Vol. 10, No. 4, pp. 430-440, 1994.
- [20] H. Zhuang and Z. S. Roth, "On Vision-Based Robot Calibration", *Proc. SOUTHCON 94*, pp. 104-109, 1994.
- [21] D. Kosić, V. Đalić, P. Marić, "Robot Geometry Calibration in an Open Kinematic Chain Using Stereo Vision", *Proc. International Scientific Conference UNITECH 2010, Gabrovo, Bulgaria*, vol. 1, pp. 528-531, 2010.
- [22] N. Sebe, M. S. Lew, *Robust Computer Vision – Theory and Applications*, Kluwer Academic Publishers, volume 26, 2003.

Land Vehicle Navigation System Based on the Integration of Strap-Down INS and GPS

Rade Stančić and Stevica Graovac

Abstract—The concept and results of integration of a strap-down inertial navigation system (INS) based on low-accuracy inertial sensors and the global positioning system (GPS) for the purposes of land vehicle navigation have been presented in this paper. The integration is made by an implementation of extended Kalman filter scheme, both for the initial alignment and navigation phases. Traditional integration schemes (centralized and cascaded) are basically held on the usage of high-accuracy inertial sensors. The idea behind suggested algorithm is to use low-accuracy inertial sensors and GPS as the main source of a navigation information, while the acceptable accuracy of INS is achieved by the appropriate damping of INS errors. The specified values of damping coefficients can have different influence depending on the fact whether the moving object is maneuvering or is moving with a constant velocity during the intervals of absence of GPS data. The analysis of integrated navigation system performances is made experimentally using the data acquired along the real land vehicle's trajectory and by artificial introduction of intervals of absence of GPS data on the parts characterized both by maneuver and by constant velocity, and by varying the values of error damping coefficients.

Index Terms—Navigation, Land vehicles, Inertial navigation, Global positioning system, Extended Kalman filter.

I. INTRODUCTION

THE integration of heterogeneous navigation systems is a frequently used approach for the increasing of overall reliability and accuracy of navigation algorithms. One of the most popular examples nowadays consists in the integration of inertial navigation and global positioning systems. The main idea behind this approach is based on the fact that the errors separately characterizing anyone of them are not mutually correlated at all. While the GPS errors are basically due to RF channel disturbances, changes in configuration of observed satellites, occultations of the receiver antenna, and atmospheric influences, the errors characterizing an INS are of a long

periodic nature and are independent on environmental conditions. According to these facts one can expect that two systems could assist and correct each other, increasing the overall navigation system reliability and accuracy, this way. Satellite based positioning system provides more accurate information regarding the moving object's position and linear velocity in comparison to INS, especially if one considers a low-cost strap-down INS suitable for the land vehicle navigation applications. The function of such type of an INS consists in providing of navigation parameters on the intervals between consecutive GPS measurements, calculation of an object's angular orientation, and providing of the overall set of navigation parameters during the intervals of absence of GPS data. In other words, only the short term accuracy of an INS is required. While the GPS measurements are available, the estimations of INS errors are calculated and these are used for INS corrections on the intervals when GPS data are absent. Besides this function, GPS data are used during the overall system initialization and calibration also.

For the cases of middle and high-accuracy INS, there are two basic methods/schemes of INS/GPS integration: centralized and cascaded. In the case of a centralized scheme, there exists a unique INS/GPS navigation algorithm with a generalized navigation parameters error model. In the cascaded integration scheme, corrections of an INS output information are done based on GPS measurements, without any changes in navigation data processing, neither in the INS nor in the GPS part of algorithm. Wherever the set of independent sources is used, the cascaded integration scheme is preferable, allowing the choice of the system providing the most complete navigation information at a particular moment. If one wants to use the low-cost/low-accuracy INS inside this integration scheme, the problem of system initialization as well as of the on-line error estimation becomes more meaningful.

The approach used here in integration of INS and GPS is based on theoretical basis exposed in [1], [3] and [13]. The optimization of an inertial navigation algorithms based on information regarding moving object's dynamics is particularly considered in [4], followed by the analysis of a scope of applicability.

One can recognize the following main items in an attempt to design and implement the integrated INS/GPS navigation system for different applications, including the cases of automatically controlled land vehicles:

This research and development work has been partially supported by Ministry of Science and Technology of Republic of Serbia, as a part of project TR-11012 "Development of an integrated navigation system applied for localization of land vehicles".

R. Stančić is with SMATSA (Serbia and Montenegro's Air Traffic Services Agency), (e-mail: rstancic@gmail.com).

S. Graovac is with School of Electrical Engineering, University of Belgrade, Signals & Systems Department, (e-mail: graovac@etf.rs)

1. Overall system configuration (basically it is GPS + “Inertial Measurement Unit”, but sometimes without rate gyros, with addition of magnetometer, angular sensors, odometer, etc.);
2. Quality of inertial instruments and their error model parameterization (the type of instrument error models and procedures of actual parameterization);
3. Duration of the initialization/alignment phases (how complex and how long these procedures are and what could be done to make them adaptive in some way in order to reduce their duration);
4. Complexity of an integration algorithm (tight or weak coupling of navigation systems, variability of measurement error model, additional navigation state error damping, how to prepare the system for the intervals of GPS data absence,...);
5. System performances during the intervals of absence of GPS data (the existence of vehicle maneuver during these intervals).

While for some applications, [5], it was proposed to use gyro-free “dead-reckoning” configuration, our choice was to use the full strap-down navigation system as the INS part of system as in [6], having in mind that moving across a terrain in general introduces meaningful pitch and roll angles and that gravity components affect the acceleration measurements. The magnetometer and two additional angular sensors have been added for the initial alignment purposes while the usage of an odometer was supposed as the alternative source of velocity data when GPS signals are absent. There is a general trend to use low-accuracy inertial instruments in these systems (mostly of MEMS type, while in our case these were of electro-mechanical type). The importance of proper error model parameterization was respected and it was done in a manner like in [7] where only strap-down INS was used for a land vehicle navigation purposes. While some authors, [8], [9], suggest the usage of adaptive Kalman filter schemes or banks of Kalman filters in order to reduce the time needed to obtain good estimates of instrument error models during the initial alignment phase, our approach was to use “regular” EKF (the same as in navigation mode) because the experiments have shown that the required duration of this process was still acceptable. Regarding the type of integration scheme we have chosen the modified cascaded one as the type allowing full separation of navigation systems. The quality of KF estimation during the integrated navigation phase is generally monitored and in some examples, [10], the adaptive tuning of EKF is suggested. In our case we have used the simple idea to monitor the values of residuals and to use the predicted values of outputs instead of actual measurements if they are of “outlier” type. The very important task of preparing of a system for the prediction phase when GPS data are absent, some authors solve by application of artificial neural networks, [11], in order to make the model of navigation state error as accurate as possible. On the other hand, we have recognized the fact of vehicle maneuvering as the most important one characterizing the dynamic environment at the moment when GPS data are

lost and suggested the usage of odometer as the alternate source of velocity data and the adaptation of velocity error damping coefficients according to the registered level of vehicle’s maneuver, [14], [15].

II. INS/GPS INTEGRATION SCHEME

The diagram shown on Fig. 1 illustrates the INS/GPS integration scheme while the Fig. 2 illustrates the system working regimes. The main working regimes of the integrated navigation system are “INITIAL ALIGNMENT” and “NAVIGATION”. Immediately after the system start-up, the initialization procedure starts by transferring GPS data regarding the geographical longitude, latitude, and height as well as of velocity. The next step consists in the initializations of Kalman filter matrices, the initialization of the block used for corrections of deterministic errors of inertial instruments (biases, scale factor errors, non-orthogonality, etc.), and in system preparation for the start of “INITIAL ALIGNMENT”.

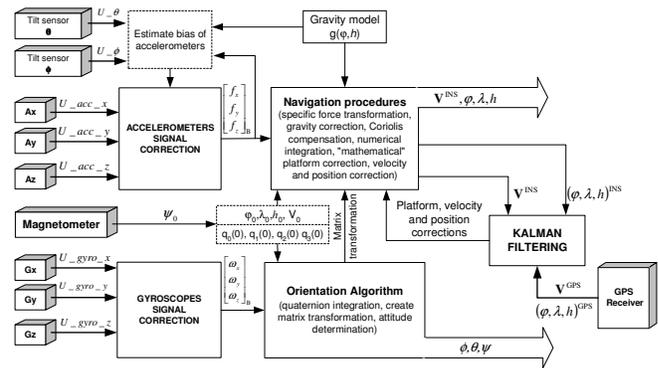


Fig. 1. Block diagram of integrated navigation system.

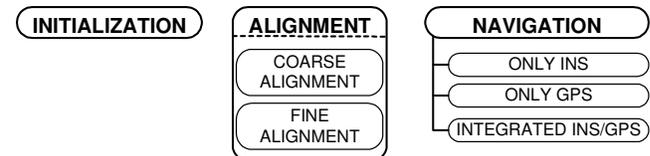


Fig. 2. Working regimes of integrated navigation system.

After the initialization is finished, the system automatically starts the first alignment step – “COARSE ALIGNMENT”. During this step the following procedures are made: coarse alignment in azimuth (using the magnetic compass), calibration of gyro drifts, horizontal alignment, initial estimation of angular attitude, determination of quaternion parameters, and calculation of transform matrix (DCM) coefficients. Using the additional angular sensors it is possible to determine the accelerometers biases also. After the “COARSE ALIGNMENT” is finished, the “FINE ALIGNMENT” starts automatically. As a result of this procedure, the following estimates are available: non-stationary components of accelerometers biases and gyro drifts, diagonal elements of covariance matrices in Kalman filter, final estimates of angular attitude, and final data for the correction block. The whole “INITIAL ALIGNMENT” phase

is done in stationary conditions and for the ground vehicles its duration was of order of 5 – 10 minutes. The extended Kalman filter (EKF) used for the estimation of INS errors uses velocity deviations $\delta\mathbf{V}$ as the measurements. If the GPS works in good geometric conditions (more than four satellites available), the system works as complete INS/GPS integrated one and the differences between velocity data obtained separately by INS and GPS are used as the measurements driving EKF algorithm. If some losses of GPS data exist during this phase, EKF works in “PREDICTION” mode using the last state vector estimates obtained before the loss as inputs.

III. STRAP-DOWN INS ALGORITHM

The sampling technique used in the implementation of INS algorithm was based on the idea that the navigation tasks can be separated into two parts: Earth rotation characterized by slow dynamics, calculated with sampling interval Δt_l , and object's motion characterized by fast dynamics (determination of inertial sensors errors, forming of vectors of velocity ($\Delta\mathbf{V}$) and attitude ($\Delta\boldsymbol{\alpha}$) increments), calculated with sampling interval Δt_h .

The attitude determination requires the forming of matrix transforms (using direction cosines or quaternions). If the orientation of a navigation coordinate frame practically is not changed during the one calculation step, it can be considered as the inertial one. Quaternion equation representing the transformation from non-inertial (body fixed) coordinate frame and the inertial one, can be represented as the first step in calculation of a matrix transform \mathbf{C}_B^N in the form [1]:

$$\mathbf{Q}_{k+1}^P = \mathbf{Q}_k^F \Delta\boldsymbol{\lambda} \quad (1)$$

where: \mathbf{Q}^P - preliminary quaternion, \mathbf{Q}^F - final quaternion, and $\Delta\boldsymbol{\lambda} = \Delta\lambda_0 + \Delta\lambda_1 i + \Delta\lambda_2 j + \Delta\lambda_3 k$ - increment of a quaternion. The transform from inertial to navigation coordinate frame is obtained on the second step represented via quaternion equation:

$$\mathbf{Q}_{k+1}^F = \Delta\mathbf{m} \mathbf{Q}_{k+1}^P \quad (2)$$

where: $\Delta\mathbf{m} = \Delta m_0 - \Delta m_1 i - \Delta m_2 j - \Delta m_3 k$ - quaternion increment representing the rotation of a navigation coordinate frame relative to the inertial one. In order to solve (1) one needs to calculate the angular increment $\Delta\boldsymbol{\Phi}$, given as

$$\Delta\boldsymbol{\Phi} = \int_{t_n}^{t_n + \Delta t_n} \boldsymbol{\omega} dt + \frac{1}{2} \int_{t_n}^{t_n + \Delta t_n} (\boldsymbol{\Phi} \times \boldsymbol{\omega}) dt \quad (3)$$

The second right hand term in (3) introduces the correction of a coning motion that should be done more frequently than the matrix transform is calculated. There are a number of suggested algorithms of coning effect correction based on (3) and for the purposes of the actual algorithm the four-step correction algorithm suggested in [1] was used.

The equations specifying the velocity and position of an

object in navigation coordinate frame (NED) can be given in the integral form as:

$$\begin{aligned} \mathbf{V}^N &= \int_0^t \mathbf{f}^N dt - \int_0^t [2\boldsymbol{\omega}_{IE}^N + \boldsymbol{\omega}_{EN}^N] \times \mathbf{V}^N dt + \int_0^t \mathbf{g}^N dt \\ \mathbf{S}^N &= \int_0^t \mathbf{V}^N dt \end{aligned} \quad (4)$$

where: $\mathbf{V}^N = [V_N \ V_E \ V_D]^T$ - velocity vector, $\mathbf{S}^N = [S_N \ S_E \ S_D]^T$ - position vector, $\mathbf{g}^N = [0 \ 0 \ g]^T$ - gravitation vector, $\boldsymbol{\omega}_{IE}^N = [\omega_e \cos \varphi \ 0 \ -\omega_e \sin \varphi]^T$, vector representing Earth rotation relative to the inertial space ($\omega_e = 7.292115 \times 10^{-5}$ rad/s, φ - geographic latitude),

$\boldsymbol{\omega}_{EN}^N = \begin{bmatrix} \frac{V_E}{R_P + h} & -\frac{V_N}{R_M + h} & \frac{V_E \tan \varphi}{R_P + h} \end{bmatrix}^T$ - vector representing

the rotation of a navigation coordinate frame relative to the Earth, where:

$$R_P = \frac{a}{(1 - e^2 \sin^2 \varphi)^{1/2}} \quad R_M = \frac{a(1 - e^2)}{(1 - e^2 \sin^2 \varphi)^{3/2}} \quad (5)$$

for ellipsoid WGS-84: $a = 6,378.137$ km, $e^2 = 0.00669438$.

The first integral in (4) represents the sum of velocity increments along one calculation step as a result of transform of a specific force vector from body-fixed frame to the navigation frame (NED). The simultaneous existence of linear and angular oscillations along two perpendicular axes introduces so called sculling effect with negative influence onto the overall accuracy if it is not corrected with the frequency not high enough.

Having in minds that the Coriolis and gravitational accelerations are slowly changing, their corrections can be made with lower frequency of calculation,

$$(\mathbf{V}_{l+1}^N)_{ccg} = [2\boldsymbol{\Omega}_{IE}^N \cdot \Delta t_l + \boldsymbol{\Omega}_{EN}^N \cdot \Delta t_l] \cdot (\mathbf{V}_l^N)_{ccg} + \mathbf{g} \cdot \Delta t_l \quad (6)$$

where: $\boldsymbol{\Omega}_{IE}^N = [\boldsymbol{\omega}_{IE}^N \times]$ - skew-symmetric matrix with elements of vector $\boldsymbol{\omega}_{IE}^N$, $\boldsymbol{\Omega}_{EN}^N = [\boldsymbol{\omega}_{EN}^N \times]$ - skew-symmetric matrix with elements of vector $\boldsymbol{\omega}_{EN}^N$, \mathbf{V}_l^N - velocity correction vector onto the l -th calculation step.

Corrected value of velocity vector is obtained as the output of INS. The object's position in NED is obtained via its integration.

IV. THE INTEGRATION ALGORITHM

The suggested INS/GPS integration algorithm was based on the assumption that inertial sensors are of low accuracy, not allowing the implementation of an autonomous INS and application of traditional centralized or cascaded integration schemes. As a result of this, the algorithm based on INS error damping is performed. This error damping is based on velocity information supplied by GPS, in order to decrease amplitudes of their oscillations, using the frequency of updating which is higher than Schuler's. Besides this, the velocity and position

corrections are produced based on the velocity and position estimates obtained by EKF. The integration scheme is shown on Fig. 3.

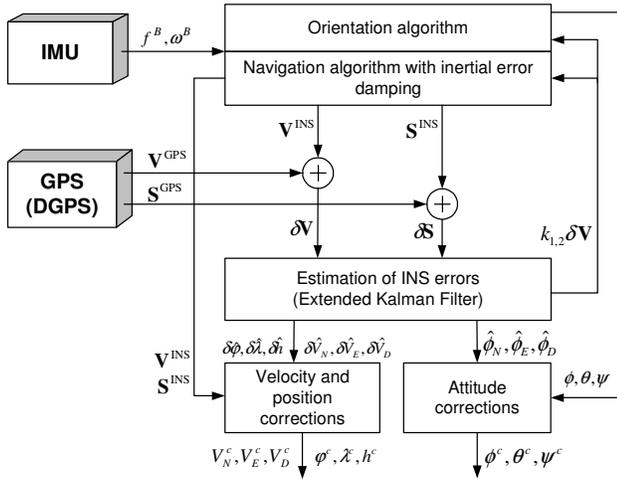


Fig. 3. Integration algorithm.

The suggested INS/GPS integration algorithm was based on the assumption that inertial sensors are of low accuracy, not allowing the implementation of an autonomous INS and application of traditional centralized or cascaded integration schemes. As a result of this, the algorithm based on INS error damping is performed. This error damping is based on velocity information supplied by GPS, in order to decrease amplitudes of their oscillations, using the frequency of updating which is higher than Schuler's. Besides this, the velocity and position corrections are produced based on the velocity and position estimates obtained by EKF. The integration scheme is shown on Fig. 3.

The choice of constant coefficients k_1 and k_2 in control signals is based on the compromise between two opposite requirements: the small navigation system static error on one side, and the required system bandwidth relative to gyro errors, on the other. Large values of control signals enable the fast damping of velocity errors but makes the system bandwidth higher introducing the higher influence of high-frequency components of gyro drift onto the overall system accuracy. This fact is more serious in cases of low-accuracy inertial sensors. As a result of this, the natural frequency ω_0 of undamped oscillations in INS should be chosen as low as acceptable. The lowest one is obtained from the condition specified in [1]:

$$\omega_0 > 1.46 \frac{\omega_{E,N}^{dr}}{|\Phi_{N,E}|} \quad (7)$$

where: $|\Phi_{N,E}|$ - orientation error of "computed" platform in linear zone ($|\Phi_N|, |\Phi_E| < 2^\circ - 3^\circ$) and $\omega_{E,N}^{dr}$ - projections of gyro drifts onto the east and north axes of NED frame.

Based on INS error analyses it was shown in [1] that the following values of coefficients k_1 and k_2 are suggestible:

$$k_1 = 2\xi\omega_0 \quad k_2 = \omega_0^2 / g, \quad (8)$$

where: ξ - damping coefficient and ω_0 - natural frequency from (7). According to this, the error model is now different from the standard one and can be represented as in (9), where: $\delta V_N, \delta V_E, \delta V_D$ - INS velocity errors, $\delta\phi, \delta\lambda, \delta h$ - INS position errors, f_N, f_E, f_D - specific forces projected on NED axes, B_N, B_E, B_D - accelerometer biases, $\omega_N^{dr}, \omega_E^{dr}, \omega_D^{dr}$ - slow varying components of gyro drifts approximated by first order Markov processes, β and A - shaping filter coefficients.

$$\begin{aligned} \delta \dot{V}_N &= -V_E \cos \varphi [2\omega_e + \dot{\lambda} \sec^2 \varphi] \cdot \delta\phi \\ &+ \left[\frac{V_E \dot{\lambda} \sin \varphi}{R_p + h} - \frac{V_D \dot{\phi}}{R_M + h} \right] \cdot \delta h + \frac{V_D}{R_M + h} \delta V_N \\ &- 2(\omega_e + \dot{\lambda}) \sin \varphi \cdot \delta V_E + \dot{\phi} \delta V_D \\ &- f_D \phi_E + f_E \phi_D + B_N + u_N^v \\ \delta \dot{V}_E &= [2\omega_e (V_N \cos \varphi - V_D \sin \varphi) + \dot{\lambda} V_N \sec \varphi] \cdot \delta\phi \\ &- \frac{\dot{\lambda}}{R_p + h} [V_D \cos \varphi + V_N \sin \varphi] \cdot \delta h \\ &+ (2\omega_e + \dot{\lambda}) \sin \varphi \cdot \delta V_N + \frac{1}{R_p + h} [V_D + V_N \tan \varphi] \cdot \delta V_E \\ &+ (2\omega_e + \dot{\lambda}) \cos \varphi \cdot \delta V_N + f_D \phi_N - f_N \phi_D + B_E + u_E^v \\ \delta \dot{V}_D &= 2\omega_e V_E \sin \varphi \cdot \delta\phi \\ &+ \left[\frac{V_N}{R_M + h} \dot{\phi} + \frac{V_E}{R_p + h} \dot{\lambda} \cos \varphi + (k-2) \frac{g}{R_e} \right] \cdot \delta h \\ &- 2\dot{\phi} \cdot \delta V_N - 2(\omega_e + \dot{\lambda}) \cos \varphi \cdot \delta V_E \\ &- f_E \phi_N + f_N \phi_E + B_D \\ \dot{\phi}_N &= -\omega_e \sin \varphi \cdot \delta\phi - \frac{\dot{\lambda}}{R_p + h} \cos \varphi \delta h + \frac{1}{R_p + h} \delta V_E \\ &- (\omega_e + \dot{\lambda}) \sin \varphi \cdot \phi_E + \dot{\phi} \phi_D - \omega_N^{dr} + u_N^\phi \\ \dot{\phi}_E &= \frac{\dot{\phi}}{R_M + h} \delta h - \frac{1}{R_M + h} \delta V_N + (\omega_e + \dot{\lambda}) \sin \varphi \cdot \phi_N \\ &+ (\omega_e + \dot{\lambda}) \cos \varphi \cdot \phi_D - \omega_E^{dr} + u_E^\phi \\ \dot{\phi}_D &= -(\omega_e \cos \varphi + \dot{\lambda} \sec \varphi) \cdot \delta\phi + \frac{\dot{\lambda}}{R_p + h} \sin \varphi \cdot \delta h \\ &- \frac{\tan \varphi}{R_p + h} \delta V_E - \dot{\phi} \phi_N - (\omega_e + \dot{\lambda}) \cos \varphi \cdot \phi_E - \omega_D^{dr} \\ \dot{\omega}_N^{dr} &= -\beta \omega_N^{dr} + A \sqrt{2\beta} \cdot w(t) \\ \dot{\omega}_E^{dr} &= -\beta \omega_E^{dr} + A \sqrt{2\beta} \cdot w(t) \\ \dot{\omega}_D^{dr} &= -\beta \omega_D^{dr} + A \sqrt{2\beta} \cdot w(t) \\ \delta \dot{\phi} &= -\frac{\dot{\phi}}{R_M + h} \delta h + \frac{1}{R_M + h} \delta V_N \\ \delta \dot{\lambda} &= \dot{\lambda} \tan \varphi \cdot \delta\phi - \frac{\dot{\lambda}}{R_p + h} \delta h + \frac{1}{(R_p + h) \cos \varphi} \delta V_E \\ \delta \dot{h} &= -\delta V_D \\ \dot{B}_N &= 0 \\ \dot{B}_E &= 0 \\ \dot{B}_D &= 0 \end{aligned} \quad (9)$$

Control signals $u_N^v, u_E^v, u_N^g, u_E^g$ are represented as:

$$\begin{aligned} u_N^v &= -k_1 \delta \hat{V}_N & u_E^v &= -k_1 \delta \hat{V}_E \\ u_N^g &= -k_2 \delta \hat{V}_E & u_E^g &= k_2 \delta \hat{V}_N \end{aligned} \quad (10)$$

where: $\delta \hat{V}_N, \delta \hat{V}_E$ - the estimates of velocity errors in north and east directions.

Using the error model (9), the EKF state vector is of order fifteen. The differences in velocities and positions obtained via INS and GPS separately are used as the measurements supplied to the EKF algorithm. The estimates of velocity and position errors are obtained from:

$$\begin{aligned} \hat{\mathbf{x}}_k^+ &= \Phi_k \hat{\mathbf{x}}_{k-1}^+ + \mathbf{L} \mathbf{u}_k + \\ &+ \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \Phi_k \hat{\mathbf{x}}_{k-1}^+ - \mathbf{H}_k \mathbf{L} \mathbf{u}_k) \end{aligned} \quad (11)$$

where: \mathbf{L} - gain matrix multiplying control signals, \mathbf{u} - vector of control signals.

In order to overcome the possible problems during the measurement process, the monitoring of following parameter was introduced:

$$Shock = (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^-)^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k) (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^-) \quad (12)$$

where: \mathbf{x}_k^- - extrapolated value of the state vector. If the value of *Shock* parameter is greater than the specified threshold the extrapolated value of system state vector is used as the filter output instead of using the actual measurements for the estimation. The estimates of velocity errors, position errors, and of the angular orientation of the "computed" platform are introduced into the navigation algorithm step by step. Velocity corrections are made in two steps. The first one consists in "correction of a computed platform":

$$\begin{aligned} V_N^c &= V_N^{INS} - g \hat{\Phi}_E \\ V_E^c &= V_E^{INS} + g \hat{\Phi}_N \end{aligned} \quad (13)$$

where: $\hat{\Phi}_N, \hat{\Phi}_E$ - the estimates of angular errors of "computed" platform. Velocity components are corrected on the second step using the estimates of velocity errors:

$$\begin{aligned} V_N^{cc} &= V_N^c - \delta \hat{V}_N \\ V_E^{cc} &= V_E^c - \delta \hat{V}_E \\ V_D^{cc} &= V_D^{INS} - \delta \hat{V}_D \end{aligned} \quad (14)$$

where: $\delta \hat{V}_N, \delta \hat{V}_E, \delta \hat{V}_D$ - velocity error estimates as EKF outputs.

Corrections of position components are done as:

$$\begin{aligned} \varphi^c &= \varphi^{INS} - \delta \hat{\varphi} \\ \lambda^c &= \lambda^{INS} - \delta \hat{\lambda} \\ h^c &= h^{INS} - \delta \hat{h} \end{aligned} \quad (15)$$

where: $\delta \hat{\varphi}, \delta \hat{\lambda}, \delta \hat{h}$ - estimates of geographical latitude, longitude, and height, obtained as the outputs of EKF.

During the intervals of absence of reliable GPS data the system works in "PREDICTION" mode when the last state vector estimates before the loss of GPS data are used. Kalman

filter gain matrix is equal to zero matrix then ($\mathbf{K} = 0$). Error damping control signals are produced using the measurements of odometer as an alternate sensor.

V. PROTOTYPE OF AN INTEGRATED INS/GPS SYSTEM

The prototype of an integrated INS/GPS system is based on usage of three mechanical rate gyros (Sfim I1426, of a range up to 20°/s) and three linear accelerometers (Sfim JT21, of a range up to 20m/s²). Sampling frequency of inertial sensor data is 100 Hz. Two additional angular sensors (Sfim JC30, ±30°) are used for initial alignment purposes. GPS receiver is of "u-blox GPS-PS1E" type (S/A code, working frequency L1, updating frequency 1Hz, declared accuracy - 5m). Components of prototype are shown on Fig. 4.

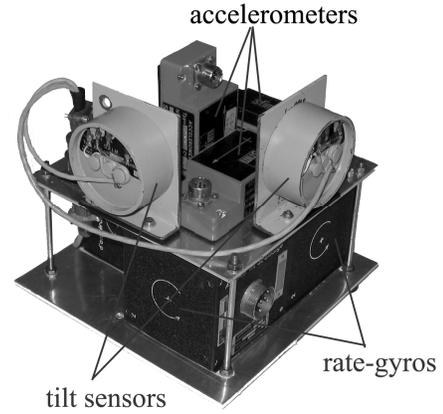


Fig. 4. Prototype of inertial measurement unit (IMU).

The specialized software support has been made [12] for the purposes of INS/GPS system development. It enables the work of system both in "INITIAL ALIGNMENT" and "NAVIGATION" phases. The basic display is shown on Fig. 5.

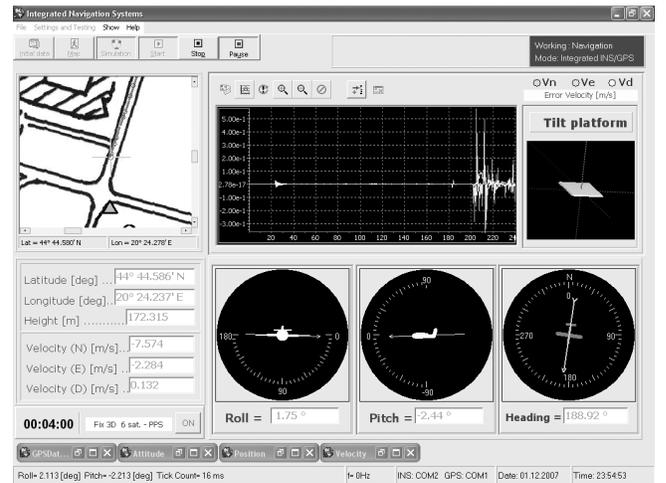


Fig. 5. The basic screen.

VI. EXPERIMENTAL VERIFICATION OF INS/GPS SYSTEM

Two types of experiments have been arranged in order to verify the accuracy of integrated INS/GPS system. The first

one regards to laboratory tests of inertial sensors and the whole INS part of system.

The second phase consisted in tests where the prototype was mounted on the car moving along accurately pre-specified trajectory.

The tests of inertial sensors and whole INS part used in order to estimate gyro and accelerometer errors as well as parameters of sensors stochastic models. These data are used for inertial sensor calibration and the initialization of covariance matrices of EKF. These tests were consisting from three procedures. The biases and scale factor errors of inertial sensors are determined during the first one, while the platform orientation errors are determined during the second one. The third procedure was related to the estimation of sensors stochastic models using Allan dispersion and autocorrelation function. It was possible this way to determine the dispersions of separate noise components as well as to distinguish the dominant ones.

Test platform "CARCO" T-922" was used for these purposes, and Table 1. summarizes the results of error parameter estimations.

The accelerometer error parameters have been obtained by averaging on the 15 min interval. The rate gyro parameters are determined during oscillatory motion of the test platform with $\pm 1^\circ/\text{s}$ amplitude.

The analysis of stochastic models of inertial sensor errors have been done in stationary regime. Fig. 6. illustrates the square root of Allan variance for accelerometers, and dispersions of different noise types for all used inertial sensors are summarized in Table 2.

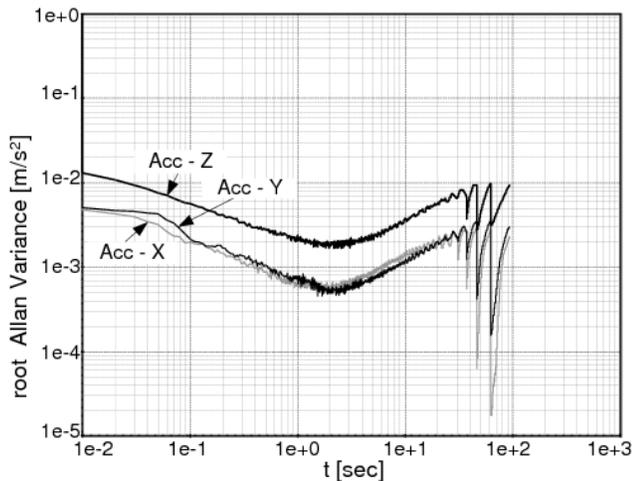


Fig. 6. Square root of Allan variance for accelerometers.

During the second phase of testing, the prototype of INS/GPS navigation system have been mounted on a moving car. The position providing continuous GPS availability was chosen for these field tests. There were a number of check-points (CP) along the trajectory where geographical coordinates had been previously determined via DGPS. The test trajectory is shown on Fig. 7. Initial alignment was done on CP-4 using the magnetometer as the external azimuth sensor.

TABLE I

INERTIAL SENSORS' BASIC ERROR MODEL PARAMETERS

accelerometer	bias [m/s^2]	scale factor error
x	-7.17×10^{-3}	-1.03×10^{-3}
y	7.506×10^{-3}	-4.49×10^{-3}
z	8.25×10^{-2}	-5.88×10^{-3}
rate gyro	drift [rad/s]	scale factor error
x	9×10^{-3}	-6.493×10^{-3}
y	5.075×10^{-3}	-4.99×10^{-3}
z	6.515×10^{-3}	1.72×10^{-2}

TABLE II

INERTIAL SENSORS' STOCHASTIC ERROR MODEL PARAMETERS

accelerometer	white noise [$\text{m}/\text{s}/\sqrt{\text{s}}$]	random walk [$\text{m}/\text{s}^2/\sqrt{\text{s}}$]
x	6.14×10^{-4}	6.48×10^{-4}
y	7.34×10^{-4}	5.73×10^{-4}
z	2.18×10^{-3}	1.94×10^{-3}
rate gyro	white noise [$^\circ/\sqrt{\text{s}}$]	random walk [$^\circ/\text{s}/\sqrt{\text{s}}$]
x	9.79×10^{-4}	8.4×10^{-3}
y	1.3×10^{-2}	8.71×10^{-3}
z	7.9×10^{-4}	-

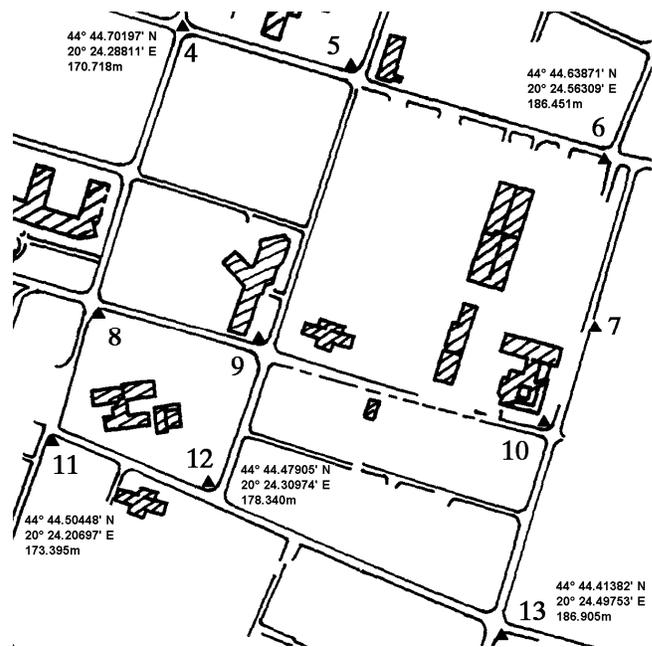


Fig. 7. Test trajectory with check-points.

One among these tests was chosen for further illustrations here. Acquisition of the inertial sensor data was made with $f_{si} = 100\text{Hz}$ while GPS updating frequency was $f_{ss} = 1\text{Hz}$. The number of available satellites have been more than four, except in interval 320 – 340 s when only two of them were available. The value of satellite geometric factor PDOP was less than 4. Angular receiver mask was set to 5° , while abrupt changes in GPS data have been prevented by the procedure for their filtering.

During the overall 9.5 min interval, the system worked 180s in the "INITIAL ALIGNMENT" regime while the remaining 6.5 min worked in "NAVIGATION" regime. Car dynamics are illustrated on Fig. 8. via diagrams of specific force profiles along x and y axes of body fixed coordinate frame.

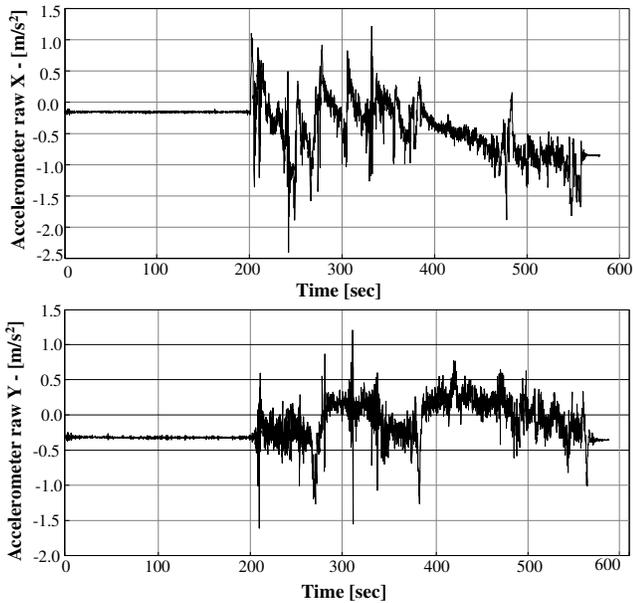


Fig. 8. Specific forces along x and y axes of body fixed coordinate frame.

Fig. 9. illustrates object's trajectory. Circular symbols are used to show GPS position data, while the linear curve represents the outputs of INS/GPS system. North and East velocity profiles are shown on Fig. 10.

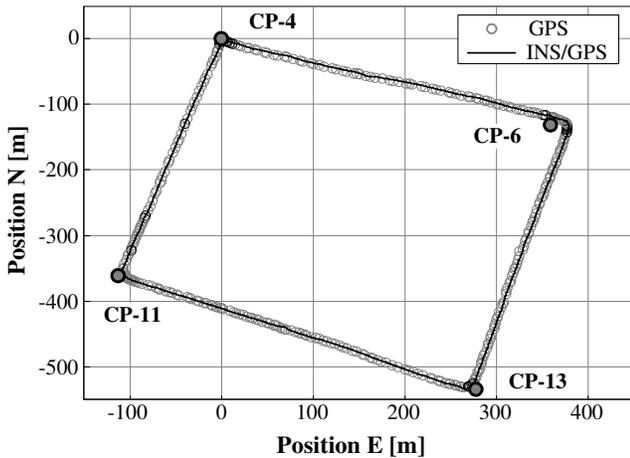


Fig. 9. GPS and INS/GPS test trajectory.

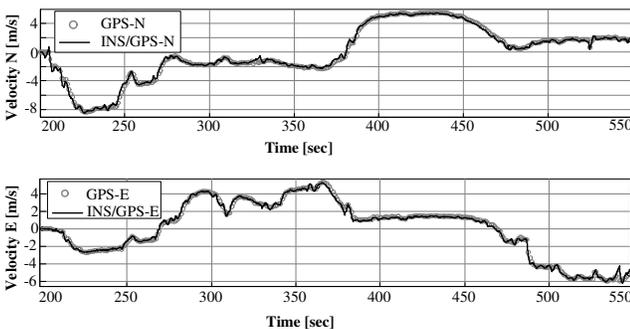


Fig. 10. GPS and INS/GPS velocity profiles.

Having in minds that the main advantage of the navigation integration concept consists in the availability of reliable navigation data during the intervals of absence of GPS

information, the following tests have been arranged in order to verify system performances in such cases. The intervals of absence of GPS data have been introduced artificially in off-line analyses. The GPS data existing in reality during these intervals were used as a reference. Two particular intervals have been chosen in order to analyze the system performances depending on the fact whether the car is moving with constant velocity or is accelerating. During the first interval ([285 – 305] s) the car was moving with variable velocity, while during the second one ([418 – 438] s) it was moving with approximately constant one.

The integrated INS/GPS system worked in “PREDICTION” mode during these two intervals. North velocity profiles have been shown now on Fig. 11. It is obvious that the accuracy of integrated system is affected by the car acceleration in the first interval (maximal position error of 36 m) in comparison to the second interval when it was moving with approximately constant velocity (maximal position error of 15 m). The main reason for these differences is in the fact that non-stationary error components (non-alignment in azimuth, accelerometer scale factor error, non-orthogonality of sensors) are strongly dependent on object's dynamics.

Higher values of velocity error damping coefficients can be used now in order to improve the overall accuracy during the intervals of GPS data absence. It reduced maximal position error to 25 m in the first interval and to 10 m during the second one. The trajectories shown on Fig. 12. illustrate this case. The choice of higher values for coefficients k_1 and k_2 in (8) reduces the velocity and position errors for such cases but makes the system bandwidth larger with some disadvantages regarding the overall influence of gyro drift high frequency contents in general.

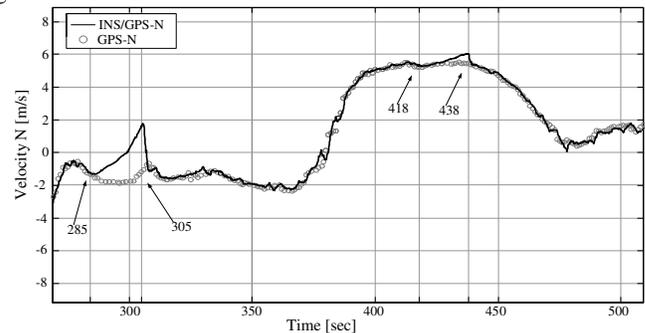


Fig. 11. GPS and INS/GPS velocity profiles for the cases of absence of GPS data.

In order to keep the acceptable accuracy for both cases (with or without GPS data) the values of damping coefficients should change according to this information. Table 3. illustrates position and velocity RMS errors in the intervals of absence of GPS data for two different values of damping coefficient k_2 .

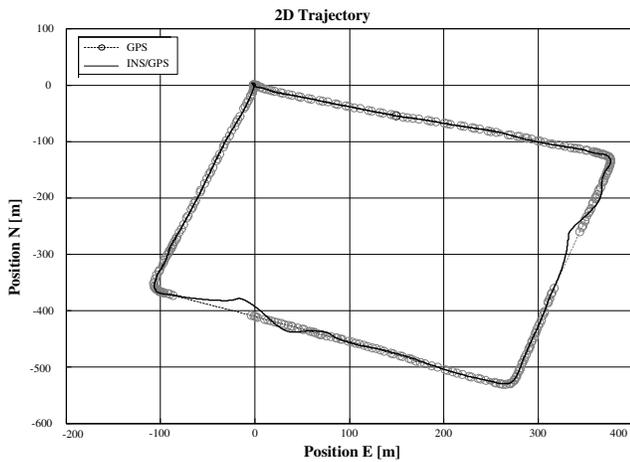


Fig. 12. GPS and INS/GPS trajectories for the cases of absence of GPS data and for strong damping of INS errors

TABLE III
POSITION AND VELOCITY RMS ERRORS

position (RMS) [m]		velocity (RMS) [m/s]		κ_1	κ_2	interval [s]
N	E	V_N	V_E			
6.917	10.279	1.054	0.593	20.5	12.35	285 – 305
15.411	12.668	1.309	0.653	20.5	10.35	
4.662	1.368	0.280	0.366	20.5	12.35	418 – 438
5.7494	5.07	0.428	0.494	20.5	10.35	

VII. CONCLUSION

The main purpose of this work was to demonstrate possibility of implementation of an integrated low-cost strap-down INS and the GPS for the land vehicle application. An extended Kalman filter scheme for the optimal estimation of navigation state errors as well as for inertial instrument errors was used, based on the measurements consisting in differences in velocity and position vectors obtained from INS and GPS separately. Additional error damping was introduced in order to overcome the fact that the inertial instruments were of low accuracy. After the laboratory work consisting in IMU error model parameterization the prototype of integrated navigation system was used in ground experiments. Duration of the initialization phase was of order of three minutes. The navigation phase have shown that the integrated system worked very well (i.e. that 1 Hz GPS measurements have made the corrections of the INS output very precisely). However, the main advantage of integration scheme should be verified when the GPS data are absent in some longer intervals. The main results of these tests for 20s long intervals of GPS data absence have indicated that the values of damping coefficients should be adapted according to the fact whether the vehicle is maneuvering or not at the moment of losing the valid GPS

data. Generally, these coefficients should keep the values suggested in literature as the optimal ones whenever the GPS data are present. If the vehicle is maneuvering during the intervals of absence of GPS data their values should be increased. In spite of the fact that the overall bandwidth is increased this way and that high frequency components of IMU errors would have the higher influence, this is beneficial because the errors dependent on vehicle dynamics are dominant then and their influence should be attenuated basically. The work in progress consists in miniaturization of INS system using MEMS inertial sensors and in further modification of the integration algorithm based on more precise adaptation of error damping coefficients according to the measured value of vehicle acceleration during the intervals of GPS data absence.

REFERENCES

- [1] O. S. Salychev, "Applied Inertial Navigation Problems and Solutions", The Bauman Moscow State Technical University, 2004.
- [2] O. N. Anutchin, G. I. Emeliantsev, "Integration of orientation and navigation systems for marine moving vehicles", (in Russian), GNC RF – CNII "Elektropribor", 2003.
- [3] D. H. Titterton, J. L. Weston, "Strapdown inertial navigation technology", AIAA, Vol. 207, 2004.
- [4] O. A. Stepanov, "Application of nonlinear filtration for the navigation information processing", (in Russian), GNC RF – CNII "Elektropribor", 2003.
- [5] J. H. Wang, Y. Gao, "GPS-based Land Vehicle Navigation System Assisted by a Low-Cost Gyro-Free INS Using Neural Network", *Journal of Navigation*, (57), 417-428, 2004.
- [6] M. E. Cannon, R. Nayak, G. Lachapelle, O. S. Salychev, V. V. Voronov, "Low-Cost INS/GPS Integration: Concepts and Testing", *Journal of Navigation*, (54), 119-134, 2001.
- [7] E. Nebot, H. Durrant-White, "Initial Calibration and Alignment of Low-Cost Inertial Navigation Units for Land Vehicle Applications", *Journal of Robotic Systems*, 16(2), 81-92, 1999.
- [8] C. Hide, T. Moore, M. Smith, "Adaptive Kalman Filtering for Low-Cost INS/GPS", *Journal of Navigation*, (56), 143-152, 2003.
- [9] Y. Yang, W. Gao, "Comparison of Adaptive Factors in Kalman Filters on Navigation Results", *J. of Navigation*, (58), 471-478, 2005.
- [10] D. Loebis, R. Sutton, W. Chudley, W. Naeem, "Adaptive Tuning of a Kalman Filter via Fuzzy Logic for an Intelligent AUV Navigation System", *Control Engineering Practice*, Vol. 12, 1531-1539, 2004.
- [11] B. H. Kaygisiz, I. Erkmen, A. M. Erkmen, "GPS/INS Enhancement for Land Navigation Using Neural Network", *Journal of Navigation*, (57), 297-310, 2005.
- [12] R. Stančić, S. Graovac, "Software Support for a Development of an Integrated INS/GPS System", 12th ICINS Conf., St. Petersburg, 2005.
- [13] J. F. Wagner, T. Wieneke, "Integrating satellite and inertial navigation – conventional and new fusion approaches", *Control Engineering Practice*, Vol. 11, 543-550, (2003)
- [14] R. Stančić, S. Graovac, "The integration of strap-down INS and GPS based on adaptive error damping", *Robotics and Autonomous Systems* 58(10): 1117-1129 (2010)
- [15] R. Stančić, "Integrated INS/GPS navigation algorithm", (in Serbian), PhD thesis, University of Belgrade, School of Electrical Engineering, Belgrade, 2010.

NHibernateMapper - A Tool for Rapid Development of Data Access Layer for Interoperable GIS Solutions

Stevica S. Cvetković, Miloš D. Bogdanović and Leonid V. Stoimenov

Abstract—In this paper we presented architecture and implementation details of an object-to-relational-database mapping specification tool (NHibernateMapper) based on NHibernate ORM framework. The tool is able to automatically traverse provided database schema and domain class library to be matched and to suggest correspondences. Interactive GUI provides ability to accept or reject suggested correspondences as well as to manually define mappings in order to prune it into a final form. By analyzing database schema of spatially enabled DBMSs, this tool supports the development of Data Access Layer for GIS solutions. If provided with a domain class library generated on the basis of application ontology used by GIS solution, NHibernateMapper becomes a tool that enables rapid development of Data Access Layer for Interoperable GIS solutions.

Index Terms—Schema mapping, Object relational mapping.

I. INTRODUCTION

MOST of the modern information systems use databases for storage of persistent information, mostly because of their reliability and standardized query language. On the other hand, object-oriented programming languages such as Java, C# and C++ are most commonly applied for the system programming. Conventional software development approaches for database access within object-oriented languages raises a problem known as “object-relational impedance mismatch”. It occurs because the object-oriented and relational database paradigms are based on different principles [3] and manifests itself in several specific differences: inheritance implementation, association’s implementation, data types, etc.

Object Relational Mapping (ORM) tools have been developed in an attempt to overcome mentioned problem. As described in the Agile methodology for software and databases development [3], ORM tools automatically create Data Access Layer based on database model. The most successful representatives of ORM tools are Hibernate [11], Oracle TopLink [12] and recently introduced Microsoft Entity Framework (EF) [1,2]. With these tools, application developer is encouraged to think in terms of data layer objects and their relationships. The ORM takes all the control of handling

objects and relationships at runtime. It automatically tracks updates made to the objects and performs the necessary SQL insert, update, and delete statements at commit time.

Hibernate is widely used, comprehensive and stable open source ORM framework. Furthermore, it is available for both Java and .NET environments (NHibernate). Although there are Hibernate tools for automatic generation of object model and mapping files, based on existing database, there are only few commercial tools which are capable to map already created object model to existing database. This is of crucial importance when trying to map standardized domain knowledge (application programming classes) to legacy database systems. For this purpose NHibernateMapper tool is developed, which is able to generate NHibernate mapping file (.hbm file) based on the already available domain knowledge and existing database. These three components (application programming classes, database and mapping file) are sufficient premises for generation of complete Data Access Layer using NHibernate ORM mechanism [8].

The aim of this paper is to present a tool that can be used for rapid development of data access layer for interoperable GIS solutions. Its primary purpose is to generate objects-to-database mapping specifications according to NHibernate ORM framework. It is used to assist in semi-automatically determining correspondences between GIS solution domain class model and the underlying database. In particular, correspondences are determined between domain class member attributes and database columns. The tool contains the visual front-end which suggests mapping using “Levenshtein distance” and “Data type correspondence” heuristics. In order to define final mapping definitions, these suggestions must be pruned by user interaction. Defined mapping definitions are then used to generate XML-based mapping file.

The paper is organized as follows: Section 2 describes the purpose of created tool and gives brief overview of previous research in schema mapping tools. Section 3 presents overall software architecture and user interface of the tool. Detailed description of implementation including technology specific details is given in Section 4. We concluded and discussed about future work in Section 5.

II. MOTIVATION AND RELATED WORK

Our motivation in this work was to apply NHibernateMapper for development of Interoperable Geographic Information System (GIS) [4]. Specifically, its purpose is for support of rapid generation of Data Access Layer using existing ORM solutions. In order to achieve interoperability of developed GIS, domain ontologies could be used as a baseline model in the GIS construction process. Ontologies give us an opportunity to define a meaning of the information and are mainly used during communication of different GIS in order to solve the meaning of the available data. However, ontologies could be also treated as a starting model for the GIS development [13]. By representing ontology in the form of object-oriented classes, ontology concepts can be observed as domain objects that should be mapped to existing relational database. In order to provide object-to-database mapping, we developed NHibernateMapper tool. It is intended for mapping of object-oriented GIS ontology classes into existing databases.

Mapping data between different representations is a common problem which has been addressed from either an algorithmic or a visualization point of view. Much work has been done in this field, differing mainly in the degree of automation achieved. In [5] is given an overview of a variety of approaches to schema mapping and automatic mappings generation. Attempt to automatically determine a mapping between two DTDs is described in [6]. In [7] and [10] are presented approaches for specifying data mapping between two XML schemas using a combination of automated schema analysis agents and selective user interaction. There are also commercial tools like Altova MapForce [8] and Microsoft BizTalk Mapper [9] which could be applied for different schema mapping scenarios including Xml-to-Xml, Database-to-Xml and Database-to-Database mapping. However, they proved to be too generic and complex for learning of specific domain application.

III. NHIBERNATEMAPPER ARCHITECTURE

Starting point in the process of *NHibernateMapper* mapping file generation is existing relational database as well as specific domain knowledge in the form of application programming classes. This can be observed in Fig. 1 which represents the NHibernateMapper overall architecture. In order to generate mapping file, the tool utilizes publicly available schema definition of NHibernate mapping file (*nhibernate-mapping.xsd*). This file is the third input element

When using NHibernateMapper tool, the user is required to select an existing database as well as domain object-oriented classes in form of Dynamic Linked Library (DLL). NHibernateMapper will automatically extract database schema of previously selected database and class member attributes of selected DLL.

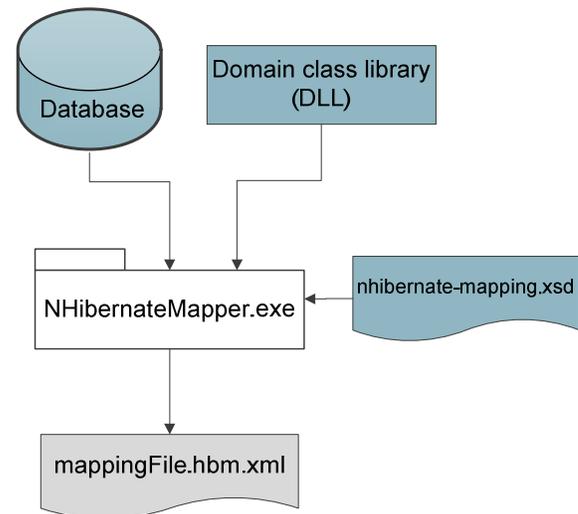


Fig. 1. NHibernateMapper Architecture.

User has to manually select the specific data table and an object-oriented class to be mapped. Then, the database schema and class attributes are automatically analyzed and suggestions of candidate mappings are generated and displayed to the user. Candidate mappings are generated for each column (column-to-attribute mapping), based on string matching algorithm as well as data type comparison. The user is able to accept or reject suggested correspondences. Also, the users can take advantage of the interactive GUI in which suggestions could be pruned into a final mapping. Beside accepting or rejecting proposed mapping, there is also possibility to redefine mappings manually. When the user is finally satisfied with mapping definition, it can be saved to a NHibernate XML mapping file. Illustration of NHibernateMapper GUI is given in Fig. 2.

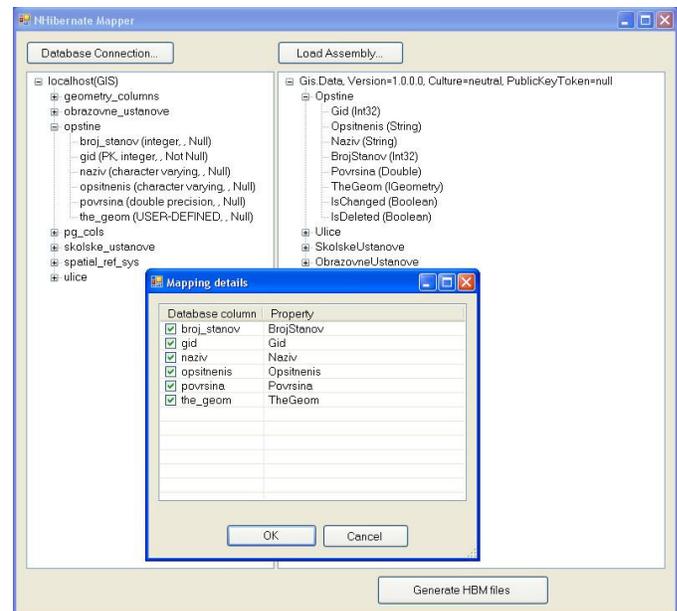


Fig. 2. NHibernateMapper GUI.

IV. IMPLEMENTATION DETAILS

NHibernateMapper has been implemented entirely in C# using Microsoft .NET Framework 4.0. As a result, *NHibernateMapper.dll* class library is created. All programming classes in the library could be divided into two major groups: classes for database schema retrieving (Fig. 3.) and classes for mapping file manipulation (Fig. 5).

Class *DbInfo* is a central class of the database schema retrieving part (Fig. 3). It represents object model of database schema. *DbInfo* class obtains and connects schema information including: tables (*DbTableInfo*), columns (*DbColumnInfo*) and constraints (*DbConstraintInfo*).

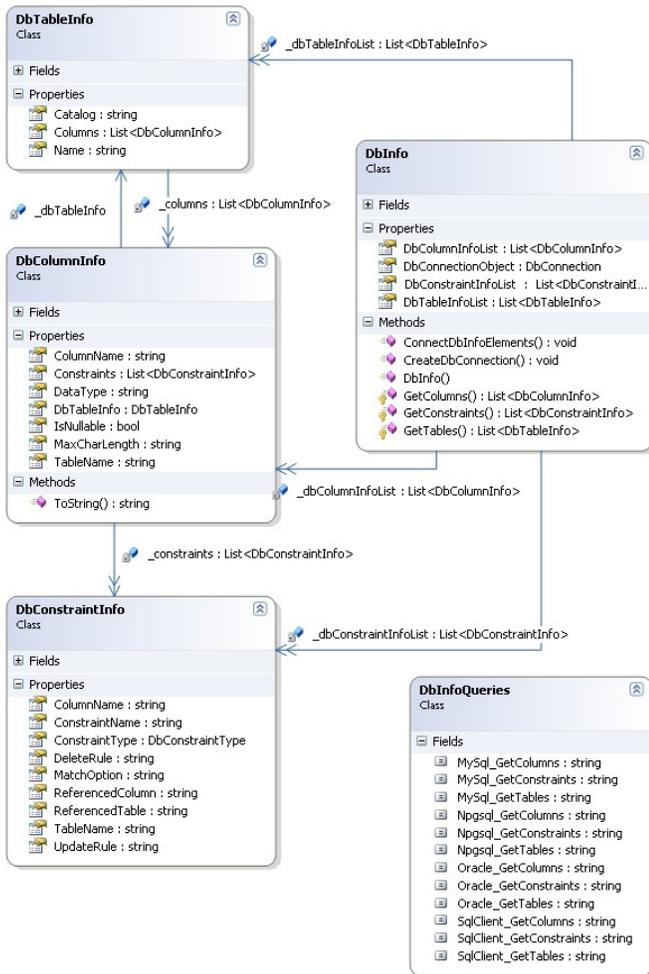


Fig. 3. Diagram of classes for database schema retrieval.

Current implementation supports following DBMS-s: MS SQL Server, MySQL and PostgreSQL. However, configurable design offers possibility to introduce support for any other DBMS which has data provider for .NET Framework. Introduction of support for new DBMS could be done in three successive steps:

- Define three additional queries that retrieve database schema information. These queries (GetTables, GetColumns and GetConstraints) are part of a class *DbInfoQueries*.
- Copy data provider (in form of DLL) offered by DBMS

producer. We have already used *MySQL.Data.dll* for MySQL and *Npgsql.dll* for PostgreSQL.

- Change configuration file so it includes added data provider declaration.

Example of existing configuration file is given in Fig. 4.

```

<system.data>
...
<DbProviderFactories>
  <add name="PostgreSQL Data Provider"
    invariant="Npgsql"
    description=".Net Framework Data Provider for PostgreSQL"
    type="Npgsql.NpgsqlFactory, Npgsql, Version=2.0.6.0, Culture=neutral, PublicKeyToken=5d8b90d52f46fda7">
  </add>
  <add name="MySQL Data Provider"
    invariant="MySql.Data.MySqlClient"
    description=".Net Framework Data Provider for MySQL"
    type="MySql.Data.MySqlClient.MySqlClientFactory, MySql.Data, Version=6.1.3.0, Culture=neutral, PublicKeyToken=c5687fc88969c44d">
  </add>
...
</DbProviderFactories>
...
</system.data>
    
```

Fig. 4. Example of database configuration file.

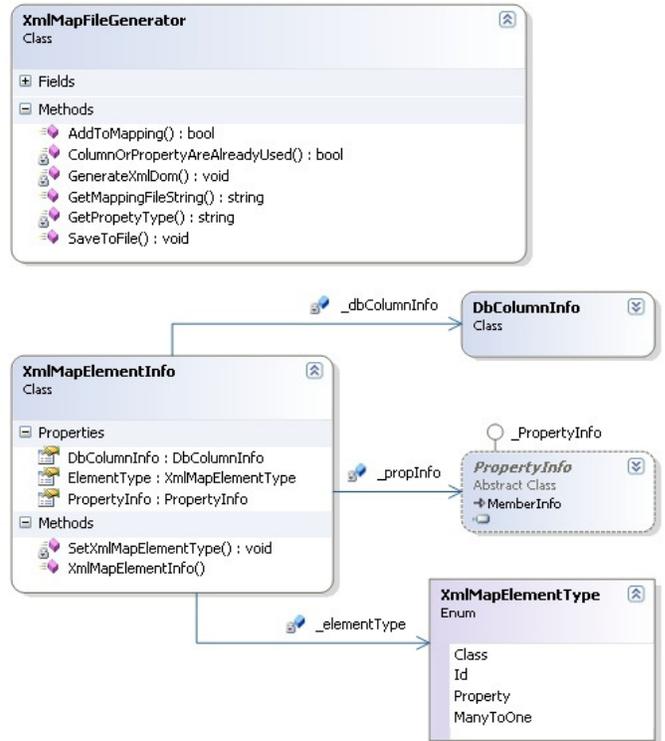


Fig. 5. Diagram of classes used for mapping files manipulation.

Classes for mapping file manipulation are responsible for populating of object model of XML mapping file as well as its storing into file system (Fig. 5). Class *XmlMapElementInfo*

manages mapping of one database column with a corresponding class attribute. Central class of this part is *XmlMapFileGenerator* which connects all mapped elements into a whole and enables saving into XML mapping file. It also implements string matching algorithms including “Levenshtein distance” and “Data type correspondence”. Architecture is flexible enough to allow incorporation of new heuristic algorithms.

On a programming level, mapping file is represented by its object model generated from schema definition file. From the developer’s point of view, using schema definition enables validation of mapping file as well as its storing into XML file. Complete process of mapping file generation using previously described programming classes could be summarized in code snippet in Fig. 6.

```
public string GenerateXmlMappingFile()
{
    // Database connection and db schema extraction
    DbInfo dbInfo = DbInfoFactory.GetDbInfo(
        DbProviderInvariantNames.Npgsql, connString);

    //Extract table structure information
    List<DbTableInfo> tableInfoList = dbInfo.DbTableInfoList;

    // Mapping file initialization
    XmlMapFileGenerator xmlFileGen = new
    XmlMapFileGenerator();

    // Iteration through tables and columns
    foreach (DbTableInfo tableInfo in tableInfoList){
        foreach (DbColumnInfo colInfo in tableInfo.Columns){
            //Generation of a mapping file element
            PropertyInfo propInfo =
            typeof(TestMappingClass).GetProperty("Gid");
            XmlMapElementInfo xmlMapEl = new XmlMapElementInfo(
                colInfo, propInfo);

            //Insert element into mapping file
            xmlFileGen.AddToMapping(xmlMapEl);
        }
    }

    // Return mapping file as string
    return xmlFileGen.GetMappingFileString();
}
```

Fig. 6. Code snippet for mapping file generation process.

As already mentioned, the purpose of developed tool is its application in the field of Interoperable GIS solutions. Therefore we have tested NHibernateMapper with an existing GIS database implemented in PostgreSQL. Example of generated mapping file is presented in Fig. 7.

```
<xml version="1.0"?>
<hibernate-mapping
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns="urn:hibernate-mapping-2.2">
```

```
...
<class name="Gis.Data.Ulice, Gis.Data" table="ulice">
  <id name="Gid" column="gid" type="System.Int32" />
  <property name="UliceName" type="System.String"
column="name" />
  <property name="TheGeom"
type="GeoAPI.Geometries.IGeometry" column="the_geom" />

  <property name="Tip" type="System.String" column="tip" />
</class>

  <class name="Gis.Data.ObrazovneUstanove, Gis.Data"
table="obrazovne_ustanove">
  <id name="IdObrazovne" column="id_obrazovne"
type="System.String" />
  <property name="Adresa" type="System.String"
column="adresa" />
  <property name="Povrsina" type="System.Double"
column="povrsina" />
  <property name="TheGeom"
type="GeoAPI.Geometries.IGeometry" column="the_geom" />
  </class>
...
</hibernate-mapping>
```

Fig. 7. Example of generated NHibernate mapping file of test GIS database.

V. CONCLUSION

NHibernateMapper belongs to the group of schema mapping tools. It is specifically designed for semi-automatic object-to-database mapping specification based on NHibernate ORM. Developed tool allows mapping of information retrieved from existing database with domain model classes. As a starting point for the domain classes definition, ontologies could be applied. Ontologies offer uniform mechanism to define a meaning of the information and represent convenient way for domain model definitions. By representing ontology in the form of object-oriented classes, ontology concepts can be observed as domain objects that will be mapped to existing relational database. NHibernateMapper is primarily developed for application in the field of Geographic Information Systems. This tool could be coupled with an already existing tool that transforms ontology concepts into object-oriented classes [4]. In this way, NHibernateMapper should support generation of Data Access Layer using existing ORM tools. By its semi-automatic mapping generation and coupled with a tool described in [4], NHibernateMapper can significantly contribute to rapid development of Data Access Layer for Interoperable GIS solutions.

Advantage of NHibernateMapper is its configurable design that allows support for intuitive introduction of new DBMS. Currently NHibernateMapper supports only Hibernate mapping file specification. Improvement plan assumes extension to other popular ORM mapping file definitions, like Microsoft Entity Framework. There is also plan for integration of the tool into complete solution for WebGIS application source code generator [14].

REFERENCES

- [1] A. Adya, J. Blakeley, S. Melnik and S. Muralidhar, "Anatomy of the ADO.NET Entity Framework". In Chan, C. Y. Ooi B. C. & Zhou, A. (eds), Proceedings of the ACM SIGMOD International Conference on Management of Data, June 11-14, Beijing, China, 2007, pp. 877-888.
- [2] P. Castro, S. Melnik and A. Adya, "ADO.NET entity framework: raising the level of abstraction in data programming." In Chan, C. Y. Ooi B. C. & Zhou, A. (eds), Proceedings of the ACM SIGMOD International Conference on Management of Data, Beijing, China, 2007, pp. 1070-1072.
- [3] S. Ambler, *Agile Database Techniques*, Wiley, 2003.
- [4] A. Stanimirović, M. Bogdanović and Stoimenov, "Data Access Layer Generation for Interoperable GIS Environments," 12th AGILE International Conference on Geographic Information Science, 02-05 June, Hannover, Germany, ISSN 2073-8013, 2009
- [5] E. Rahm, P. A. Bernstein, *A survey of approaches to automatic schema mapping*, The VLDB Journal, Springer Verlag, 10: 334-350, 2001
- [6] H. Su, H. Kuno and E. A. Rudensteinern, "Automating the Transformation of XML Documents," Proc Workshop on Web Information and Data Management, 2001.
- [7] L. L. Yan, R. J. Miller, L. M. Haas, and R. Fagin, "Data-driven understanding and refinement of schema mappings." In SIGMOD Conference, 2001.
- [8] Altova Mapforce, http://www.altova.com/products_mapforce.html
- [9] M. A. Goulde, *Microsoft's BizTalk Framework adds messaging to XML*, E-Business Strategies & Solutions, 1999, pp.10-14.
- [10] S. Bossung, H. Stoeckle, J. Grundy, R. Amor and J. Hosking, "Automated Data Mapping Specification via Schema Heuristics and User Interaction," 19th International Conference on Automatic Software Engineering (ASE'04), 2004, pp. 208—217.
- [11] C. Bauer, G. King, *Java Persistence with Hibernate*, Manning Publications, 2006.
- [12] OracleTopLink, <http://www.oracle.com/technology/products/ias/toplink>
- [13] F. T. Fonseca, M. J. Egenhofer, "Ontology-Driven Geographic Information System," 7th ACM Symposium on Advances in Geographic Information System, Kansas City, 1999.
- [14] M. Bogdanović, A. Stanimirović and L. Stoimenov, "Ginis MvcGen – a Prototype of WebGIS Application Source Code Generator," ICEST 2010, Ohrid, Makedonija, 2010, pp. 307 – 310.

Localization Techniques in Wireless Sensor Networks using Measurement of Received Signal Strength Indicator

M. Srbinovska, V. Dimcev, C. Gavrovski and Z. Kokolanski

Abstract—The presented paper describes different localization techniques in wireless sensor networks (WSN) and corresponding localization algorithms. Correct location is required for many WSN applications, but it is often too expensive to include GPS adapters in each sensor node. Instead of requiring every node to have GPS installed, all localization methods assume only a few nodes equipped with GPS hardware which know their positions without communicating with other nodes. The sensors with unknown location information and their coordinates will be estimated by the sensor network localization algorithms. The main goal of localization procedures is to deduce, as accurately as possible, the location of a node from the partial information obtained from a set of nodes, which already know their location.

Measurement of the received signal strength indicator (RSSI) parameter is used for distance determination between sensor nodes. The developed algorithm is implemented in a real sensor network.

Index Terms—Wireless sensor networks, Localization, Sensor nodes

I. INTRODUCTION

Wireless networking has become a very popular research topic, over the last decade. Recent advances in wireless communications and electronics have enabled the development of low-cost, low-power and multi-functional sensors that are small in size and communicate in short distances. Cheap, smart sensors, networked through wireless links and deployed in large numbers, provide enormous opportunities for environmental monitoring, medical treatment, agriculture, military affairs but also in households and traffic pollution areas [1]. Beside these, WSN are also used in some dangerous working environment like mines and nuclear power stations in order to monitor their safety status.

The progress of both sensing and information technology is making the physical world measurement data more available

and cheap in terms of acquisition, processing, storage and retrieval. Intelligent sensors, which are possibly arranged in wired or wireless networks, can acquire raw data from the environment for long periods of time and those data can be safely stored [2,3].

In sensor networks, nodes are deployed into an unplanned infrastructure where there is no a priori knowledge of location. The problem of estimating spatial – coordinates of the node is referred as localization. The main practical objective is to locate each node as accurately as possible with a certain amount of error about the distances between a subset of nodes in wireless sensor networks. To identify the coordinates of sensor nodes (also called unknown nodes) requires measuring a distance e.g., measuring time of arrival (ToA) or time difference of arrival (TdoA) [4]. Difficulties concerning time measurement result from synchronization of involved devices as well as the high mathematical effort to calculate the position. The measurement of the received signal strength (RSS) offers a possibility to determine distance with minimal effort.

Perhaps the simplest method of providing localization is to equip every sensor node with a GPS receiver. However, a GPS receiver is expensive in terms of money, size and energy. Another alternative is hand-placing each sensor and manually recording its position. This is unpractical and error prone approach unsuitable for large sensor networks and many of the proposed WSN applications [5].

Most of the existing works focus on increasing the accuracy in position estimation by using different mathematical techniques such as triangulation, multilateration, multidimensional scaling, etc.

The rest of the paper is organized as follows. In section II measurement techniques in WSN localization are discussed; these include distance related measurements. In section III radio signal measurements are explained, received signal strength (RSS) – based distance measurements, time of arrival (ToA) and time difference of arrival (TDoA).

Section IV discusses different radio signal propagation models for wireless communications that predict signal strength loss with distance – path loss; include free space propagation model, two-ray ground model and log – distance model. Experimental and measurement results are given in Section V. Finally a summary is provided in Section VI.

M.Srbinovska is with the Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia (e-mail: mares@feit.ukim.edu.mk)

V.Dimcev is with the Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia (e-mail: vladim@feit.ukim.edu.mk)

C.Gavrovski is with the Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia (e-mail: cvetang@feit.ukim.edu.mk)

Z. Kokolanski is with the Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia (e-mail: zivko.kokolanski@feit.ukim.edu.mk)

II. LOCALIZATION TECHNIQUES

The goal of the location estimation methods is to calculate the position (coordinates) of sensor nodes in wireless sensor networks. Localization in WSN is a complex problem that can be solved in different ways [6].

Some localization techniques allow to estimate nodes position using information transmitted by a set of nodes that are aware of their own locations (GPS for example). It can be distinguished two classes of methods depending from the nodes hardware capabilities:

- Connectivity based methods that use only connectivity information to locate the entire sensor network;
- Distance based methods that use inter-sensor distance or angle measurements to determine location of the nodes.

Distance based methods require additional equipment but with this method it can be reached much better resolution than in case of connectivity based ones. The proximity information between nodes often helps to determine their geometric relationship. The distance between them or angle of a singular triangle can be easily estimated.

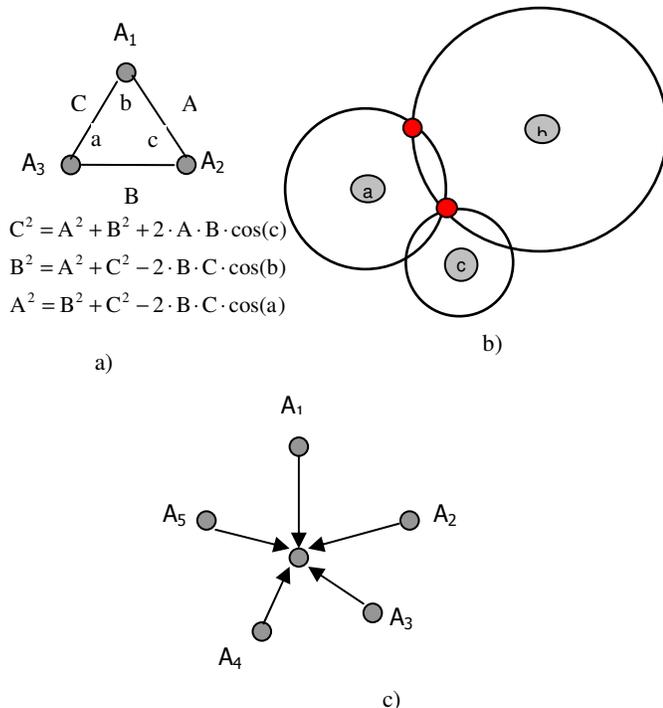
When the distances between an object and a number of nodes are measured, it is also required to find the actual position of nodes by combining the distance measurements. The most used methods for the distance measurements from three or more nodes are: triangulation, simple trilateration and multilateration.

Triangulation is a method for finding the position of a node, when the angles are measured by Angle of Arrival technique. An example of such a procedure is shown in Figure 1.a. The object X measures its angles with respect to the beacons A1, A2 and A3. The measured angles form three straight lines along the directions XA1, XA2 and XA3. The intersection between the three lines defines the location of the node X. The accuracy of this technique is heavily dependent upon the accuracy of the employed angle measurement technique

Simple trilateration is used when we have an accurate estimate of distances between a node and at least three beacon nodes. This simple method finds the intersection of three circles centered at beacons as the position of the node. The scenario is shown in Figure 1.b.

Multilateration is accepted as the most appropriate way to determine the location of a sensor node based on locations of beacons. An example is shown in Figure 1.c, where the nodes A1, A2, A3, and A4 are beacons, with known estimates of their locations, while the node X estimates its location using a multilateration procedure. The procedure attempts to estimate the position of a node by minimizing the error and discrepancies between the measured values.

The characteristics of wireless communication are partially determined by the distance between sender and receiver, and if these characteristics can be measured at the receiver they can serve as an estimator of distance. The most important characteristics are Time of Arrival (ToA), Time Difference of Arrival (TdoA), and Received Signal Strength Indicator (RSSI).



III. RADIO SIGNAL MEASUREMENTS

Time of Arrival (ToA) uses the relationship between distance and transmission time when the propagation speed is known. If the sender and the receiver know the time when the transmission starts the time of signal propagation can be used to estimate the distance. To provide acceptable accuracy a very high accurately synchronized clocks are required for sender and receiver. To overcome the need for explicit synchronization, the roundtrip time is often utilized. Roundtrip propagation time measurements measure the difference between the time when a signal is sent by a sensor and the time when the signal returned by a second sensor is received at the original sensor. But, this method is affected by the internal delay required for handling the signal in a second sensor [7].

Another technique that is used to overcome the need of explicit synchronization is time difference of arrival (TDoA) method, which utilizes implicit synchronization by directly providing the start of transmission information to the receiver. It is possible if two transmission mediums of different propagation speeds are used. For example the sender can send and ultrasound and radio signal simultaneously. When the receiver gets the radio signal it can start measuring the time until arrival of the ultrasound transmission, safely ignoring the propagation time of the radio communication [7]. The main disadvantage of this technique is the need of two types of senders and receivers on each node.

Fig. 1. Localization Basics a) Triangulation b) Trilateration c) Multilateration.

Technique which estimates the distances between neighboring nodes is the method based on a standard feature found in most wireless devices, a received signal strength indicator (RSSI). Based on the known transmit power, the effective propagation loss can be calculated. Theoretical and empirical models are used to translate this loss into a distance estimate. This method has been used mainly for RF signals.

In practical scenarios, the ideal distribution of RSSI is not applicable, because the propagation of the radio signal is interfered with a lot of influencing effects e.g:

- Reflection of objects;
- Diffraction at edges;
- Refraction by media with different propagation velocity.

The low complexity and the fast calculation recommend this localization algorithm as very popular and often used in wireless sensor networks.

IV. THE RADIO SIGNAL PROPAGATION MODELING

There are many radio propagation models [8] known for wireless communications that predict signal-strength loss with distance – path loss. There are three models widely used for wireless sensor networks.

Free space propagation model is built on the assumption that the transmitter and receiver are in the line of sight, and there are no obstacles between them.

The receiver signal power, P_R is related to distance, d , by the inverse square law:

$$P_R = C_f \frac{P_t}{d^2} \tag{1}$$

Two-Ray ground model adds reflection to the previous model. A two-ray ground model receiver receives two rays, direct communication ray and reflected ray:

$$P_R = C_f \frac{P_t}{d^4} \tag{2}$$

In the real world the signal often decays at a faster or slower rate, known as log-distance model

$$P_R \propto \frac{P_t}{d^n} \tag{3}$$

Where: n is the loss exponent, C_f is a constant that depends from the transceiver characteristics, P_t is the power of transmitter signal. An excepted form of the relation between distance and receive power simplified for the case of a one meter reference distance is:

$$RSSI[dBm] = 10 \cdot n \cdot \log_{10}(d) + A \tag{4}$$

where: d denotes the transmitter-to-receiver distance, A represents the RSS value measured by a receiver that is located 1m away from a transmitter, and RSS is the actual measured value at distance d . The signal propagation

coefficient n shows the damping of the signal. Both parameters must be determined empirically [9]. By measuring the RSSI at different distances the values of n and A can be determined (4). A linear least square approximation to solve n and A should be implemented.

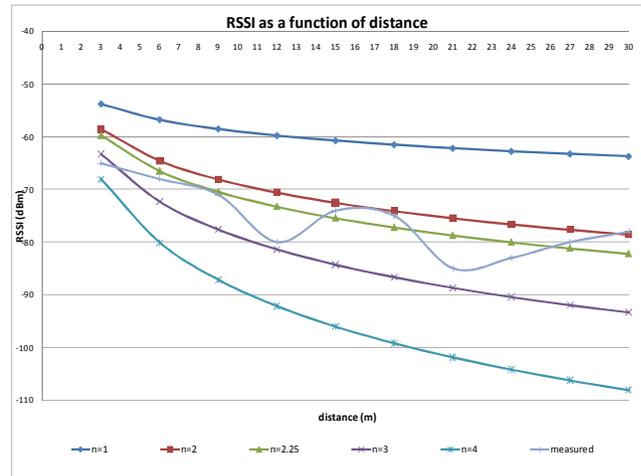


Fig.2 Measured vs calculated RSSI with different n.

To determine n , RSSI values within 30m are measured and compared with theoretical curves according to equation 4 whereby n is varied from 1 to 4 (figure 2).

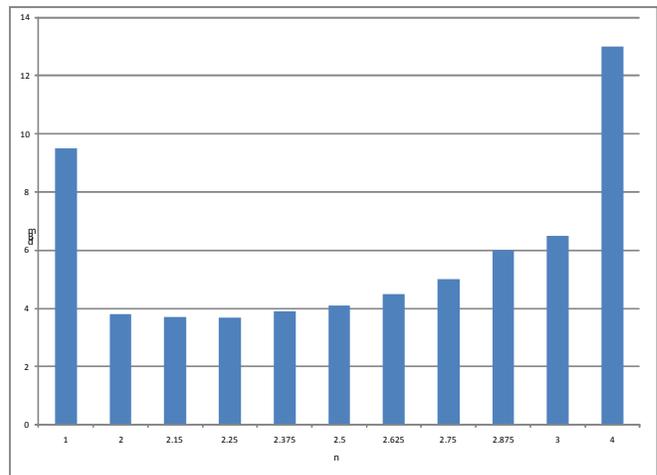


Fig. 3 Root mean square error between theoretical and measured RSSI with different n.

The majority of the RSSI values are in the area between theoretical curves $n=2$ and $n=3$, thus more n values in interval [2.0, 3.0] are researched.

The root mean square error between theoretical and measured RSSI values is calculated and compared with each other. Due to the minimal root mean square error $n=2.25$ is chosen as the appropriate damping coefficient.

Figure 3 shows the typical errors of distance measurements with RSSI.

V. EXPERIMENTAL RESULTS

In order to create an experimental setup for a ZigBee network, 2.4GHz 802.15.4 development kit belonging to the Silicon Laboratories is used. The board is shown in Fig. 4.



Fig. 4. Sensor node.

Each board features a silicon laboratories C8051F121 microcontroller and a Chipcon CC2420 [10] 2.4 GHz 802.15.4 transceiver. Support components include a USB interface, JTAG programming interface, a variety of pushbuttons and a voltage regulator.

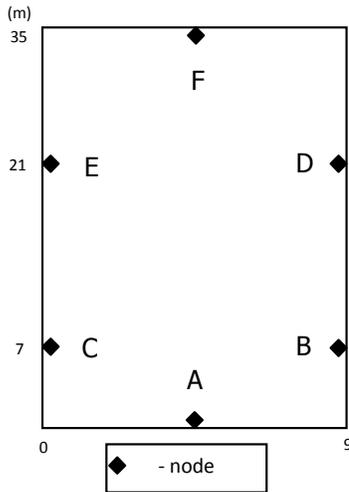


Fig.5 Positions of the nodes in free space.

It is conducted an experiment to investigate the relationship between the measured RSSI and the distance between nodes.

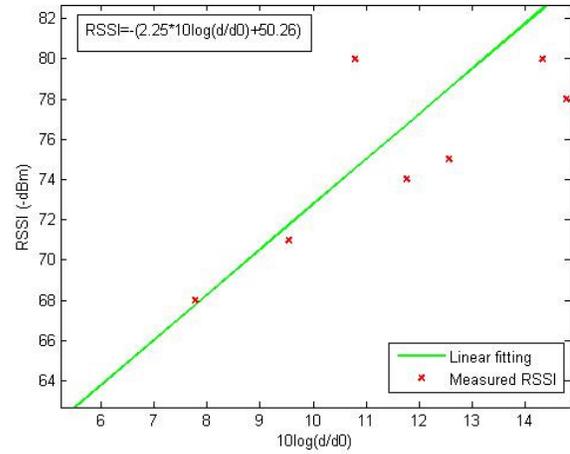
The locations of three main nodes are known (in the figure named with D, E and F), the received signal strength to this nodes are known, the location of the three nodes named with A, B and C should be determined (Fig.5)

Ten measurements for each position and took the average value are performed. The relationship between the received signal strength and the distance is shown in fig.6.

RSSI – based ranging methods require the knowledge of the RSS-to-distance curve as follows:

$$d = 10^{\left[\frac{RSSI - A}{10 \cdot n} \right]} \quad (5)$$

where d denotes the transmitter-to-receiver distance, n is the propagation path-loss exponent. Due to the minimal root mean square error n=2.25 is chosen as the appropriate damping factor.



The RSS-to-distance reference curve in (5) is obtained via a least-square linear fitting from several RSS measured values. The obtained result is shown in Fig.6 along with real measurements. In Fig. 6 i) the RSS values are represented in dBm as provided by the receiver nodes, ii) the distance d in x-axis is normalized to the reference distance of 1m, and iii) the fitting parameters are A=50.26 and n=2.25. Obtained result for the path-loss exponent compared to the theoretical value (n=2) is due to the fact that receiver nodes are located very closed to huge buildings, which provides constructive reflected propagation path in addition to the direct one.

In general, the result of a measurement is only estimate of the value of the measurand and is complete only when accompanied by a statement of the uncertainty of that estimate [11].

Thus the ideal method for evaluating and expressing measurement uncertainty should be capable of readily providing an interval, in particular, one with a coverage probability or level of confidence that corresponds in a realistic way to that required.

The measurement uncertainty $u(x)$ [11] is defined as standard deviation of arithmetic mean \bar{x} :

$$u(x) = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

For the uncertainty of x_T is valid:

$$x_T = \bar{x} \pm s_x \quad (7)$$

The equation (7) is defined as interval in which the accurate value will appear with some probability, known as confidence interval.

The measurement uncertainty u_y of the unknown parameter is defined with:

$$u_y = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \cdot u(x_i)^2} \quad (8)$$

where u_x are the uncertainties of the known parameters;

In this case, for the distance uncertainties between nodes, parameter d depends from the uncertainties of transmitted u_{P_T} and received power u_{P_R} :

$$d = \sqrt{C_f \frac{P_T}{P_R}} \quad (9)$$

Distance uncertainty is calculated with the relation:

$$u_y = \sqrt{\left(\frac{\partial d}{\partial P_T} \cdot u_{P_T}\right)^2 + \left(\frac{\partial d}{\partial P_R} \cdot u_{P_R}\right)^2} \quad (10)$$

The distance uncertainties of the sensor nodes are presented in Figure 7. The localization error (fig. 7) is increasing, when the receiving nodes depart from the transmitter [12]. The positions of the nodes are shown in figure 5. The largest distance between nodes is 30 meters, for this distance the uncertainty is 1m.

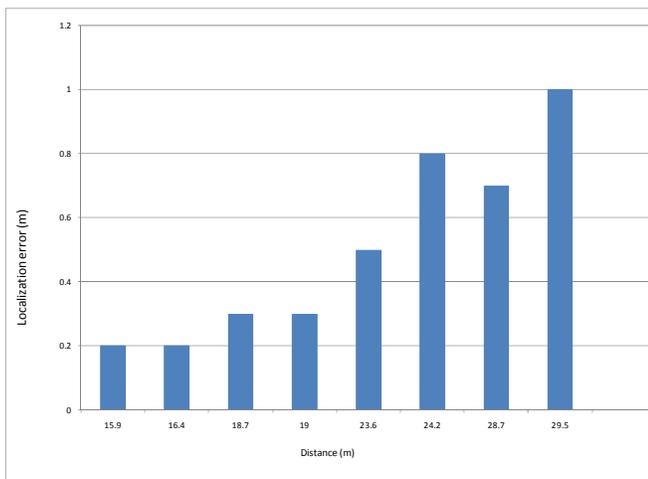


Fig. 7 Localization error.

For the largest distance of 30 meters the localization error is 3%.

This inaccuracy is due to the fact that receiver nodes are located very close to huge buildings, which provides constructive reflected propagation path in addition to the direct one. Multipath arises when more than one path is available for radio signal propagation. Metallic materials as well as dielectrics (or electrical insulators) cause reflections. When multiple signal propagation paths exist, the actual signal level received is the vector sum of all the signals incident from any direction or angle of arrival.

A main challenge with RSS ranging is that the effect of reflecting and attenuating objects in the environment can have larger effects on RSS than distance.

The presented algorithm in combination with Zigbee offers lots of advantages. The most important advantage is the simplified implementation process due to already defined fundamental functions within the provided protocol suite of Zigbee.

VI. CONCLUSIONS

Wireless sensor network localization is attracting significant research interest. WSN are widely applicable to many practical applications including environmental monitoring, military applications, etc. in which sensors may need to know their geographical locations. Accurate and low-cost localization is a critical requirement for the deployment of wireless sensor networks in a wide variety of applications. The simplicity of RSS is especially appealing for the localization in WSN because of their cost, size and power consumption.

This paper has provided a review of the measurement techniques in WSN localization and the corresponding localization algorithms. In this experimental work a RSS-based localization approach for outdoor wireless sensor networks is presented. The RSSI was used to determine the distances between the sensor nodes.

This experiment has summarized theoretical and practical investigations concerning the analysis of RSSI measurements. Localization system that uses RSSI in a sensor network based on the Zigbee standard is implemented. The distance measurement accuracy of the used technique through actual experimental results is evaluated.

In the future research, it is planned to apply measurement of RSSI using different antenna polarization. Rotation of the received antenna from vertical to horizontal will affect the receive energy from multiple signals. It is important to investigate the position of an antenna in order to decrease the scattering of the received signal.

REFERENCES

- [1] Anna Hac, "Wireless Sensor Networks Designs", University of Hawaii at Manoa, Honolulu, USA, 2003
- [2] E. Callaway, P. Gorday, L. Hester, J. A. Gutierrez, M. Naeve, B. Heile, and V. Bahl, "Home Networking with IEEE 802.15.4: A Developing Standard for Low-Rate Wireless Personal Area Networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 70–77, August 2002
- [3] Vlado Handziski, Andreas Köpke, Holger Karl, Adam Wolisz, "A common wireless sensor networks architecture", Technical Report TKN-03-012, Telecommunication Networks Group, Technische Universität Berlin, 2003
- [4] Paolo Santi, "Topology Control in Wireless Ad hoc and Sensor Networks, Istituto di Informatica e Telematica del CNR- Italy, John Wiley, 2005
- [5] Kazem Sohraby, Daniel Minoli, Taieb Znati, "Wireless Sensor Networks: Technology, Protocols and Applications, 2007
- [6] Jason Lester Hill, "System Architecture for Wireless Sensor Networks", PhD Thesis, University of California, Berkeley, 2003
- [7] W. Turin. *Digital Transmission Systems – Performance Analysis and Modeling*. McGraw-Hill Telecommunications. McGraw-Hill, New York, 1998
- [8] Holger Karl and Andreas Willig, "Protocols and Architectures for Wireless Sensor Networks", John Wiley & Sons, 2005
- [9] Bernhard H. Walke, Stefan Mangold, Lars Berlemann, "IEEE 802 Wireless Systems: Protocols, Multi-hop Mesh/Relaying, Performance and Spectrum Coexistence, 2006
- [10] Datasheet for chipcon CC2420 2.4 GHz IEEE 802.15.4/ZigBee RF Transceiver
- [11] Guide to the expression of uncertainty in measurement, European committee for standardization, Brussels, 1999
- [12] S. Slijepcevic, S. Megerian, and M. Potkonjak. Location Errors in Wireless Embedded Sensor Networks: Sources, Models, and Effect on Applications. *ACM Mobile Computing and Communications Review*, 6(3): 67–78, 2002

Continual One Point Auto-Calibration Technique in Simple Sensor-Microcontroller Interface

Zivko Kokolanski, Cvetan Gavrovski and Vladimir Dimcev

Abstract—The modulating resistive and capacitive sensors can be directly interfaced with a microcontroller. The sensor and one reference element (resistor or a capacitor) form a RC circuit that is excited and measured with the microcontroller. In this way the modulating sensor acts like a quasi-digital and allows direct sensor to microcontroller interface without use of a classical A/D converter. There are several known direct sensor to microcontroller calibration techniques. One of the simplest is the one point calibration technique. This calibration technique has nonlinear transfer characteristics and for wide measurement range several calibration resistors have to be employed. The continual one point auto-calibration technique expands the measurement range without the use of additional calibration resistors and in the same time keeps the simplicity of the measurement. In this paper the continual one point auto calibration technique is presented.

Index Terms—Modulating sensors, Continual one point auto calibration, Microcontroller interface.

I. INTRODUCTION

DIRECT sensor to microcontroller interface is an alternative approach for conditioning of modulating resistive and capacitive sensors without the use of A/D converter and signal conditioning circuits. The microcontroller uses the built in timer to measure the charging or discharging time of RC circuit formed by the sensor and reference resistor/capacitor. In this way, the microcontroller and the sensor form a relaxation oscillator causing the modulating sensor to act like a quasi-digital sensor. Hence, this approach is attractive to use for a low cost and small size applications.

Two measurement methods are proposed: a method based on charging [1] or discharging time [2] of the RC circuit. The two methods differentiate by the crossing of the upper or the lower threshold voltage (V_{th} or V_{tl}) of the Schmitt Trigger port to create an interrupt. The method based on discharging time gives better measurement results [3] because the lower threshold voltage V_{tl} has better rejection of the power supply interference and because usually the microcontroller ports can

drain more current than they can generate. In this paper the analysis are restricted to the interfaces based on measurement of the discharging time but the same methodology can be applied to the interfaces based on measurement of the charging time. The most basic direct sensor-microcontroller interface can be realized by using two microcontroller pins, one output and one input pin (Fig. 1).

The measurement contains two phases: charging phase and discharging phase. The wave shape of the capacitor voltage in the two phases is shown in Fig. 2.

At the beginning the pin P_i is set as output with logical state “1” and the pin P_o is set as input (high impedance state). The capacitor charges through R_p to V_{dd} in an interval $t_1 \div t_2$. In the next step the pin P_o is set as output with logical state “0”, the timer starts and the pin P_i is set to high impedance state. This time the capacitor discharges through R_x until the voltage reaches the lower threshold voltage V_{tl} . Crossing of the threshold voltage V_{tl} initiates interrupt that stops the timer. The time needed for the capacitor to discharge from V_{dd} to V_0 is expressed with the equation

$$t_x = (t_3 - t_2) = \tau \ln \left(\frac{V_0 - V_{dd}}{V_0 - V_{tl}} \right) \quad (1)$$

where $\tau = R_x C$ is the discharging time constant.

Having in mind that V_0 , V_{dd} , V_{tl} and C are constant, from (1) can be seen that the time interval t_x is proportional to the measuring resistance R_x . This time interval (t_x) is measured with the built in timer in the microcontroller. The result of the time to digital conversion in clock cycles is

$$N = kR_x \quad (2)$$

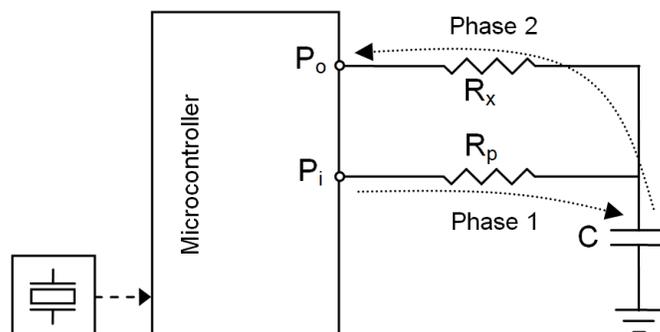


Fig. 1. Direct sensor-microcontroller interface.

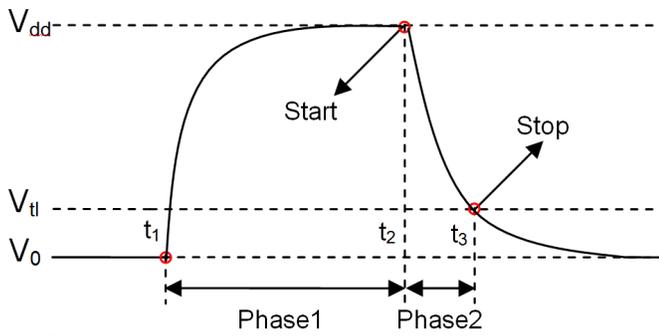


Fig. 2. Wave shape of the capacitor voltage in the two phases.

where k is constant dependent on V_0 , V_{dd} , V_{tl} , C and the time base of the timer. In practice the input/output resistances and leakage currents of the microcontroller ports cause gain, offset and nonlinearity errors [4]. Moreover, the constant (k) in the equation (2) is not very stable. Therefore, usually direct sensor to microcontroller interface is realized by using some calibration technique [5] that cancels the contribution of V_0 , V_{dd} , V_{tl} and C .

II. SINGLE POINT CALIBRATION TECHNIQUE

The simplified electrical circuit of the direct sensor to microcontroller interface by using single point calibration is shown in Fig. 3.

The measurement contains two phases: measurement of the unknown resistance R_x and measurement of the calibration resistance R_c . The wave shape of the capacitor voltage in the two phases is shown in Fig. 4.

The respective times needed to discharge the capacitor through R_x and R_c , t_x and t_c are

$$t_x = R_x C \ln \left(\frac{V_0 - V_{dd}}{V_0 - V_{tl}} \right), \text{ and} \quad (3)$$

$$t_c = R_c C \ln \left(\frac{V_0 - V_{dd}}{V_0 - V_{tl}} \right). \quad (4)$$

Since the capacitance C and the parameters V_0 , V_{dd} and V_{tl} are the same for both measurements, calculating the ratio of t_x and t_c cancels their contribution. The measurement result is

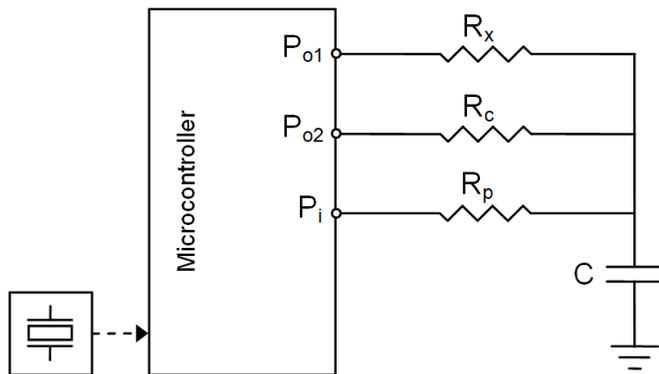


Fig. 3. Electrical circuit of direct sensor to microcontroller interface by using single point calibration.

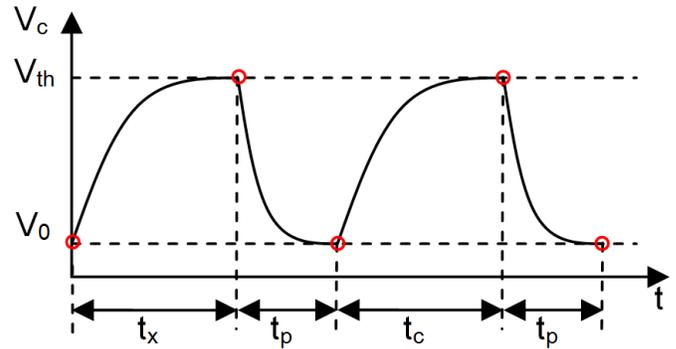


Fig. 4. Electrical circuit of direct sensor to microcontroller interface by using single point calibration.

$$R_x^* = \frac{t_x}{t_c} R_c. \quad (5)$$

The one point calibration technique of this kind is analyzed in [1,2], where a resolution from 6 to 10 bits is achieved. The resolution is limited by the input and output resistances and leakage currents of the microcontroller ports. These limiting factors are considered in [4] and the dependence of the measured value and the estimate of the actual sensor resistance is expressed with

$$R_x^* = (R_{o1} + R_x) k \frac{\ln A}{\ln B} \left[\frac{1}{1 + \frac{R_x}{R_{o1} + R_{e1}}} \right] \quad (6)$$

where R_{o1} is the output resistance of the port P_{o1} , and A , B and k are constants dependent on the resistances and the leakage currents of the microcontroller ports. From (6) it can be seen that the input and output resistances and leakage currents cause offset, gain and nonlinearity errors. If the ports P_{o1} and P_{o2} were identical these errors would have being zero for $R_x=R_c$. This leads to conclusion that the value of the calibration resistor should be chosen to be $R_c=R_x$ in the middle of the measurement range.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments were realized by using microcontroller PIC16F877 [6] with clock frequency of 4 MHz, effective instruction cycle speed of 1 MHz and period of $1\mu s$. The falling edge of the input signal was registered with the RB0/INT Smith Trigger pin. This pin initiates interrupt that stops the 16-bit timer - Timer1. To reduce the noise effects affecting the voltage comparison between V_c and V_{tl} several design solutions were applied:

- Decoupling capacitor of 100 nF was placed as close as possible to the microcontroller pins as recommended from the manufacturer
- The board ground plane was carefully designed for low electromagnetic interference
- Only the microcontroller was supplied from the power supply to eliminate other interference effects
- The microcontroller didn't execute any other task while waiting for the interrupt.

The measurements were performed by using a variable resistor in the range from 1000 to 3000 ohms. These values are typical for resistive temperature sensors. The resistance of the sensor was measured with measurement instrument with maximal error of $\pm 0.1\% + 5$. According to (6), for minimal errors the calibration resistance was chosen to be 2000 ohms, which is in the middle of the measurement range. The absolute and the relative errors of the measurements are shown in Fig. 5 and Fig. 6.

From the results reported in Fig. 5 and Fig. 6 it can be seen that the static errors caused by the input and output resistances and leakage currents of the microcontroller ports are minimal in the middle of the measurement range. Having in mind (6) this completely confirms our expectations. However, the errors are maximal for the minimal value of the sensor resistance ($\approx 23\Omega$ or 2.3%). The results suggest that in this case a lower value for the calibration resistance have to be chosen in order to reduce the relative error for the whole measurement range. However, this approach cannot be generalized because according to (6) the errors are nonlinear and grater for lower values of the sensor resistance.

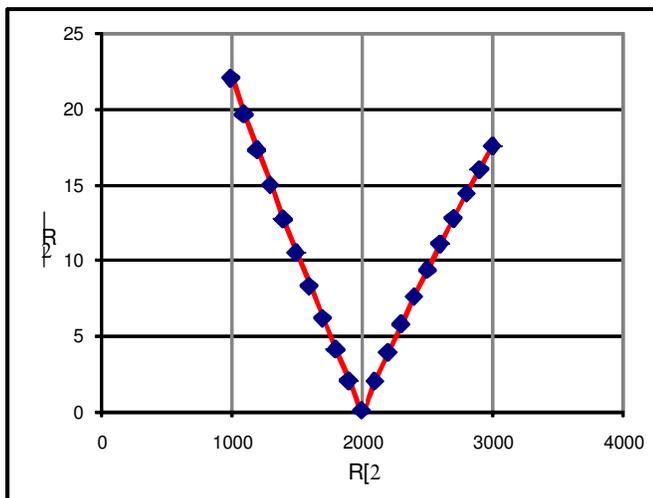


Fig. 5. Absolute errors of the measurements.

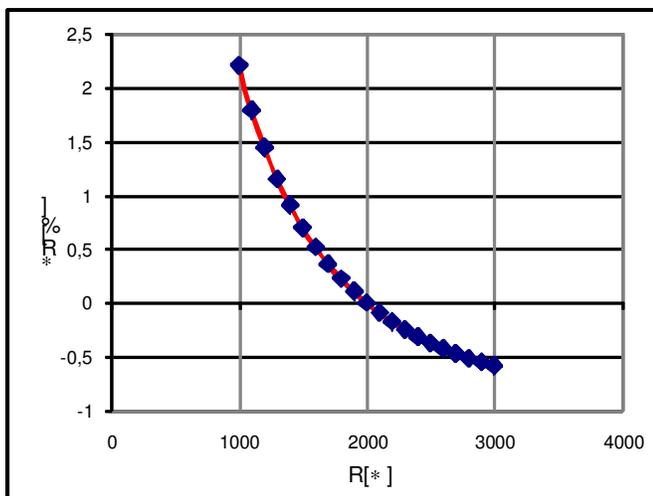


Fig. 6. Relative errors of the measurements.

The dependence of the measured value and the actual sensor resistance was approximated by using the least square method. The approximation with linear function is shown in Fig. 7.

The experimental results for the one point calibration method showed that the measurement range is limited by the maximal error. Therefore, for a given maximal error the measurement range must be divided into appropriate sub ranges. Each sub range must have its own calibration resistance in the middle of the range. This proportionally increases the used microcontroller pins and the number of calibration resistances. Hence, in some cases where the sensor resistance varies over wide dynamic range and where “high” measurement resolution must be achieved this approach tends to lose its economical justification.

IV. CONTINUAL ONE POINT AUTO-CALIBRATION TECHNIQUE

One of the disadvantages of the one point calibration technique applied in direct sensor to microcontroller interface is the limited measurement range. The measurement range is limited by the nonlinear dependence of the measured and the actual sensor resistance when approximated by a linear function. To overcome this problem, the measurement range must be divided into smaller segments and for each segment different calibration resistor have to be used. Consequently, the nonlinearity errors will be smaller than the half of the least significant bit for the desired resolution. Hence, this approach can be used for expanding the measurement range at the cost of additional calibration resistances and additional microcontroller pins.

The same benefit can be achieved when using the continual one point auto calibration technique but using significantly less microcontroller pins. This approach uses digital variable resistor that replaces a lot of calibration resistances and in the same time uses fixed number of microcontroller pins. Today, there are a lot of commercially available digital variable resistors with different characteristics. Most of them are capable to trim the resistance over at least 256 points. Such

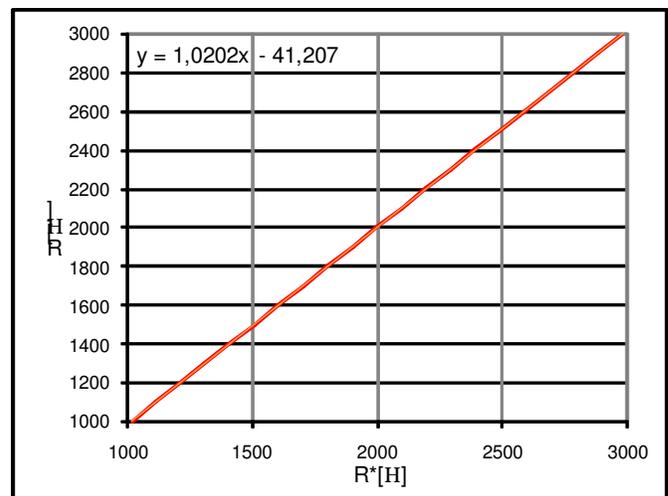


Fig. 7. Approximation by using least square method.

performances are good enough to achieve relatively high measurement resolution and wide measurement range when using the continual one point auto calibration technique. However, the values of the calibration resistances have to be accurate enough to maintain the relative uncertainty, otherwise the calibration resistances have to be measured with precise measurement instrumentation. The measured values are stored in the microcontroller EPROM in a form of lookup table.

The electrical scheme of the continual one point auto calibration technique is shown in Fig. 8.

The digital resistor in Fig. 8 is used as a calibration resistor with variable resistance in 2^n points, where n is the resolution of the digital resistor. Today, there are a lot of available commercial digital variable resistors with different resolutions and different range of values. In most of the cases the communication with the microcontroller is realized by using some serial communication interface such as SPI, RS232, I²C.

The implementation of the one point auto calibration technique is performed by division of the measurement range into several sub ranges. The nonlinearity error of each sub range must be lower than half of the least significant bit for the desired resolution of the measurements. The condition that must be fulfilled for each sub range (reported in [7]) is

$$\frac{\Delta R_x}{2} < \frac{c - R_o}{1 - 2^{-(n+1)}} - c, R_o - \frac{\Delta R_x}{2} < R_x < R_o + \frac{\Delta R_x}{2} \quad (7)$$

where the constant c depends of the input and output resistances and leakage currents of the microcontroller ports and the calibration resistance value. The graphical representation of the expanded measurement range by using the continual one point auto calibration technique is shown in Fig. 9.

The continual one point auto calibration technique is performed in three phases: rough measurement of the unknown resistance, choice of optimal calibration resistance and precise measurement of the sensor resistance.

In the first phase, the value of the calibration resistance is chosen in the middle of the measurement range, that is $R_c = \Delta R_x / 2$. In this way a rough measurement of the unknown resistance is performed, and the optimal sub range is

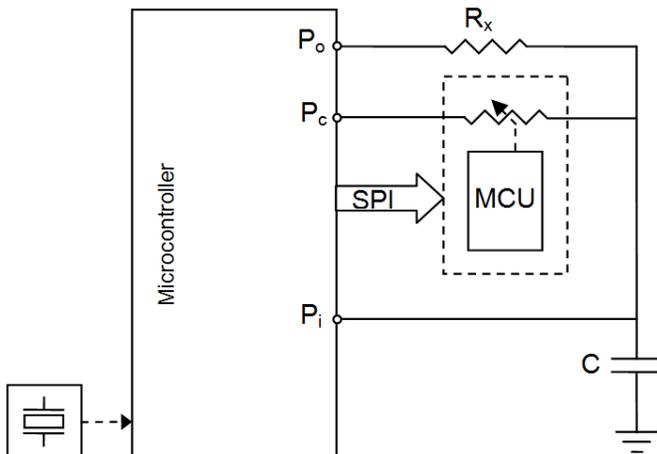


Fig. 8 Continual one point calibration technique.

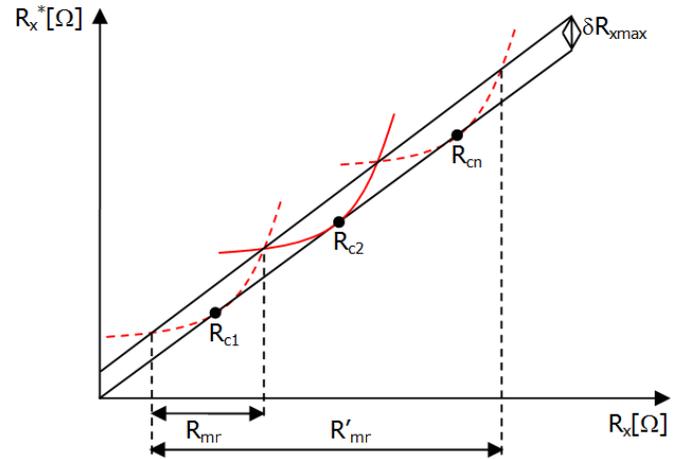


Fig. 9. Continual one point calibration technique.

determined. In the second phase the digital resistor is set to a value in the middle of the chosen sub range. In the third phase a precise measurement and calculation of the sensor resistance is performed.

From the graphic shown in Fig. 9 it can be seen that the measurement range is drastically expanded, and in the same time the maximal nonlinearity error remains within the predefined limits. However, the implementation of this calibration technique requires more complex algorithm compared to the one point calibration. This slightly decreases the speed of the measurements (1/3 slower than the one point calibration technique).

V. IMPLEMENTATION RESULTS AND DISCUSSION

The experiments were realized by using the same microcontroller as in section IV, PIC16F877 with clock frequency of 4 MHz. The falling edge of the input signal was registered with the RB0/INT Smith Trigger pin. The noise effects affecting the voltage comparison between V_c and V_{il} was reduced by applying the same design solution as in section IV.

The measurements were performed by using a multi-turn variable resistor in the range from 1kΩ to 100kΩ with a step of 1kΩ. The resistance was measured by using the four wire method, with instrument with maximal error of $\pm 0.1\% + 5$.

Having in mind the equation (6), if we use only one calibration resistor in the middle of the 100kΩ measurement range, according the theoretical calculations the input/output resistances and leakage currents will produce a static errors less than 6% at the beginning and at the end of the measurement range. Hence, if we divide the measurement range, for example on ten sub-ranges then the relative error is expected to be decreased ten times. To test these analyses the one point auto-calibration technique was implemented by using ten calibration resistors. According to (6), for minimal errors the values of the calibration resistances was chosen in the middle of each sub-range.

The calibration resistances were obtained by using a digital

variable resistor, type MCP41100 [8]. This device is a 256-position digital potentiometer with maximal value of 100k Ω . The calibration resistances were set as close as possible to the middle of the predefined sub-ranges and were measured with a measuring instrument with maximal error of $\pm 0.1\% + 5$. The measured values were stored in the microcontroller EPROM forming a lookup table. The ordinal number of the sub-range was used as a pointer in the lookup table.

The MCP41100 communicates with the microcontroller by using the SPI serial interface. The communication was realized through the RC0 microcontroller pin. To reduce the noise effects, the communication between the microcontroller and MCP41100 was disabled during the measurement phases.

The practical experiments were performed twice, when using continual one point auto calibration and when using one point calibration. Each measurement was performed ten times over the whole measurement range of 100k Ω with a step of 1k Ω . The absolute and relative errors of the measurements when using continual one point auto calibration are shown in Fig. 10 and Fig. 11 and when using one point calibration in Fig. 12 and Fig. 13. In Fig. 14 a comparison of the measured and the actual sensor resistance when using the continual one point auto calibration technique (green line) and the one point calibration (red line) is given.

From the results reported in Fig. 10÷Fig. 13 it can be seen that the static errors caused by the input and output resistances and leakage currents of the microcontroller ports are minimal in the middle of the measurement range or sub-range. Having in mind the equation (6) this again confirms our expectations. However, the errors are maximal near the borders of each sub-range and they are less than 1.3% when using the continual one point auto calibration technique. The relative errors when using one point calibration are less than 11%. It can be seen that the continual one point auto calibration technique reduces the static errors nearly ten times comparing to the one point calibration. The relative error reduction factor is proportional to the number of the calibration resistors used. The digital

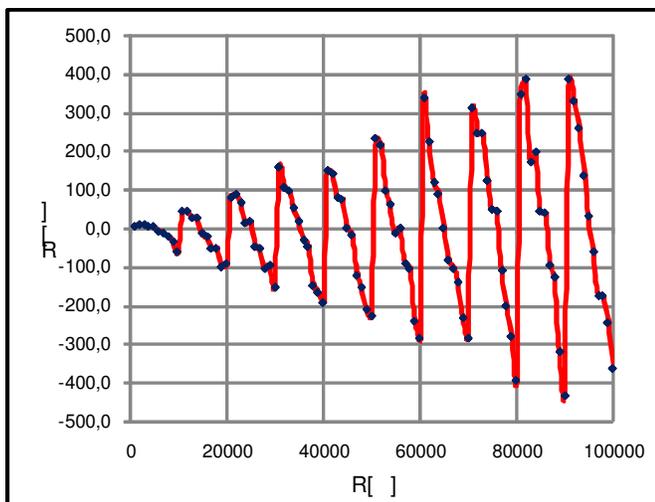


Fig. 10. Absolute errors of the continual one point auto calibration technique.

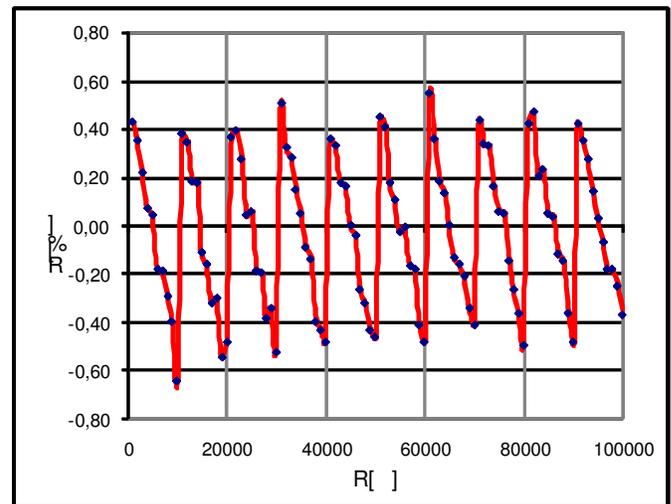


Fig. 11. Relative errors of the continual one point auto calibration technique.

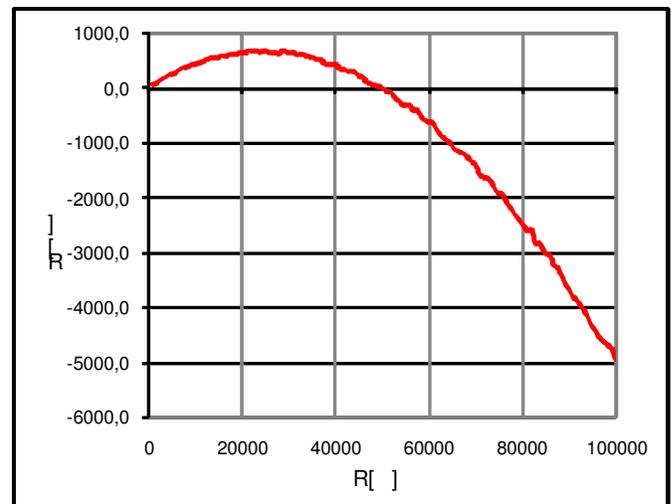


Fig. 12 Absolute errors of the one point calibration technique.

variable resistor MCP41100 can trim the calibration resistance over 256 discrete points. Thus, if all the resistances are used then the relative errors caused by the input and output resistances and leakage currents will be reduced nearly 256 times comparing to the one point calibration technique. However, there is no meaning of using such large number of calibration resistances because in that case the quantization errors and noise interference are much more dominant in the total measurement uncertainty. Moreover, using a large number of calibration resistances complicates the implementation algorithm of this technique and reduces the speed of the measurement. The optimal number of calibration resistances can be calculated by using the relation (7), where the relative errors caused by the input and output resistances and leakage currents should be lower than half of the least significant bit for the desired resolution. For example, if we want to achieve 10 bit resolution over 100k Ω measurement range, then the relative error should be lower than 0.05%. This can be achieved by using 104 calibration resistors. For this particular example, if the microcontroller PIC16F877 is used,

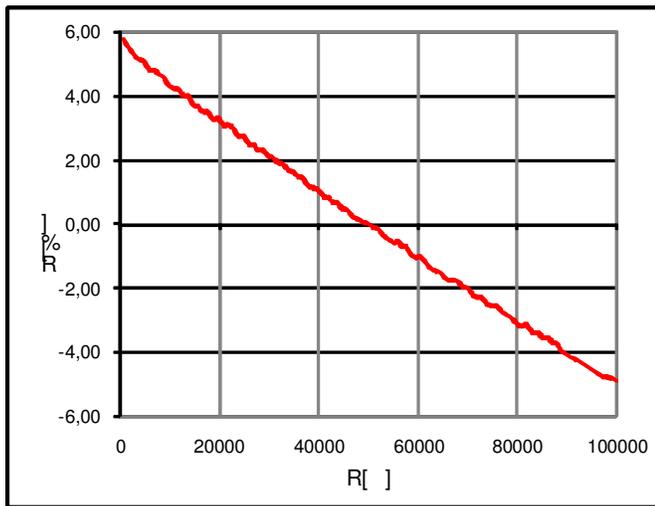


Fig. 13. Relative errors of the one point calibration technique.

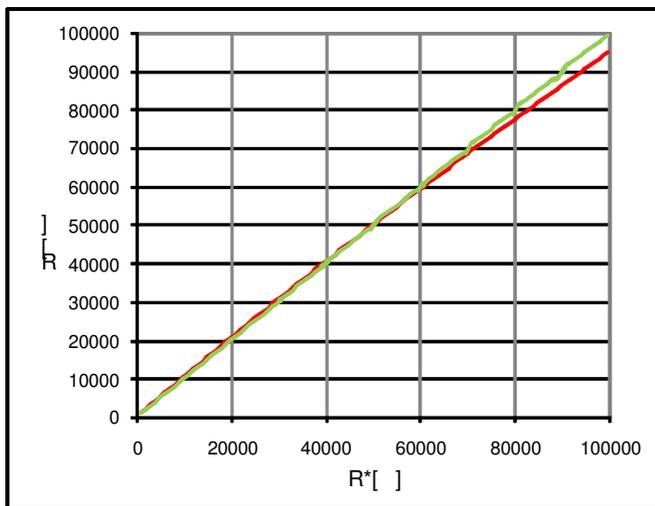


Fig. 14. Comparison of the measurement results.

the value of the constant c in (7) for the second sub-range would be $c=2541$.

VI. CONCLUSION

The paper is organized in five sections. In the first section a brief introduction of the direct sensor-microcontroller interface is presented. The second section elaborates the one point calibration technique. The implementation of this calibration technique is experimentally verified by using the PIC 16F877 microcontroller and the results are reported in section three. In the next two chapters the continual one point auto-calibration technique and its implementation by using the PIC 16F877 microcontroller and the digital variable resistor MCP41100 is presented.

The one point calibration in direct sensor to microcontroller interface is very simple and effective calibration technique.

However, the input and output resistance and leakage currents cause gain, offset and nonlinearity errors. Thus, the measurement range is limited by the maximal nonlinearity error. For a defined maximal error, the measurement range must be divided into several sub ranges. The disadvantage of this approach is that it uses additional calibration resistances and microcontroller pins.

The same benefit can be achieved when using the continual one point auto calibration technique but using significantly less microcontroller pins. This technique is characterized by a very simple implementation and usage of little microcontroller pins. However, the use of digital variable resistor increases the cost of the measurement system and slightly decreases the speed of the measurements (by 1/3) because of more complex implementation algorithms. Therefore, this technique is reasonable to apply in cases where a higher measurement resolution and wider measurement range needs to be achieved.

In this paper a directions and conditions that must be fulfilled for implementation of the continual one point auto calibration technique were presented.

In the paper, practical experiments when using both calibration techniques were performed. The measurements were realized by using a multi-turn variable resistor in the range from $1\text{k}\Omega$ to $100\text{k}\Omega$ with a step of $1\text{k}\Omega$. For implementation of the continual one point auto calibration technique a digital variable resistor MCP41100 was used.

The measurement errors when using the one point auto calibration technique were maximal near the borders of each sub-range (around 1.3%). When the one point calibration was used, the measurement errors were around 11%. It can be seen that the continual one point auto calibration technique reduced the nonlinearity errors nearly ten times comparing to the one point calibration. The relative error reduction factor was proportional to the number of the calibration resistors used.

REFERENCES

- [1] D. Cox, "Implementing ohmmeter/temperature sensor", AN512 Microchip Technology Inc., 1997.
- [2] L. Bierl, "Precise measurements with the MSP430", Report Texas Instrument, 1996.
- [3] F. Reverter, M. Gasulla, R. P. Areny, "Analysis of Power-Supply Interference Effects on Direct Sensor-to-Microcontroller Interfaces", IEEE Trans. Instr. and Meas., vol.56, no.1, Feb. 2007.
- [4] A. Custodio, R. P. Areny, R. Bragos, "Error Analysis and Reduction in Simple Sensor-Microcontroller Interface", IEEE Trans. Instrumentation and Measurement, Vol.50, No.6, Dec. 2001.
- [5] F. Reverter, R. P. Areny, "Direct Sensor to Microcontroller Interface Circuits", Marcombo, Barcelona, 2005.
- [6] Microchip Technology Inc., "PIC16F8XX Data Sheet - 28/40-Pin 8-Bit CMOS FLASH Microcontrollers", DS30292C, 2001.
- [7] Z. Kokolanski "Improving the Uncertainty of Reconfigurable Microprocessor-Based Measuring Systems", Master Thesis, Faculty of Electrical Engineering and Information Technologies, Skopje, March 2010.
- [8] Microchip Technology Inc., "Single Digital Potentiometer with SPI Interface", DS11195C, 2003.

Erratum

After the publication of the previous “Electronics” issue (Vol. 14, No. 2, December 2010) we have noticed that there was a mistake in the paper title of authors T. B. Šekara and M. R. Mataušek (pp. 77–81). Instead of title “Self-Oscillating Fluxgate Current Sensor with Pulse Width Modulated Feedback”, there should stand “Comparative Analysis of the Relay and Phase-Locked Loop Experiment Use to Determine Ultimate Frequency and Ultimate Gain”. We use this opportunity to apologize to the authors and readers of the “Electronics” journal for any inconvenience caused by this technical mistake.

“Electronics” journal

Instruction for Authors

Editorial objectives

In the review "Electronics", the scientific and professional works from different fields of electronics in the broadest sense are published. Main topics are electronics, automatics, telecommunications, computer techniques, power engineering, nuclear and medical electronics, analysis and synthesis of electronic circuits and systems, new technologies and materials in electronics etc. In addition to the scientific and professional works, we present new products, new books, B. Sc., M. Sc. and Ph.D. theses.

The main emphasis of papers should be on methods and new techniques, or the application of existing techniques in a novel way. Whilst papers with immediate application to particular engineering problems are welcome, so too are papers that form a basis for further development in the area of study.

The reviewing process

Each manuscript submitted is subjected to the following review procedures:

- It is reviewed by the editor for general suitability for this publication;
- If it is judged suitable two reviewers are selected and a double review process takes place;
- Based on the recommendations of the reviewers, the editor then decides whether the particular article should be accepted as it is, revised or rejected.

Submissions Process

The manuscripts are to be delivered to the editor of the review by the e-mail: electronics@etfbl.net.

Upon the manuscript is accepted for publication, the author receives detailed instructions for preparing the work for printing.

Authors should note that proofs are not supplied prior to publication and ensure that the paper submitted is complete and in its final form.

Copyright

Articles submitted to the journal should be original contributions and should not be under consideration for any other publication at the same time. Authors submitting articles for publication warrant that the work is not an infringement of any existing copyright and will indemnify the publisher against any breach of such warranty. For ease of dissemination and to ensure proper policing of use, papers and contributions become the legal copyright of the publisher unless otherwise agreed.

ELECTRONICS, VOL. 15, NO. 1, JUNE 2011

SPECIAL SECTION DEDICATED TO PROF. DR MILIĆ STOJIĆ 70 YEARS OF LIFE (INVITED PAPERS)	
PROF. DR MILIĆ STOJIĆ BIOGRAPHY	1
EDITORIAL	3
<hr/>	
EXTRACTION OF EXTERNAL TORQUE DISTURBANCE IN POSITIONING SERVOMECHANISM	5
Milić R. Stojić and Đorđe M. Stojić	
HIGH FREQUENCY POWER SUPPLY FOR ELECTROSTATIC PRECIPITATORS IN THERMAL POWER PLANTS	11
Slobodan Vukosavić	
REALIZATION OF AN ELECTRONIC LOAD FOR TESTING LOW POWER PEM FUEL CELLS	21
Djordje Šaponjić, Vojislav Arandjelović, Milica Marčeta Kaninski and Aleksandar Maksić	
ROBUST RST CONTROLLER DESIGN BY CONVEX OPTIMIZATION	24
Milan S. Matijević, Ranko Sredojević and Vladimir M. Stojanović	
PI-LIKE OBSERVER STRUCTURES IN DIGITALLY CONTROLLED DC SERVO DRIVES: THEORY AND EXPERIMENTS	30
Milica B. Naumović	
FREQUENCY DOMAIN DESIGN OF A COMPLEX CONTROLLER UNDER CONSTRAINTS ON ROBUSTNESS AND SENSITIVITY TO MEASUREMENT NOISE	40
Tomislav B. Šekara, Miloš B. Trifunović and Vidan Govedarica	
CHOICE OF WINDOW SIZE IN CALIBRATING THE GEOMETRY OF MANIPULATORS BASED ON THE REGIONS CORRELATION	45
Petar Maric and Velibor Djalic	
LAND VEHICLE NAVIGATION SYSTEM BASED ON THE INTEGRATION OF STRAP-DOWN INS AND GPS	54
Rade Stančić and Stevica Graovac	
<hr/>	
PAPERS	
NHIBERNATEMAPPER - A TOOL FOR RAPID DEVELOPMENT OF DATA ACCESS LAYER FOR INTEROPERABLE GIS SOLUTIONS	62
Stevica S. Cvetković, Miloš D. Bogdanović and Leonid V. Stoimenov	
LOCALIZATION TECHNIQUES IN WIRELESS SENSOR NETWORKS USING MEASUREMENT OF RECEIVED SIGNAL STRENGTH INDICATOR	67
M. Srbinovska, V. Dimcev, C. Gavrovski and Z. Kokolanski	
CONTINUAL ONE POINT AUTO-CALIBRATION TECHNIQUE IN SIMPLE SENSOR-MICROCONTROLLER INTERFACE	72
Zivko Kokolanski, Cvetan Gavrovski and Vladimir Dimcev	
