FACULTY OF ELECTRICAL ENGINEERING
UNIVERSITY OF BANJA LUKA

ELECTRONICS

VOLUME 13, NUMBER 2, DECEMBER 2009

F A C U L T Y   O F   E L E C T R I C A L   E N G I N E E R I N G
U N I V E R S I T Y   O F   B A N J A   L U K A

Address: Patre 5, 78000 Banja Luka, Bosnia and Herzegovina
Phone: +387 51 211824
Fax: +387 51 211408

E L E C T R O N I C S

Web: www.electronics.etfbl.net

E-mail: electronics@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

Number of printed copies: 100

Guest Editorial

SYMPOSIUM INFOTEH®-JAHORINA is continuation of the International symposium JAHORINA that was held last time on April 1991. The main organizer of the Symposium is the Faculty of Electrical Engineering East Sarajevo and the co-organizer is the Faculty of Electrical Engineering Banja Luka. The Symposium is supported by The Faculty of electrical engineering, Belgrade, Serbia, the Faculty of electronics, Niš, Serbia, the Faculty of technical sciences, Novi Sad, Serbia.

The goal of the Symposium is multidisciplinary survey of the actual state in the information technologies and their application in the industry plants control systems, the communication systems, the manufacturing technologies, power system, as well as in the other branches of interest for the successful development of our living environment.

During the first Symposium, that was held on 12-14 March 2001, 53 works have been presented, six companies presented their development and manufacturing programs in telecommunications, power electronics, power systems and process control systems. More than hundred participants took part in the Symposium working. Round table about potentials and possibilities of economic cooperation between Republic of Srpska and FR Yugoslavia has been held during the Symposium, regarding successful appearance at domestic and foreign market.

During the second Symposium, that was held on 25-27 March 2002, 76 works have been presented, and five companies presented their development and manufacturing programs. More than hundred and thirty participants took part in Symposium working. Round table entitled "Reforms in high education – step forward to the European University" has been held during the Symposium.

The third Symposium was held on 24-26 March 2003. More than hundred and fifty participants took part in the Symposium working. At the Symposium 73 papers have been presented and nine student papers. Four companies presented their development and manufacturing programs. Round table entitled "New Technologies and Industrial Production Capabilities for small Countries in the Transition Process" has been held during the Symposium.

The fourth Symposium was held on 23-25 March 2005. At the Symposium two invited papers, eighty and three papers and three student's papers have been presented. Three companies presented their development and manufacturing programs. All papers are presented in symposium proceedings, CD version (ISBN-99938-624-2-8).

The fifth Symposium was held in 2006, March 23 – 25. The Symposium was dedicated to the 150th anniversary of Nikola Tesla's birth. It was the first official manifestation in the Republic of Srpska dedicated to this jubilee. The university character of this traditional international scientific and professional gathering was reflected in a complete and multidisciplinary approach to the discussion of themes within the discussion forums, lectured and presentations. There were exposed 137 papers which passed the review, and 172 guests of the symposium were registered.

The sixth Symposium was held in 2007, March 28 – 30. and it was dedicated to global warming and its side effects. The opening lecture by call was held by Ana Pavlović, physicist of meteorology and living environment modeling from Faculty of Technical Sciences, University of Novi Sad. 160 papers were written for the Symposium, and 131 of them were accepted, including 6 students' works.

The seventh Symposium was held in 2008, March 26 – 28. The Symposium was dedicated to the life and work of Milutin Milanković, and Professor Milić Stojić, Ph. D. from the University of Belgrade had an extraordinary introductory lecture about the life and achievements of this great Serbian scientist. The Symposium was opened by the inspiring speech of the academician Rajko Kuzmanović, the President of the Republic of Srpska. Over 220 papers were written for the Symposium, and 169 of them were accepted.

The eighth fourth Scientific – Professional Symposium INFOTEH®-JAHORINA 2009 was held on 18-20 March 2009 in the hotel Bistrica at Jahorina. The Symposium was dedicated to the year of astronomy. The opening lecture by call "The Origin and the Fate of the Universe" was held by Professor Pavle Kaluđerčić, Ph. D. from the Faculty of Electrical Engineering, University of East Sarajevo. The main topics of the Symposium were: Computer science application in control systems, Information-communication systems and technologies, Information system in manufacturing technologies, Information technologies in power systems, Information technologies in other branches of interest. Over 260 papers were written for the Symposium, and 199 of them were accepted and exposed, including 19 students' works, and 265 guests of the symposium were registered.

I would like to invite all readers of the "Electronics" journal to take active participation at the next Symposium INFOTEH®-JAHORINA. Updated information can be obtained from the Symposium web page: http://www.infoteh.rs.ba.

Professor emeritus Slobodan Milojković. Ph. D.
Chairman of the Programming Commitee,
INFOTEH®-JAHORINA 2009

Short Biography of Guest Editor

Slobodan M. Milojković was born in Belgrade, Serbia on 26 June 1941. he received the B.Sc. and M.S.E.E. degrees from the University of Belgrade in 1963, 1975, respectively, and Ph.D. degree from the University of Sarajevo in 1978. In January 1964, he joined the Institute of Thermo Technique and Nuclear Technique, Energoinvest Sarajevo, where he worked in modeling and numerical simulation in the various physical processes: nuclear technique, thermo technique, electrical engineering, electrochemistry, electro thermal engineering, electromagnetic. In September 1964, he began academic career as an honorary lecturer in the Faculty of Electrical Engineering. In May 1970, he was an honorary docent, teaching courses in theoretical electrical engineering. In January 1981, he joined the Faculty of Electrical Engineering, University of Sarajevo, where he is presently Full Professor, teaching courses in theoretical electrical engineering. From October 1994 to April 1996, he worked as a Guest Professor at Technical University of Munich in High Voltage Institute. During that time, he was working on the development of the software modules for modeling, calculation and visualization of electromagnetic fields in high voltage equipment. The Asea Brown Boveri AG Research Center, Heidelberg, used those newly created modules in a software package POLOPT®. He led research projects supported by international organizations and commercial enterprises. His current research interest involves the computer-aided design of electromagnetic devices. Professor S. Milojković is one of the authors of monograph “Integral Methods for the Calculation of Electric Fields, for Application in High Voltage Engineering”, Scientific Series of the International Bureau Research Center Juelich – Germany, 1992, ISBN 3-89336-084-0. He is also author of 6 books and over 10 scientific journal papers, 35 conference papers and 76 research projects. He was the supervisor of nearly 85 Diploma engineering thesis, 18 master thesis and 5 Ph.D. thesis. Professor S. Milojković was awarded the Prize of the enterprise Energoinvest Sarajevo, the Gold medal of Technical Military Academy Rajlovac, the Gold diploma of the Faculty of Electrical Engineering Sarajevo, the Diploma of the Faculty of Electrical Engineering Tuzla, the Gold diploma of The University of Sarajevo. In May 2009 the Senate of the University of East Sarajevo elected Professor Milojković into the title of Professor Emeritus.

Upgrade of Conventional Positional Systems into High-Precision Tracking Systems Using Sliding Mode Controlled Active Digital Compensators

Boban Veselić, Branislava Peruničić, and Čedomir Milosavljević

***Abstract*—The paper offers a possibility of upgrading conventional PD controlled positional systems into high-precision tracking systems using active compensators. For improving of tracking as well as disturbance rejection capabilities of these systems, two digital active compensators are used. The first one is feedforward improvement of tracking, whereas the second one represents feedback compensation of disturbances. The introduced compensators contain active sliding mode controlled subsystems. The proposed solution does not require any additional sensors. The proposed control extension is described as well as digital sliding mode controller design procedures. Also, simulation results in case of dc motor servo-system are presented.**

***Index Terms*— Servo-systems, Active compensators, Sliding mode control, Digital controllers.**

I. Introduction

POSITIONING and tracking are the two basic control tasks that can be met in motion control. In positioning the input or referent signal is step function. It is required to provide as fast and accurate response as possible, preferably without overshoot, whereas the transient trajectory is not specified. In tracking it is necessary to enforce the system output to continuously and accurately follow the referent signal, which may represent very complex trajectory. Modern production technologies impose on control systems more rigorous demands. One of them is flexibility, meaning that the same positional servo-system equally successfully execute the both afore mentioned control tasks, under action of parameter variations and external disturbances.

Most of positional systems in mechatronics, robotics and various industrial applications are realized by using conventional PD controllers. Such systems can ideally track only constant signals, but already under action of constant external disturbance the positioning error occurs. In the applications where an accurate tracking of complex trajectories is required under action of disturbances, these systems give unsatisfactory results. Then some other control technique must be applied that provides simultaneously both accurate tracking and great robustness. One approach may be use of the two degree of freedom controllers, which allow the problems of tracking and disturbance rejection to be treated separately. Moreover, it is possible to independently tune the responses with respect to the referent signal and to the disturbances, [1]-[3]. Further improvement is suggested in [4,5] by multirate sampling.

An appropriate solution to the described control task is implementation of variable structure control systems (VSCS) [6], whose theoretical invariance to disturbances in ideal sliding mode [7] is reduced to excellent robustness in practical realizations. That is the reason why VSCS found their largest application exactly in this field. As a state space technique, VSCS need information of all state coordinates for ideal tracking of arbitrary referent signals. This practically means the knowledge of the tracking error signal and its successive derivatives, and therefore the knowledge of referent signal derivatives. Accordingly, ideal tracking is possible only for the analytically known or known in advance references. Since this is not the case in servo-systems, tracking accuracy depends on a number of available derivatives of the tracking error signal [8]. Second order sliding mode control is suggested in [10] for the servo-system synthesis, where sliding mode based differentiator is used for evaluation of the error signal derivative [11,12]. Differentiators are practically useful only for the first and second order derivatives of the signal, whereas high order derivatives are completely inapplicable due to severe noise contamination.

In order to further improve system accuracy additional disturbance compensation is often carried out. Extraordinary improvements were achieved in various servo-systems by so called active disturbance estimator (ADE) [9,13], which contains a sliding mode controlled active subsystem. Also, there is a possibility of introduction of supplemental integral action into VSCS that additionally increases system

B. Veselić is with Universisty of Niš, Niš, Serbia (e-mail: boban.veselic@elfak.ni.ac.rs).

B. Peruničić, is with the Faculty of Electrical Engineering, University of Sarajevo, Sarajevo, Bosnia and Herzegovina (e-mail: brana_p@hotmail.com).

Č. Milosavljević is with the Faculty of Electrical Engineering, University of Istočno Sarajevo, Istočno Sarajevo, Bosnia and Herzegovina (e-mail: cedomir.milosavljevic@elfak.ni.ac.rs).

accuracy [14].

This paper proposes a way to upgrade the conventional positioning systems into a high-accuracy robust tracking systems by using active compensators (ACs). Since a conventional system needs to be improved in tracking capability as well as in disturbance rejection, two digital ACs are introduced. The first AC represents feedforward improvement of tracking. The second AC is actually the ADE [9,13] that compensates system disturbances and is located in a local feedback loop. These digitally implemented ACs involve an active control substructure based on discrete-time sliding mode control (DSMC). In the paper the proposed control extension is described in details, DSM controller design procedure is explained and simulation tests on DC motor are presented.

II. Ideal Tracking System

The well-known control structure with feedforward and disturbance compensations is shown in Fig.1 in digital realization. Under certain conditions this structure can ensure the output signal to ideally track the reference.

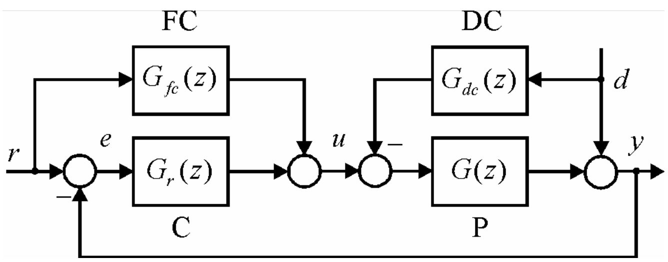


Fig. 1. Block scheme of ideal tracking system: P-plant; C-main controller; DC-disturbance compensator; FC-feedforward compensator.

According to the structure in Fig. 1, the error signal may be easily expressed in complex domain with respect to reference and disturbance :

$$E(z) = \frac{[1 - G(z)G_{fc}(z)]R(z) + [G(z)G_{dc}(z) - 1]D(z)}{1 + G_r(z)G(z)} . \quad (1)$$

Ideal tracking occurs when the tracking error is annulated $(e(k) = 0)$, which is the case when it holds

$$G_{fc}(z) = G^{-1}(z) \wedge G_{dc}(z) = G^{-1}(z) . \quad (2)$$

Hence, in order to achieve ideal tracking it is necessary that the transfer functions of the feedforward and disturbance compensators represent plant inverse dynamics. This requirement inevitably raises the following questions:

- how to obtain the information about disturbance if it is not available for direct measurement?

- how to overcome plant parameters uncertainty and variations as well as unmodeled dynamics.?

- how to realize plant inverse dynamics, since it is not a causal system?

The answers to these questions are offered in the following consideration.

A. Disturbances Estimation and Compensation

Information about external disturbances is practically impossible to obtain by direct measurement. Therefore it is necessary to estimate the disturbance for its compensation. One possible structure for disturbance estimation is presented in Fig. 2a. In this digital realization extraction of the equivalent disturbance $q(k)$ is done using the nominal plant model $G_n(z)$. Mismatch between the nominal model and real plant inevitably exists due to parameter variations and unmodeled dynamics. Hence, plant dynamics may be described as

$$G(z) = G_n(z)(1 + \delta G(z)) , \quad (3)$$

where perturbations are limited by the multiplicative bound of uncertainty $\left|\delta G(e^{j\omega T})\right| \le \gamma(\omega),\ \omega \in [0, \pi / T]$. The extracted equivalent disturbance is obtained in the form

$$q(k) = d(k) + G_n(z)\delta G(z)u_k(k) , \quad (4)$$

indicating that the equivalent disturbance carries information about external disturbance, which can always be mapped onto plant output, and parameter perturbations and unmodeled dynamics, i.e. internal disturbances.

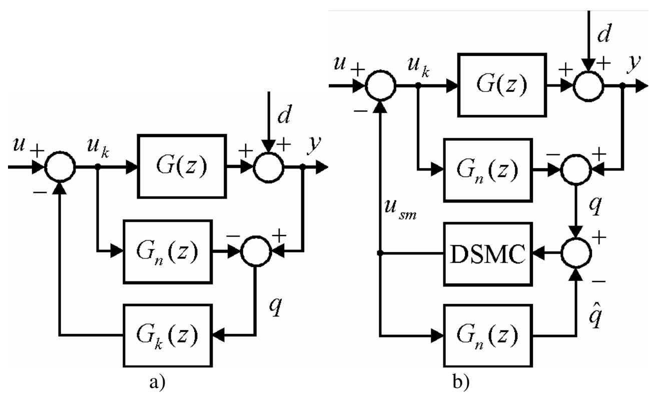


Fig. 2. a) Disturbance estimator; b) ADE based on DSMC.

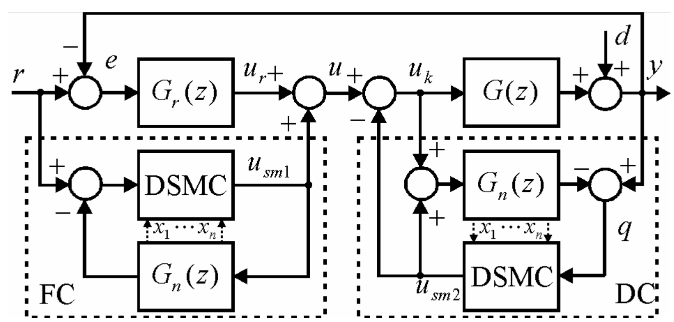


Fig. 3. The proposed control extension of the conventional system.

According to Fig. 2a, plant output as a function of control and disturbance may be expressed as

$$Y(z) = \frac{G_n(z)(1 + \delta G(z))}{1 + G_k(z)G_n(z)\delta G(z)} U(z) + \frac{1 - G_k(z)G_n(z)}{1 + G_k(z)G_n(z)\delta G(z)} D(z) \quad (5)$$

If the compensation filter $G_k(z)$ represents nominal plant inverse dynamics, i.e. $G_k(z) = G_n^{-1}(z)$, output becomes $U(z) = G_n(z)U(z)$, which shows that all disturbances are completely eliminated and that the nominal plant behavior is ensured. Unfortunately, such filter is non-causal and cannot be realized.

Solution is proposed in [9] through the concept of ADE, Fig. 2b, where passive filter is replaced by an actively controlled subsystem. If DSM controller within ADE provides $\hat{q}(k) = q(k)$, i.e. ensures an ideal DSM regime, then controller output may be described as $U_{sm}(k) = G_n^{-1}(z)Q(z)$, showing that this subsystem acts as nominal plant inverse dynamics. Thus, complete disturbance rejection is achieved nominal plant behavior is secured. This way transforms the disturbance compensation problem into tracking problem of the referent signal $q(k)$. In the tracking subsystem of ADE, DSM controller governs nominal model, not the real plant, so there are not any uncertainties and all state coordinates are available. Generally, due to the not known in advance referent signal $q(k)$ it is possible to establish only quasi-sliding regime [15], resulting in nonideal disturbance rejection. However, since DSMC systems provide high-accuracy tracking, an excellent compensation may be expected, i.e. near nominal behavior.

B. Active Compensators

The notion from ADE may also be used in realization of inverse dynamics that is required by the feedforward compensator FC in Fig. 1, which should improve reference tracking capacity of the system. Theoretically designed structure in Fig. 1 may be practically realized as shown in Fig. 3.

Disturbance compensator DC is actually an observer variant of ADE, which is formed by optimization of the structure in Fig. 2b. If DSM controller within FC establishes discrete sliding mode, then it holds $U_{sm1}(z) = G_n^{-1}(z)R(z)$, which shows that FC acts as nominal model inverse dynamics. Since DC ensures plant nominal behavior, the resulting system exhibits ideal tracking of arbitrary referent signal. This control extension is suitable for the upgrade of already existing systems, whose main conventional controller G_r, usually PD type, gives modest performance in tracking of complex signals in the presence of disturbances. Retuning of the main controller is not necessary, since the proposed control structure requires the main controller to be designed for the nominal plant, which is already the case in the practice.

Although it is known that SMC systems need measurement of all state coordinates, such system expansion does not require any additional sensors. Only input/output measurements are required, since ACs contain nominal models, which provide necessary state information for the DSM controllers. Stability of the overall system is secured by the occurrence and existence of the sliding regimes in the DSMC subsystems.

III. DSMC DESIGN

Both DSM controllers within KP and PK govern the nominal model and may be identical. The priority is to ensure as accurate tracking as possible in order to gain the precise nominal model inverse dynamics. Good results were obtained using DSMC algorithm [14], which guaranteed ideal tracking of parabolic signals. This control algorithm is based on the algorithm [16] enriched with the introduction of additional integral action with respect to the switching variable. Integration is activated only in the predefined vicinity of the sliding surface. Emergence of chattering, extremely undesirable phenomenon in SMC systems, is eliminated by imposing a linear control zone near the sliding surface [16]. Since the exposed control algorithm has been thoroughly elaborated in [14] and [13], it will be briefly described hereafter.

Let the nominal plant model be in the form

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} u_{sm} \Rightarrow \dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u_{sm},\ \hat{q} = x_1. \quad (6)$$

It is a second order model representing the mechanical subsystem dynamics of an electromechanical positional system. The dynamics of the electrical part, i.e. electric drive, is neglected since it is much faster than the mechanical counterpart. Tracking error may be calculated as $e(t) = q(t) - \hat{q}(t) = q(t) - x_1(t)$. The model (6) may be transformed into the tracking error space

$$\dot{\mathbf{e}} = \mathbf{A}\mathbf{e} - \mathbf{b}u_{sm} + \mathbf{p},\ \mathbf{e} = \begin{bmatrix} e_1 = e \\ e_2 = \dot{e} \end{bmatrix}, \mathbf{p} = \begin{bmatrix} 0 \\ a\dot{q} + \ddot{q} \end{bmatrix}. \quad (7)$$

Unlike the previous model, a disturbance vector **p** occurs in this model as a consequence of variability of the input signal q. The first component $a\dot{q}$ scan be easily compensated since forming of e_2 already needs knowledge of $\dot{q}$, which may be obtained using a differentiator. However, $\ddot{q}$ cannot be reliably obtained by twofold differentiation due to drastic amplification of noises. Hence, vector $\mathbf{p} = [0\ \ddot{q}]^{\mathrm{T}}$ may be regarded as a disturbance vector. The discrete-time model of the system (7) for the given sampling period T is obtained in the form

$$\begin{aligned} &\mathbf{e}(k+1) = e(k) + T\mathbf{A}_\delta \mathbf{e}(k) - T\mathbf{b}_\delta u_{sm}(k) + T\mathbf{d}_\delta(k), \\ &\mathbf{A}_\delta = (\mathbf{A}_d - \mathbf{I})/T, \mathbf{b}_\delta = \mathbf{b}_d/T, \mathbf{d}_\delta = \mathbf{d}_d/T \\ &\mathbf{A}_d = e^{\mathbf{A}T}; \mathbf{b}_d = \int_0^T \mathbf{e}^{\mathbf{A}t}\mathrm{d}t\mathbf{b}; \mathbf{d}_d = \int_0^T \mathbf{e}^{\mathbf{A}t}\mathbf{p}(kT + T\text{-}t)\mathrm{d}t. \end{aligned} \quad (8)$$

Control task is to annul the tracking error, i.e. the trajectories of the system (8) should reach state space origin. Using the concept of DSMC this would mean that system trajectories from an arbitrary initial point should reach in finite time the sliding line *s(k)*=0, defined by the switching function

$$s(k) = \mathbf{c}_\delta \mathbf{e}(k), \mathbf{c}_\delta = [c_{\delta 1}\ \ c_{\delta 2}], \quad (9)$$

and continue to slide along the line into the origin, which would result in ideal tracking. System dynamics in the sliding mode is strictly defined by the sliding line vector $\mathbf{c}_\delta$, which should be chosen according to the desired dynamics.

Sliding line reaching dynamics in [16] is proposed as

$$s(k+1)-s(k)=-\Phi(s),\ \Phi(s)=\min\{|s(k)|,\sigma T\}\operatorname{sgn}(s(k)) \quad (10)$$

which is accomplished according to (9) and (8) by the following control

$$u_{sm}(k)=\mathbf{c}_\delta \mathbf{A}_\delta \mathbf{e}(k)+\mathbf{c}_\delta \mathbf{d}_\delta(k)+T^{-1}\Phi(s) \quad (11)$$

under normalization $\mathbf{c}_\delta \mathbf{b}_\delta = 1$. This control is not feasible due to unknown $\mathbf{d}_\delta$, that is $\ddot{q}$. A feasible control

$$u_{sm}(k)=\mathbf{c}_\delta \mathbf{A}_\delta \mathbf{e}(k)+T^{-1}\min\{|s(k)|,\sigma T\}\operatorname{sgn}(s(k)) \quad (12)$$

gives the following reaching dynamics.

$$s(k+1)-s(k)=-\min\{|s(k)|,\sigma T\}\operatorname{sgn}(s(k))+T\mathbf{c}_\delta \mathbf{d}_\delta(k)\,. \quad (13)$$

It is evident that control (12) has two modes: nonlinear and linear. Nonlinear control

$$u_{sm\_n}(k)=\mathbf{c}_\delta \mathbf{A}_\delta \mathbf{e}(k)+\sigma \operatorname{sgn}(s(k)) \quad (14)$$

acts outside the zone $|s(k)|<\sigma T$, which produces the reaching dynamics given by

$$s(k+1)-s(k)=-\sigma T\operatorname{sgn}(s(k))+T\mathbf{c}_\delta \mathbf{d}_\delta(k)\,. \quad (15)$$

To ensure the reaching of the sliding line, the condition $[s(k+1)-s(k)]s(k)<0$ must be satisfied. Under assumption that the reference is a smooth function, its second derivative is bounded $|\ddot{q}|\le M_r$ and therefore the disturbance is also bounded $\|\mathbf{d}_\delta\|\le M$. Reaching is secured if the switching gain σ fulfills inequality $\sigma > \|\mathbf{c}_\delta\| M$. It means that the system trajectories will enter zone $|s(k)|<\sigma T$ in finite number of steps.

Inside this zone, the control signal is linear

$$u_{sm\_l}(k)=\mathbf{c}_\delta \mathbf{A}_\delta \mathbf{e}(k)+T^{-1}s(k)\,, \quad (16)$$

which provides $s(k+1)=T\mathbf{c}_\delta \mathbf{d}_\delta(k)$, indicating that a quasi-sliding mode arises in a single step within a domain described by

$$S_{qs}=\{\mathbf{e}\,\big|\,|s(\mathbf{e})|\le T\,\|\mathbf{c}_\delta\|\,M\}\,. \quad (17)$$

For small sampling periods T the width of the quasi-sliding domain is also small, which guarantees high-precision tracking. If the referent signal is q is a constant or a ramp function, the second component of the disturbance is zero, $\ddot{q}=0$, which gives $\mathbf{p}=\mathbf{b}_d=\mathbf{b}_\delta=M=0$. It yields $s(k+1)=0$, that is, ideal discrete sliding mode occurs in one step that provides ideal tracking of ramp references. The control (16) is the so called equivalent control u_{eq}.

Further tracking improvement was suggested in [14] by introduction of the supplemental integral action with respect to switching variable *s(k)*. Namely, integration is activated only inside the linear control zone, only when the tracking error is small, i.e. in the sliding mode final stage. Activation of the integral action within the nonlinear control zone or distant from the origin is completely unnecessary and may can produce an unwanted overshoot. This idea is described by the following expression

$$u_I(k)=\begin{cases}0, & \|\mathbf{e}(k)\|>\rho>0,\\ hs(k)+u_I(k-1), & \|\mathbf{e}(k)\|\le\rho,\end{cases} \quad (18)$$

where ρ is a small positive constant. Integral gain h should satisfied condition $0<h<1/T$ to preserve system stability.

The resulting control signal as an output of the designed DSM controller, created by merging the described control components, is summarized by

$$u_{sm}(k)=\begin{cases}u_{sm\_n}(k), & |s(k)|>\sigma T,\\ u_{sm\_l}(k), & |s(k)|\le\sigma T,\\ u_{sm\_l}(k)+u_I(k), & |s(k)|<\sigma T\cap\|\mathbf{e}(k)\|<\rho.\end{cases} \quad (19)$$

Because of the introduced additional integral action, which increases tracking accuracy, the designed DSM controller provides ideal tracking of parabolic signals.

Since the main controller is already tuned by some conventional method, it remains to define AC sliding mode dynamics, which is prescribed by the selection of vector $\mathbf{c}_\delta$. In case of a second order system in sliding mode, due to the order reduction of the differential equation that describes sliding mode dynamics, a single eigenvalue $z_1=e^{-\alpha T}$ determines desired system dynamics. The desired slope α of the sliding line is established if it holds

$$\mathbf{c}_\delta \mathbf{b}_\delta=1 \quad\wedge\quad c_{\delta 1}/c_{\delta 2}=\alpha\,. \quad (20)$$

The procedure for the calculation of vector $\mathbf{c}_\delta$ in case of higher order systems is given in [16].

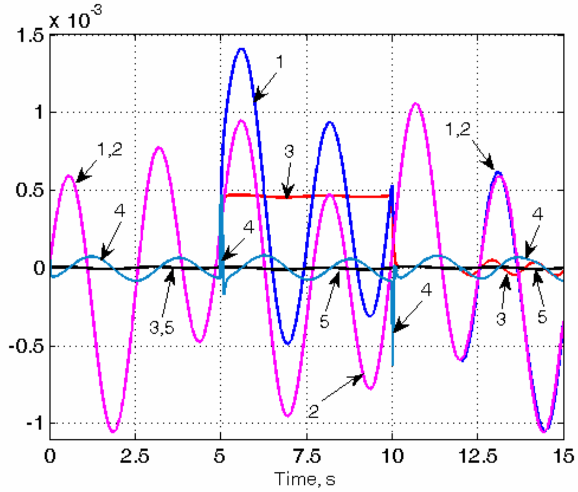
IV. Simulation Example

Permanent magnet direct current motor is considered as a plant, whose nominal mode is given by

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}0 & 1\\ 0 & -26.5\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}+\begin{bmatrix}0\\ 654\end{bmatrix}u+\begin{bmatrix}0\\ 1\end{bmatrix}f,\quad y=x_1\,. \quad (21)$$

Sampling period is $T=0.4$ ms. Parameters of the DSM controllers within ACs are set as: $\alpha=50$; $h=100$ in DC and $h=1000$ in FC; $\sigma=10$; $\rho=0.01$; $\mathbf{c}_\delta=-[7.557248\ 0.151449]\cdot 10^{-2}$. Differentiator [12] is employed to obtain the derivatives of the input signals r and q. The main controller is a PD controller that is tuned by the following selection of the well-known parameters: $K_r=25$ and $T_d=1/26.5$ s. The input signal that represents angular position reference is described by $r(t)=5[\cos(t)-\cos(2.5t)]$. Load torque that acts as an external disturbance is expressed by $f(t)=200[h(t-5)-h(t-10)]+$ $+20\sin(5t)h(t-12)$, where $h(t)$ represents step function. Tracking errors obtained by simulations in case of different configurations are given in Fig. 4.

Fig. 4 shows that the performance of the main controller only (line 1) is unsatisfactory both with respect to reference tracking and disturbance rejection. Activation of active DC (line 2) eliminates disturbances whereas tracking remains unchanged. Only the inclusion of active FC (line 5), which ensures almost ideal tracking, results in a superior response comparing with the conventional structure.



Line	Main controller	DC	FC
1	On	Off	Off
2	On	On.	Off
3	On	Off	On
4	Off	On	On
5	On	On	On

Fig. 4. Tracking errors of the proposed system for various configurations.

V. CONCLUSION

The paper proposes a way to upgrade conventional servo-systems by introduction of digital ACs. Adjoining of DC and FC improves system performances in reference tracking as well as disturbance rejection. Both compensators contain active DSM controlled subsystems, whose controllers are designed for the nominal plant. Simulation results evidently show that the proposed control extension ensures superior performance comparing to the initial system, confirming that a conventional positioning system becomes a robust high-performance tracking system.

REFERENCES

[1] T. Umeno, T. Hory, "Robust speed control of DC servomotors using modern two degrees-of-freedom controller design," IEEE Trans. Ind. Electron., Vol. 38, No. 5, pp. 363-368, 1991.
[2] T. Umeno, T. Kaneko, Y. Hory, "Robust servosystem design with two degrees of freedom and its application to novel motion control of robot manipulators, " IEEE Trans. Ind. Electron., Vol. 40, No. 5, pp. 473-485, 1993.
[3] Y. Fujimoto, A. Kawamura, "Robust servo-system based on two-degree-of-freedom control with sliding mode", IEEE Trans. Ind. Electron., Vol. 42, No. 3, pp. 272-280, 1995.
[4] H. Fujimoto, Y. Hori, A. Kawamura, "Perfect tracking control based on multirate feedforward control with generalized sampling periods," IEEE Trans. Ind. Electron., Vol. 48, No. 3, pp. 636-644, 2001.
[5] H. Fujimoto, Y. Hori, "High-performance servo systems based on multirate sampling control," Control Engineering Practice, Vol. 10, pp. 773-781, 2002.
[6] V.I. Utkin, Sliding modes in control and optimization, Springer-Verlag, 1992.
[7] B. Draženović, "The invariance conditions in variable structure systems", Automatica, Vol. 5, pp. 287-295, 1969.
[8] Č. Milosavljević, N. Mihajlović, G. Golo, "Static accuracy of the variable structure system," Proc. 6th Int. SAUM Conference on Systems Automatic Control and Measurements, Niš, Yugoslavia, pp. 464–469, 1998.
[9] B.Veselić, Č. Milosavljević, B. Peruničić-Draženović, D. Mitić, "Digitally controlled sliding mode based servo-system with active disturbance estimator," Proc. IEEE 9th Int. Workshop on Variable Structure Systems (VSS2006), Algero, Italy, pp. 51-56, 2006.
[10] A. Damiano, G.L. Gatto, I. Marongiu, A. Pisano, "Second-order sliding-mode control of DC drives", IEEE Trans. Ind. Electron., Vol. 51, No. 2, pp. 364-373, 2004.
[11] A. Levant, "Robust exact differentiation via sliding mode technique", Automatica, Vol. 34, No. 3, pp. 379-384, 1998.
[12] J.X. Xu, J. Xu, R. Yan, "On sliding mode derivative estimators via closed-loop filtering," Proc. IEEE 8th Int. Workshop on Variable Structure Systems (VSS2004), Vilanova i la Geltru, Spain, 2004.
[13] B. Veselić, B. Peruničić-Draženović, Č. Milosavljević, "High-performance position control of induction motor using discrete-time sliding-mode control," IEEE Trans. Ind. Electron., Vol. 55, No. 11, pp. 3809-3817, 2008.
[14] Č. Milosavljević, B. Peruničić-Draženović, B. Veselić, D. Mitić, "A new design of servomechanisms with digital sliding mode," Electrical Engineering, Vol. 89, No. 3, pp. 233-244, 2007.
[15] Č. Milosavljević "General conditions for the existence of a quasi-sliding mode on the switching hyperplane in discrete variable structure systems," Automatic Remote Control, Vol 46, pp. 307-314, 1985.
[16] G. Golo, Č. Milosavljević, "Robust discrete-time chattering free sliding mode control," Systems & Control Letters, Vol. 41, pp. 19-28, 2000.

Efficiency Optimized Control of High Performance Induction Motor Drive

Branko D. Blanuša, Branko L. Dokić, and Slobodan N. Vukosavić

***Abstract*—Algorithms for efficiency optimized control of induction motor drives are presented in this paper. As a result, power and energy losses are reduced, especially when load torque is significant less compared to its rated value. According to the literture, there are three strategies for dealing with the problem of efficiency optimization of the induction motor drive: Simple State Control, Loss Model Control and Search Control. Basic characteristics each of these algorithms and their implementation in induction motor drives are described. Moreover, induction motor drive is often used in a high performance applications. Vector Control or Direct Torque Control are the most commonly used control techniques in these applications. These control methods enable software implementation of different algorithms for efficiency improvement. Simulation and experimental tests for some algorithms are performed and results are presented.**

***Index Terms*—Induction motor, efficiency optimization, dynamic programming.**

I. Introduction

UNDOUBTEDLY, the induction motor is a widely used electrical motor and a great energy consumer. Three-phase induction motors consume more than 60% of industrial electricity and it takes a lot of effort to improve their efficiency [1]. The vast majority of induction motor drives are used for heating, ventilation and air conditioning (HVAC). These applications require only low dynamic performance and in most cases only voltage source inverter is inserted between grid and induction motor as cheapest solution. The classical way to control these dives is constant V/f ratio and simple methods for efficiency optimization can be applied [2]. From the other side in applications like electric vehicle energy has to be consumed in the best possible way and use of induction motors in such application requires an energy optimized control strategy [3]. Also, there are many high performance industrial drives which operate in periodic cycles. In these cases implementation of efficiency optimization algorithm are more complex.

In a conventional setting, the field excitation is kept constant at rated value throughout its entire load range. If machine is under-loaded, this would result in over-excitation and unnecessary copper losses. Thus in cases where a motor drive has to operate in wider load range, the minimization of losses has great significance. It is known that efficiency improvement of induction motor drive (*IMD*) can be implemented via motor flux level and this method has been proven to be particularly effective at light loads and in a steady state of drive. Moreover, induction motor drive is often used in servo drive applications. Vector Control (*VC*) or Direct Torque Control (*DTC*) are the most commonly used control techniques in such applications and these methods enable software implementation of different algorithms for efficiency improvement.

Functional approximation of the power losses in the vector controlled induction motor drive is given in the second Section. Strategies for efficiency optimization of IMD and their basic characteristics are described in third section. Qualitative analysis and comparison of interesting algorithms for efficiency optimization with simulation and experimental results are presented in fourth section. Brief description of efficiency optimized control of high performance IMD is described in fifth section.

B. D. Blanuša and B. L. Dokić are with the Faculty of Electrical Engineering, University of Banja Luka, Banja Luka, Bosnia and Herzegovina.

S. N. Vukosavić is with the Faculty of Electrical Engineering, University of Belgrade, Belgrade, Serbia.

II. Functional Approximation of the Power Losses in the Induction Motor Drive

The process of energy conversion within motor drive converter and motor leads to the power losses in the motor windings and magnetic circuit as well as conduction and commutation losses in the inverter [4].

Converter losses: Main constituents of converter losses are the rectifier, DC link and inverter conductive and inverter commutation losses. Rectifier and DC link inverter losses are proportional to output power, so the overall flux-dependent losses are inverter losses. These are usually given by:

$$P_{INV} = R_{INV} \cdot i_s^2 = R_{INV} \cdot \left(i_d^2 + i_q^2\right), \qquad (1)$$

where i_d, i_q are components of the stator current i_s in d,q rotational system and R_{INV} is inverter loss coefficient.

Motor losses: These losses consist of hysteresis and eddy current losses in the magnetic circuit (core losses), losses in the stator and rotor conductors (copper losses) and stray losses. The main core losses can be modeled by:

$$P_{Fe} = c_h \Psi_m^2 \omega_e + c_e \Psi_m^2 \omega_e^2, \qquad (2)$$

where ψ_d is magnetizing flux, ω_e supply frequency, c_h is hysteresis and c_e eddy current core loss coefficient.

Copper losses are due to flow of the electric current through the stator and rotor windings and these are given by:

$$p_{Cu} = R_s i_s^2 + R_r i_q^2 , \tag{3}$$

The stray flux losses depend on the form of stator and rotor slots and are frequency and load dependent. The total secondary losses (stray flux, skin effect and shaft stray losses) usually don't exceed 5% of the overall losses [4].

III. Strategies for Efficiency Optimization of IMD

Numerous scientific papers on the problem of loss reduction in *IMD* have been published in the last 20 years. Although good results have been achieved, there is still no generally accepted method for loss minimization. According to the literature, there are three strategies for dealing with the problem of efficiency optimization of the induction motor drive [5]:

- Simple State Control - SSC ,
- Loss Model Control - LMC and
- Search Control- SC.

The first strategy is based on the control of one of the variables in the drive [5-7] (Fig. 1). This variable must be measured or estimated and its value is used in the feedback control of the drive, with the aim of running the motor by predefined reference value. Slip frequency or power factor displacement are the most often used variables in this control strategy. Which one to chose depends on which measurement signals is available [5]. This strategy is simple, but gives good results only for a narrow set of operation conditions. Also, it is sensitive to parameter changes in the drive due to temperature changes and magnetic circuit saturation.

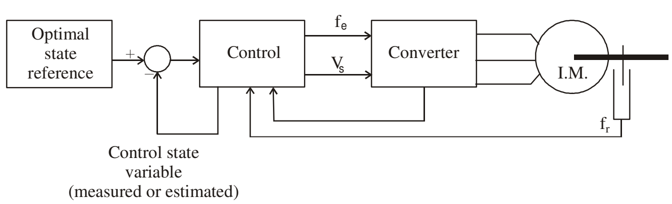


Fig. 1. Control diagram for the simple state efficiency.

In the second strategy, a drive loss model is used for optimal drive control [4,8] (Fig. 2). These algorithms are fast because the optimal control is calculated directly from the loss model. But, power loss modeling and calculation of the optimal operating conditions can be very complex. This strategy is also sensitive to parameter variations in the drive.

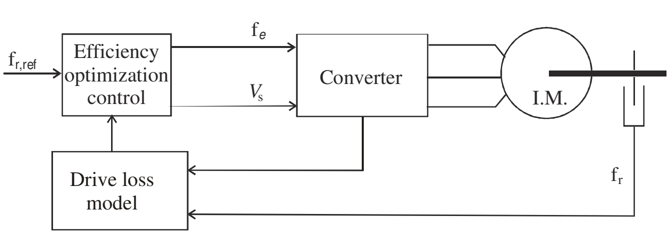


Fig. 2. Block diagram for the model based control strategy.

In the search strategy, the on-line procedure for efficiency optimization is carried out [9-11] (Fig. 3). The optimization variable, stator or rotor flux, increases or decreases step by step until the measured input power is at a minimum. This strategy has an important advantage over others: it is insensitive to parameter changes.

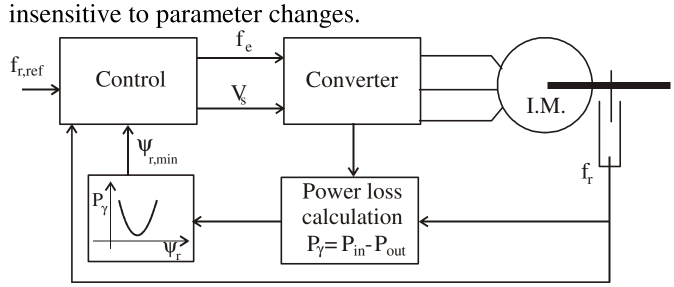


Fig. 3. Block diagram of search control strategy.

Also, there are hybrid methods which include good characteristics of different strategies for efficiency improvement [10].

The published methods mainly solve the problem of efficiency improvement for constant output power. Results of applied algorithms highly depends from the size of drive (Fig. 4) [2] and operating conditions, especially load torque and speed (Figs. 5 and 6). Efficiency of IM changes from 75% for low power 0,75kW machine to more than 95% for 100kW machine. Also efficiency of drive converter is typically 95% and more.

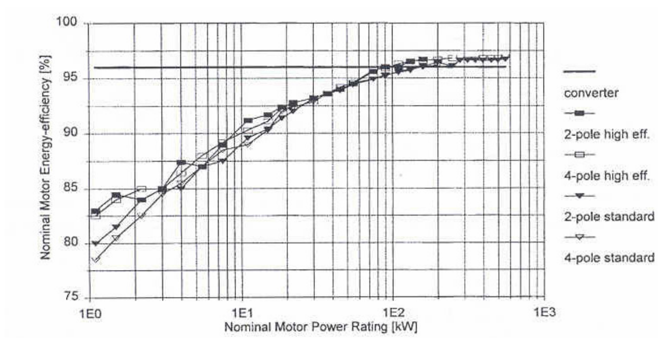


Fig. 4. Rated motor efficiences for ABB motors (catalog data) and typical converter efficiency.

That's obvious, converter losses is not necessary to consider in efficiency optimal control for small drives. Also, these algorithms for efficiency optimization give best result in power losses reduction for a light loads and in a steady state.

IV. Comparison of Some Algorithms for Efficiency Optimization of IMD

Selection of algorithm for efficiency optimization depends from many factors, drive features, operating conditions, measuring signals, drive control and etc.

If the losses in the drive were known exactly, it would be possible to calculate the optimal operating point and control of drive in accordance to that. For the following reasons it is not possible in practice [9]:

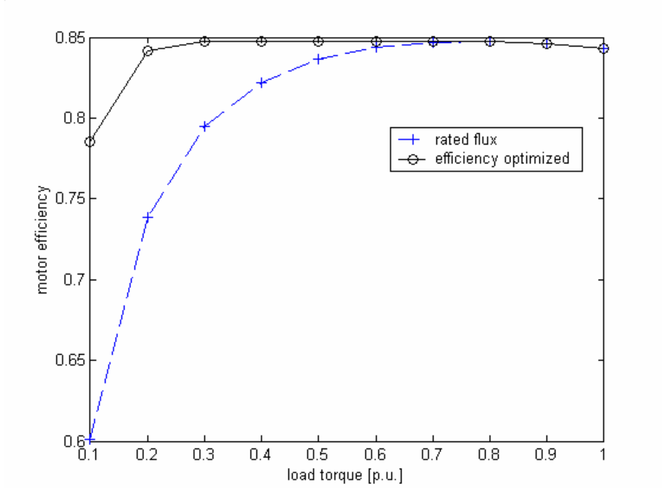


Fig. 5. Measured standard motor efficiencies with both rated flux and efficiency optimized control at rated mechanical speed (2.2 kW rated power).

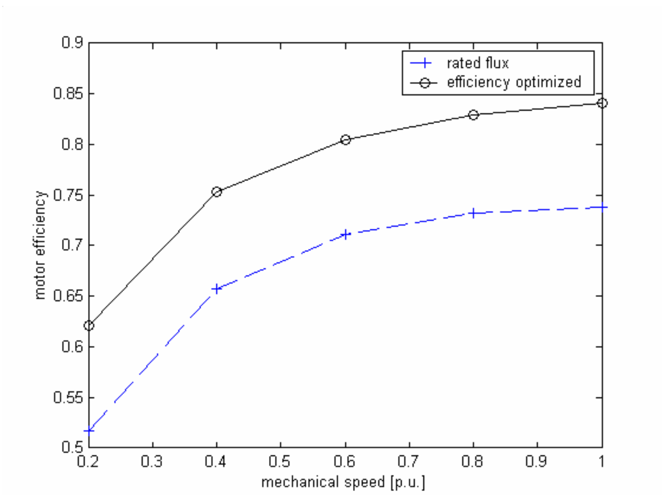


Fig. 6. Measured standard motor efficiencies with both rated flux and efficiency optimized control at light load (20% of rated load).

– A number of fundamental losses are difficult to predict: stray load, iron losses in case of saturation changes, copper losses because of temperature rise etc.
– Due to limitation in costs all the measurable signals cannot be acquired. It means that certain quantities must be estimated which naturally leads to an error.

Two interesting algorithms *SC* and *LMC* are discussed and their results in efficiency optimization are compared for different operating conditions. Also, power losses for these algorithms are presented together with a case when motor is excited by rated magnetizing flux. Operation of drive has been tested under following operating conditions. There are three intervals: acceleration from 0 to ω_{ref}, interval [0, t_1], constant speed $\omega=\omega_{ref}$, interval [t_1, t_2], deceleration from ω_{ref} to 0, interval [t_2, t_3]. Load torque changes at the moment t_4=5s from 0.4 p.u. to 1.05 p.u. and vice versa at the moment t_5=10s for a constant reference speed of ω_{ref}=0.6 p.u. (Fig.7). The steep change of load torque appears with the aim of testing the drive behavior in the dynamic mode and its robustness within sudden load perturbations.

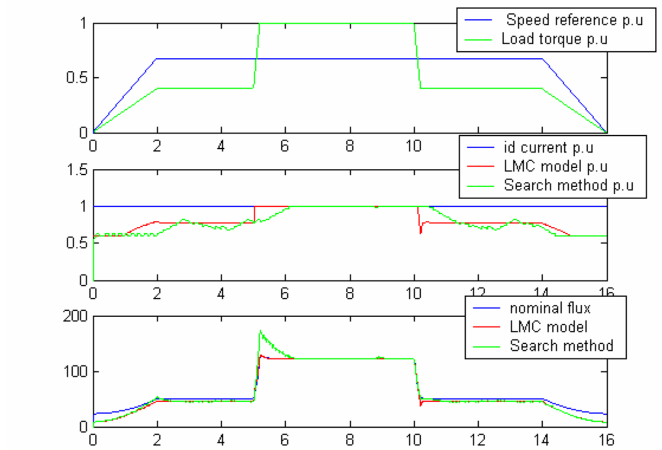


Fig. 7. Comparison of SC and LMC algorithms for efficiency optimization in IMD.

Simulation tests show that LMC algorithm is faster than SC algorithm and gives better result in power loss reduction than SC algorithm. Optimal magnetizing flux is derived directly from the loss model of IMD. Loss modeling, optimal flux calculation and especially its sensitivity to parameter changes are problems which limits implementation of this control strategy. But LMC algorithm with on-line parameter identification in the loss model and hybrid models make this strategy very actual [4].

From the other side search strategy optimization does not require the knowledge of motor parameters and the algorithm is applicable universally to any motor. Besides all good characteristics of search strategy methods, there is an outstanding problem in its use. Flux in small steps oscillates around its optimal value. Torque ripple appears each time the flux is stepped. Sometimes convergence to its optimal value is to slow, so these methods are not applicable for high performance drives. There are numerous papers which treats problem of step size in the magnetization flux for SC algorithms. Fuzzy or neuro-fuzzy controllers are often used to obtain smooth and fast flux convergence during optimization process [9,10].

V. Efficiency Optimization in Dynamic Operation

There is an interesting question to ask, how algorithms for efficiency optimization can be applied in the dynamic mode and what are problems and constrains. There are two distinctive cases: when the operation conditions are not known in advance and when they are.

In the cases when the operating conditions are not known in advance (e.g. electrical vehicles, cranes, etc.), it is important to watch for the electromagnetic torque margin and energy saving presents a compromise between power loss reduction and dynamic performances of the drive [12].

There are two common approaches when operation conditions are known in advance:

a) Steady state modified [13,14] and

b) Dynamic programming [13-15].

In the first case, the same methods, LMC or SC controllers, are used for steady state as well. Magnetizing flux is set to its nominal value during the dynamic transition [13], or a fuzzy controller is used to adjust the flux level in a machine by operation conditions [13, 14]. This can be realized in cases when torque or speed response is not so important (e.g. elevators or cranes).

If the both high dynamic performance and losses minimization are required dynamic optimization is necessary. By using the dynamic programming approach, optimal control is computed so that the drive runs with minimal losses. Torque and speed trajectories have to be known in advance and flux trajectory has to be computed off-line, which requires a lot of processing time.

Also, an interesting problem is how to minimize energy consumption of IMD when it works in a periodic cycles. Closed-cycle operation is often for robots and other high performance industry machines. Efficiency optimized control for closed-cycle operation of high performance IMD, based on dynamic programming approach is applied.

Following dynamic programming approach, performance index, system equations, constraints and boundary conditions for a vector controlled IMD in the rotor flux oriented reference frame, can be defined as follows:

a) The performance index is [12, 16]:

$$J=\sum_{i=0}^{N-1}\left[ai_d^2(i)+bi_q^2(i)+c_1\omega_e(i)\psi_D^2(i)+c_2\omega_e^2(i)\psi_D^2(i)\right], \quad (1)$$

where i_d, i_q: d and q are components of the stator current vector, ψ_D is rotor flux and ω_e is supply frequency. The a, b, c_1 and c_2 are parameters in the loss model of the drive. These parameters are determined through the process of parameter identification [4,12]. Rotor speed ω_r and electromagnetic torque T_{em} are defined by operating conditions (speed reference, load and friction).

b) The dynamics of the rotor flux can be described by the following equation:

$$\psi_D(i+1)=\psi_D(i)\left(1-\frac{T_s}{T_r}\right)+\frac{T_s}{T_r}L_m i_d(i), \quad (2)$$

where $T_r=L_r/R_r$ is a rotor time constant.

c) Constraints:

$$\begin{aligned}
&ki_d(i)i_q(i)=T_{em}(i),\ k=\frac{3}{2}\frac{p}{2}\frac{L_m^2}{L_r},\ (for\ torque)\\
&i_d{}^2(i)+i_q{}^2(i)-I_{s\max}^2\le 0,\ (for\ stator\ current)\\
&-\omega_{rn}\le\omega_r\le\omega_{rn},\ (for\ speed)\\
&\psi_D(i)-\psi_{Dn}\le 0,\ (for\ rotor\ flux)\\
&\psi_{D\min}-\psi_D(i)\le 0.
\end{aligned} \quad (3)$$

I_{smax} is maximal amplitude of stator current, ω_{rn} is nominal rotor speed, p is number of poles, Ψ_{Dmin} is minimal and Ψ_{Dn} is nominal value of rotor flux.

Also, there are constraints on stator voltage:

$$0\le\sqrt{v_d^2+v_q^2}\le V_{s\max}, \quad (4)$$

where v_d and v_q are components of stator voltage and V_{smax} is maximal amplitude of stator voltage. Voltage constraints are more expressed in *DTC* than in field-oriented vector control.

d) Boundary conditions:

Basically, this is a boundary-value problem between two points which are defined by starting and final value of state variables:

$$\begin{aligned}
&\omega_r(0)=\omega_r(N)=0,\\
&T_{em}(0)=T_{em}(N)=0,\\
&\psi_{Dn}(0)=\psi_{Dn}(N)=free,\\
&considering\ constrains\ in\ (3)
\end{aligned} \quad (5)$$

Presence of state and control variables constrains generally complicates derivation of optimal control law. On the other side, these constrains reduce the range of values to be searched and simplify the size of computation [17].

In a purpose to determine stationary state of performance index, next system of differential equations are defined:

$$\begin{aligned}
&\lambda(i)=\lambda(i+1)\frac{T_r-T_S}{T_r}+2\left(c_1\omega_e(i)+c_2\omega_e^2(i)\right)\psi_D(i)\\
&2bi_q(i)+\mu(i)ki_d(i)=0\\
&2ai_d(i)+\mu(i)ki_q(i)+\lambda(i+1)\frac{T_S}{T_r}L_m=0\\
&ki_d(i)i_q(i)=T_{em}(i),\ \omega_e(i)=\omega_r(i)+\frac{L_m}{T_r}\frac{i_q(i)}{\psi_D(i)}\\
&i=0,1,2,..,N-1,
\end{aligned} \quad (6)$$

where λ and μ are Lagrange multipliers.

By solving the system of equations (6) and including boundary conditions given in (5), we come to the following system [16]:

$$\begin{aligned}
&2ai_d^4(i)+\lambda(i+1)\frac{T_S}{T_r}i_d^3(i)=\frac{2b}{k^2}T_{em}^2(i)\\
&\psi_D(i)=\frac{T_r}{T_r-T_s}\psi_D(i+1)-\frac{T_s}{T_r-T_s}L_m i_d(i)\\
&i_q(i)=\frac{T_{em}(i)}{ki_d(i)},\omega_e(i)=\omega_r(i)+\frac{L_m}{T_r}\frac{i_q(i)}{\psi_D(i)},\\
&\lambda(i)=2\left(c_1\omega_e(i)+c_2\omega_e^2(i)\right)\psi_D(i)+\lambda(i+1)\frac{T_r-T_s}{T_r}\\
&i=0,1,2,..,N-1.
\end{aligned} \quad (7)$$

Every sample time values of $\omega_r(i)$ and $T_{em}(i)$ defined by operating conditions is used to compute the optimal control ($i_d(i)$, $i_q(i)$, i=0,..,N-1) through the iterative procedure and applying the backward procedure, from stage i =N-1 down to stage i =0. For the optimal control computation, the final value of ψ_D and λ have to be known. In this case, $\psi_D(N)=\psi_{Dmin}$ and

$$\lambda(N)=\frac{\partial\varphi}{\partial\psi_D(N)}=0. \quad (8)$$

Expressed problem in efficiency optimization methods are its sensitivity to steep increase of load or speed reference, especially for low flux level. Therefore, some experiments are made to appraise speed response on steep increase of load for LMC and optimal flux control method (Fig. 8).

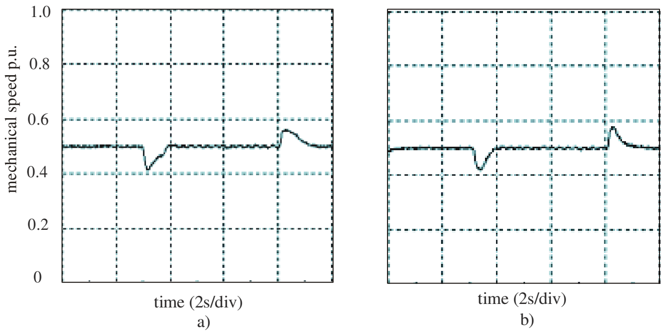


Fig. 8. Speed response on steep load change for a) LMC method, b) optimal flux.

The method for efficiency optimization based on the dynamic programming approach should show good results regarding the loss reduction during transient processes. Thus, it is very important to measure power losses in the drive for this method during the transient process and compare it with other efficiency optimization methods. The graphic of power losses for steep increase of load torque for optimal flux and LMC method is shown in Fig. 9.

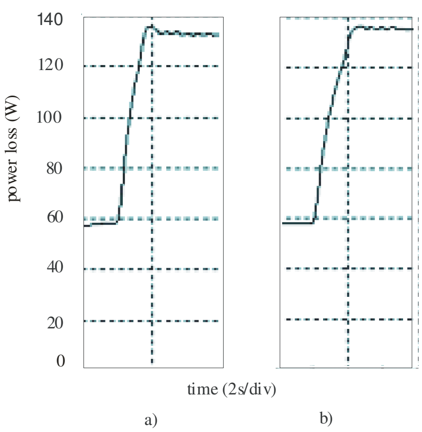


Fig. 9. Graph of power losses during dynamic operation for a) LMC method, b) optimal control.

Simulation and experimental tests are performed for typical closed-cycle operation, although this algorithm can be applied regardless of IMD operating conditions.

VI. Conclusion

Algorithms for efficiency optimization of IMD are briefly described and some comparison between LMC and SC strategies are made. Also, one procedure for efficiency optimization in dynamic operation based on dynamic programming approach has been applied. According to the performed simulations and experimental tests, we have arrived at the following conclusions:

1. If load torque has a value close to nominal or higher, magnetizing flux is also nominal regardless of whether an algorithm for efficiency optimization is applied or not. For a light load algorithm based on optimal flux control gives significiant power loss reduction when drive works with its nominal flux (Figs. 5, 6 and 7).

2. For a steady state, power losses are practically same for both methods, SC and LMC, but SC algorithms give faster convergence of magnetizing flux during transient prosess and consequently less energy consumption. (Fig. 7). From the other side SC algorithms do not require knowledge of motor parameters and not sensitive to motor parameters changes.

3. Optimal flux control based on dynamic programming gives better dynamic features and less speed drops on steep load increase, then LMC methods (Figs. 8 and 9). The obtained experimental results show that this algorithm is applicable. It offers significant loss reduction, good dynamic features and stable operation of the drive. One disadvantage of this algorithm is its off-line control computation.

References

[1] S. N. Vukosavic," Controlled electrical drives - Status of technology" Proceedings of XLII ETRAN Conference, No. 1, pp. 3-16, June 1998.

[2] F. Abrahamsen, F. Blaabjerg, J.K. Pedersen, P.Z. Grabowski and P. Thorgensen," On the Energy Optimized Control of Standard and High Efficiency Induction Motors in CT and HVAC Applications", IEEE Transaction on Industry Applications, Vol.34, No.4, pp.822-831, 1998.

[3] M.Chis, S. Jayaram, R. Ramshaw, K. Rajashekara:" Neural network based efficiency optimization of EV drive", IEEE-IECIN Conference Record, pp. 454-457, 1997.

[4] S.N. Vukosavic, E Levi:"Robust DSP-based efficiency optimization of variable speed induction motor drive", IEEE Transaction of Ind. Electronics, Vol.50, No.3, pp. 560-570, 2003.

[5] F. Abrahamsen, J.K. Pedersen and F. Blaabjerg: "State-of-Art of optimal efficiency control of low cost induction motor drives" Proceedings of PESC'96, pp. 920-924, 1996.

[6] T. Hatanaka, N. Kuwahara: Method and apparatus for controlling the supply of power to an induction motor to maintaining in high efficiency under varying load conditions, U.S. Patent 5 241 256, 1993.

[7] M.E.H. Benbouzid and N.S. Nait Said, "An efficiency-optimization controller for induction motor drives", IEEE Power Engineering Review, Vol. 18, Issue 5, pp. 63 –64, 1998.

[8] F. Fernandez-Bernal, A. Garcia-Cerrada and R. Faure: "Model-based loss minimization for DC and AC vector-controlled motors including core saturation", IEEE Transactions on Industry Applications, Vol. 36, No. 3, pp. 755 -763, 2000.

[9] G. C. D. Sousa, B. K. Bose, J. G. Cleland, "Fuzzy Logic Based On-Line Efficiency Optimization of an Indirect Vector-Controlled Induction Motor Drive", IEEE Trans. Ind. Elec., Vol.42, No.2, 1995.

[10] D.A. Sousa, Wilson C.P. de Aragao and G.C.D. Sousa: "Adaptive Fuzzy Controller for Efficiency Optimization of Induction Motors", IEEE Transaction on Industrial Electronics, Vol. 54, No.4, pp. 2157-2164, 2007.

[11] Ghozzy S., Jelassi K., Roboam X.:" Energy optimization of induction motor drive". International Conference on Industrial Technology, Conference Record of the 2004 IEEE, pp. 1662 -1669, 2004.

[12] B. Blanusa, P. Matić, Z. Ivanovic and S.N. Vukosavic: " An Improved Loss Model Based Algorithm for Efficiency Optimization of the Induction Motor Drive", Electronics, Vol.10, No.1, pp. 49-52, 2006.

[13] E. Mendes, A.Baba, A. Razek:" Losses minimization of a field oriented controlled induction machine", Electrical Machines and Drives, Seventh International Conference on (Conf. Publ. No. 412), pp. 310 -314, 1995.

[14] J. Moreno, M. Cipolla, J. Peracaula, P.J. Da Costa Branco:"Fuzzy logic based improvements in efficiency optimization of induction motor drives" , Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, Vol. 1, pp. 219 -224, 1997.

[15] R. D. Lorenz, S.-M. Yang:" Efficiency–optimized flux trajectories for closed-cycle operation of field-orientation Induction Machine Drives" IEEE Transactions on Industry Applications , Vol.28, No.3, pp. 574-580, 1992.

[16] B. Blanusa and S. N. Vukosavic:"Efficency Optimized Control for Closed-cycle Operations of High Performance Induction Motor Drive", Journal of Electrical Engineering, Vol. 8 Edition: 3, pp. 81-88, 2008.

[17] Brayson A. E., Applied Optimal Control,Optimization, Estimation and Control, John Wiley & Sons, 1975.

Acquisition System for Static Torque Characteristics Measuring

Goran Vuković, Srđan Ajkalo, Srđan Jokić and Petar Matić

***Abstract*—In this paper developement of an acquisition system (AS) for static torque characteristics of electrical machines measuring is described. AS is implemented on the device „Dr.Staiger, Mohilo+CoGmbH" using acquisition card with appropiate software. Performances of the new system are illustrated by few representative measurements.**

***Index Terms*—Static Torque Characteristics, Machine Testing, Measuring and acquisition.**

I. Introduction

MEASUREMENT and acquisition of the torque of electric machines is one of the most difficult machines inspections. Test system must have the capability to drive machine in investigated operating mode (burden state) while a torque transducer, for direct measurements is sophisticated, delicate and expansive device. Regarding that, torque measuremets of machine in burden state are, in most all events, performed in specialized test laboratories, mostly as a standard testing [1-3].

True torque value is required information for electric machine construction in order to determine whether the tested machine reaches its projected performance, to determine operating characteristics, as well as in standardization to identify if tested machine satisfies designated standard. In modern electric motor drive, torque is estimated (calculated) with accessible gauge measurements (current and voltage), but the only way to determinate exact value of torque on motor spindle is direct measurement. Torque measurement domain is covered by the appropriate regulations. [1-4]

Torque characterisitcs are represented with torque conditionality on machine speed, on machine spindle (or indirectly, conditionality of time).

Existing equipment for recording torque characteristics is due to the price and complexity, for many years held in use. Improvement technologies, primarily measurement - acquisition systems, the existing devices are exceeded. In such cases, it is very cost-effective to renovate measurement-acquisition systems, by the retention of burden system. In this way, the existing set of equipment (torque encoders and burden system) are still in use, and allows the direct processing of measured data on the computer.

In this paper, just such a procedure is described, in which the measurement system of outdated technological device for torque measurements is changed with contemporary system. First described is the torque sensor with burden system, and hardware and software of a new measurement - acquisition system. After that, follow the pictures of important torque characteristics that serve to illustrate the performance of projected system. At the end are the guidelines for further work.

G. Vuković, S. Ajkalo, and S. Jokić are with the Faculty of Electrical Engineering, University of East Sarajevo, East Sarajevo, Bosnia and Herzegovina.

P. Matić, is with the Faculty of Electrical Engineering, University of Banja Luka, Banja Luka, Bosnia and Herzegovina.

II. Description of the Device and Purpose

Analyzing device "Dr. Steiger, Mohilo+CoGmbH" is intended for inspection and verification torque characteristics of electric machines and other rotation devices [5]. Block scheme of the device "Dr.Steiger,Mohilo+CoGmbH" with measurement-acquisition system is represented on Fig. 1.

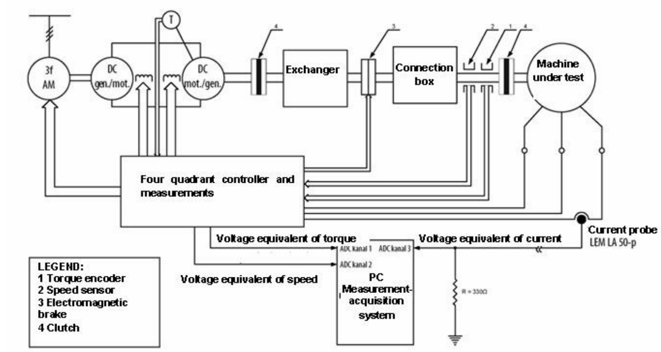


Fig. 1. Block scheme of the device "Dr. Steiger, Mohilo+CoGmbH".

Components of the device are measurement desk, burden system, and command-measurement desk. On the measurement desk is fixed machine under test and connected with one of three measurements spindles, depending on motor speed and power. Coupling between device and tested machine has been accomplished with special claw clutch. For each of the three spindles there are a couple of sensors for speed and torque measurements. Torque encoders are inductive and speed measurement is implemented with optical increment encoders. Spindles are getting out from a specific transmission box that allows various combinations of simultaneous or particular use of spindles with several speed ratios. Ranges of torque and speed values that can be measured with observe equipments have been mentioned in Table I.

TABLE I
TORQUE AND SPEED RANGE

	Rotating speed	Torque
1	by 3600 rpm	by 30 Nm
2	by 10000 rpm	by 10 Nm
3	by 26000 rpm	by 3.5 Nm

The transmission box is connected through the clutch with electromagnetic brake, which blocks spindle by the starting torque measurement. The clutch connects transmission and changing box with four speed ratios: 1:1, 1:1.4, 1:2.2 and 1:4. Changing box is connected with the burden system by the rubber clutch. Burden system is controlled in all of four quadrants and has measurements tacho generator.

Burden system has been realized as the Ward Leonard group. Operating and direct mechanical coupled asynchronous machine are placed in mount of measurement desk. DC machine is electric connected with the burden DC machine. When machine under test is tested in motor mode regime, then burden machine is generator and returns power back to network. In case that machine under test is tested in generator mode regime, the role of other three machines will also be changed. In this way, operating in four quadrants with recuperating is enabled.

Signals from each of torque sensors (each spindle has one sensor) are transmitted in appropriate analog amplifiers with small delays, which on output generate signal of range ±10V. Analog amplifier has potentiometers for gauging, cancel offsets and fine calibration. Each torque sensor has been calibrated in steady state, thus the spindle is blocked with electromechanical brake, and on spindle adjust torque with known weight and length on measurement bar. With known torque value on blocked spindle and voltage measurement on sensor, (assuming that sensor characteristic is linear) it is possible to determine torque constant in Nm/V.

Equipment is electronically (analogue) controlled from the command-measurement desk. Burden system is controlled by speed, manually or automatically. In manual mode, wanted speed of burden machine is directly determined by potenciometer. Manual mode includes also starting torque measurement, when spindle is blocked with eletromagnetic brake. Blocked duration can be adjusted in the range of 1-5s. During blocked period, at the same time machine is started under test and electromagnetic brake, and after that period both of them, machine under test and brake are turned off.

In automatic mode burden machine linearly accelerates or slows down to or from setting speed, at which the accelerate slope reference can be set through periode of acceleration or periode of slow down. It is possible recording torque characteristics of the linear acceleration to the given reference and instantly by linear slow down, as well as by only linear acceleration or slow down. The duration of slow down or acceleration (slope of characteristic) are set in the range of 1-10s. Based on that, recording of group of dynamic torque characterisitcs is possible.

Measurement-acquisition system consists of analog instruments (voltmeters) on wich the read is voltage proportional to measurement's torque registered, respectively voltage proportional to speed of taho generator and digital display with rotating speed of incremental encoder on wich experiment is based. Aquisition system contains grafic plotter with torque or speed signal on vertical axis, and time or speed proportional signal on horisontal axis.

Regarding that torque sensors as well as energetics part of equipment in good functional state, instead of outdated measurement-aquisitions system, new system, based on PC and aquisition card, is projected and adjusted. From electronically parts of old system, only control of burden system is withold.

III. MEASUREMENT-ACQUISITION SYSTEM

For the purposes of innovations in the previous chapter described measurement system is using multifunctional card – Humusoft MF 624. The card is especially adapted to the application of the control systems in real-time and systems for acquisition of measuring data. This PC card in addition to standard personal computer, offers a variety of opportunities. Because of its relatively low price, its application is not limited only to industrial use, but has wide application in research and education institutions [6]. Humusoft MF624 has a digital inputs, digital outputs, timers, counters, Pulse Wide Modulation, encoders inputs, measuring of frequency, as well as 8 analog inputs with 14-bit A/D converter with simultaneous circuits for selection and delay and very short conversion time, which is in case of use one of the channels up to 1.6 μs. Voltage range of analog signal, which card accepts is ± 10 V with the protection of the inputs up to ± 18 V. The important feature of card, except management in real time, is as well the possibility of data acquisition and waves forms analysis with application MATLAB Simulink Real-Time Windows Target and xPC-Target.

Software communication with the card is enabled with a large number of libraries for the programming language C or MATLAB. Card access is possible, from the command windows in MATLAB, as well as using standard blocks of Real-Time Toolbox from Simulink. Part of developed aquisition measurement system, which is realized in the MATLAB Simulink, is shown in Fig. 2. The picture also shows preview for the setting parameters of its card and, in particular, channel selection for A/D conversion.

For development of aplication in Real Time it is necessary to implement a specific procedure in MATLAB Simulink environment. Creation of the models into the Simulink develop environment, represents first step in preparation of simulation in real time and after that folows the input of parameters of models required for simulation and input of parameters for graphic performance of signal. After this step, it is necessary to configure the Real Time Windows Target and the input of

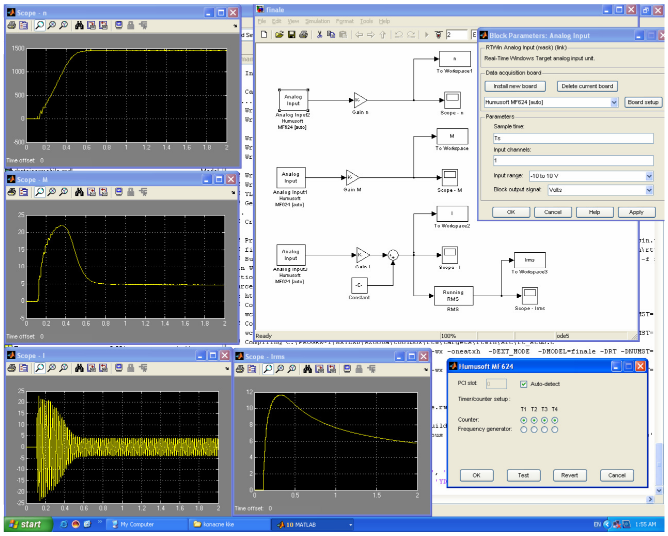


Fig. 2. Real-Time Windows Target blocks and developed measurement-acquisition system.

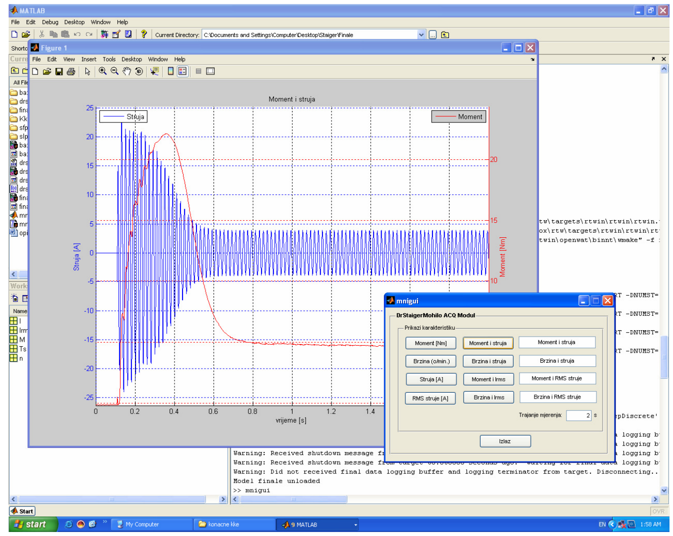


Fig. 3. GUI for the presentation of measuring characteristics.

simulation parameters for the Real Time Workshop. This Parameters are used to generate C code for the purpose of Real Time aplication. Namely, the Real Time Workshop generates C code for Simulink model, then the Open Watcom C/C ++ Compiler compiles and links C code in the Real Time executable application. After these steps folows connection of Simulink with acquisition card, while executable aplication starts from the local computer [7]. This is one of the differences between Humusoft MF624 card and more expensive I/O cards such as dSpace 1104 with DSP and that after connecting Simulink to acquisition card forwards the HEX executive program. In this way it carries out the relief of local computer and allows working of more complex aplication in real-time. Nevertheless, a compromise price and features of MF624 card in the system are completely met.

Voltage signals from a measurement torque and speed converters machine under test from the front panel of the "Steiger" device, as well as the measurement currents signal are accepted on the channels 1, 2 and 3 of A/D converter. Voltage dependence of the value of the appropriate physical size (torque constant) is determined by the calibration. Value mappings are entered in the Simulink model in the form of a gain block and possible DC offset, so that the output data measurements corresponds to values in the appropriate measurement units. Scope component allows the current view thus scaled values of measurement signal in real time. Also, it is possible to record the characteristics in the files, in order to further process, store or present measured data.

Part of the program for the representation of measurement signal characteristics and the development of measurements documentation was made in the form of graphical user panels (GUI). The appearance of the user panel to display measuring characteristics is given in Fig. 3.

It is necessary to note that depending on the complexity of the Simulink models, the Real Time Applications may have time selection of signals between 1 ms and 100 µs, so they are suitable for most laboratory experiments. In this concrete example the chosen time of the selection is 1 ms.

Fig. 4. gives an appearance of measurement apparatus. On the left side of Fig. 4. can be seen electro-mechanical part of the equipment, in the middle are adaption circuits and on the right side is a personal computer with acquisition card.

Fig. 4. Measurement apparatus.

IV. EXPERIMENTAL RESULTS

Below are a few clips and analysis of important torque characteristics, that were recorded in the developed equipment. The modes of interest are regime states from the point of view as with the exploitation, as well as projected (nominal) size. Device under test is standard low voltage three-phase asynchronous motor with nominal power 1.1 kW, the nominal speed of 1400 rpm and 2.85 A nominal current. Nominal torque value of the machine is 7.5 Nm. The first measured value is from starting torque of machine Motor is at the same time blocked with the electromagnetic brake and connected to the nominal voltage value. Recording of the starting motor torque, in the duration of 5 seconds, is shown in Fig. 5.

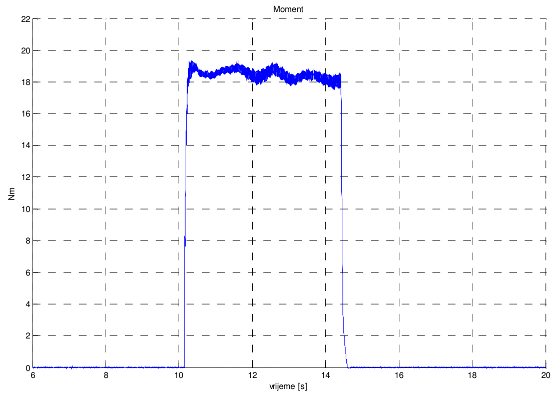


Fig. 5. Initial torque of the nominal power.

Measured value of starting torque is 18.5 Nm, and starting current 14.2A, so the ratio of starting and nominal torque is 2.5 and ratio of starting and nominal current is 5. Noise in measuring torque on Fig. 5 is the consequence of vibrating of claw clutch between machine and burden system.

Second measurement is a verification of nominal working point. Asynchronous motor is bound to burden system and connected to nominal voltage. Speed rotation of machine spindle is set to nominal value which is confirmed by reading from the encoder. Measured values are current (effective value) and machine torque, and they're displayed on Fig. 6. Fig. 6. shows that when machine has nominal torque, it conducts nominal current (in tolerance interval [1-3]), so that measured machine has real values like on name plate. If this experiment would have additional temperature measurement, and extended length, it would provide verification of nominal power in heating experiment.

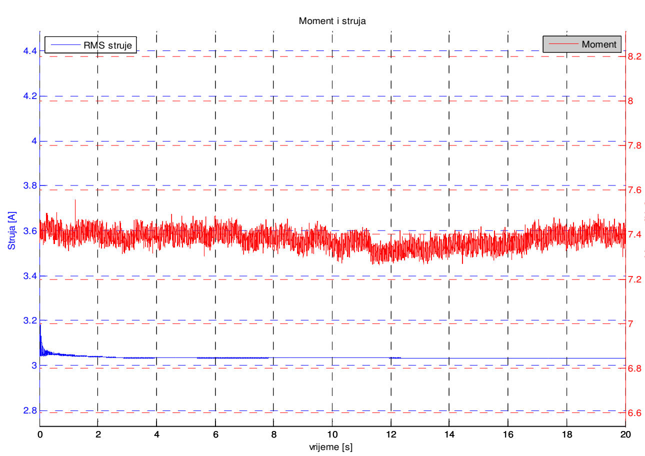


Fig. 6. Torque and the effective value of current when the motor has nominal speed and nominal power.

To illustrate possibility of recording dynamic torque characteristics, machine was linearly accelerated from idle to nominal speed, and then turned off. Acceleration time was 5s. Oscillations of torque are again the result of clutch vibrations in idle state and in starting. Fig. 7.a shows recorded torque characteristics, and Fig. 7.b rotation speed. This measurement confirms the value of starting torque recorded in experiment from Fig. 5.

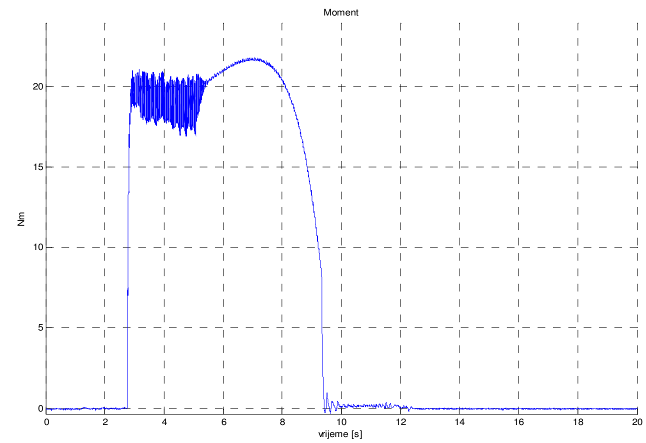


Fig. 7.a. Moment of the linear motor accelerating.

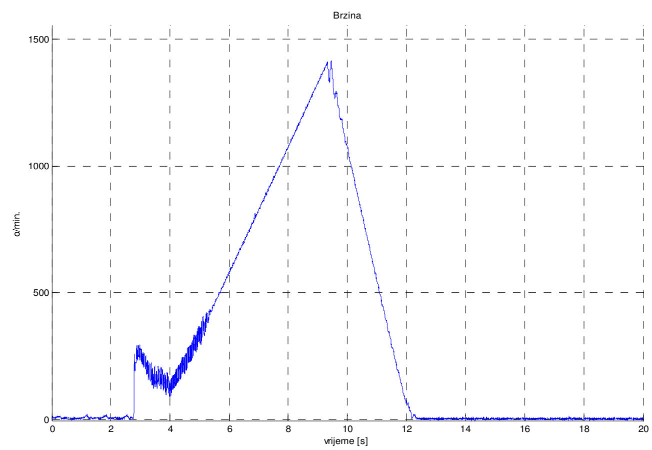


Fig. 7.b. Motor speed by determining the pull-out torque.

To illustrate performance of measuring equipment in different operating regime states, experiments are done with burdening the measured machine in motor and in generator regime state. Fig. 8.a shows machine torque and Fig. 8.b matching speed when machine is burdened from ideal idle state in motor regime state. Experiment is done by having speed of burdened machine equal to synchronous (ideal idle state), and after that speed of rotation was manually changed, so the torque also changed. Machine was burdened in four points in total.

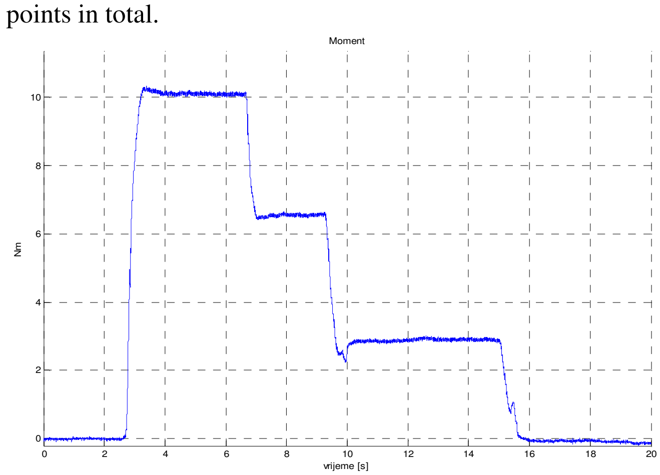


Fig. 8.a. Burdened machine in motor mode - torque.

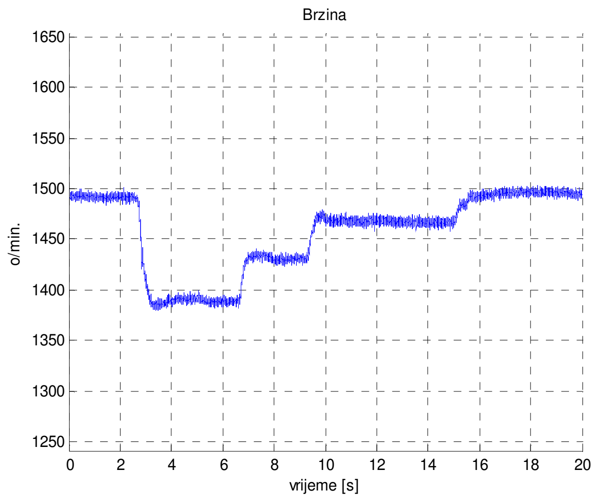


Fig. 8.b. Burdened machine in motor mode - speed.

Experiment similar to measurement in Fig. 8. has been done in generator regime state and displayed on Fig. 9. Results of the torque measuring are shown in Fig. 9.a, and the appropriate speed in Fig. 9.b.

On the basis of burden tested machines can be set slope (gradient) of torque characteristics of the linear part of static characteristics.

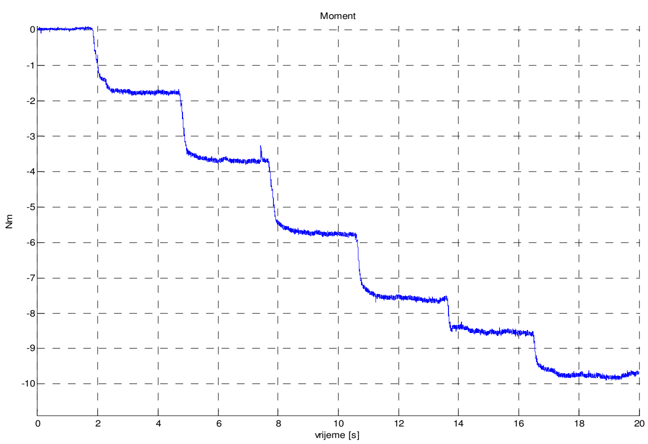


Fig. 9.a. Burdened machine in generator mode - torque.

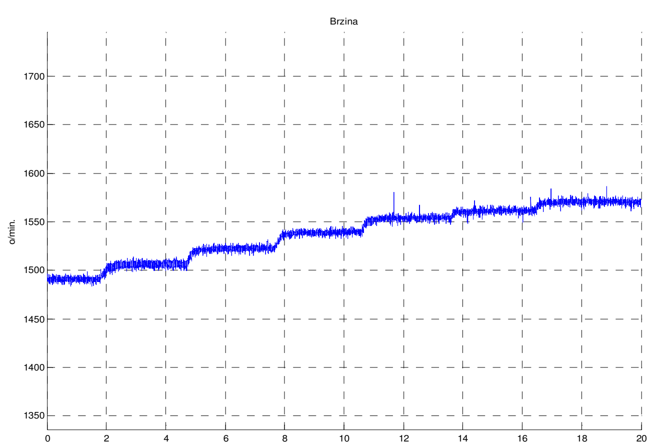


Fig. 9.b. Burdened machine in generator mode - speed.

V. Conclusion

Installation of measurement – acquisition system is obtained a new quality in the accuracy and speed of processing of the measuring data. When to use the existing analog and digital outputs of the acquisition card for the control of the experiment, then the complete recording procedures of torque characteristics could be automate with increasing accuracy of infliction references. Preferably, the modified control system, so that testing can be managed by the moment, and not just by speed.

By details calibration system, with compliance of applicable standards, measurement equipment could be certified to become a legal measurement means (gauge). For the purposes of calibration, it is necessary to determine the exact characteristics of sensor, the linearity, and possible delays in the adaptation instances, according to legal regulations. In this case, the equipment, except in the scientific and teaching purposes, coulde be used as a gauge of the issuing of third person attests.

Small variations of speed are a consequence of speed being regulated by feedback coupled with the encoder, and graphics shows speed recorded with taho generator. To avoid these deviations, it is necessary to input signal from encoder in the acquisition card.

Electromechanical Ward Leonard groups do not have the possibility of faster regulation time or speed. Therefore, this test equipment is used to record relatively slow dynamic processes which, in reality appropriate operation in practice, while recording a fast process (acceleration in idle state or reverse) depends not only on the limit load, but the speed of processing torque signal to the acquisition system. Specific robust, simple four-quadrant work, as well as the need for recording static torque characteristics justify further exploitation.

Measurement equipment can be used for the purposes of examination of modern electromotor drives, by the verification techniques of torque estimation. Comparing estimated and measured values of the torque in stationary state, shortcomings of mathematical models or used estimation techniques can be noticed.

Developed measurement - acquisition system has showed very good in the implementation, and it was free to build to the other seven measuring system for recording of torque characteristics, such as Faculty of Electrical Engineering in East Sarajevo has. In this way it will also improve testing of different machines , ie. larger forces.

Acknowledgment

The authors give great thanks to prof. Dr Milan Zečević and prof. Dr Radovan Radosavljević for numerous suggestions that have contributed to the quality of this work.

References

[1] F. Avchin, P. Jereb: Ispitivanje električnih strojeva, Tehnička založba Sovenije, Ljubljana 1968.

[2] B. Mitrakovic: Ispitivanje električnih mašina, Naučna knjiga, Beograd, 1969.
[3] M. Petrovic: Ispitivanje električnih mašina, Akademska misao, Beograd, 2000.
[4] V.Vučkovic: Električni pogoni, Akademska misao, Beograd, 2002.
[5] Uputstvo za rukovanje uređajem Dr.Staiger,Mohilo+CoGmbH, Elektrotehnički fakultet, Sarajevo, 1985.
[6] User’s Manual, MF 624 MULTIFUNCTION I/O CARD, HUMUSOFT 2006.
[7] The Language of Technical Computing, Version R2006a, Mathworks MATLAB Manual.

Voltage Sag Effects on High Performance Electric Drives

M. Petronijević, N. Mitrović, and V. Kostić

***Abstract*—This paper researches symmetrical and unsymmetrical voltage sag influence on speed and torque deviation in rotor field oriented (RFO) and direct torque controlled (DTC) drives. To overcome appeared drop in speed and torque ripple it was proposed a field-weakening algorithm during voltage sag and/or disturbance estimator application in torque control loop. Prompt DTC flux recall enables efficiency overcoming torque deviation.**

***Index Terms*—Voltage sag, field orinted control, direct torque control, torque ripple, power quality.**

I. Introduction

FREQUENCY converters present sophisticated and nonlinear power electronics devices, which are used in manifold industry branches, in wide range of installed power. Mainly, these AC voltage converters consist of input rectifier (in most cases three-phase diode bridge), DC link circuit, output converter (three-phase inverter with power switching transistors) and additional system for protection, measuring and control. Each of the mentioned subsystem can particularly respond in case of power supply disturbance appearing or in coupling with other sub systems. Complex converter sensitivity of power quality parameters are achieved as a unique response of the overall subsystems.

Frequency converters voltage sag sensitivity is led by various studies and experimental researches (e.g. [1]), in which determined sensitivity limits for equipment operation is related only to disconnection/tripping. Voltage sags may cause significant adjustable speed drives performance degradation: symmetrical voltage sags lead to maximum available torque reduction ([2]) and unsymmetrical ones consequently initiate DC link voltage ripple. Problem of the influence of unsymmetrical voltage sags on adjustable speed drives was analytically examined in [3] and [4] where the importance of specific DC bus capacitance value (μF/kW) is stressed. Reference [5] presents some experimental results referring to unsymmetrical voltage sag influence and recommends under voltage protection settings to overcome power quality disturbance.

Recently researches in voltage sags sensitivity have not taken into consideration control algorithm influence in order to maintain operation without loss of performance. This paper experimentally verifies control algorithms influence on symmetrical (type A) and unsymmetrical (type B) voltage sags outcomes, especially in high performance electric drives. In case of an input supply voltage reduction, which does not trigger under voltage protection, usual converter control brings along torque reduction, meanwhile to the drop in speed. Based on analytical relation and detailed computer simulation we presented the new method, which overcome conventional control method limits. The algorithm was successfully tested on dSpace prototyping system DS1104 applied in electric drives with rotor field oriented control (RFOC) and direct torque and flux control (DTC).

Unsymmetrical voltage sags B, C, and D types are the results of the most common fault – single-phase short circuit ([6]). Numerous sags of the previous mentioned types (over 70% of all sags in network) and the fact that DC bus voltage usually remains over undervoltage protection limit provoke the idea of undesired torque ripple elimination.

In case of RFOC induction motor drive, it was proposed internal current control loop modification by adding a disturbance observer. It was revealed that DTC drives are less sensitive to this sag type, even in case of modified control method with constant switching frequency and PI (proportional–integral) flux and torque controller. Experimental results follow up achieved results and present the applied methods effectiveness for disturbance elimination.

II. RFO and DTC Induction Motor Drives

Nowadays, two basic industrial solutions are designated as methods of instant torque control in high performance induction motor drives:

- **Vector control:** based upon stator currents control in synchronous reference frame, using pulse width modulation (PWM);
- **Direct torque control:** based upon stator flux phasor position control, in the basic method with hysteresis controllers for stator flux magnitude and instantaneous torque value.

Basic control structure, which uses indirect rotor field orientation control (IFOC), was shown in Fig. 1. Two inner current control loops for d and q stator currents component are noted and synchronous speed (ω_e) estimator based on

M. Petronijević, N. Mitrović, and V. Kostić are with the Faculty of Electronic Engineering, University of Niš, Niš, Serbia.

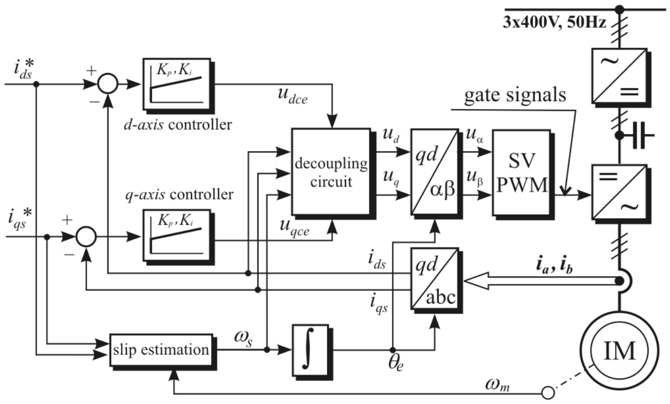


Fig. 1. Basic IFOC control scheme.

reference stator currents components. Linear PI controllers are mainly used. Basic equations for proportional (K_p) and integral (K_i) gain adjusting for q and d current loops, considering decoupling circuit influence, are shown in [7]. In the experiments accomplished in this paper, currents control loops bandwidth is set on 1250rad/s, which is a consequence of noise presence in measured currents. For induction motor data given in Table I digital PI controllers for d and q loops are with

TABLE I
MOTOR DATA (PER PHASE) AND LIST OF USED SYMBOLS

Symbol	Motor data
R_s	Stator resistance, 2.7 Ω
R_r	Rotor resistance, 2.22 Ω
L_{ls}, L_{lr}	Stator and rotor leakage inductance, 10.5mH
L_m	Mutual inductance, 0.27 H
u_{sd}, u_{sq}	d and q axis stator voltages
i_{sd}, i_{sq}	d and q axis stator currents
i_{rd}, i_{rq}	d and q axis rotor currents
P	Pole pairs, 1
P_n	Induction motor rated power, 2200W
U_n	Public distribution network rated voltage, 400V
ω_n	Angular rated speed, 297 rad/s
σ	Leakage coefficient, 0.0733
T_r	Rotor time constant, 0.126s
m	Modulation index

K_p=21.75 and K_i=0.768 when sample time is 100µs.

Adequate current control loop adjusting is presented in Fig. 2. where in torque control regime (pulse torque train reference ±2.5Nm) is speed response presented.

In the simplest variant direct torque control, consist of three level hysteresis comparator for torque control and two-level hysteresis comparator for flux. To achieved the acceptable torque ripple it is necessary to calculate appropriate switching states executing in time which is about 10 times shorter (≈25µs) then switching frequency. The same hardware

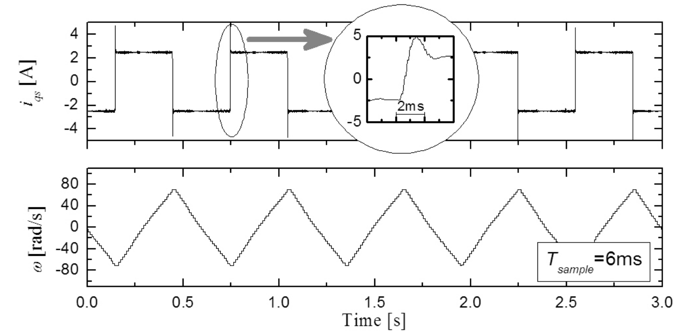


Fig. 2. IFOC drive dynamic performance (top – estimated torque T_e [Nm], bottom – speed ω_m [rad/s]).

prototype platform usage needed modified DTC method realization with PI torque and flux controllers with identical sample and calculation time as in IFOC control algorithm. Basic modified control structure (PI-DTC) was shown in Fig. 3. Constant switching frequency was realized by space vector PWM (SV-PWM). Initial PI controllers' parameters adjusting are accomplished by symmetric optimum method given in [9]. During the experiment parameters where additionally adjusted to achieve torque control loop bandwidth equal to 1250rad/s, which is the same with IFOC current control loops bandwidths.

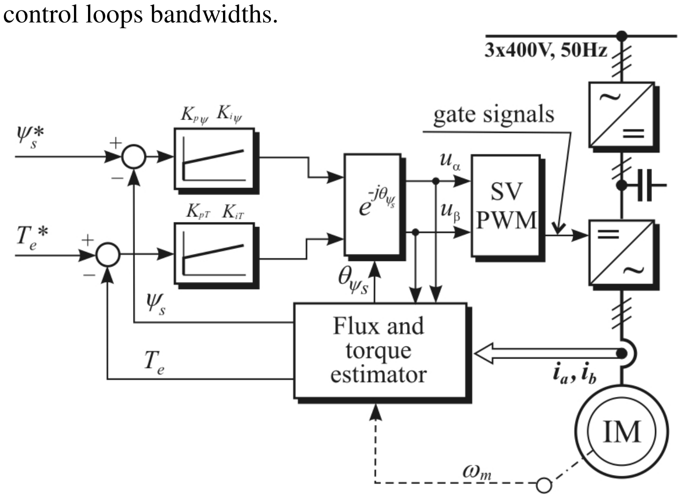


Fig. 3. Modified PI-DTC control scheme.

Illustration of dynamical performances of DTC drives is shown on Fig. 4 with identical control requirements as for IFOC.

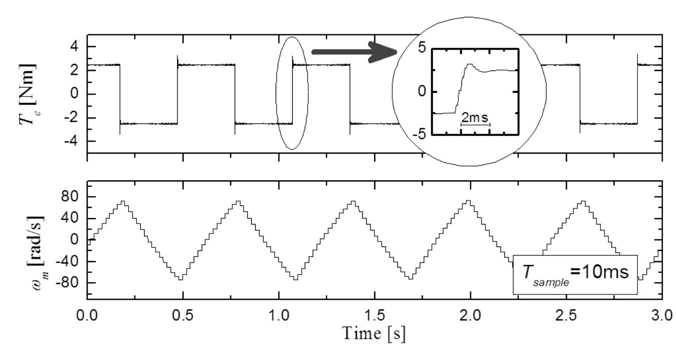


Fig. 4. PI-DTC drive dynamic performance (top – estimated torque T_e [Nm], bottom – speed ω_m [rad/s]).

III. Symmetrical Voltage Sag Effects

Under the assumptions of linear magnetic circuit and balanced operating conditions, induction motor can be described by the following set of equations in two-phase synchronous reference frame:

$$\begin{bmatrix} u_{ds} \\ u_{qs} \end{bmatrix} = \begin{bmatrix} R_s & 0 \\ 0 & R_s \end{bmatrix} \begin{bmatrix} i_{ds} \\ i_{qs} \end{bmatrix} + \begin{bmatrix} p & -\omega_s \\ \omega_s & p \end{bmatrix} \begin{bmatrix} \lambda_{ds} \\ \lambda_{qs} \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} R_r & 0 \\ 0 & R_r \end{bmatrix} \begin{bmatrix} i_{dr} \\ i_{qr} \end{bmatrix} + \begin{bmatrix} p & -\omega_r \\ \omega_r & p \end{bmatrix} \begin{bmatrix} \lambda_{dr} \\ \lambda_{qr} \end{bmatrix} \quad (2)$$

In the previous equation p represents the differential operator, ω_r - slip angular velocity ($\omega_r = \omega_s - \omega$), ω_s - synchronous reference frame speed and ω is rotor angular speed ($\omega = P\omega_m$).

Basic equations for flux linkage are:

$$\begin{aligned} \lambda_{ds} &= L_s i_{ds} + L_m i_{dr}, \ \lambda_{qs} = L_s i_{qs} + L_m i_{qr} \\ \lambda_{dr} &= L_r i_{dr} + L_m i_{ds}, \ \lambda_{qr} = L_r i_{qr} + L_m i_{qs} \end{aligned} . \quad (3)$$

Electromagnetic torque can be calculated by using the following formula:

$$T_e = \frac{3}{2} P \frac{L_m}{L_r} (i_{qs}\lambda_{dr} - i_{ds}\lambda_{qr}) . \quad (4)$$

If the reference frame d-axis is aligned with rotor flux linkage phasor (this presumption does not determine control method), the flux component will be:

$$\lambda_{qr} = 0, \lambda_{dr} = \lambda_r . \quad (5)$$

Limits as consequences of the maximum converter output current (I_{max}) and the maximum PWM voltage at motor stator terminals (U_{max}) referring to appropriate qd motor quantities can be written:

$$i_{qs}^2 + i_{ds}^2 \le I_{\max}^2 \qquad \text{(current limit)} \quad (6)$$

$$u_{qs}^2 + u_{ds}^2 \le U_{\max}^2 \qquad \text{(voltage limit).} \quad (7)$$

Voltage limit equation, respecting the presumption (5) and basic motor equation, is transformed to:

$$A i_{ds}^2 + C i_{qs}^2 + B i_{ds} i_{qs} \le U_{\max}^2 \quad (8)$$

where: $A = R_s^2 + \omega_s^2 L_s^2$; $B = 2R_s\omega_s \dfrac{L_m^2}{L_r}$ and $C = R_s^2 + \omega_s^2\sigma^2 L_s^2$.

If we suppose that IFOC drive operation in basic speed range is with constant rotor flux, reference d-axis stator current at steady state will be:

$$i_{ds}^* = \lambda_r / L_m \quad (9)$$

Equations (6), (8) and (9) represent IFOC drive static characteristics during voltage sag operation and presented in dq reference frame in Fig. 5.

DTC drive posses a control requirement to remain constant stator flux magnitude in basic speed range. Because of this d stator current component will be:

$$i_{ds} = \sqrt{(\lambda_s^*)^2 - \sigma^2 \cdot L_s^2 \cdot i_{qs}^2} / L_s . \quad (10)$$

The last formula accompanied by current (6) and voltage (8) limit equations explains DTC drive steady state characteristics in voltage drop regime, which is also shown in Fig. 5. Obvious fact is that IFOC and DTC drive control differences can lead to various torque limits during voltage sag conditions.

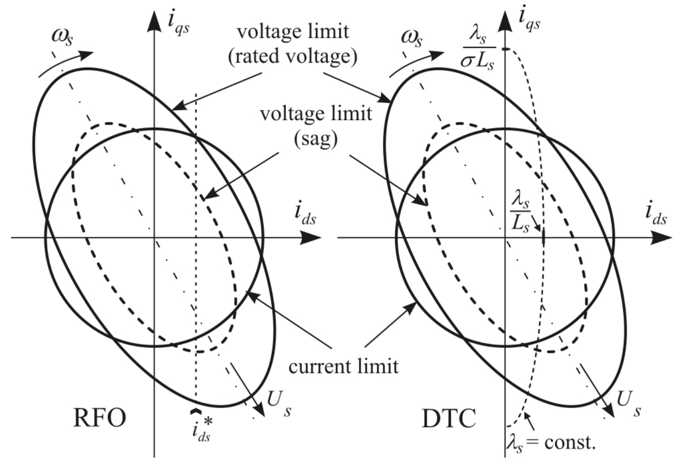


Fig. 5. Voltage and current limit under nominal and voltage sag conditions.

Maximum output PWM voltage reduction during voltage sag provokes decreasing of the available electromagnetic torque with the possible consequence of drop in speed and loss of control accuracy. The main idea of the proposed algorithm is to readjust flux reference value to available voltage at motor terminals. More analytical and simulation results can be found in [2]. This paper in chapter 5 only presents IFOC and DTC drive responses in symmetric three-phase voltage sag as an illustration.

Having in mind that rotor flux recall is restrained by rotor time constant T_r influence it is necessary to enhance stator current component i_{ds} to prevent droop in speed. DTC drive reference flux λ_s^* prompt recall does not require to enhance flux-producing component.

IV. Unsymmetrical Voltage Sags Effects

Voltage sags B, or C and D types influence input rectifier to be transformed into single-phase operation having as a consequences input current distortion, and DC bus voltage ripple increasing with 100Hz dominant component [3]. Output inverter, properly controlled, transforms DC bus voltage to PWM voltage, which is applied at motor stator terminals.

Modulation signals are generally presented as:

$$u_i(t) = u_i^*(t) + e_i(t) \quad (11)$$

where $e_i(t)$ are injected harmonics (also represents direct transformation SVPWM into carrier based PWM [10]), and $u_i^*(t)$ are called fundamental signals. Fundamental components of line to neutral output PWM voltages are:

$$u_{an}(t) = \frac{1}{2} v_{dc}(t) \cdot [\overbrace{m \cdot \sin(\omega_{out} t + \varphi)}^{u_{ai}^*} + e_i(t)]$$

$$u_{bn}(t) = \frac{1}{2} v_{dc}(t) \cdot [m \cdot \sin(\omega_{out} t + \frac{2\pi}{3} + \varphi) + e_i(t)]$$

$$u_{cn}(t) = \frac{1}{2} v_{dc}(t) \cdot [m \cdot \sin(\omega_{out} t + \frac{4\pi}{3} + \varphi) + e_i(t)] \qquad (12)$$

where ω_{out} - inverter output fundamental frequency with modulation index m. Phase angle φ corresponds to initial phase voltage angle respecting to d axis. DC bus voltage $v_{dc}(t)$ in single-phase operation having in mind [3] can be written as:

$$v_{dc}(t) = V_{DC} + V_{DC2} \cos(2\omega_i t + \theta_2) \qquad (13)$$

where V_{DC} presents DC voltage mean value, V_{DC2} is second harmonic voltage amplitude, ω_i is power supply frequency ($2\pi 50$ s^{-1} or $2\pi 60\,$s^{-1}) and θ_2 is second harmonic angle regarding to reference d-axis which defines points on wave at the sag initiation.

Combining equations (12) and (13), and applying coordinate transformations induction motor stator voltages in dq reference frame are:

$$u_{ds}(t) = \frac{1}{2} m V_{DC} \sin\varphi + \frac{1}{2} m V_{DC2} \cos(2\omega_i t + \theta_2) \sin\varphi$$

$$u_{qs}(t) = \frac{1}{2} m_1 V_{DC} \cos\varphi + \frac{1}{2} m_1 V_{DC2} \cos(2\omega_i t + \theta_2) \cos\varphi\,. \qquad (14)$$

In equations above second term is direct consequence DC link voltage ripple because of rectifier single-phase operation. According to [3] sag type, DC link components arrangement and load value influence on voltage pulsation. Having in mind that inverter and motor behave as active load it is hard to calculate second harmonic voltage V_{DC2} value and to predict undesirable torque component value. Beside this, in recent analysis the influence of the fast inner current or torque control loop was not considered.

Combining equations (14) with (1) and (2) and (3) we can, based on (4), calculate in closed form the torque value:

$$T_e = T_{e0} + T_{e2} \cos(2\omega_i t + \phi_2) + T_{e4} \cos(4\omega_i t + \phi_4) \qquad (15)$$

and T_{e0} presents average DC component, T_{e2} is second harmonic amplitude, and T_{e4} is fourth toque harmonic amplitude. More details about calculation can be found in [3], but the achieved results must be accepted carefully, valid only for scalar controlled drives. Illustration from (15) is given in Fig. 6 where single-phase voltage sag consequences are presented. Clearly are noted dominant second harmonics in DC voltage and motor torque. Torque pulsating components can, beside noise increasing, exciting resonance oscillation in mechanically coupling multimotor drives as paper production lines.

In high performance drives, especially in DTC drives, can be expected the undesirable torque pulsation will be significantly suppressed if the inner control loops bandwidth is greater than the dominant second harmonic frequency (100Hz). Usual adjusting of the q-axis stator current component or torque control loops satisfies this requirement. In the next Chapter, we present experimental results regarding to IFOC and DTC drives under single-phase voltage sag.

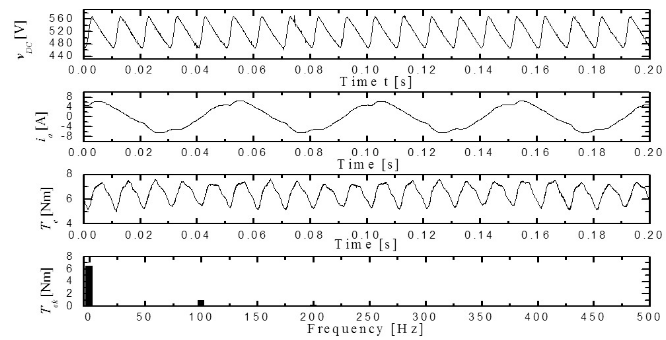


Fig. 6. B type voltage sag effect in drive with V/f control method (load torque= 80% rated, output frequency= 20Hz).

V. Experimental Results

Electric drive control algorithms are implemented by using rapid prototyping system based on dSpace DS1104 control board and Matlab/Simulink software. The experimental verification of the theoretical results was carried out at drive system, which consists of the modified industrial frequency converter, with nominal power of 3.1kVA, an induction motor designated nominal power 2.2kW and a TTL pulse generator with 1024 pulse/revolution was mounted at drive shaft end. Induction motor mechanical loading was done by AC servo drive, which was directly coupled with motor under test. Simple voltage sag generator was made using three-phase power transformer rated power 15kVA with the tap changer under load.

IFOC drive current control subsystem, coordinate transformation blocks, decoupling circuit and slip calculation estimator was realized digitally with the sample time equal to 100μs. Slower speed control loop was implemented with 10ms sample time. Switching frequency of symmetrical SVPWM modulation is set at 5kHz, which also presents simultaneous sampling frequency for two stator currents. In DTC drive torque and flux control loops calculation time period set to 100μs. Sampling time regarding to speed control loop and SVPWM switching frequency are equal to IFOC drive settings.

Fig. 7 illustrates three-phase voltage sag influence for both types of high performance drives. The results of the application enhance of rotor flux readjusting, also stator flux recalls are shown in the same figure. Efficiency of the proposed algorithm in details was discussed in separate paper [2].

To avoid the influence of the speed control loop on torque pulsation we carried out an experiment in torque control regime where rated motor load was applied. At converter terminals we cut off one phase at t=0.24s which respond to the B type voltage sag. In Fig. 8, electromagnetic torque time diagram was shown where it could be clearly noted torque pulsation, which is significantly suppressed regarding to *V/f* drive. These torque harmonics also exist at same characteristics frequencies.

One of the simplest methods for this periodic pulsation reduction is disturbance observer application in current control loop, only for q axis. In Fig. 9 can be seen undesirable torque components reduction.

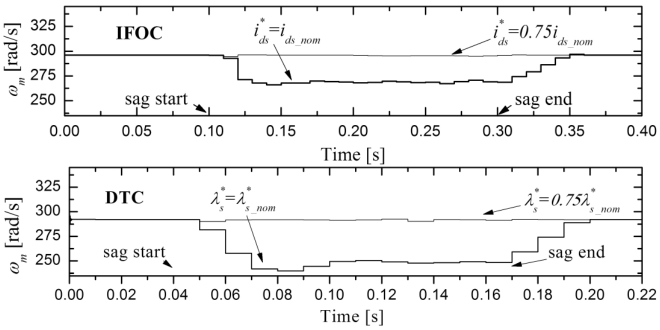


Fig. 7. IFOC (top) and DTC (bottom) drive drop in speed (load torque= 100% rated, U_{sag}=75%U_n).

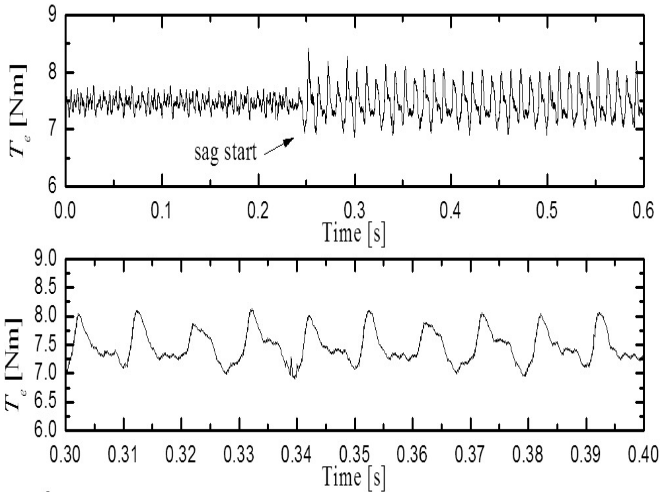


Fig. 8. B type voltage sag effect in drive with IFOC control (load torque= 100% rated).

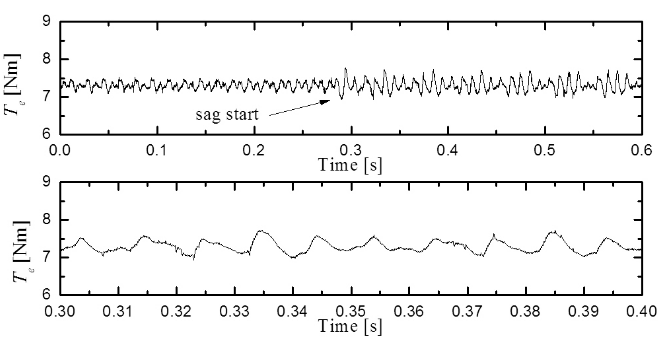


Fig. 9. B type voltage sag effect in drive with IFOC control (load torque= 100% rated, Q observer application).

DTC controlled drive was also tested on B type voltage sag which was initiate at time t=0.22s. In Fig. 10, it can be seen minor voltage sag sensitivity, which is represented by small additional torque harmonic component.

VI. Conclusion

In high performance adjustable speed drives can be expected the undesirable torque pulsation and speed reduction if voltage disturbance occur at input converter terminals.

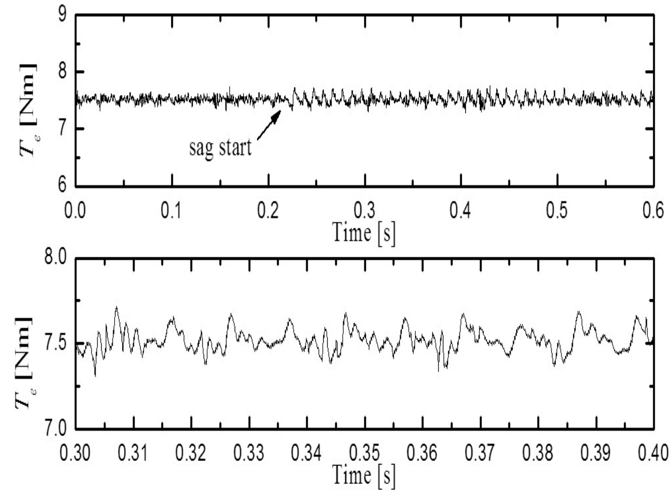


Fig. 10. B type voltage sag effect in drive with DTC control (load torque= 100% rated).

Simplified analysis, without control algorithm taking into account, yields inaccurate results. Numerous simulation and experimental results identifies significant differences in drive behaviors. In high performance drives, especially in DTC drives, can be expected the undesirable torque pulsation will be significantly suppressed. Disturbance observer application in IFOC drives significantly reduces torque pulsation. Further research will be based on application of advanced disturbance estimators. This is significant for converters with lower value of DC link capacitance.

References

[1] S. Ž. Djokić, K. Stockman, J. V. Milanović, J.J. M. Desmet, and R. Belmans, “Sensitivity of AC adjustable Speed drives to voltage sags and short interruptions”, IEEE Trans. Power Delivery, vol. 20, no. 1, pp. 494–505, Jan. 2005.

[2] Petronijevic, M.P.; Jeftenic, B.I.; Mitrovic, N.M.; Kostic, V.Z., "Voltage sag drop in speed minimization in modern adjustable speed drives," Industrial Electronics, 2005. ISIE 2005. Proceedings of the IEEE International Symposium on , vol.3, pp. 929-934 vol. 3, 20-23 June 2005.

[3] K. Lee, T. M. Jahns, W. E. Berkopec and T. A. Lipo, "Closed-form analysis of adjustable speed drive performance under input voltage unbalance and sag conditions", IEEE Trans. Ind. Appl., vol.42, no.3, May/June, 2006, pp. 733-741.

[4] M. H. J. Bollen, and L. D. Zhang, “Analysis of voltage tolerance of AC adjustable-speed drives for three-phase balanced and unbalanced sags”, IEEE Trans. Ind. Appl., vol.36, no.3, May/June, 2000, pp. 904-910.

[5] K. Stockman, F. D’hulster, K. Verhaege, M. Didden, and R. Belmans, “Ride-through of adjustable speed drives during voltage dips” Electric Power System Research, vol 66, pp. 49-58, 2003.

[6] M. H. J. Bollen, Understanding power quality problems: Voltage sags and interruptions, IEEE Press series on Power Engineering, New York, 2000.

[7] D. Telford, M. W. Dunnigan, and B. W. Williams, “Online Identification of Induction Machine Electrical Parameters for Vector Control Loop Tuning,” IEEE Trans. Ind. Electron., vol. 50, no. 2, pp. 253-261, April 2003.

[8] Y. S. Lai, and J.H. Chen, “A New Approch to Direct Torque Control of Induction Motor Drives for Constant Inverter Switching Frequency and Torque Ripple Reduction,“ IEEE Trans. On Energy Conversion, vol.16, no.3, Sept., 2001, pp. 220-227.

[9] M. P. Kazimierkowski, F. Blaabjerg, and R. Krishnan, Control in Power Electronics – Selected problems, Academic Press, New York, 2002.

[10] K. Zhou, and D. Wang, “Relationship between space-vector modulation and three-phase carrier-based PWM: a comprehensive analysis”, IEEE Trans. Ind. Electr. vol.49, no.1, February, 2002, pp. 186-196.

Two Distant Cross-Coupled Positioning Servo Drives: Theory and Experiment

Milica B. Naumović and Milić R. Stojić

Abstract—**The special structure of two distant speed-controlled or positioning servo drives with cross-coupling control, proposed in previous papers of the authors, enables synchronous motion of the servomechanisms. The considered cross-coupling control is based upon the idea of simulation the effects that appear in a virtual mechanical link between the shafts of drives. In this paper, the suggested structure was implemented and experimentally verified using the dSPACE R&D DS1104 system, which works within MATLAB®/Simulink environment. Both numerical simulations and real-time experimental results show good properties of the proposed structure of the cross-coupled observer-based control system.**

Index Terms—**Cross-coupling digital control, Synchronous motion of drives, Rapid control prototyping.**

I. Introduction

DIFFERENT types of cross-coupling controllers are met in many practice applications in which motions of two or more drives are to be coordinated in a certain sense. The synchronous operation of the drives is particularly important in paper making and processing machines, in numerically-controlled machines for metal, in nonlinear path control in automated vehicle and robot guidance, as well as in flexible systems in general.

In order to achieve a cooperation control, that since the eighties became more attractive, many different control structures have been developed. Two basic approaches can be distinguished. The first approach is the conventional distributive control where each robot is controlled separately by its own local controller, while the interactions between the robots are measured by sensors. The second approach is based on the idea of so-called cross-coupling of control systems of two distant independent servomechanisms. The first system of cross-coupling type was proposed by Sarachik and Ragazzini (1957, [1]) and had a "master-slave" structure (y follows x). Namely, the error in the y-axis affects both the x- and y-control loops, but the error signal in the x-axis is not generated. Such nonsymmetrical cross-coupled structure, however, requires a substantial difference in the gains for each axis, resulting in a contour error.

The concept of cross-coupling control with a symmetrical structure was primarily introduced by Koren (1980, [2]) for the machine tool control. In machine tool servo control, the main idea of cross-coupling control is based on calculation of the actual contour error, multiplying it by a controller gain, and feeding the result back to the individual loops. Later, many structures based on the application of cross-coupling control scheme have been proposed and tested in numerous engineering applications: biaxial control of machine tools [3], automatic guidance system of self-controlled vehicles [4], computer-controlled mobile platform for nursing or household robots [5], cross-coupling motion controller for mobile robots [6], etc.

Quite original structures of cross-coupling speed- and position-control of shafts of two distant electromechanical drives based on the application of the concept of electrical shaft are proposed in papers [7] and [8], respectively. Namely, the cross-coupling control is accomplished by the digital simulation of stiffness and friction of a virtual twisting mechanical connection between the output shafts of the drives. The schematic diagram presented in Fig. 1 shows the relationship between design and testing activities in the case of considered structure of servo system with cross-coupling control. This is the well-known V-cycle that is an internationally recognized development standard for IT systems. The results of a research lasting many years at the level of modeling, design and simulation, in order to improve the structure of cross-coupling, are presented in the papers of authors and contributors [9]-[18]. Thus, an effective procedure of parameter setting in digital position regulator is formulated in the paper [9]. The special observer-based structures of the system with two digitally controlled and cross-coupled servomechanisms, with induction and DC motors as actuators, are considered in papers [10]-[12]. A novel structure of the observer-based system for two manipulators which are cooperatively handling the same object in the presence of slow varying load torque disturbances is proposed in [13]. Note that two servomechanisms with cross-coupled control can be treated as a multivariable control system and then designed using different approaches [10], [15].

In this paper, the synthesis of the structure of coordinated control in a two-axis positioning system is presented. The proposed structure is implemented and experimentally verified

This work was supported in part by the WUS-Austria under the Grant C.E.P. No. 076/2002-2003.

M. B. Naumović is with the Faculty of Electronic Engineering, University of Niš, Niš, Serbia.

M. R. Stojić, is with the Faculty of Electrical Engineering, University of Belgrade, Belgrade, Serbia.

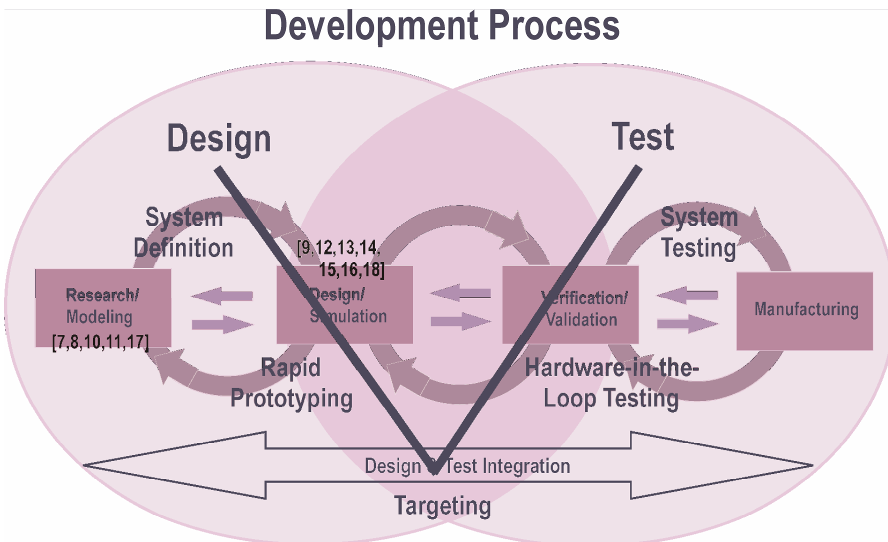


Fig. 1. Implementation of design functions of cross-coupled servosystem according to V-Cicle.

using the system for fast development and implementation of control algorithms dSPACE R&D DS1104 that provides comfortable work in the MATLAB®/Simulink environment.

During the creation of the experimental platform, in the absence of two drives with similar characteristics, one drive is modeled in Simulink environment together with the control part of the cross-coupled servo system, while a real low power permanent magnet DC motor served as the other drive. The performances of the proposed structure of the cross-coupled observer-based control system are verified by digital computer simulation, as well as experimentally in real conditions.

II. Structural Synthesis of Cross-Coupling Control System

The functional block diagram of a positioning servo system with the proposed cross-coupling control is shown in Fig. 2 [13]. Angular positions $\theta_1(t)$ and $\theta_2(t)$ of two distant electrical drives represent controlled variables. In the steady-state, angular positions of drive shafts are to be the same and equal to the common reference θ_{ref}. In addition to the set point $\theta_{ref}(t) = \theta_{ref} \cdot \mathrm{h}(t)$, the system is subjected to two kind of disturbances: load torques T_{L1} and T_{L1} acting on first and second drive, respectively, and disturbance in the form of the initial angular displacement $\Delta\theta(0) = \theta_1(0) - \theta_2(0)$.

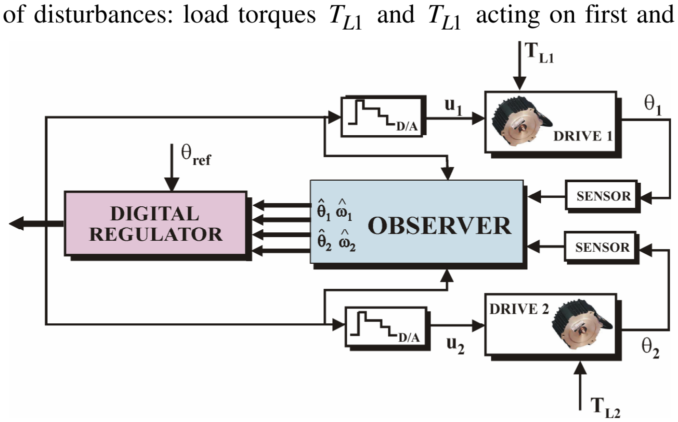


Fig. 2. Block diagram of a servo system with cross-coupling control.

Fig. 3 visualizes in details the structure of digital regulator, which consists of common position regulator and elements of cross-coupling mechanism. The angular positions $\theta_i(k)$, as well as the shaft speed signals of both drives $\omega_i(k)$, $i = 1, 2$ are adopted as coordinates of the state vector $\mathbf{x}(k)$.

Each control channel of two drive system given in Fig. 2 has its main position feedback loop. The channels are coupled by two minor local feedback loops with proportional and derivative actions that electrically simulate the stiffness and friction of a virtual mechanical link (electrical shaft) between shafts of the drives. In such kind of link, the torsion tension, that is manifested in the form of difference between the steady-state values of angular positions of the distant servos in the presence of different load torque disturbances, is relaxed by the additional digital PI regulator in the local feedback loop that simulates the stiffness of the link.

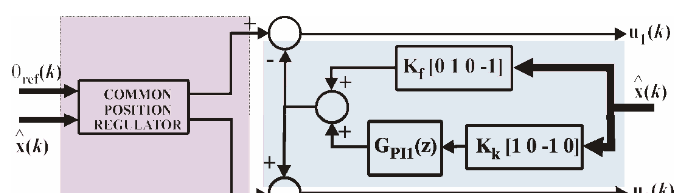


Fig. 3. Structure of digital regulator in the positioning servo system with cross-coupling control in Fig. 2.

The angular positions in system given in Fig. 2 are only measurable. It is suitable to apply PI^2 observers [19], bearing in mind effects of full observer implementation from the viewpoint of filtering of measurement noise, as well as in order to enable the correct estimation of state variables in the presence of constant or slow varying load torque disturbances.

III. Parameter Setting in Digitally Controlled Positioning System with Cross-Coupling Control

Note that the considered system with cross-coupling control has ten adjustable parameters – six control parameters and four gain values of the reduced-order observers.

In the case of identical drives, the pole spectrum of the considered control system is decoupled; it consists of two pole pairs that can be placed separately. This means that the cross-coupling control parameters (K_f , K_s) can be adjusted independently from the position regulator parameters. This property of decoupled effect of some parts of digital regulator given in Fig. 3, which has been described in detail in earlier papers [8]-[12], allows a relatively simple 3-step setting parameters of the regulator, while the observer gains are determined in the fourth step. First, we calculate the parameters of common position regulator, afterwards the coefficients of stiffness and friction of the virtual mechanical link, and then the parameters of the additional PI regulator which serves to relax the torsion of the virtual coupling in steady-state. At the end, according to the separation principle, the observer structural synthesis and its parameter tuning may be accomplished. A summary of the procedure for setting fourteen parameters

($K_p, K_I, K_f, K_k, K_{p1}, K_{I1}, g_{ij}$, $i=1,2$, $j=1,2,3,4$)

of the digital regulator and two observers in the servo system shown in Fig. 2 is given in Table I.

TABLE I
PROCEDURE OF PARAMETER SETTING FOR THE SYSTEM SHOWN IN FIG. 2

Suppose that the transfer function of both servo drives are identified in the form

$$W_i(s)=\frac{\Theta_i(s)}{U_i(s)}=\frac{K_i}{s(1+sT_{mi})}, \quad i=1,2 , \tag{1}$$

where K_i and T_{mi} , $i=1,2$ are gains factors and mechanical time constants of the considered drives, respectively.

1° Calculate the parameters of position PI regulator (K_p and K_I) using

$$\begin{bmatrix} 0 & d_1 & 0 & -1 \\ -d_1 & -d_2 & -1 & z_1+z_2 \\ d_2 & -d_3 & z_1+z_2 & -z_1z_2 \\ d_3 & 0 & -z_1z_2 & 0 \end{bmatrix}\begin{bmatrix} P_1 \\ P_2 \\ A_0^* \\ A_1^* \end{bmatrix}=\begin{bmatrix} -(z_1+z_2)+2+A+B \\ z_1z_2-1-2A-2B-AB \\ A+B+2AB \\ -AB \end{bmatrix}, \tag{2}$$

where

$$P_1=K_p , \; P_2=K_p+K_I \tag{3}$$

and

$$\frac{\Omega_1(z)}{U_1(z)}=\frac{K_1(1-A)}{z-A}, \quad \frac{\Omega_2(z)}{U_2(z)}=\frac{K_2(1-B)}{z-B}, \qquad d_1=\frac{1}{2}(b_{01}+b_{02}),$$

$$\frac{\Theta_1(z)}{U_1(z)}=\frac{b_{01}z+b_{11}}{(z-1)(z-A)}, \quad \frac{\Theta_2(z)}{U_2(z)}=\frac{b_{02}z+b_{12}}{(z-1)(z-B)}, \qquad d_2=\frac{1}{2}[(Bb_{01}-b_{11})+(Ab_{02}-b_{12})], \quad d_3=\frac{1}{2}(Bb_{11}+Ab_{12}); \tag{4}$$

2° Choose a pair of cross-coupling parameter values (K_s and K_f) that belongs to the determined stable region of (K_s, K_f) -plane;

3° Calculate the parameters of the additional PI regulator in the local feedback loop as:

$$K_{p1}=1 \quad \text{and} \quad K_{I1}=\frac{K_I}{K_p} ; \tag{5}$$

4° Set gain values g_{ij} , $i=1,2$, $j=1,2,3,4$ of digital PI^2 observers yielding the poles to be real and equal to $\sigma_z=\exp(-2\pi f_0 T)$, where f_0 is bandwidth of the observers, as follows:

$$\det\begin{bmatrix} z-1+g_{i1} & -e_{i1} & -1 & 0 \\ g_{i2} & z-e_{i2} & 0 & -1 \\ g_{i3} & 0 & z-1 & 0 \\ 0 & g_{i4} & 0 & z-1 \end{bmatrix}=(z-\sigma_z)^4 , \tag{6}$$

where

$$e_{11}=T_{m1}(1-A), \; e_{12}=A, \; e_{21}=T_{m2}(1-B) \text{ i } e_{22}=B . \tag{7}$$

IV. Experimental Setup for Verification of Cross-Coupling Control Algorithms

To illustrate and verify the usefulness of the procedure of independent setting of parameters of the proposed structure given in Fig. 3, the example of two cross-coupling servomechanisms with quite different characteristics will be considered. Fig. 4 visualizes the structure of the experimental environment for rapid control prototyping that was realized during the PhD thesis research [20].

On the basis of the experimentally recorded step response of the servo drive with low power permanent magnet DC motor given in Fig. 5, the parameters in transfer function (1) can be determined as: $K_1=22.065$ and $T_{m1}=0.0348$ s . In the MATLAB®/Simulink environment are implemented the control part of the cross-coupled system, as well as the other servo drive with the posibility of varying the parameter values in its transfer function.

A. Parameter setting in control part of the system

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

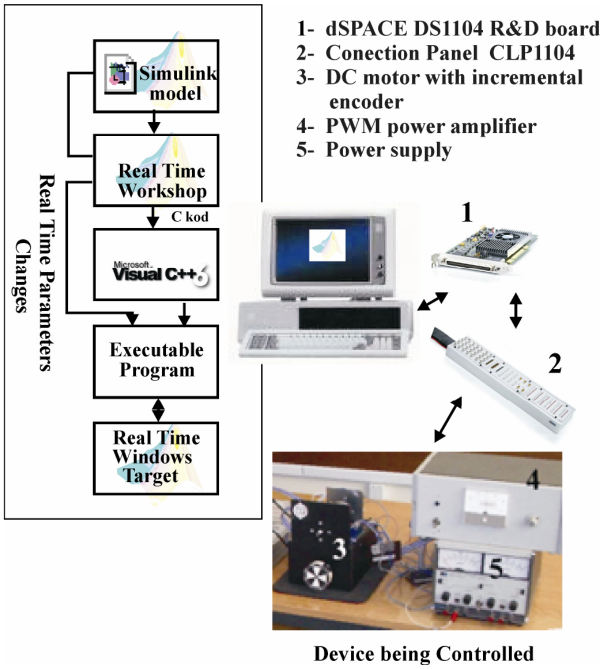


Fig. 4. Schematic representation of experimental setup.

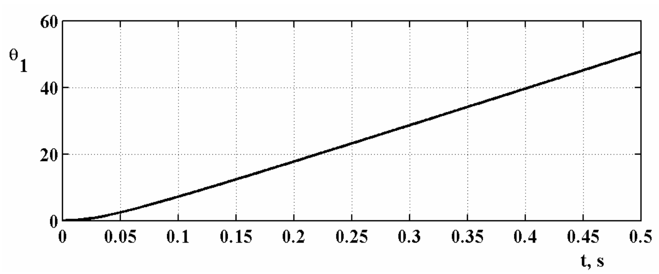


Fig. 5. Step response of servo drive angular position which is measured by using incremental encoder with 1000 pulses per revolution; input signal is $u_1(t) = 10 \cdot \mathrm{h}(t)$.

the common position PI regulator $K_p = 0.445$ and $K_I = 0.002$ obtained by solving matrix equation (2). In papers [8]-[13], [15] it has been shown that it is possible to adopt values K_s and K_f from the corresponding stable region shaded in the parameter plane and given in Fig. 6. In the considered case the values $K_s = 110$ and $K_f = 0.6$ are adopted.

According to relations (6) and (7), the gains of digital PI^2 observers were set to values $g_1 = 0.0854$, $g_2 = 0.9214$ and $g_3 = g_4 = 0.0008$ insuring bandwidth of $4.5\,\mathrm{Hz}$. Unlike ordinary identity observer, these observers will recognize effects of constant or slow varying disturbance on control plants.

B. *Experimental results*

After several digital computer simulation runs, that are used for verification of results of analytical investigation, the experimental research is carried out by using experimental setup of Fig. 4. Note that the considered control plant is the low power DC motor with dry friction problems, which are especially expressive in the tasks of positioning. Under the

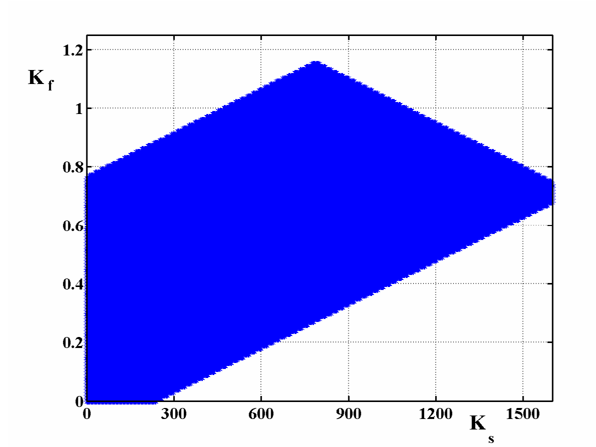


Fig. 6. Stable region in $\left(K_s, K_f\right)$ - plane.

same excitation conditions, both differences between the angular positions $\Delta\theta$ and the shaft speeds of the drives $\Delta\omega$ are recorded and shown in Fig. 7. Control variables u_1 and u_2 in Fig. 8 are without chattering and in agreement with the results of synchronous motion of drives.

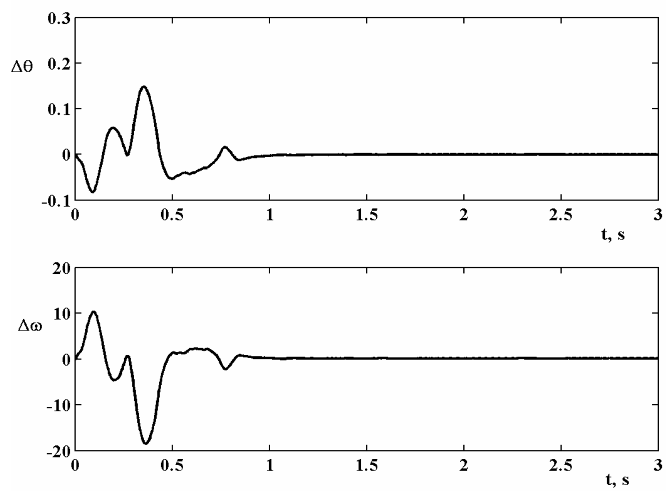


Fig. 7. Transient response and steady-state values of position and speed differences.

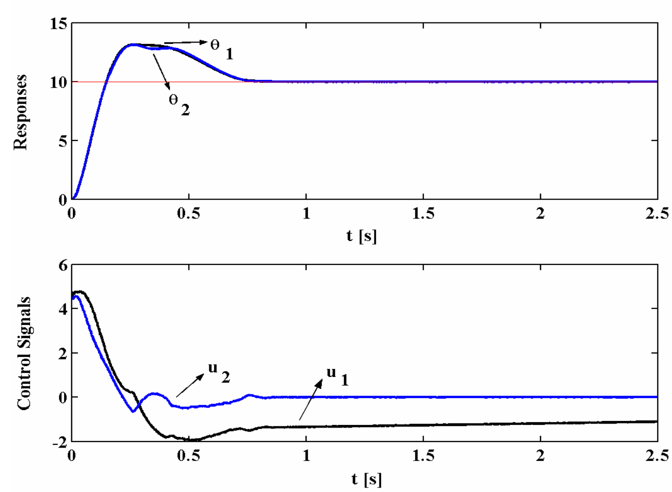


Fig. 8. Responses of angular positions and control signals of digital regulator after step reference signal.

V. CONCLUSION

The efficiency of the special structure of two positioning servomechanisms with cross-coupling control, considered in this paper, is experimentally verified. Since the certain portions of controlling mechanisms are decoupled or weakly coupled, the control parameters may be tuned by using a relatively simple procedure that can be applied in both the similar and quite different servodrives.

REFERENCES

[1] P. Sarachik, J.R. Ragazzini, A Two Dimensional Feedback Control System, Trans. AIEE 76, 1957, pp. 55-61.

[2] Y. Koren, Cross-coupled biaxial computer control for manufacturing systems, Journal of the Dynamic Systems, Measurement and Control, vol. 102, 1980, pp. 265 272.

[3] K. Srinivasan, P.K. Kulkarni, Cross-coupled control of biaxial feed drive servomechanism, ASME J. Dynamic Syst., Meas., Control, vol. 112, no. 2, June 1990, pp. 225-232.

[4] T. Hongo, H. Arakawa, G. Sugimoto, K. Tange, Y. Yamamoto, An autonomous guidance system of a self-controlled vehicle, IEEE Trans. Ind. Electron 34 (1), 1987, pp. 1772-1778.

[5] J. Borestein, Y. Koren, A mobile platform for nursing robots, IEEE Trans. Ind. Electron 32 (2), 1985, pp. 158-165.

[6] L. Feng, Y. Koren, J. Borenstein, Cross-coupling motion controller for mobile robots, IEEE Control Systems, December 1993, pp.35-43.

[7] M.R. Stojić, S.N. Vukosavić, and Đ.M. Stojić, Design of digitally drives coupled by an electrical shaft. In Proc. of the Symposium on Power Electronics Ee'95, (Novi Sad, Yu), 1995, pp. 55-63 (invited paper in Serbian).

[8] M.R. Stojić, S.N. Vukosavić, and M.B. Naumović, Positioning servodrives with cross-coupling control, Part I: Structure design. In Proc. XL ETRAN Conference, (Budva, Yu), Society for ETRAN, Beograd, 1996, pp. 533-536 (in Serbian).

[9] M.B. Naumović and M.R. Stojić, Positioning servodrives with cross-coupling control, Part II: Setting of parameters. In Proc. XL ETRAN Conerence, (Budva, Yu), Society for ETRAN, Beograd, 1996, pp. 537-540 (in Serbian).

[10] M.R. Stojić, M.B. Naumović, and S. N. Vukosavić, Cross-coupled speed-controlled drives with Tesla's induction motor. In Proc. of the 5th Inter Conference TESLA III MILLENNIUM, (Beograd, Yu), SASA, Beograd, 1996, pp. II.51-II.64 (Invited paper).

[11] M.B. Naumović and M.R. Stojić, Positioning servodrives with cross-coupling control, PartIII: Multivariable Approach in System Analysis and Design, In Proc. XLI ETRAN Conference, (Zlatibor, Yu), Society for ETRAN, Beograd, 1997, pp. 401-404 (in Serbian).

[12] Naumović, M.B. and M.R. Stojić, Design of the Observer-Based Cross-Coupled Positioning Servo-drives, In Proc. of the IEEE International Symposium on Industrial Electronics ISIE'97, (Guimarães, Portugal), IEEE, Piscataway, N.J., 1997, pp. 643-648.

[13] M.B. Naumović, Cross-Coupled Motion Controller for Two Cooperating Robot Arms. In Proc. of the IEEE International Symposium on Industrial Electronics ISIE'99, (Bled, Slovenia), IEEE, Piscataway, N.J., 1999, pp. 909-913.

[14] Z. Kalinić, Synthesis of digital control of coordinated movement for multi-axis production systems and mobile robots, M.S. Thesis, Faculty of Mechanical Engineering, Kragujevac, 2004.

[15] M.B. Naumović and M.R. Stojić, Two Distant Positioning Servodrives with Cross-Coupling Control, Transactions on Automatic Control and Computer Science-Special Issue dedicated to 6th International Conference on Technical Informatics, CONTI2004, Timisoara, May 27-28, 2004, vol.49(63), No. 1, pp. 35-40.

[16] M.B. Naumović, Tracking in digital systems with cross-coupling control, In Proc. XL ETRAN Conference, (Čačak, SCG), Society for ETRAN, Beograd, 2004, pp. 215-218 (in Serbian).

[17] M.B. Naumović and M.R. Stojić, Tracking in Digital Systems with Cross-Coupling Control, Preprints of the 4th International IFAC Workshop on Automatic Systems for Bilding the Infrastructure in Developing Countries, DECOM'04, Bansko, Bulgaria, 2004, pp. 173-178.

[18] M.B. Naumović and M.R. Stojić, Regulation and tracking in digital systems with cross-coupling control, In Proc. INFOTEH-JAHORINA Symposium,2005, Vol. 4, Ref. A-13, pp. 58-62 (in Serbian).

[19] M.B. Naumović, M.R. Stojić, Velocity Estimation in Digital Controlled DC Servo Drives, in Proc. of the 24th Annual Conference of the IEEE Industrial Electronics Society IECON'98, Aachen, 1998, pp. 1505-1508.

[20] B.R. Veselić, Synthesis of robust systems for coordinated tracking of complex trajectories by using discrete-time sliding modes, Ph.D. Thesis, Niš, 2006 (in Serbian).

The Simulation Model of Optical Transport System and Its Applications to Efficient Error Control Techniques Design

Predrag Ivaniš and Dušan Drajić

Abstract—**Modeling of the all effects that appear during the transmission of optical signal represents the basic condition for performance evaluation of the optical communication system. The new generation optical transport networks are based on the wavelength division multiplexing (WDM). In this paper, the simulation model that includes a typical effects that appear in the every channel of WDM system, when data rate in each of then is not greater than 10 Gb/s. Using the simulation model, the capacity of the channel is numerically evaluated, and the spectral efficiency of the techniques that combine multilevel modulation and error control coding is compared with this theoretical limits.**

Index Terms—**Channel capacity, channel modeling, error control coding, multilevel modulation, optical transport system, wavelength multiplex division.**

I. Introduction

It is well-known that the error control coding represents an efficient technique that enables the reliable data transmission under difficult conditions in transmittion channel. Furthermore, combining the error control codes with multilevel modulations (e.g. trellis coding modulation), the error rate can be reduced without decreasing the data rate or increasing spectrum bandwidth [1].

Although the error control techniques are successfully applied in digital communication systems for decades, it is almost neglected in optical system design for a long time. Reliable transmission is provided by applying a sufficient number of amplifiers and increase of the data rate is provided by increasing the bandwidth, the resource that seemed to be almost inexhaustible in optical fiber.

However, it has been shown recently that reliability of data transmission in optical transport systems can be extensively increased using the forward error correction (FEC). It is especially significant when throughput in every particular channel is heavy and when long-haul transmission has to be performed using optical fiber channels [2].

This paper is financially supported in part by Serbian Ministry of Science (Technology development project No. 11036 "Multiservice SDH/Ethernet/CWDM/OADM platform for 2,5Gbps/1000baseT/X traffic transport" and Innovation project "Controllable Layer2/Layer3 Ethernet Switch with integrated EoE1 (Ethernet over E1) interface converters ".

P. Ivaniš and D. Drajić are with the Faculty of Electrical Engineering, University of Belgrade, Belgrade, Serbia.

In such a system, it is usually considered that the data rates are at least 10 Gb/s per one channel of Wavelength Division Multiplexing (WDM) optical system, and distances between amplifiers are greater than 40km, when the transmission is usually performed in spectral window about 1,55 μm. This window is about 25 THz wide (1450nm-1650nm) [3], but the greater demands for high speed communication and some actual services (video signal transmission, HDTV, high speed Internet access) require additional increase of data rates. It can be performed using several approaches [4]:

a. *Increase of the optical bandwidth of every particular channel.* Beside the bandwidth of optical fiber itself, this quantity is determined with range of frequencies where the amplifiers transfer functions are linear, and it cannot be increased infinitely.

b. *Increase of the channel number*, i.e., number of the different wavelengths in WDM multiplex. In such a case, the every channel bandwidth is previously limited, and throughput in every particular channel is determined by speed of used electronic components. In every wavelength, the separate set of amplifiers and the other equipment is used. Number of the channels in WDM system is limited according to optical fiber bandwidth and system complexity.

c. *Increase of the spectral efficiency*, e.g. the throughput that can be achieved in every separate WDM channel for the fixed bandwidth. This quantity is theoretically determined by the channel capacity (fundamental quantity in Information theory, defined by Shannon in his famous paper [5]), and practically determined by using the modulation type and error correction code, applied in the considered communication system.

During the previous development of the transport optical systems, increase of the information data rate is performed by increasing the bandwidth or adding of the new channels in WDM multiplex system. In this moment, it appears that information data rate cannot be sufficiently increased using these methods, so techniques that inherently provide high spectral efficiency became more important. Two basic techniques of this type are multilevel modulations and error correction codes (ECC) [1,3].

In the next section, two the most common solutions for transmission of packet traffic through optical transport network are presented. In newer standard, the obligatory protection of transmitted information using the linear block code is proposed. In such a system, Rid Solomon code (255,239) is chosen to reduce the error rate in transmission and increase the length between the regenerators (very important in undersea optical transport systems). This fact motivated the researchers to consider the combination of multilevel modulation and error control codes that leads to the increased spectral efficiency. In this paper, we describe a simulation model and apply it to estimate capacity of the corresponding optical system. Furthermore, we will determine performances (bit error rate and spectral efficiency) of some error correction codes, previously proposed in the literature.

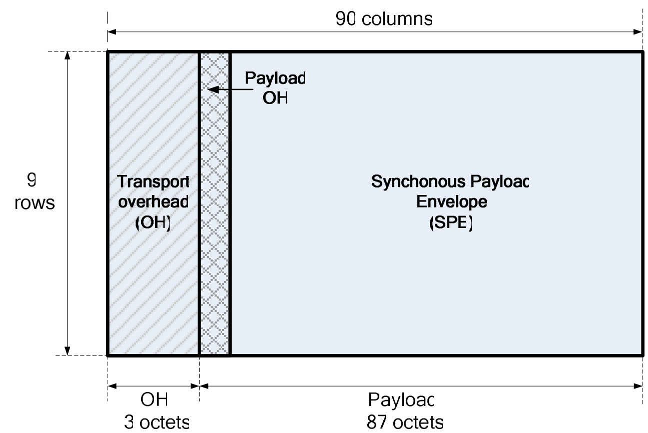


Fig. 1. The STM-0 frame structure.

II. Overview of the Optical Transport Network Technology

Synchronous digital hierarchy (SDH) today represents an dominant technology that provide reliable and wide-spread transmission of digital signals over optical fiber channels. Multiplexing protocol in synchronous optical networks (SONET - *Synchronous Optical Networking*) is designed to support transmission of Time Division Multiplex (TDM) traffic. These networks are connection-oriented, and its synchronous nature requires fixed frame size and constant data rate [6]. Typical throughputs in SDH/SONET network are presented in Table I. The STM-0 frame structure is consisted of 90 columns and 9 rows, where basic cell of the frame is one octet, as it is presented in Fig. 1. STM frames with greater data rates have more columns (variable size of the frame) but its duration is always 125µs.

TABLE I
SDH/SONET Throughputs

SONET	SDH	Throughput (Mb/s)
OC-1	STM-0	51.84
OC-3	STM-1	155.52
OC-12	STM-4	622.08
OC-48	STM-16	2488.32
OC-192	STM-64	9953.28
OC-768	STM-256	39813.12

Is expected that the Ethernet will be the serious competitor to the other technologies applied in the transport networks, and combination of Ethernet and SDH seems to be the most economic solution for design of the regional communications networks (WAN – *Wide Area Network*). However, as most of the Ethernet packets have variable size, the SDH based transport networks are not the most suitable for its transmission. Furthermore, the networks based on Ethernet technology are connectionless oriented and number of packets transmitted in the time unit is not fixed in these networks.

In the previous period, the several practical solutions for optimization of the Ethernet packet transmission over SDH networks appear. One of them is known as OTS 166/622 IRITEL transport system, where GFP (*Generic Framing Procedure*), VCAT (*Virtual Concatenation*) and LCAS (*Link Capacity Adjustment Scheme*) mechanisms are implemented. This techniques makes possible efficient encapsulation of Ethernet packets in SDH frames, segmentation into the smaller pieces (containers) and dynamical allocation of available channel capacity inside the SDH "transport path" [7]. The more sophisticated solution, that support transmission of six independent STM-16 (2.5Gb/s) optical interfaces for distances up to 100km, based on WDM technology, is recently described in [8].

In a few previous years, a lot of effort is made to develop the interface that could support transmission of packet traffic of various nature (SDH/SONET frames, ATM, IP, Ethernet packets) over the optical transport network (OTN). The standard is accepted as IEEE recommendation G709 [9], and it is also known as Optical Transport Hierarchy (OTH) standard, that basically has to support transmission of the different services over WDM optical system.

As an example of the frame structure in OTH system, in the following sentences the structure of OTU-2 frame, with the throughput of approximately 10Gb/s (practically the same as in 10Gb-Ethernet networks) will be explained. The frame is consisted of one header octet, followed by the transmitted data (*payload*), with length of k-1=238 octets, obtaining the information word that enter the Reed Solomon code RS(255,239). This way, the subframes with length of n=255 octets are generated (one codeword corresponds to one subframe). Using the concatenation of six successive subframes, as it is shown in Fig. 2, one row of frame is generated. As it is shown in Fig. 3, one OTU frame is consisted of the four rows with length of 4090 octets.

The overview of the remaining hierarchy levels and its comparation with SDH hierarchy are given in Table II. Although the information data rate (throughput of data located in *payload* part) is identical as the line date rate in SDH/SONET network, line data rate in the corresponding OTH network is multiplied by factor 255/239, since the Reed-Solomon code (255,239) is applied. This enable the simple conversion of SDH frames into the OTU frames. For the case of OTU-2 frames, the similar operation can be performed for

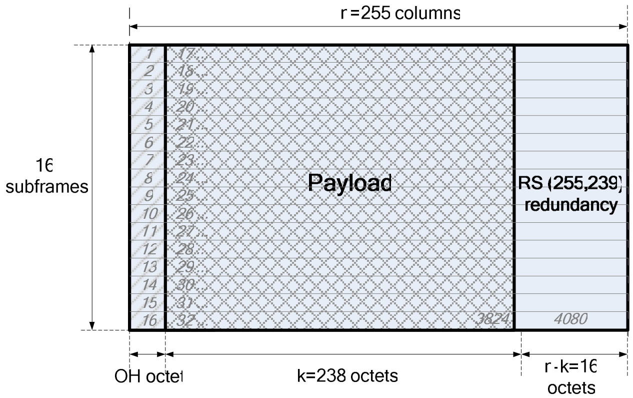


Fig. 2. The structure of one row of the OTU frame.

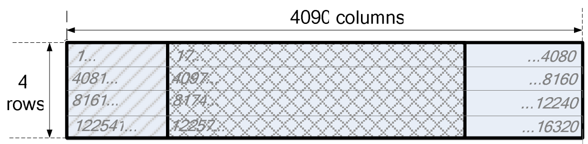


Fig. 3. The structure of the OTU frame.

10Gb Ethernet packets, although the full compatibility is not provided. On the other side, in contrary to SDH networks, with the increase of the throughput frame duration is decreased as the transmission frame size remain unchanged in all OTH hierarchy levels (4080 columns x 4 rows).

TABLE II
THE OTH THROUGHPUTS

SDH (IEEE G707)	OTH (IEEE G709)	Line data rate in OTH (Mb/s)	Information data rate in OTH (Mb/s)	OTU frame duration (μs)
STM-16	OTU-1	2666.06	2488.32	48.97
STM-64	OTU-2	10709.23	9953.28	12.19
STM-256	OTU-3	42836.90	39813.12	3.04

III. THE OPTICAL TRANSPORT SYSTEM CAPACITY

As it is already mentioned, capacity of WDM system can be calculated as a summation of individual channels capacities (in different wavelengths). In this section, we will calculate the capacity of every individual channel denoted by C, and present it normalized to bandwidth B.

In the next sections, we will assume that the symbol at the input of any individual channel in WDM multiplex can take values from the finite set, described by the constellation points. If the constellation points are denoted by $a^{(k)}$, k=1,2,...M, and x represents the send signal in the time instant of interest, we can write

$$x \in \left\{ a^{(1)}, a^{(2)}, \dots, a^{(M)} \right\}. \tag{1}$$

On the other side, it will be assumed that the decision in the receiver will be made according to the "soft decision" principle. Therefore, although the signal at the channel input has a discrete constellation, the constellation of the signal at the channel output is continual by nature and it is not compared with fixed region bounds.

Let $P(a^{(k)})$ denotes the a priory probability for k-th elementary signal at the channel input (that corresponds to one of the constellation points), while $p(y/a^{(k)})$ denotes the conditional probability density function (PDF). The last defined quantity determines the probability that send signal $a^{(k)}$ produce received signal y, that can take all possible (continual) values in range (-∞,∞). It is well known that the capacity of this channel, for the case when the channel is memoryless, can be calculated according to formula [10]

$$C/B = \max_{P(a^{(1)}),\dots,P(a^{(M)})} \sum_{k=1}^{M} P(a^{(k)}) \int_{-\infty}^{\infty} p(y/a^{(k)}) \times \mathrm{ld}\left\{ \frac{p(y/a^{(k)})}{\sum_{i=1}^{M} P(a^{(i)})\, p(y/a^{(i)})} \right\} dy. \tag{2}$$

Problem of the above expression optimization for the case of one-dimensional constellations are analyzed in details by Wozencraft and Jacobs [10], and final solution for the case of two-dimensional constellations is given by Ungerboeck in his famous paper [11]. For the practical purposes, it is extremely important that a priory probability that corresponds to constellation point are equal $\{P(a^{(k)})=1/M, k=1,2,\dots,M\}$.

The first realized optical systems used the *Intensity Modulation with Direct Detection* (IMDD), also known as *On-Off Keying* (OOK). Although this solution results in reduced spectral efficiency, it is still very popular in low-cost optical networks. This modulation type, described by one-dimensional constellations, had not shown enough efficiency for high-speed transmission (>10Gb/s), where non-linear effects highly degrade amplitude of the received signal. Therefore, the contemporary optical systems usually use coherent phase modulations. Simpler solutions are applied in systems based on differential phase modulation (DPSK) and differential quaternary phase modulation (DQPSK), with non-coherent detection [3, 12]. Recently, systems with multilevel modulation and coherent detection are proposed [13], as a tool for the significant increase of the spectral efficiency.

If we assume that only the noise resulted from *Amplified Spontaneous Emission* (ASE) effect in optical amplifiers is present in the optical channel, in the first approximation we can assume that the probability density function $p(y/a^{(k)})$ is Gaussian. In this special case, the channel capacity for two-dimensional constellations is determined by expression [11]

$$\frac{C^*}{B} = \log_2 M - \frac{1}{M}\sum_{k=1}^{M} E\left\{ \log_2 \sum_{i=1}^{M} \exp\left[-\frac{\left|a^{(k)}+n-a^{(i)}\right|^2 - |n|^2}{2\sigma^2} \right] \right\}. \tag{3}$$

The measurements in available optical systems show that, with throughput of 10 Gb/s, the intersymbol interference due to chromatic aberration also exist in the channel. With throughputs greater than 40 Gb/s, due to high nonlinearity if optical fiber it cannot be neglected cross-talk inside the channel and between the channels (IXPM - *Interchannel Cross-Phase Modulation between pulses*, IFWM -

Interchannel Four-Wave Mixing), and *Polarization Mode Dispersion* (PMD) [2,4,12].

In the available literature, it can be found several approaches for the design of simulation model that could take into the account all mentioned channel effects. In paper [14] these effects are jointly modeled using the channel with memory (impact of the intersymbol interference), and in paper [15] the more sophisticated simulation model that include most of the mention effects is developed.

In this paper, we concentrate to the channel model that could be used for efficient calculation of the channel capacity and Monte Carlo simulation on the basis of modulation channel. This model could be used for the estimation of the bit error probability, when the modulation and error control code parameters are defined. For this purpose, it is necessary to know a-posteriori probability density function for the corresponding channel, denoted by $p(y/a^{(i)})$ and defined in eq. (2). The measurement results in previously realized optical systems have shown that $p(y/a^{(i)})$ in this case does not perfectly fit to normal (Gaussian) distribution. For the data rate less than 10 Gb/s, the dominant interference is due to the chromatic dispersion, and channel can be described using asymmetric Gaussian distribution [16]. More precise results could be obtained if the received signal is described by chi-quadrate distribution [17].

On the other side, as a quality measure in optical channel, it is usually used the Q-factor (instead of signal-to-noise ratio). This quantity takes into account noise as well as the other effects that degrade the optical transmission. For various conditional PDFs $p(y/a^{(k)})$, that can even be different for individual constellation points $a^{(k)}$, we can always determine corresponding mean values μ_k and standard deviations σ_k. For the case of binary intensity modulation, the Q-factor is defined by expression [2]

$$Q = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2}, \quad (4)$$

and similar identities can be written for the case of the multilevel modulations. In paper [14], it is described how the variables necessary for calculation of Q-factor can be experimentally determined, when sum of the noise and the overall and interference is chi-square distributed. The first step in the capacity calculation is generation of the complex Gaussian noise with enough size. The summation of quadrate values of the every sample real and imaginary parts is chi-quadrate distributed, and it can be used for statistical averaging of the channel output signal.

In Fig. 5, the capacity that corresponds to the system where M-PSK constellation is applied (for the cases M=2, 4, 8 and 16). It can be noticed that for the bit error rate BER=10^{-6}, using 16-PSK, the spectral efficiency 4b/s/Hz can be achieved for Q-factor of 21dB, while the intensity modulation with sixteen levels (16-IMDD) require Q factor of 29dB, for the same spectral efficiency.

On the other side, the Shannon formula anticipate that the same spectral efficiency (with arbitrary low bit error rate during transmission) can be achieved for Q=12dB. In Fig. 4, it

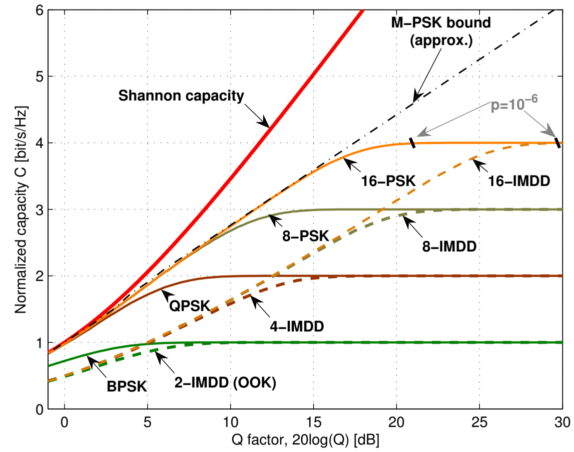


Fig. 4. Channel capacity in WDM system with intensity and M-PSK modulation.

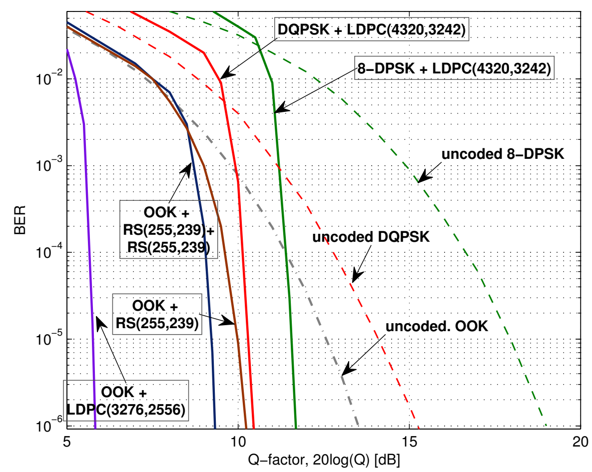


Fig. 5. Performance of WDM system with binary intensity and DPSK modulation.

is also shown the bound that corresponds to approximate expression for system capacity, when number of the M-PSK constellation points grows toward the infinity. It is known that, for the greater values of Q-factor (or the signal-to-noise ratio) capacity can be approximated with expression [12]

$$C/B \le ld(Q)/2 + 1.1 . \quad (5)$$

It is clear that the gap between the Shannon capacity (without a limitation to constellations with the same average power) and capacity for M-PSK given with the above expression became wider for the greater spectral efficiency. For the case of the intensity modulation, this effect became more noticeable.

IV. The Efficiency of Error Control Codes in the Optical Transport Systems

In the previous section, estimation of the optical transport system capacity is performed for the case when the noise and chromatic dispersion is present. It is also assumed when the transmission is based on (binary or multilevel) intensity modulation or on M-PSK modulations with constant transmit power. In this section, performances of the OTH optical transport systems are considered for the specified modulation formats.

The bit error rate represents typical performance measure, and in this section it is estimated using Monte Carlo simulation. In simulation, the information blocs with size of 810 bytes (corresponds to payload and header in STM-0 frame) are generated, as it is shown in Fig. 1. For the system with the binary intensity modulation (OOK) with applied error control code (n,k), this sequence is repacked in the structure shown in Fig. 2. We assume that it is not necessary to keep parameters n=255 and k=239, but all the other parameters of the OTU frame remain unchanged. Finally, the binary sequence is transformed into unipolary signal without return to zero (*Not Return to Zero* - NRZ). Then, according to described simulation model, the additive Chi-Square distributed interference (that model the influence of ASE and the chromatic dispersion) is generated and added to NRZ signal. At the receiver, the "direct detection" principle is applied [2].

The numerical results are obtained using the estimation in N=1000 transmitted STM-1 frames and presented in Fig. 5. If we apply RS(255,239) code (recommended by standard [9]), the coding gain of 4dB is detected for on-off keying and the error rate BER=10^{-6}. However, for the code rate R=0.93 and OOK modulation, this result is more than 5dB away from Shannon bound!

Additional improvement can be achieved using the concatenation of RS code, and the LDPC(3276,2586) code achieve coding gain of 9dB, about 2dB away from the Shannon bound that correspond to the code rate R=0.79.

We further considered the combination of the error control code LDPC(4320,3242) and M-PSK modulations with non-coherent detection, proposed in paper [12]. The corresponding numerical results are presented in Fig. 5. The most interesting results are obtained for coded 8-DPSK, where, for error rate BER=10^{-6}, the coding gain of 7dB is achieved compared to the uncoded 8-DPSK, and about 3dB compared to the uncoded DQPSK. In this system, for the reason of non-coherent detection, spectral efficiency of 2b/s/Hz is achieved for Q-factor about 6dB greater then minimal theoretical value, determined by the Shannon limit.

V. CONCLUSION

In this paper, the overview of technologies suitable for transmission of packets through optical transport network, and the special attention is dedicated to impact of error control coding techniques use in current systems. We proposed the simulation model, which can be used for modeling of typical imperfections in data transmission through the considered optical WDM system. Using the chosen simulation model, estimation of channel capacity is performed for systems with intensity and phase modulation. Using the simulation model, it has been shown that applying the considered error control codes can simultaneously increase spectral efficiency and reduce bit error rate, improving the service quality in optical transport networks.

REFERENCES

[1] S. Lin, D. J. Costello, Error Control Coding, Second Edition, Prentice Hall, New Jersey, 2004.
[2] B. Vasic, I. Djordjevic, “A forward error correction scheme for ultra long haul optical transmission systems based on low-density parity-check codes”, In Proc. IEEE ICC 2003, pp. 1489 - 1493, May 2003.
[3] M. Jinno, Y. Miyamoto, Y. Hibino: “Optical-transport networks in 2015”, Nature photonics (Technology focus), pp. 157-159, Mar. 2007.
[4] D. M. Gvozdić, “Trendovi razvoja optičkih telekomunikacionih sistema”, Telekomunikacije – naučno stručni časopis Republičke agencije za telekomunikacije, Vol 1, Novembar. 2008.
[5] C. E. Shannon, “A Mathematical Theory of Communication”, Bell Syst. Tech J., Vol. 27 (1948), pp. 379-423, 623-656.
[6] ITU-T recommendation G.707, “Network node interface for the synchronous digital hierarchy (SDH)”, Ženeva 2001.
[7] D. Pešić, V. Kostić, P. Mićović, M. Ilić, P. Knežević, N. Radivojević, Z. Čiča, P. Ivaniš “OTS155/622 IRITEL – Prenos Ethernet paketa preko SDH mreže”, TELFOR 2005, Beograd, Nov. 2005.
[8] P. Mićović, D. Pešić, M. Ilić, V. Kostić, N. Radivojević, R. Đenić, “Multiservisna SDH/Ethernet/CWDM/OADM platforma ODS2G5 IRITEL”, TELFOR 2008, Nov. 2008.
[9] ITU-T preporuka G.709, “Interfaces for the Optical Transport Network (OTN)”, Ženeva 2001.
[10] J. M. Wozencraft, I. M. Jacobs, Principles of Communication Engineering, John Wiley & Sons, New York, 1965.
[11] G. Ungerboeck, “Channel Coding with Multilevel/Phase Signals”, IEEE Trans. Inf. Theory, Vol. 28, pp. 55-67, 1982.
[12] I. B. Djordjevic and B. Vasic, “Multilevel coding for spectrally efficient noncoherent optical transmission”, Proc. IEEE ICC 2006, pp. 1211 - 1216, Jun. 2006.
[13] P. J. Winzer, R. J. Essiambre: “Advanced Modulation Formats for High-Capacity Optical Transport Networks”, IEEE/OSA Journal of Lightwave Technology, Vol. 24, pp. 4711-4728, 2006.
[14] I. Djordjevic, B. Vasic, “An Advanced Direct Detection Receiver Model”, Journal of Optical Communications, vol. 25, no. 1, pp. 6-9, Feb. 2004.
[15] M. Ivkovic, I. Djordjevic, P. Rajkovic, B. Vasic, “Modeling errors in long-haul optical fiber transmission systems by using instantons and edgeworth expansion”, In Proc. IEEE ICC 2007, pp. 2352-2356, June 2007.
[16] Y. Cai, N. Ramanujam, J.M. Morris, T. Adali, G. Lenner, A.B. Puc, A. Pilipetskii, “Performance limit of forward error correction codes in optical fiber communications”, In Proc. OFC 1999, pp. TuF2-1-3.
[17] I. Djordjevic, B. Vasic, M. Ivkovic, I. Gabitov, “Achievable Information Rates for High-Speed Long-Haul Optical Transmission”, J. Lightwave Technology. Vol. 23, pp. 375.

Automatic Emotion Recognition in Speech: Possibilities and Significance

Milana Bojanić and Vlado Delić

Abstract—**Automatic speech recognition and spoken language understanding are crucial steps towards a natural human-machine interaction. The main task of the speech communication process is the recognition of the word sequence, but the recognition of prosody, emotion and stress tags may be of particular importance as well. This paper discusses the possibilities of recognition emotion from speech signal in order to improve ASR, and also provides the analysis of acoustic features that can be used for the detection of speaker's emotion and stress. The paper also provides a short overview of emotion and stress classification techniques. The importance and place of emotional speech recognition is shown in the domain of human-computer interactive systems and transaction communication model. The directions for future work are given at the end of this work.**

Index Terms—**Emotional speech recognition, stress, ASR.**

I. Introduction

THE advent of new technologies and the introduction of interactive systems have increased the demand for more natural man-machine communication − in the way people communicate among themselves. Using voices people express their emotions. In order to achieve more natural verbal communication between humans and machines, it is necessary to recognize the mood of the speaker during automatic speech recognition (ASR) as well as to generate emotionally "colored" speech during its synthesis (text-to-speech systems, TTS). During the ASR, the recognition of emotion can be useful for proper handling of a man-machine dialogue. For example, detection of speaker's impatience, irritability or frustration, will help to appropriately redirect the dialogue (ticket reservation systems, call centers) [1].

Emotional speech recognition (ESR) strives towards automatic identification of emotional or psychological state of the individual based upon analysis of individual's speech. The mood and emotional condition of the speaker belong to paralinguistic aspects of verbal interaction. Analysis of these elements of verbal man-machine communication is necessary for successful implementation of ASR, regarding spontaneous dialogue between humans and machines [2].

This paper represents a part of a research which is supported through the technological project of the Ministry of Science of Republic Serbia, "Human–machine Speech Communication" TR11001.

M. Bojanić and V. Delić are with the Faculty of Technical Sciences, Trg Dositeja Obradovića 5, Novi Sad, Serbia.

Even though the emotional state can be manifested at the semantic level, the emotional content of the speech is contained in prosodic features to a considerable extent. While the classic ASR is based on correct recognition of word strings, for natural language processing and dialogue systems it is necessary to understand the context of the speech. In that case, the prosody and the emotional content can play an important role.

First studies in this field were conducted in the mid 80's of the last century using statistical properties of certain acoustic features. Subsequent progress of computer architectures allowed the use of more complex algorithms for ESR implementation. Today's studies are focused on combining classifiers in order to improve the efficiency of emotion classification within real applications [3].

There are three types of speech in emotional speech data collections [4]: natural, simulated (acted) and elicited speech. Natural speech is spontaneous speech where all emotions are real. Simulated speech is speech most often expressed by professional actors in order to invoke certain emotional states. In elicited speech the emotions are induced, for example by showing an adequate audiovisual material to the examinees. Simulated speech is most reliable for ESR because the professional actors express emotions vividly, with large amplitude and great power. Additional signals recorded during the data collection are (most often) laryngograph, heart beat rate, blood pressure, and the facial expressions of the speaker [3]. The majority of emotional speech data collections contains 5-6 emotional states alongside with neutral speech. The most frequent emotions are: anger, fear, joy, sadness, disgust, surprise, boredom and similar. Data collections can contain radio and TV excerpts and also recorded conversations with psychologists and phoneticians.

The acoustic features which reflect emotions in speech signal will be presented in the next section. The third section will then briefly explain the techniques of emotion classification, while the fourth section will cover characteristics of speech under stress. The fifth section shows the role and significance of emotion detection with regard to the transaction communication model, including the appropriate example of human–machine dialogue. The sixth section represents the conclusion of this paper and proposes directions for further research.

II. Acoustic Features of Emotional Speech

Emotions in speech are expressed through variation of speech characteristics on three levels: (1) prosodic or suprasegmental level through specific frequency, intensity and duration changes, (2) segmental level (changes in articulation quality) and (3) intrasegmental level (general voice quality, whose acoustic correlates are glottal pulse shape and distribution of its spectral energy, amplitude variations (shimmer), frequency variations (jitter)) [4]. Special attention will be paid to short-term acoustic features used in ESR. Short-term features are derived on frame basis:

$$f_s(n;m) = s(n)w(m-n) \tag{1}$$

where *s(n)* is the speech signal and *w(m-n)* is the window of length $N_{w.}$

Furthermore, several such features will be described, of which, according to many authors, the most important are pitch (fundamental frequency), energy of speech signal, vocal tract features and harmonics in the spectrum.

Pitch, fundamental frequency of phonation F_0, is the vibration rate of vocal cords during phonation. The emotional state of speaker affects the tension of vocal cords and the subglottal air pressure, which ultimately affects the pitch. This is why many authors consider it as the most important prosodic feature for ESR [5]. Numbers of algorithms have been developed for pitch estimation. Here will be presented the widely spread autocorrelation method [6]. First, the signal is low filtered at 900Hz, and then it is segmented to short time frames of speech which are later clipped. The clipping is a nonlinear procedure which should prevent the influence of the first formant on the pitch. The clipped signal looks like this:

$$\hat{f}_s(n;m) = \begin{cases} f_s(n;m) - C_{thr} & if \ |f_s(n;m)| \geq C_{thr} \\ 0 & if \ |f_s(n;m)| < C_{thr} \end{cases} \tag{2}$$

C_{thr} is set at 30% of the maximum value of $f_s(n;m)$. After calculation of autocorrelation value

$$r_s(\eta;m) = \frac{1}{N_w} \sum_{n=m-N_w+1}^{m} \hat{f}_s(n;m)\,\hat{f}_s(n-\eta;m) \tag{3}$$

the fundamental frequency (*pitch*) of the frame which ends at moment m can be estimated by

$$\hat{F}_0(m) = \frac{F_s}{N_w} \ \arg\max_\eta \left\{ |r(\eta;m)| \right\}_{\eta = N_w(F_l/F_s)}^{\eta = N_w(F_h/F_s)} \tag{4}$$

where $F_s = 8000Hz$ is the sampling frequency, while F_l and F_h are the lowest and the highest pitch frequencies which humans can perceive. Their typical values are 50 Hz and 500 Hz, respectively. It is necessary to note that this is a relatively wide bandwidth which includes several octaves. This fact additionally complicates the task of automatic detection of fundamental frequency i.e. the pitch and allows mistaken detection of twice as high, or twice as low F_0. Except by the pitch, glottal waveform is characterized by air velocity through glottis, whose measurement is based on the maximum value of autocorrelation of clipped signal frames. One of the insufficiently studied topics is the shape of glottal waveform which is evidently associated with emotional coloring of the speech.

TEO (*Teager Energy Operator*) – on the occasions of emotional states of anger or speech under stress, fast and nonlinear air flow causes vortices located near vocal cords providing additional excitation signals beside the pitch. These additional excitation signals are present in the spectrum as harmonics and cross-harmonics. TEO [7] for signal frame is calculated as follows:

$$\Psi[f_s(n;m)] = (f_s(n;m))^2 - f_s(n+1;m)f_s(n-1;m)\,. \tag{5}$$

When applied on AM-FM sinewave it gives a squared product of instantaneous amplitude and instantaneous frequency of the signal:

$$\Psi[f_s(n;m)] = \alpha^2(n;m)\sin(\omega_i^2(n;m))\ . \tag{6}$$

In the case when the signal has a single harmonic, TEO operator gives a constant number, otherwise it is a function of discrete time n. Since the speech signal contains more than one harmonic in the spectrum, it is more convenient to split the bandwidth into smaller bands and then observe each one independently. The polynomial coefficients which describe the TEO autocorrelation envelope area could be used for classifying the emotional speech [8]. This method achieves 89% of accuracy when classifying neutral speech versus speech under stress. The pitch frequency affects the number of harmonics in the spectrum, so that more harmonics are present in the spectrum when the pitch frequency is low. This effect, as well as observations that the harmonics from additional excitation signals are more intense than those caused by the pitch, can be a subject of further research.

Vocal tract features – The shape of the vocal tract is changed under the influence of the emotional state of the speaker. The features which describe the shape of the vocal tract during the emotional speech production are: formants, cross-section areas of the tubes modeling the vocal tract, and coefficients derived from frequency transformations of the speech signal. Formants, as emphasized parts of the spectrum (spectral peaks), reflect locations of the vocal tract resonances. Their position in the spectrum (formant center frequency) and their bandwidth depend upon the shape and dimensions of the vocal tract, and those change depending on emotional state of the speaker. Experimental analysis [3] has shown that the emotional state largely influences the first and the second formant. MFCC (*Mel-frequency Cepstral Coefficients*) represent the signal spectrum in frequency bands which correspond to human auditory frequency response (Mel frequency scale). These coefficients are used in some emotion classifications, but better results are achieved with LFPC (*Log-frequency Power Coefficients*) which include the pitch information.

Energy of speech signal – the short-term speech energy is directly related to level of emotions in speech, therefore it can be efficiently applied in algorithms for emotion recognition. The short-term energy of the speech frame ending at m is:

$$E_s(m) = \frac{1}{N_w} \sum_{n=m-N_w+1}^{m} |f_s(n;m)|^2 \tag{7}$$

Contours of short-term acoustic features – The contours and their trends (rising and falling slopes, plateaux) also provide information useful for emotion recognition. These contours are formed when the value of some feature, calculated on the frame level, is assigned to all samples belonging to that frame. For example, the energy contour is calculated as follows:

$$e(n) = E_s(m), \qquad n = m - N_w + 1, \ldots, m \tag{8}$$

Frequently used statistics for extracted features and their contours are: mean value, variance and pitch contour trends, mean and range of the intensity contour, rate of speech and transmission duration between utterances. For example, in the state of anger it is characteristic that speech possesses high energy (especially male voices) and the high pitch level (F_0). Table I shows that compared to men, and under the similar circumstances women express anger with higher speech rate. Emotional state of sadness corresponds to lower pitch frequency F_0, lower intensity in relation to neutral speech as well as falling slope in pitch contour. Men express sadness with higher speech rate as opposed to women.

TABLE I
INFLUENCE OF SEVERAL EMOTIONS ON SELECTED PROSODIC FEATURES IN RELATION TO NEUTRAL SPEECH (ADAPTED FROM [3])

	Pitch				Intensity		Duration	
	Mean	Range	Variance	Contour	Mean	Range	Speech rate	Transmission Duration
Anger	>>	>	>>		>>M >W	>	<M >W	<
Disgust	<	>M <W			<		<<M <W	
Fear	>>	>		↘	≥			<
Joy	>	>	>	↗	>	>		<
Sadness	<	<	<	↘	<	<	>M <W	>

>: greater, <: less than neutral, ↗: increasing and ↘: decreasing contour trend, m: men and w: women.

III. EMOTION CLASSIFICATION TECHNIQUES

The classification techniques can be divided into two groups:

- Techniques which use prosody contours (sequence of short-time prosody features)
- Techniques which use statistics of prosody contours (the mean, the variance, etc.)

The classifier output should evaluate, or give an assumption of emotion content of some statement, word, phrase.

The classification techniques which use prosody contours are: (1) the technique of artificial neural networks (*ANNs*), (2) the multi-channel hidden Markov model (*multi-channel HMM*), (3) the mixture of HMMs. In the technique based on ANNs, the short-time features are extracted for the frames grouped into phoneme groups. ANN is being trained on the *k*th emotional state of the *m*th phoneme group. The output node gives the likelihood of the part of some frame given specific emotional state and phoneme group. The multi-chanell HMM actually represents a group of several Markov chains which independently model the speech for certain emotional state. Transitions are possible within the single chain as well as between the chains which correspond to certain emotions. This technique achieved a correct speech recognition rate of 94.4%, and 57.6% for stress classification (this low result is attributed to insufficient training data collection) [3]. The third technique, so called the mixture of HMMs consists of two phases. The first phase uses iterative clustering algorithm in order to obtain M clusters in the feature space of training collection. Then, the second phase involves training of C HMMs where every HMM corresponds to single emotional state $c = 1, 2, C$. The correct classification rate of 4 emotional states using mixture of HMMs was 62% using energy contours in different frequency bands [3].

The classification techniques which use statistics of prosody contours can be divided into two subgroups: (1) those which use the estimation of the probability density function (pdf) of the features, and those (2) which do not use the estimation of pdf of acoustic features. The Bayes classifier belongs to the first subgroup, using different methods for $P(y|\Omega_c)$ estimation (conditional distribution of the feature vector given the emotional state Ω_c). These class pdfs are modeled as Gaussians, mixtures of Gaussians, or via Parzen windows. The other subgroup contains classifiers: (a) the k-nearest neighbors, (b) the support vector machines and (c) the artificial neural networks *(ANN).*

IV. SPEECH UNDER STRESS

Many researchers narrow down the notion of emotions and their variety to the speech under stress. In that case the task is not to recognize the particular emotion of the speaker, but to binary classify if the speech is under stress or not. The speech under stress can be a result of the outside factors (pressure at work, danger) and/or the emotional states (fear, anger, anxiety, excitement…), and it manifests itself through speech changes in relation to neutral speech (situations without stress and pressures), through the modus of speech (stuttering, tongue-slip..), the selection and the usage of certain words, sentence duration and articulation of phonemes. These speech changes are consequence of physiological changes when a person is under stress. They consist of increased respiration rate, increased subglottal pressure, increased F_0, dry mouth, changes in muscles of larynx and vibrations of vocal cords [9]. There are numerous situations and jobs which change speaker's physical and mental condition in the way it affects the implementation of ASR. These include: police officers, fire-fighters, emergency services, air traffic controllers, pilots in noisy environment, deep sea divers, astronauts, nuclear plant operators and others.

Here will be presented the results obtained by CRSS group of researchers [9] while studying changes of speech characteristics for different speaking styles and stress conditions (fast and slow speech, quiet and loud, anger, Lombard effect, speech during moderate and high computer workload tasks, neutral speech). Within the speech characteristics analysis for the speech under stress, the following were considered: fundamental frequency (pitch), intensity, duration, formant locations, spectral slope. The fundamental frequency is a good stress indicator in the wide range of stress conditions, especially angry speech, Lombard effect and loud speech. Observed characteristics for F_0 are contour, mean, variance and distribution. The mean and the variance are significantly different for neutral speaking style in relation to other styles.

In order to improve existing ASR algorithms, it is necessary to detect the period of speech under stress. Prior to stress detection, it is necessary to estimate the acoustic features from the input signal, and then detect periods of speech under stress and periods of neutral speech. There are several ways to perform the stress detection: detection-theory-based method, methods based on distance measure and methods of statistical modeling [9]. For given input feature vector x, two conditional probability density functions are estimated, $p(x|H_0)$ and $p(x|H_1)$, that the input signal belongs to a neutral speech and that the input signal belongs to the speech under stress, respectively. Comparing the ratio of these two densities with chosen threshold (depending on particular criteria and application), according to detection theory the decision of whether the input speech is stressed or not is made. The distance measure method reflects the distance of the input feature vector x, in relation to feature distribution for neutral speech, and in relation to feature distribution for speech under stress.

V. Emotion Recognition and the Transaction Communication Model

The knowledge of the variations in acoustic features during emotional speech expression, or speech under stress, may improve human-machine communication which is deteriorated or even disabled under those conditions. The computer may have problems in speech recognition due to acoustic features affected by presence of emotions and stress. Although, from the aspect of ASR, those changes do not carry a linguistic message, they do affect the accuracy of recognition. On the other hand, employing the knowledge about variations in acoustic and prosodic features into the TTS module, it would achieve better quality and more natural synthetic speech.

For the purpose of better understanding, the machine model in human-machine speech interaction [2] is envisioned as a combination of information from two sources: (1) by processing the speaker picture, and (2) by processing his/her speech, as shown in Fig. 1.

Audiovisual nature of speech perception in the human-machine dialogue is reflected in two phenomena. First, watching the speaker, especially speaker's face and lip movement which are synchronized with articulated speech,

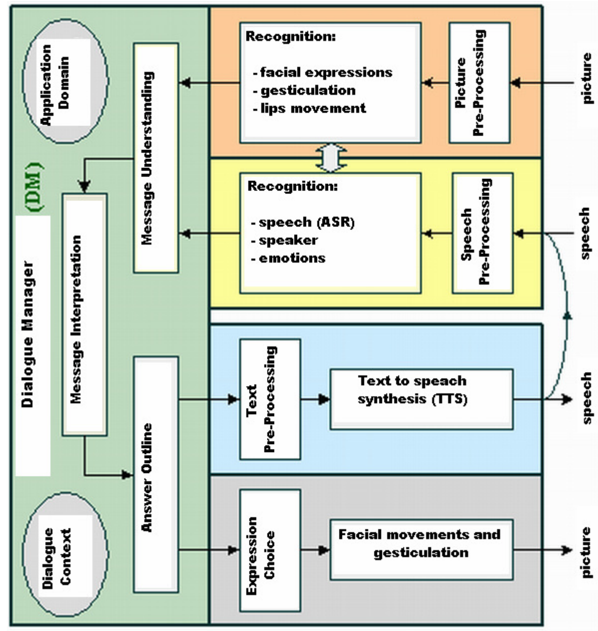


Fig. 1. Functional model of the machine in the verbal dialogue. (taken from [1]).

enhances and facilitates the perception of speech, especially in the noisy environment or in situations of lower speech intelligibility.

The other example is the McGurk effect, which denotes changes in auditory speech perception in case when the lip movement does not match spoken words [10]. Besides its verbal expression, emotions have their form of characteristic facial expressions. For example, happiness is followed by a smile on face. Within the framework of multimodal human-machine communication, the machine listens with the assistance of ASR module. The implementation of ESR algorithm (emotion recognition) into ASR module will increase the reliability of speech recognition and ensure natural and spontaneous human-machine dialogue. The module which processes the picture of human face (lips movement, facial expressions and gesticulations) provides the information which are video correlates of spoken phonemes, words and expressed emotions as well. The synergy of these two modules complements the semantics of spoken words, while verbal interaction is complemented with paralinguistic and nonverbal elements of dialogue, as shown in Fig. 1.

Based on extensive prior knowledge about application and language, the machine performs postprocessing with purpose of spoken language understanding (SLU), that is, the integration of speech recognition and natural language understanding (NLU). This would also be valid in the opposite direction, from machine to the human, if the TTS module would be able to synthesize emotionally "colored" speech, and along with the adequate animation of the speaker's face in output picture, it would lead to a specific illusion of having a conversation with other "person", i.e. having a natural communication.

Transaction communication demands from its participants to recognize the influence of one message on the other and broaden the field of mutual experience through active process of language understanding [2]. For better understanding of meaning and intentions of the utterance, it is necessary to consider not only its verbal elements but also its nonverbal elements. In that sense, the automatic emotional speech recognition could have an important role because the emotional state of the speaker points out his/her needs as the conversational partner and his/her reactions towards current dialogue, and so helps proper dialoge handling or accommodates the initialized communication.

Here is the example of human-machine dialogue where the machine recognizing the key words in the user's answers as well as the level of his satisfaction (based on recognition and emotion classification), directs the dialogue and gives answers in order to successfully complete communication. This example is shown in Table II. The machine has understood the need of the person after the first sentence – the ticket reservation to city of Niš, and then directed the question to obtain more data (the date and the time of the trip). While answering, the person has shown some degree of uncertainty and confusion, and then has corrected himself/herself. The machine "understood" this using ESR algorithm, accepted different timing as requested, and then provided the answer. The machine has correctly connected the time specification "day after tomorrow" to a concrete date in order to avoid misunderstanding. The next user's answer, "yes, yes, yes" was detected by using emotion detection and classification as a delight and satisfaction of the person. Comfort and natural dialogue was achieved with additional pronunciation instructions to the machine. The effect of emotionally "colored" speech was achieved by varying prosodic features within the synthesized speech.

TABLE II
THE EXAMPLE OF THE DIALOGUE BETWEEN THE MACHINE (M) AND THE HUMAN (H)

M 1	Hallo, **good day!** [a bit faster] This is service of inter-city bus station Novi Sad. [short pause, slower] **How can I help you?**
H 1	*I would like to make a reservation to city of Niš.*
M 2	[clear and slowly] Please, give me the **date** and the **time** of your travel.
H 2	*Day after tomorrow around 6 o'clock, er, 16 o'clock.* // recognized confusion and the subsequent correction by the user
M 3	[clear] For March 10^{th} there is a departure at 16:15, Nišekspres. [polite and inquiring] **Does it** suit you?
H 3	*Yes, yes, yes….* // recognized user's satisfaction
M 4	[] All **right.** Your reservation has been confirmed, to city of Niš, [slow], **March 10^{th}** at **16:15.** [inquiring] Can I help you with anything else?
H 4	*No*, thanks.
M 5	[joyfully] Have a nice day and **bon voyage**!

Bold text should be emphasized. Text in the brackets gives instructions for speaking style and intonation. Words in *italics* are the keywords for dialogue flow.

With the correct word recognition and their semantics, the recognition of user's emotional state allows correct understanding of the user's message and needs, confirming the success of the dialogue and achieving the goal of the conversation.

VI. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

This paper deals with several topics which are related to automatic emotional speech recognition. Firstly, the extraction of relevant acoustic features within speech has been presented as well as space of their variations oposed to neutral speech. In order to do such task it is necessary to have formed speech databases whose diversity and size are still unsatisfactory, especially for Serbian language.

The review of some widespread classification techniques is given, but the problem which remains is that their accuracy cannot be directly compared because the results were obtained using different databases and experiment protocols. It is necessary to detect the speech under stress, as a way of expressive speech, in order to improve existing ASR algorithms. One of the general approaches is the equalization of stress, that is, normalization of parameters variability due to presence of the stress within the speech signal.

The importance and application of emotion recognition has been emphasized within the human-computer communication. It contributes to higher accuracy of recognition within the ASR module, and within the human-machine dialogue it helps better understanding of the meaning of the message, and also needs and intentions of individuals. For TTS module knowledge of characteristic prosodic variations will give the possibility that machine speech can be as close to the human speech.

One of the future research directions refers to ESR algorithm implementation within the ASR module as well as their integration with the module for processing face pictures. This would be a step forward in aspirations to model machine as an audiovisual "conversational partner" within the multimodal human-machine communication.

REFERENCES

[1] L. Bosch, Emotions, speech and the ASR framework, Speech Communication 40, pp. 213-225, 2003.
[2] V. Delić, M. Sečujski, "Transaction model of human–machine verbal interaction", DOGS, Kelebija, 2-3.10.2008, pp. 8-15.
[3] D. Ververidis, C. Kotropoulos, I. Pitas, Automatic emotional speech classification, In. Proc. 2004 IEEE Int. Conf. Acoustics, Speech and Signal Processing, vol. 1, pp. 593-596, Montreal, 2004.
[4] S. T. Jovičić, Z. Kašić, M. Đorđević M. Vojnović, M. Rajković, J. Savković, Forming corpus of emotion and attitude expression in Serbian language-GEES, XI Telekomunikacioni forum TELFOR 2003, Beograd, 25-27.11.2003.
[5] Y. Li, Y.Zhao, Recognizing emotions in speech using short-term and long-term features, In: Proc. ICSLP 1998, pp. 2255-2258, 1998.

[6] M. M. Sondhi, New methods of pitch extraction, IEEE Trans. Audio and Electroacoustics 16, pp. 262-266, 1968.

[7] H. M. Teager, S. M. Teager, Evidence for nonlinear sound production mechanisms in the vocal tract, NATO Advanced Study Institute, Series D, vol. 15, Boston, MA: Kluwer, 1990.

[8] G. Zhou, J. H. L. Hansen, J. F. Kaiser, Nonlinear feature based classification of speech under stress, IEEE Trans. Speech and Audio Processing 9 (3), 201-216, 2001.

[9] C. Műler (editor), Speaker Classification I, LNAI 4343, pp. 108-137, Springer-Verlag Berlin Heidelberg, 2007.

[10] H. McGurk, J. MacDonald, Hearing lips and seeing voices, Nature 264, pp. 746-748, 1976.

Analysis of Geometry Influence on Performances of Capacitive Pressure Sensor

Mirjana Maksimović and Goran Stojanović

Abstract—**In this work analysis of performances of a capacitive pressure sensor is performed, using COMSOL software tool. The basic sensor structure is composed of two cavities separated with a thin membrane. One cavity contains reference pressure (vacuum) and the other is connected with the measured pressure. When this pressure is changed, membrane is deformed depending on several factors such as applied pressure, used materials mechanical properties and geometry of the sensor structure. Influence of these parameters on sensor characteristics is analyzed and capacitance as a function of applied pressure plots is presented. It is demonstrated that COMSOL is a powerful software tool for successful prediction of performances of capacitive pressure sensors with different geometries, even in the cases with dynamics such as when sensor membrane is deformed.**

Index Terms—**Capacitive pressure sensor, COMSOL, geometry influence.**

I. Introduction

MOST electrical sensors usually use membranes as primary elastic element. Applied membranes are suitable for measuring the pressure from its lowest values to the highest. Measurement range, operating frequency and the sensor sensitivity depend on the characteristics of the primary element. Deformation of elements that occurs due to effects of pressure (differential pressure), is further converted into electrical output signal. According to that, sensors can be divided into: electromagnetic, capacitive, resistive and piezoelectric sensors.

Principle of the capacitive sensor operation is that it uses a membrane as movable capacitor electrode. Typical measurement range of this sensor is from 100 Pa to 10^8 Pa, and the accuracy is ± 0.25-0.05%. Some disadvantages of capacitive pressure sensors are: the capacity and the movement of outer lines affect the output signal distortion; high output impedance must be balanced, the sensitivity on the change of temperature and required shielding of the connection cables. On the other hand, good features are: high frequency permeability, manufacturing process simplicity, low costs, the possibility of measuring static and dynamic changes, the small membrane mass, small volume and continual resolution [1].

Pressure sensors are used in various automotive, biomedical and other industrial applications [2]. Since 1980, a great number of publications has appeared having silicon capacity pressure sensors as the topic. Thus, for example, blood pressure measurement sensors were developed [3, 4], a different forms of membrane geometry were discussed (circular, rectangular, square) [5], and also the characteristics of differential pressure sensors were considered [6].

In this work, the analysis of the capacitive pressure sensor that measures the static pressure in the range from zero to atmospheric was performed. Performance analysis of the proposed sensor was carried out using COMSOL [7] software tool for different geometry of the sensor structure and different membrane thicknesses.

II. Capacitive Pressure Sensor Structure

The basic sensor structure is composed of two cavities separated with a thin membrane. One cavity contains reference pressure (vacuum) and the other is connected with the measured pressure. When the pressure changes, the membrane is deformed and the magnitude of the deformation depends on several factors: the pressure amount, the mechanical properties of material, and the structure shape [7].

Any initial stress in the material also affects the deformation. Therefore, the manufacturing process and the selected materials directly affect the sensor operation. In some structures the membrane and cavities are engraved into silicon and sealed with glass layers. Because the materials are bonded together at a high temperature, cooling them down to the sensor's normal operating temperature produces undesirable stress in the material that affects device performance.

A common way of detecting the membrane deformation is the capacitance measuring. The surface of the deforming membrane and the opposite side of one of the cavities are coated with metal. Thus they form a capacitor whose value depends on the distance of the plates and on the system geometry. Hence, deflection of the membrane results capacitance changes between movable and fixed electrodes.

COMSOL software tool has been used for electric parameters and simulation analysis as a function of the applied pressure. COMSOL Multiphysics is a powerful interactive

This work was financially supported by the Provincial Secretariat for Science and Technological Development within the project "Realization of high-performance micro sensors for operation in extreme environmental conditions", number of project: 114-451-01009/2008-02.

M. Maksimović is with the Faculty of Electrical Engineering, University of East Sarajevo, East Sarajevo, Bosnia and Herzegovina.

G. Stojanović is with the Faculty of Technical Sciences, Novi Sad, Serbia.

environment for modeling and solving all kinds of scientific and engineering problems based on partial differential equations (PDEs), and its use does not require a deep knowledge of mathematics or numerical analysis, but the models are built on the basis of adequate physical characteristics equations. For a capacitive pressure sensor analyzed in this work the model first computes the initial stresses from device construction; then it accounts for the structure's mechanical deformation resulting from an applied pressure. It finally calculates the sensor's capacitance for the deformed shape from the electric field. Often the deforming membrane is a circular or rectangular diaphragm fixed at all boundaries. But viewing this structure in 2D, a bridge type structure results, which is fixed only at the two edges. Fig. 1 and Fig. 2 illustrate the analyzed model's geometry.

As shown on Fig. 1 and Fig. 2 the model consists of three layers. The active silicon structure sits between two blocks of glass. The following list provides descriptions of the different structures in the sensor:

• Top and bottom layers
  - Rectangular
  - Material: Glass, HOYA, SD-2
  - Width: 2.5 mm
  - Height: 0.5 mm

• Middle layer
  - Complex structure: a rectangle in which cavities needed for sensor operation are engraved
  - Material: Silicon
  - Width: 2.5 mm
  - Height: 0.5 mm
  - Membrane width: 1.5 mm or 1.9 mm
  - Membrane height (thickness of membrane): 10, 20 or 30 μm

• Cavity with a vacuum
  - Symmetric trapezoid or rectangular
  - Material: Vacuum
  - Width at the top: 1.9 mm
  - Width at the bottom: 1.5 mm
  - Height: 0.475 mm

• Cavity with ambient pressure
  - Rectangular
  - Material: Air
  - Width: 1.5 mm or 1.9 mm
  - Height: 5 μm

• Capacitance measurement
  - Done with two metal plates at the top and bottom of the cavity with ambient pressure
  - Top plate potential: 1 V
  - Bottom plate potential: Ground
  - Plate width: 1.0 mm

The thickness of all parts is 2.5 mm.

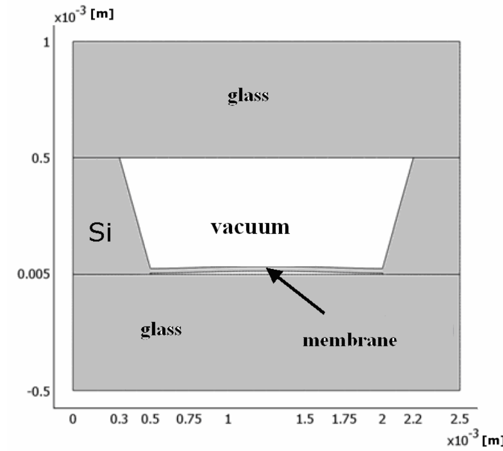


Fig. 1. 2D view of a pressure sensor (geometry: symmetric trapezoid).

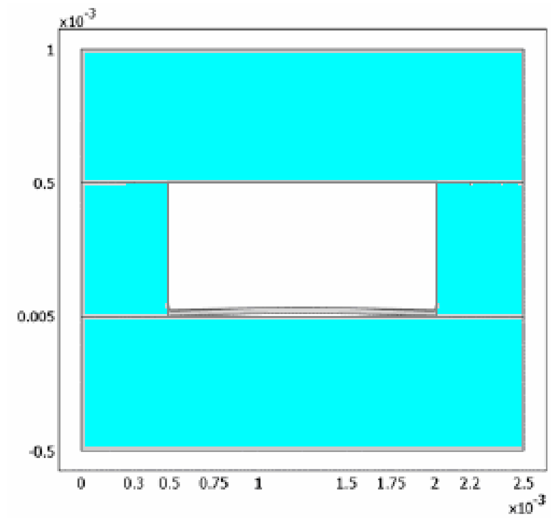


Fig. 2. 2D view of a pressure sensor (geometry: rectangular).

III. Stress and Deformation

Mechanical deformation is a change of shape and volume of the body under the action of external forces (or pressure).

During manufacturing, the sensor is bonded together in a vacuum and at a high temperature and is then cooled down. Therefore, during this process no external forces act on the sensor's boundaries, but internal stresses appear because the two materials have different coefficients of thermal expansion. This process also produces a vacuum in the upper cavity, and it serves as the reference pressure.

During normal operation, the sensor is fixed on a solid surface, and ambient pressure pushes on all outer boundaries. The temperature also changes, which produces extra stresses due to thermal expansion.

For a linear elastic material, the stress-strain relationship including the initial stress (σ_0), initial strain (ε_0) and thermal effects (ε_{th}), is:

$$\sigma = D\varepsilon_{el} = D\left(\varepsilon - \varepsilon_{th} - \varepsilon_0\right) + \sigma_0 \tag{1}$$

where D is the elasticity matrix.

Initially only thermal expansion is active, and it comes from the relationship:

$$\varepsilon_{th} = \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{xz} \end{bmatrix} = \alpha_{vec}\left(T - T_{ref}\right) \tag{2}$$

where α_{vec} are the coefficients of thermal expansion, T is the ambient temperature, and T_{ref} is the reference temperature.

This model assumes that after manufacturing the sensor is close to its initial geometry and thus the initial strain is zero.

For calculating large deformations, strain values come from the expression:

$$\frac{\gamma_{ij}}{2} = \varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} + \frac{\partial u_k}{\partial x_i} \cdot \frac{\partial u_k}{\partial x_j}\right) \tag{3}$$

For the case of large deformations, the model solves the problem using the principle of virtual work, which states that the sum of virtual work from internal strain is equal to work from external loads.

IV. Modeling in COMSOL Multiphysics

In COMSOL Multiphysics this problem can be solved using four application modes: two Plane Stress application modes, one Moving Mesh (ALE) application mode, and one Electrostatics application mode. The latter two are defined in a frame to allow the mesh to move.

Because the structure's deformation can be large, a large deformation analysis for both Plane Stress application modes must be used. The electric field was solved only in the small air gap where the ambient pressure is applied to the sensor (Fig. 3).

The solution process takes place in four steps:

1. The first Plane Stress application mode represents the sensor's fabrication, and it computes the initial stresses that result from thermal expansion using static solver.
2. The second Plane Stress application mode solves the deformation and stresses that result when the sensor is exposed to ambient temperature and pressure. It uses the initial stresses and deformation from the first plane stress application mode. This is solved with a parametric solver for different values of ambient pressure.
3. The ALE mesh for each ambient pressure can be solved using a parametric solver.
4. Finally, electric field for each ambient pressure can be solved using a parametric solver.

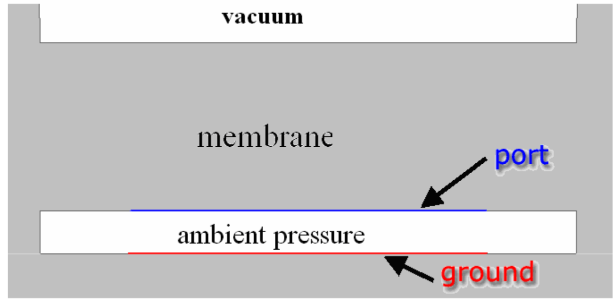


Fig. 3. Position of capacitors plate.

V. Results and Discussion

Fig. 4 shows the results after the bonding phase, where bonding took place at 400 °C and the sensor is then cooled down to 22 °C. In the image the x- and y- axes have different scales, and the structural deformation is scaled by 20.

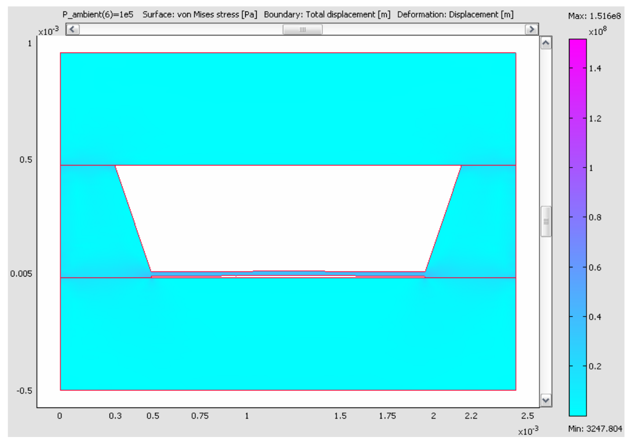


Fig. 4. Initial stresses of the materials in the pressure sensor (geometry: symmetric trapezoid).

It appears that the membrane slightly pulls towards the larger cavity even though there are no applied loads. Stresses appear near the boundaries of the different materials and in the silicon membrane, which is narrower than other parts of the sensor. The maximum appears at the lower left corner of the smaller cavity.

Fig. 5 shows the results when the sensor is in operation: it is exposed to a pressure of one atmosphere at 22 °C. The figure is arbitrarily scaled and is focused on the left half of the lower cavity. The membrane deforms toward the vacuum with maximum deformation at the middle. Maximum stresses appear at the upper corners of the lower cavity where the membrane attaches to the silicon boundaries. The streamlines show the electric field in the lower cavity. The lines are vertical between the two electrodes. Some field lines appear outside of the electrode region, but the field strength is very small there (dark blue color).

The capacitance change as a function of the ambient pressure for different membrane thicknesses is presented in Fig. 6, from the model shown in Fig. 4.

As can be seen from Fig. 6 capacitance change as a function of the pressure is slower for thicker membrane. This is a consequence of lower resilience of higher membrane and

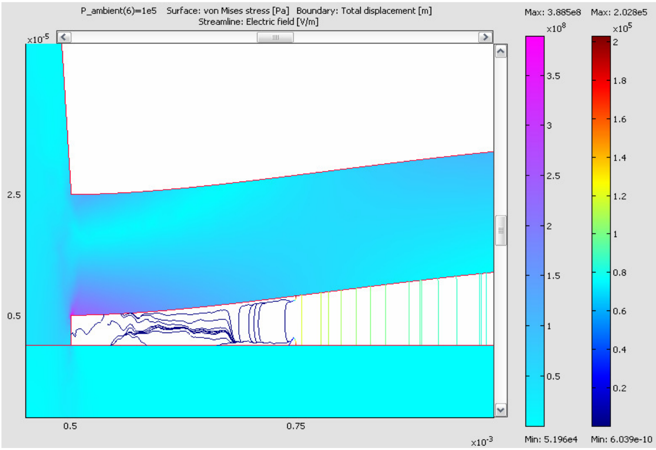


Fig. 5. Sensor deformation, stresses (left color bar: von Mises Stress) and electric field (right color bar: Electric field strength) when exposed to ambient pressure.

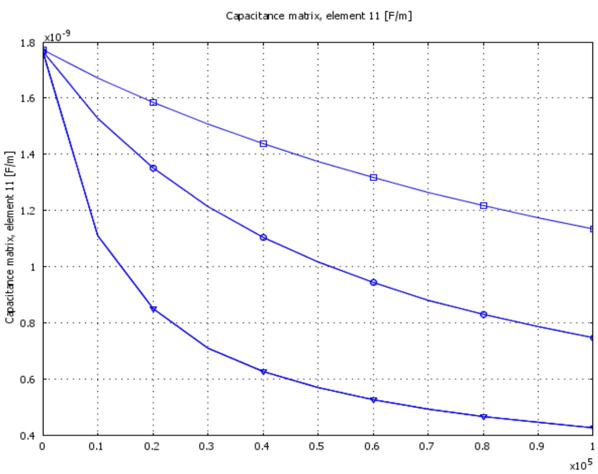


Fig. 6. Computed capacitance vs. ambient pressure for three different values of membrane thicknesses: (membrane thickness =10 μm, ∇; membrane thickness =20 μm, ○; membrane thickness =30 μm, □).

smaller influence on the capacitance change at the same applied pressure.

Fig. 7 illustrates geometry of the upper cavity of the middle layer: rectangular (width 1.5 mm).

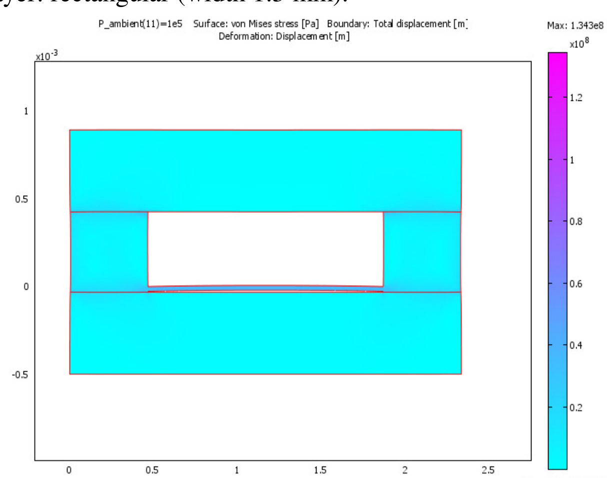


Fig. 7. Initial stresses of the materials in the pressure sensor (geometry: rectangular (width 1.5 mm)).

For this geometry, computed capacitance vs. ambient pressure for three different values of membrane thicknesses is almost the same as in Fig. 6. So, it can be noticed that the change of cavity geometry from the symmetric trapezoid into rectangular, with a width equal to the shorter base of the symmetric trapezoid (membrane width is not changed) is not significantly influenced on computed capacitance.

In the next step, geometry of upper cavity and the geometry of lower cavity have been changed. Geometry is now rectangular with width of 1.9 mm, as it shown in Fig. 8.

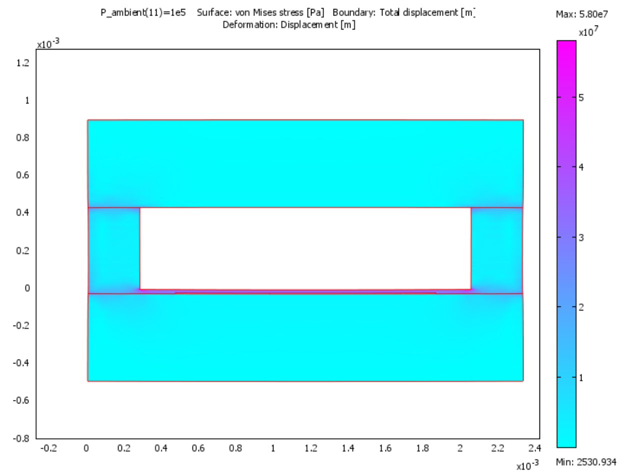


Fig. 8. Initial stresses of the materials in the pressure sensor (geometry: rectangular (width 1.9 mm)).

Fig. 9 presents the capacitance change as a function of applied pressure, C=f(P), from model shown in Fig. 8 for different membrane thicknesses.

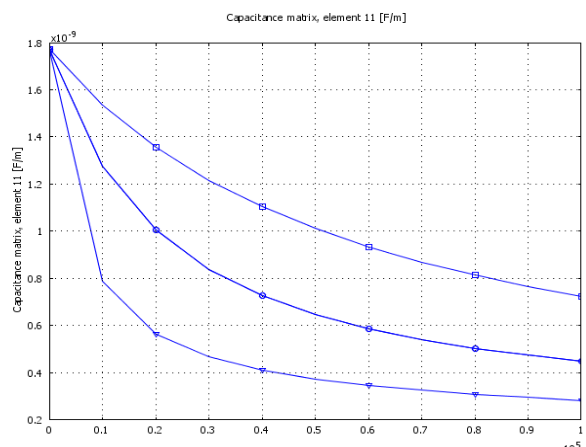


Fig. 9. Computed capacitance vs. ambient pressure for same temperature conditions and three different values of membrane thicknesses: (membrane thickness =10 μm, ∇; membrane thickness =20 μm, ○; membrane thickness =30 μm, □). Membrane width is 1.9 mm.

Fig. 9 plots C=f(P) for sensor structure with constant membrane width (1.9 mm) and variable membrane thickness (10 μm, 20 μm, 30 μm). Fig. 9 as the previous plots shows that thicker membrane causes slower change of capacitance in a function of applied pressure. However, increasing the membrane width from 1.5 mm to 1.9 mm resulted lower capacitance values at the same values of pressure compared with results shown in Fig. 6.

The rate of capacitance change in a function of applied pressure can be changed choosing a different membrane material (with more or less elasticity). Material with higher modulus of elasticity (Young's modulus - E) has a higher rigidity, and lower elasticity. In Fig. 10, except for silicon, is shown computed capacitance vs. ambient pressure for materials with twice and four times higher Young's modulus of silicon membrane.

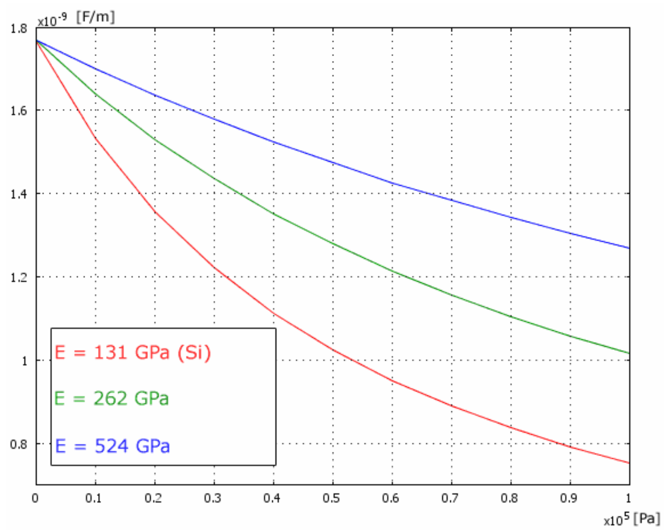


Fig. 10. Computed capacitance vs. ambient pressure for three different values of Young's modulus of membrane material.

VI. Conclusion

Capacitive deformation detector can be realized as the differential capacitive sensor with movable membrane as electrode. The membrane is engraved in the middle layer of sensor structure. Fixed electrode is the metal coated opposite side of one of the cavities. Deflection of the membrane due to pressure differences results in changes in capacitance between movable and fixed electrodes. Results showed that if the membrane was thicker, capacitance values were higher and capacitance change as a function of applied pressure was slower. The capacitance as a function of pressure is not significantly changed when the symmetric trapezoid geometry at the larger cavity of the middle layer is replaced with rectangular (membrane width and geometry of smaller cavity are not changed). But, when the symmetric trapezoid geometry at the larger cavity of the middle layer is replaced with rectangular whose width is equal to wider base of symmetric trapezoid and with wider membrane in smaller cavity, obtained capacitance values at the same values of applied pressure were lower compared with previous results.

References

[1] M. Popović, "Senzori i merenja", Viša elektrotehnička škola, Beograd, 2000.

[2] G. Fragiacomo, "A micromachined capacitive pressure sensor with signal conditioning electronics", Master's thesis, Università degli studi di Trieste, Italy, 2008.

[3] H. Chau and K. Wise, "An ultra-miniature solidstate pressure sensor for a cardiovascular catheter", Proc. 4th Int Conf Solid-State Sensors and Actuators (Transducers ´S87), Tokyo, Japan, June 2-5, 1987, pp 344-347.

[4] B Puers, A Vanden Bossche, E Peeters and W Sansen, "An Implantable pressure sensor for use in cardiology", Sensors and Actuators, A21 - A23 (1990) 944-947.

[5] W.H. Ko and Q. Wang, "Touch mode capacitive pressure sensors", Sensors and Actuators A, vol. 75, no. (1999), 242-251.

[6] Dimitropoulos, Kachris, Karampatzakis and Stamoulis, "A new SOI monolithic capacitive sensor for absolute and differential pressure measurements", Sensors and Actuators A, vol. 123-124 (2005), 36-43.

[7] www.comsol.com.

Design and Realisation of Over-voltage Protection in Push-pull Inverters

Milomir Šoja, Slobodan Lubura, Dejan Jokić, Milan Đ. Radmanović

Abstract—**In this paper are presented our research results about possibility of use different types over-voltage protection circuits in push-pull inverters. We first analyzed the conventional passive type RC and RCD over-voltage protection circuits and gave experimental results. After that we analyzed active over-voltage protection circuit, made design of protection circuit components and provided experimental results. Final investigation has shown that active over-voltage protection is better solution than passive protection circuits with respect to efficiency and reliability.**

Index Terms—**Passive RC I RCD over-voltage circuit, active protection circuit, push-pull inverters.**

I. Introduction

INCREASE in share of renewable energy resources in total energy balance resulted in wider application of energy electronics inverters in power supply systems. Converters, which are part of power supply systems with renewable energy resources are power inverters in push-pull inverters. Topology of push-pull inverters is interesting for a number of reasons: existence of energy transformer for galvanic separation of input and output, simple modification of output voltage by its value and use of minimal number of switch components. Bearing in mind that design of such components always entails efficiency and reliability as performance criterium, use of minimal number of components is often crucial factor in favor of use of push-pull inverters as topology in power inverters.

Besides the above mentioned, one of the crucial factors that influences reliability of energy electronics is over power and over-voltage protection of switches used in the inverter. In the concrete case, in order to achieve reliable operation of push-pull inverter, it was necessary to design over-voltage protection that protects the switches from voltage peaks created by various parasite components of energy transformer and other elements of the energy circle of the push-pull inverter. The first step in realization of over-voltage protection is construction, and it entails minimizing the number of parasite components by minimizing switching power loops. In bridge inverters this procedure alone is in most cases sufficient if it is performed correctly, but because of topology of push-pull inverters other methods for decrease of over-voltage have to be applied. Several types of over-voltage protection have been described in literature, and they can be generally divided into three categories: passive dissipative, passive non-dissipative and active over-voltage protection. Passive RC and RCD type protections are the simplest. Basic shortcomings of this type of protection are energy dissipation and complexion of calculations because for making the right choice in protection components it is necessary to know parasite components of L_σ and C_σ energy circuit, which is very complicated to determine and it is also not uniform.

Second type of over-voltage protection is not dissipative in nature which gives it upper hand, but it demands extra accumulative components, which take over energy from parasite components and it all leads to larger dimensions of the device.

Third type of protection applied in practice is active protection which contains some active elements alongside to the passive ones. Described active over-voltage protection in push-pull inverters with adjusted trigger level turns on both energy switches at the same time which creates power switching loop in which energy that was building up in parasite components dissipates. In this manner over-voltage in switches is avoided. That means that powerful semi-conductive switches protect themselves from over-voltage.

In the first part of this paper is provided overview of some conventional over-voltage protection designs in push-pull inverters. In the second part is described active over-voltage protection. Experimental results are provided for all types of protection. Conducted research showed advantage in use of active over-voltage protection to the passive one.

II. Two Parallel Push-pull Inverters

The inverters with lower input voltage and power load above 1000 W frequently used topology of two parallel push-

This paper was created as part of the project "Development and Evaluation of Performances of PV (photo voltage) inverters as basic component of PV Micro-Distributive Network", contract number: 0660-020/961-52/07, 03.12.2007., financed by Ministry of Science and Technology in Republic of Srpska Government.

Milomir Šoja is with the Elektrotehnički fakultet, Istočno Sarajevo, Republika Srpska – BiH (phone: 057-342-788; fax: 057-342-788; e-mail: milomir.soja@etf.unssa.rs.ba).

Slobodan Lubura is with the Elektrotehnički fakultet, Istočno Sarajevo, Republika Srpska – BiH (phone: 057-342-788; fax: 057-342-788; e-mail: slubura@gmail.com).

Dejan Jokić is with the Elektrotehnički fakultet, Istočno Sarajevo, Republika Srpska – BiH (phone: 057-342-788; fax: 057-342-788; e-mail: dejan.jokic@etf.unssa.rs.ba).

Milan Đ. Radmanović is with the Elektronski fakultet, Niš, Republika Srbija (phone: 018-519-663; fax: 057-529-100; e-mail: radmanovic@elfak.ni.ac.yu).

pull converters (2PP) [6] [7]. Fig. 1. shows the scheme of the above mentioned inverter that was used for testing different configurations of over-voltage protection circuits, and Fig. 2. shows manner of forming inverter output voltage wit output

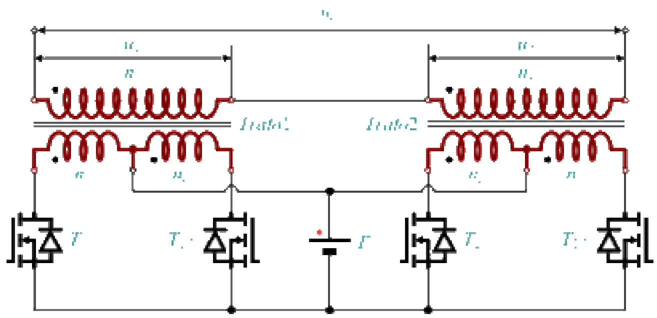


Fig. 1. Two parallel push-pull inverters.

voltage time shifts for individual push-pull inverters.

Each push-pull inverter forms square voltage by alternating conducting of corresponding switches, and "*quasi-sinus*" output voltage is formed by shifting the formed square

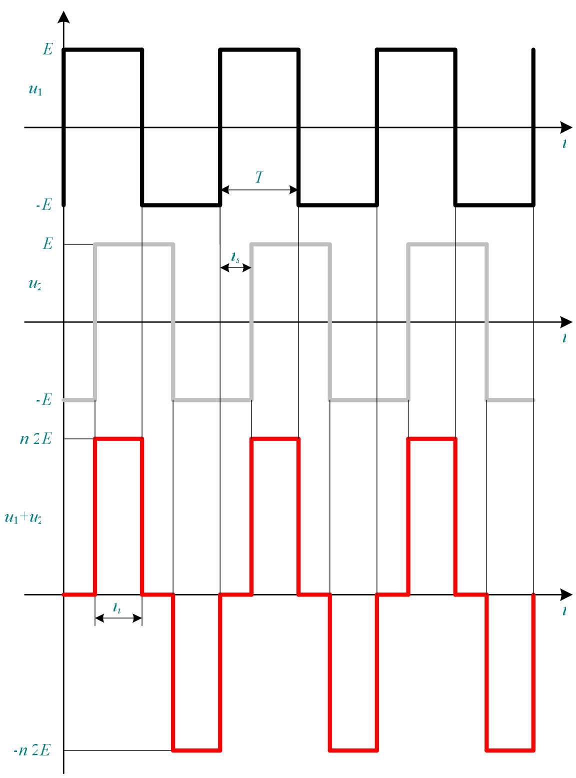


Fig. 2. Formation of output voltage ("quasi-sinus") in two parallel push-pull inverters.

voltages, as provided in Fig. 2.

It is obvious that two switches in the presented configuration are always turned on, while over-voltages are created in the other two in the moments of their switching off, and the over-voltages are superposed with double value of battery voltage because of push-pull inverter's nature of operating.

Typical wave shape of voltage on one switch of the push-pull inverter from Picture 1 in the moment of switching on without over-voltage protection with input voltage of couple of volts (2-3 V_{DC}) is shown in Fig. 3.

During the switching, power is moved from one to the other half of primary coil, which demands strong magnetic bond between primary coils in order to reduce built up energy in dissipative inductivity that causes over-voltage on switches while turning them off.

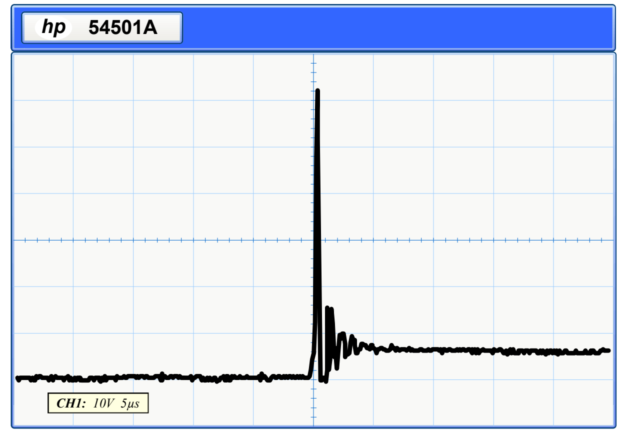


Fig. 3. Over-voltage on a switch at the moment of switching off in V_{BAT}=2-3 V_{DC}.

Standard manner of reducing over-voltage on switches is passive RC/RCD protection of dissipative or non-dissipative type.

III. OVER-VOLTAGE PROTECTION CIRCUITS

Basic function of all passive over-voltage protections is energy "absorption" of parasite components L_σ and C_σ of inverter energy circuit, which completely or partially eliminates over-voltage on switches. Capacitors that are connected in parallel with the switch are used for "absorption" of energy in these protections. If the energy of this capacitor dissipates on resistor then we say that we are talking about dissipative passive over-voltage protection. Non-dissipative passive over-voltage protection is also mentioned in [5] where energy of capacitor is transferred to input or not that frequently to output of power inverter by additional reactive components. Conducted research that is described in this paper gave answer to question of modes of application and performances of passive dissipative protections in push-pull inverters for protection of power switches.

A. RC protection

Fig. 4. shows way of connecting RC protection to switches of push-pull inverter.

Calculation of RC protection elements is rather complex because of not knowing exact values of parasite elements of individual inverter components, so in practice more simple methods are used for determination of values of components R and C protection. In order to create attenuation of oscillations in resonant circuit that is formed by parasite components of energy transformer L_σ and C_σ and energy switch (MOSFET) C_{DS} it is usually taken in push-pull inverters that $C>C_{DS}$. As initial value of capacitor C in [1] is recommended C=2-3*C_{DS}, and for R initial value can be selected according to nominal power of inverter I_O reduced to primary side of transformer and battery voltage E according to the following expression:

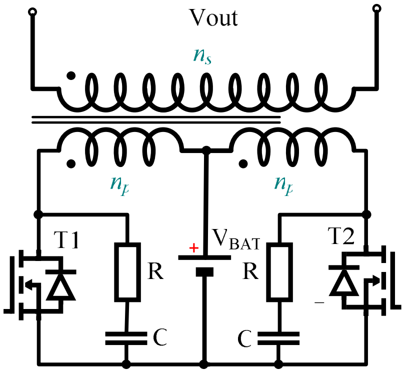


Fig. 4. Scheme of RC protection of push-pull inverter.

$$R = \frac{2E}{nI_O},\tag{1}$$

Power that is dissipated on resistor R in RC protection with maximum power value on capacitor C is:

$$P_R = 2CE^2 .\tag{2}$$

As the energy dissipates in capacitor charging and emptying medium power value of power dissipated on resistor R is provided in the following expression:

$$P_{DIS=} 4CE^2 f_s ,\tag{3}$$

where f_S is inverter switching frequency.

Wave shape of switch voltage in inverter from Fig. 1. with nominal load is shown in Fig. 5.

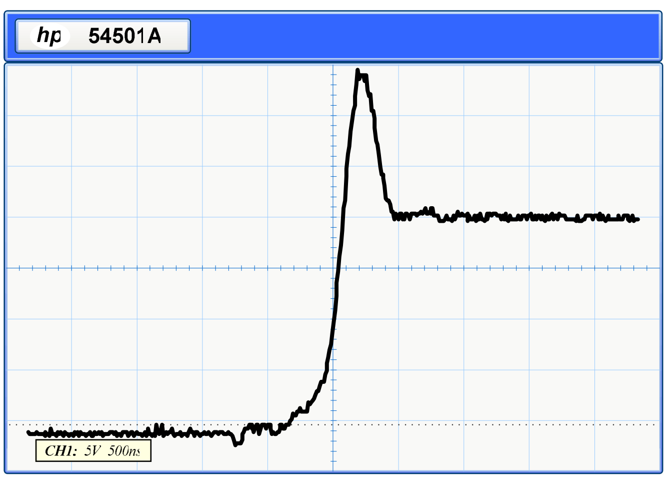


Fig. 5. Switch voltage in push-pull inverter.

B. RCD protection

It is obvious that two RC switch protections are necessary in push-pull inverters, which additionally complicates design of the device.

Unlike the above mentioned type of RC protection, this type of protection is in class of polarized protections and its mode of operating is completely different than the one we described earlier. The first step in designing RCD protection is determination of voltage increase time on the switch at its maximum current I_{Omax} as well as maximum allowed voltage value on capacitor C. Connection between current and voltage in the capacitor is provided in the following expression:

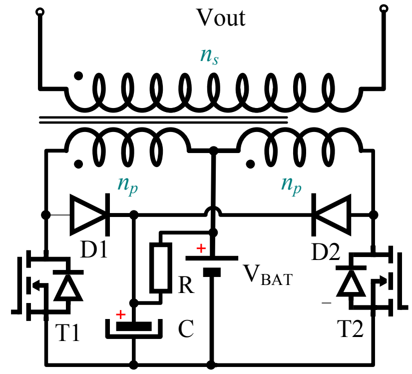


Fig. 6. RCD protection scheme.

$$I_{O\max} = C\frac{\Delta v_C}{t_r},\tag{4}$$

where: $I_{O\max}$ - maximum switch current, Δv_C - change of voltage in the capacitor, t_r - switch voltage increase time.

Necessary capacitor value in over-voltage RCD depends on the value of parasite inductivity of energy circle L_σ and it can be determined according to the energy balance:

$$W_L + W_{C1} = W_{C2},\tag{5}$$

where: W_L - built up magnetic energy on parasite inductivity L_σ, W_{C1} - initial capacitor energy in over-voltage protection, W_{C2} - total capacitor energy. Equation (4) can be written in the following form in case of push-pull inverters:

$$\frac{1}{2}LI_{O\max}^2 + \frac{1}{2}C(2E)^2 = \frac{1}{2}C(2E+\Delta V)^2 .\tag{6}$$

From the previous equation it is possible to determine capacitor C value for previously set value of over-voltage on switch ΔV :

$$C = \frac{LI_{O\max}^2}{4E\Delta V + \Delta V^2} .\tag{7}$$

Fig. 7 shows voltage wave shape on the switch of the push-pull inverter with RCD protection.

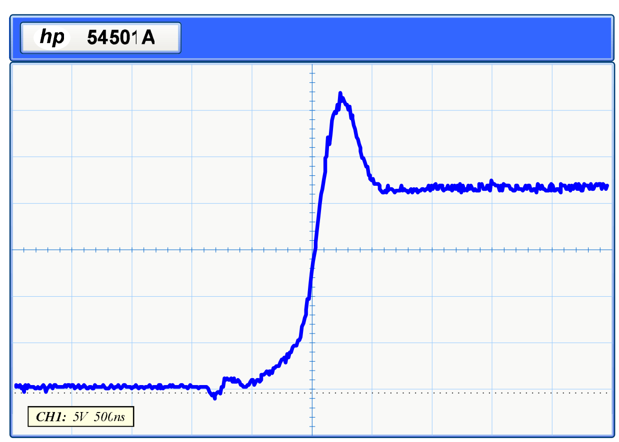


Fig. 7. Voltage on the switch of unloaded inverter.

C. Active protection

As we have mentioned earlier, basic problem in application of previously mentioned protections is complex calculus of the components that create the protection (they often have to be determined experimentally), difficulties in construction and additional losses that appear in the protection components.

All of these difficulties can be overcome by use of active protection. Scheme of active protection is provided in Fig. 8.

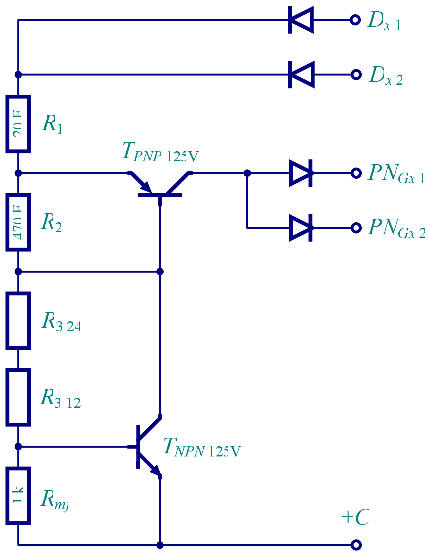


Fig. 8. Active protection scheme.

Operation of the presented protection comes down to active monitoring of the voltages between transistor drains in push-pull inverter (ports $D_{x.1}$, $D_{x.2}$) in relation to power supply voltage (voltage on input positive terminal of electrolyte capacitor +C). If the voltage between any MOSFET drain and input power supply is greater than protection trigger level voltage, it is conducted through transistors $T_{NPN.125V}$ and $T_{PNP.125V}$, switching on both powerful switches in push-pull inverter through $PN_{Gx.1}$ i $PN_{Gx.2}$. In that manner is used all the magnetic energy that accumulated in transformer parasite inductivity L_σ which caused over-voltage and power switches are hence protected.

With regard to the fact that both transistors and diodes which form active protection belong to the signaling components (block voltage should not be ≥100 V), and resistance is 0.25 W, it is obviously very cheap solution with practically no dissipation and which is without any difficulties possible to be fit in the energy part of inverter during the construction process.

Trigger level of the active protection is usually chosen by making maximum voltage on the power switch 10-25% greater than "normal" double DC input voltage.

$$V_{DS.\max} = 2 \cdot k_{VDS.\max} \cdot U_{bat.\max} = [1.1 - 1.25] \cdot 2 \cdot U_{bat.\max} . \quad (8)$$

Protection trigger level can also be determined according to the following expression:

$$\begin{aligned} V_{prag} &= V_{DS.\max} - V_{bat.\max} = (2 \cdot k_{VDS.\max} - 1) \cdot V_{bat.\max} , \\ V_{prag} &= [1.2 - 1.5] \cdot V_{bat.\max} \end{aligned} . \quad (9)$$

Once adjusted, trigger level remains constant and it does not depend on battery voltage. On the other hand, transistor voltage V_{DS} which turns on the protection depends on the battery voltage and it shifts within the boundaries of its change, which is acceptable. In order to simplify the calculation procedure for the resistor network it is agreed that the current through resistors $I_{\Sigma R.\max}$ is equal to 0.5 mA, at the moment the protection switches on. Resistor R_1 limits the current $T_{NPN.125V}$ and its value is usually 20 Ω. Voltage drop on resistor R_2 should be less than V_{be}≈0.6 V, and its value is 470 Ω.

Resistor R_{mj} determins protection trigger level and it is calculated from the following condition:

$$R_{mj} \cdot I_{\Sigma R.\max} = V_{be} \approx 0.6 \text{ V} , \quad (10)$$

which results in:

$$R_{mj} = \frac{V_{be}}{I_{\Sigma R.\max}} = \frac{0.6 \text{ V}}{0.5 \text{ mA}} = 1.2 \text{ k}\Omega .$$

Agreed R_{mj}=1 kΩ. Lower R_{mj} resistor value is agreed than the calculated one because transistor trigger level is not strictly defined and conductivity can start at slightly lower voltage.

Resistor R_3 should be adjusted in such manner that besides agreed values of other resistors it also defines current $I_{\Sigma R.\max}$ at the moment protection starts to operate (0.5 mA).

$$\frac{V_{prag}}{\Sigma R} = I_{\Sigma R.\max} = 0.5 \text{ mA}, \quad \Sigma R = R_1 + R_2 + R_3 + R_{mj}$$

$$(11) \quad \begin{aligned} \Sigma R &= \frac{(2 \cdot k_{VDS.\max} - 1) \cdot U_{bat.\max}}{I_{\Sigma R.\max}} \\ \Sigma R &= 2000 \cdot (2 \cdot k_{VDS.\max} - 1) \cdot U_{bat.\max} \end{aligned}$$

$$(12) \quad \begin{aligned} R_3 &= \Sigma R - (R_1 + R_2 + R_{mj}) \\ R_3 &= 2000 \cdot (2 \cdot k_{VDS.\max} - 1) \cdot U_{bat.\max} - (R_1 + R_2 + R_{mj}) \end{aligned}$$

(13)

Fig. 9 and 10 show dependence of change in resistor R_3 from demanded overvoltage on the transistor ($k_{VDS.max}$), for two battery voltages 12 and 24V_{DC}.

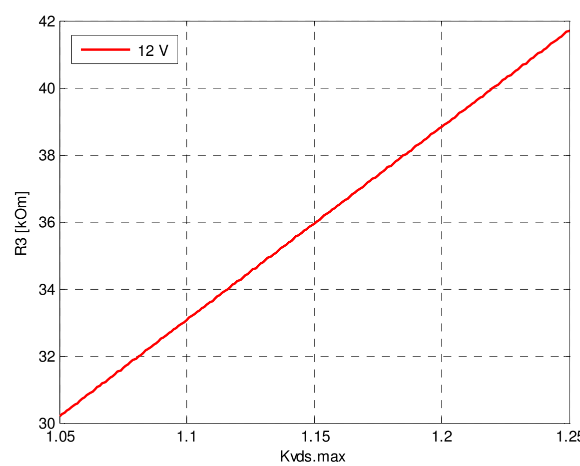


Fig. 9. Dependence of R_3 from allowed overvoltage on the switch ($k_{VDS.max}$) for U_{bat}=12 V.

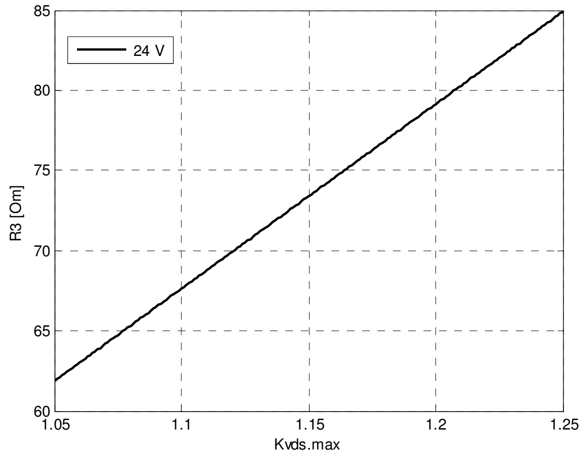


Fig. 10. Dependence of R_3 from allowed overvoltage on the switch ($k_{VDS.max}$) for $U_{bat}=24$ V.

IV. ACTIVE OVER-VOLTAGE PROTECTION EXPERIMENTAL RESULTS

Efficiency of suggested active protection was tested on power inverter realized as 2PP converters in parallel operating mode of nominal power 2000 W (Fig. 1.). Fig. 11. shows voltage on switches of one branch of push-pull inverter functioning as power supply voltage. Different voltage values on switches as result of changes of test battery voltages from 6 V to 24 V_{DC}. As it was previously mentioned, switch voltage is equal to sum of trigger level voltage of the active protection and battery voltage. Consequently, if we increase battery voltage, voltage on the switch also increases. Fig. 12 shows voltage of one branch of push-pull inverter under different inverter loads. When we compare voltage wave shapes on the switches for all types of protection we described (Fig. 5., 7. and 12.), it is obvious that active over-voltage protection gives the best results and that it provides the possibility that size of over-voltage does not depend on maximum current through the switch as it is the case with all other described types of protection.

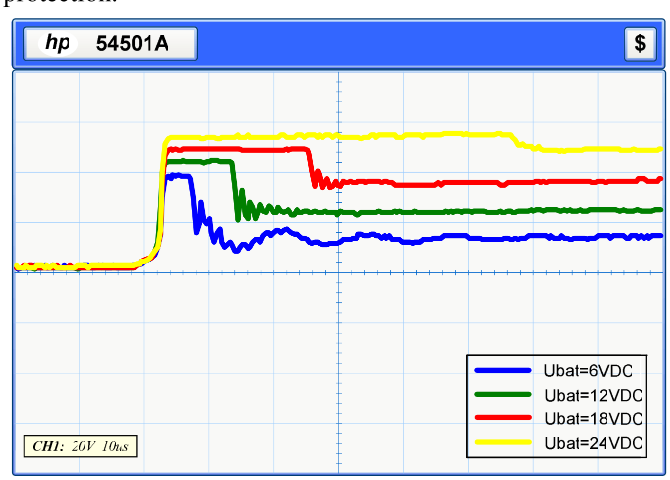


Fig. 11. Voltage wave forms on the drain of MOSFET with different input voltage values.

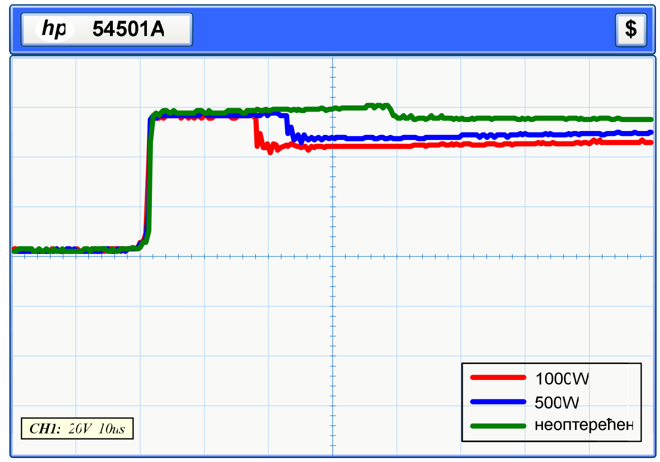


Fig. 12. Voltage wave forms on the drain of MOSFET under different inverter loads.

V. CONCLUSION

This paper presents results of research about application possibilities of different over-voltage protection types in push-pull inverters. We first analyzed conventional passive RC and RCD passive over-voltage protections and we provided experimental results. Then we analyzed active over-voltage protection and we provided calculations of protection components and experimental results. Conducted research resulted in conclusion that active over-voltage protection presents better solution than passive ones from the aspects of efficiency and reliability of the device.

REFERENCES

[1] Philip C. Todd: “Snubber Circuits: Theory, design and application”, Unitrode Corporation, may 1993.

[2] Rudy Severns: “Design of snubbers for power circuits”.

[3] Udeland, T: “Switching stress reduction in power transistors converters”, IEEE IAS annual meeting proceedings, 1976, pp. 383-392.

[4] Domb, M.: “Nondissipative turn-off snubber alleviates switching power dissipation, second-breakdown stress and Vce overshoot”, IEEE PESC proceed-ings, 1982, pp. 445-454.

[5] Finney, Williams and Green: “RCD Snubber revisited”, IEEE TRANSACTIONS ON IN-DUSTRY APPLICATIONS, VOL. 32, NO. 1, January 1996, pp. 155-160.

[6] D. Jokić, M. Šoja, S. Lubura:” Naponski invertor veće snage, napajan sa 12(24)VDC”, INFOTEH, mart 2006.

[7] D. Jokić, M. Šoja, S. Lubura:”Naponski invertor realizovan sa dva puš – pul pretvarača u paralelnom radu”, INFOTEH, mart 2007.

[8] K-INEL, Tehnička dokumentacija, 2002-2006.

[9] APC SB208 INT’L BACK - UPS.

[10] www.maxim-ic.com/an3835.

A Half Bridge Inverter with Ultra-Fast IGBT Module – Modeling and Experimentation

Dinko Vukadinović, Ljubomir Kulišić, and Mateo Bašić

Abstract—**This paper presents an operation analysis of a single-phase half bridge inverter with ultra-fast IGBTs (insulated gate bipolar transistors) and freewheeling diodes (module SKM100GB125DN, manufactured by Semikron). The Simplorer software package, which is especially suitable for electrothermal modelling of power electronics circuits, was used for inverter operation modelling. In addition, a laboratory setup of the inverter was built in order to experimentally verify simulation results. The control unit of the inverter consists of a stabilised power supply, SG3525A pulse-width modulator and SKHI22B hybrid dual driver. At the same time, the abilities to change the transistor switching frequency and dead time are provided. Good agreement of the simulation and experimental results was confirmed.**

Index Terms—**Power electronics, IGBT, Inverter, Modeling.**

I. INTRODUCTION

IT is well known that modern closed-loop electromotor drives for induction machine control are designed with three-phase inverters with IGBTs driven by pulse width modulation (PWM) principles. In these drives, IGBTs work as switches. However, transistors are not ideal switches, and they have certain turn-on and turn-off times. In this case, a dead or lock-out time should be provided between the switching of the devices in a leg of the inverter to prevent a shoot-through. Usually, the dead time is within 2-5 μs ([1]). As the dead time increases, the current ripple of the stator current of an induction motor becomes higher. One modern research topic in the field of industrial automation is development of compensation techniques to eliminate effects caused by this dead time. Advanced dead time compensation techniques include knowledge of the forward conduction voltage drop of the transistor and freewheeling diode. The difference between the target and actual phase voltage of an induction motor supplied by the voltage-source inverter can be given as ([2])

$$\Delta u = \frac{t_d + t_{on} - t_{off}}{T_c}(u_B - u_T + u_d) + \frac{u_T + u_D}{2} \qquad (1.1)$$

where
t_d – the dead time of an inverter,
t_{on} (t_{off}) – the turn on (turn off) time of the transistor,
T_C –the PWM carrier period,
u_B – the DC-link voltage,
u_T – the forward conduction drop of the transistor and
u_D – the forward conduction drop of the diode.

Dead time effect compensation techniques are complex because the conduction drops in the transistor and in the diode depend on the related current and temperature of the semiconductor. IGBTs are usually chosen by the forward conduction current, which is two times higher than the nominal motor current, for short-time overloads ([3]). In this case, the conduction drop across the transistor depends significantly on the conduction current (visible from the data sheet for module SKM100GB125DN [4]), which results in more complex dead time effect compensation techniques and in a more expensive DSP (digital signal processor) to carry out this compensation.

As a first stage in the research of the dead time effect, a laboratory setup of a half bridge inverter with IGBTs was built. The half bridge inverter is only used in power circuits for DC/AC power conversion. One leg of the three-phase inverter is identical to one leg of the half bridge inverter, and hence problems existing in the half bridge inverter are common in three-phase inverters. For this reason, the half bridge inverter will be investigated, and the results can then be applied to three-phase PWM inverter analysis.

The laboratory setup was designed with IGBTs rated at 100 A, while load currents between 0.5-2 A were observed. Hence, the power transistor is deliberately oversized. Over this current range, the turn-off and turn-on times and the temperature influence are negligible. Given that the price of the transistor module made by Semikron in the Croatian market is less than 70 EUR, the proposed technical solution is economically justified.

II. HALF BRIDGE INVERTER WITH R-L LOAD

Fig. 1 shows the half bridge inverter with a resistive-inductive (R-L) load.

The related current and voltage waveforms are shown in Fig. 2. In this section, the ideal inverter is explained, where non-idealities of the components and parasitic effects are neglected.

D. Vukadinović, LJ. Kulišić and M. Bašić are with the Mechanical Engineering and Naval Architecture Department, Faculty of Electrical Engineering, University in Split, Split, Croatia.

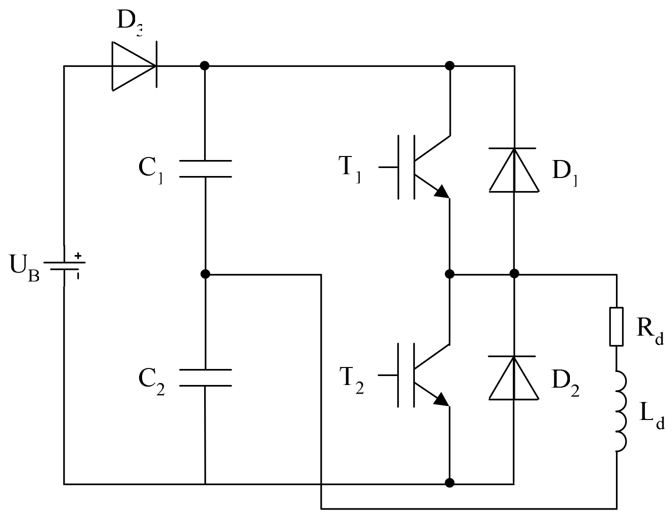


Fig. 1. Half bridge inverter with R-L load.

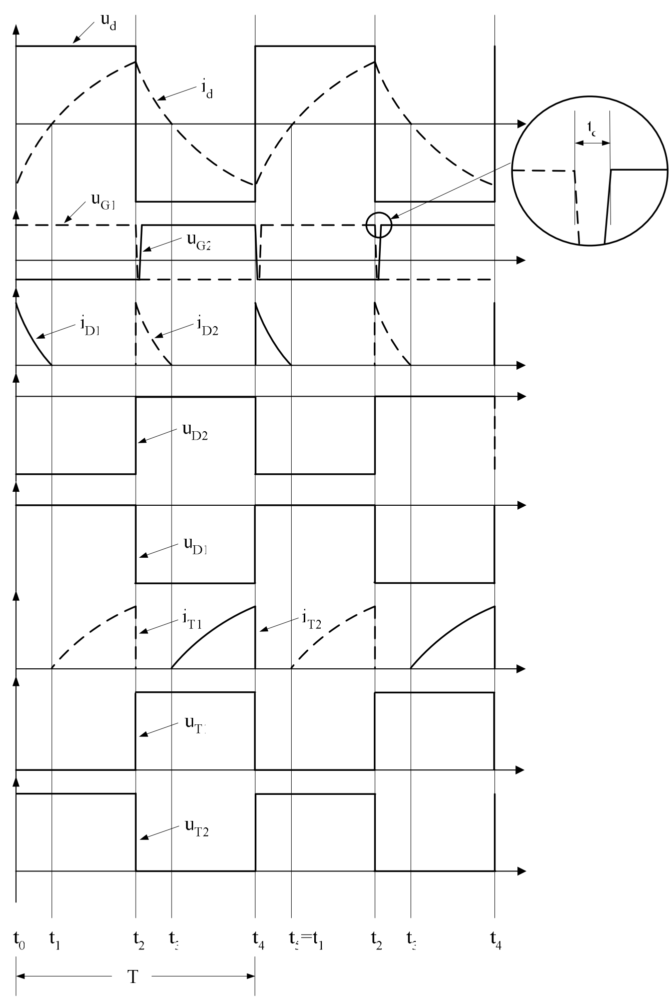


Fig. 2. Current and voltage waveforms of half bridge inverter.

The analysis of this inverter is given according to certain time intervals as the following.

The time interval (t_1, t_2): The transistor T_1 is turned on by a current pulse applied at the gate during the time interval $t < t_1$. However, the transistor T_1 will not conduct until t_1, when the load current i_d becomes zero, i.e. when the magnetic energy accumulated in the inductance L_d is discharged. Because the load circuit includes an inductance, the load current cannot instantly reach its steady state value. The current increase is defined by the load, i.e. R_d and L_d. During this time interval, the reverse voltage across the diode D_2 is U_B. The voltage across the transistor T_2 is U_B, and it blocks voltage.

The time interval (t_2, t_3): At time t_2, a driver turns off the transistor T_1, and the blocking voltage U_B lies across it until time t_4. It is clear that the pulse U_{G1} reaches negative value in order to ensure fast turn off of the transistor. At time t_2 the load current switches from the circuit C_1-T_1-R_d-L_d to the circuit C_2-D_2-R_d-L_d. During turn-off, because of the magnetic energy accumulated in the inductance L_d, the load current retains the same direction. The diode D_2 starts to conduct, because the conducting conditions are fulfilled until time t_3. The diode D_2 is reverse biased. At the end of the turn-off of the transistor T_1 it is necessary to ensure the dead time t_{dead}. The dead time t_{dead} prevents shoot-through, i.e. the incoming transistor should be delayed by a dead time from the outgoing transistor. At the end of the dead time t_{dead}, the transistor T_2 is triggered to turn on. Although the transistor fulfils conditions to forward conduct, it will not conduct until the load current falls to zero, at time t_3. As the load current i_d reaches zero value, the current changes its direction, and the next time interval starts.

The time interval (t_3, t_4): At time t_3 the load current changes its direction. As the transistor T_2 is able to conduct, the current starts to flow through the circuit C_2-L_d-R_d-T_2. The transistor T_2 conducts until time t_4 when it will be triggered to turn off.

The time interval (t_4, t_5): Immediately after turn-off of the transistor T_2 it is necessary to ensure the dead time before the incoming transistor is turned on. Then, the transistor T_2 blocks voltage. The load current flows through the diode D_1 until time t_5 when the current reaches zero value. When the current reaches zero value, it flows through the transistor T_1, and the overall process is repeated.

III. SIMPLORER SIMULATION OF HALF BRIDGE INVERTER

Simplorer is a multipurpose program for designing electrothermal high performance systems ([5]). It was developed, primarily, for automation, the airline industry, automotive design and power electronics. Simplorer is among the first simulation tools to implement electrothermal models of power semiconductors at device level. It offers three levels of the simulation model complexity.

In this paper, we have chosen the advanced dynamic of the inverter model including freewheeling diode. The simulation parameters are the following ([6]):

- The electric parameters of the circuit: supply voltage U_B = 48 V, gate resistors $R_{G1} = R_{G2}$ = 12 Ω, capacitors $C_1 = C_2$= 1000 µF.
- The trigger voltage: trapezoidal voltage waveform, with amplitudes +15 V and -7 V, frequency f = 1.2 kHz; rise time is t_r = 1 µs, fall time t_f = 1 µs, dead time t_{dead} = 3.9 µs.
- The load: R_d variable between 0-50 Ω, L_d variable between

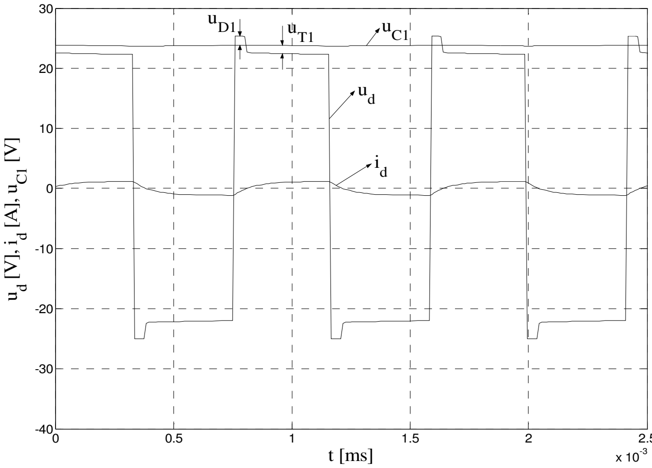


Fig. 3. Waveforms of load voltage, load current and voltage across the capacitor C_1; R_d =19 Ω, L_d = 1.5 mH.

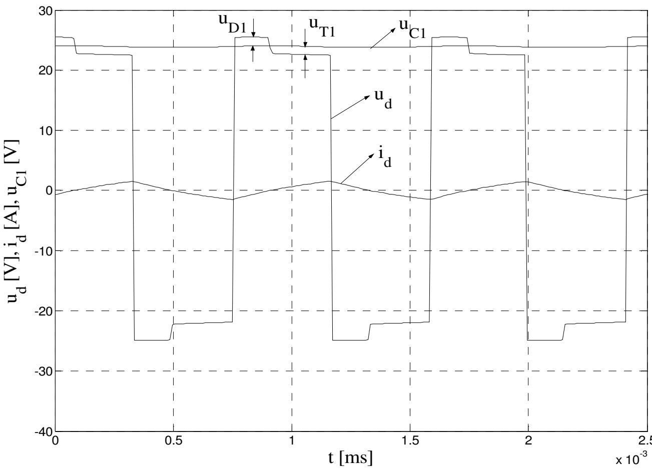


Fig. 4. Waveforms of load voltage, load current and voltage across the capacitor C_1; R_d =19 Ω, L_d = 1.5 mH.

1-52 mH.

Figs. 3 and 4 show the waveforms of the current and voltage of the variable resistive-inductive load and the voltage across the capacitor C_1 as well.

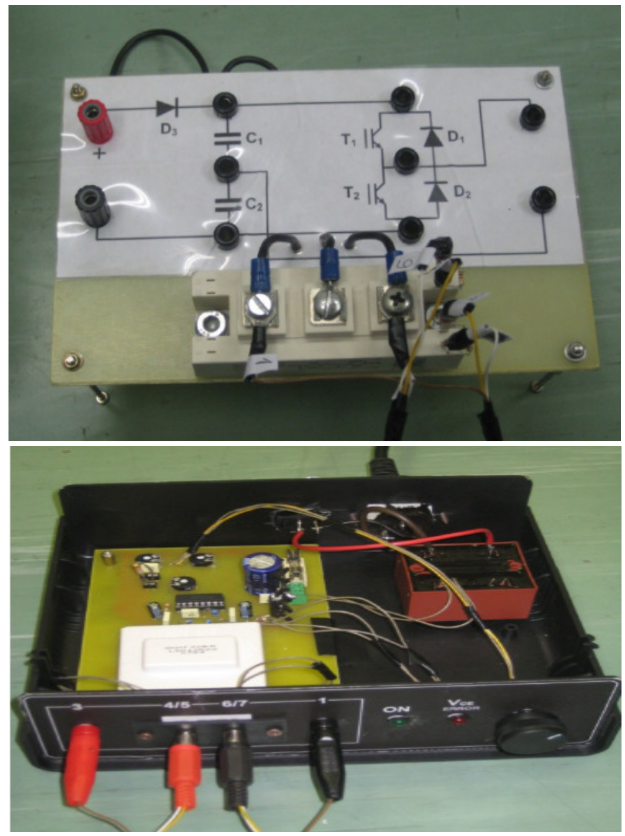
By comparison of the waveforms shown in Figs. 3 and 4 with the ideal waveforms shown in Fig. 2, it can be concluded that the load voltage does not have an ideal rectangular form. The amplitude of the load voltage u_d is not constant during the whole half-cycle. This amplitude is visibly higher when a diode conducts than when a transistor conducts. The reasons for this are the forward conduction voltage drops across diodes and transistors. The voltage drop across real IGBTs is usually between 1-3 V, and across diodes it is usually 1 V ([1]). When the transistor T_1 conducts, the load voltage is equal to the voltage across the capacitor C_1 ($\approx U_B/2$) minus the voltage drop across the transistor T_1 (Figs. 3 and 4). In the observed cases, the forward conduction voltage drop of the transistor u_T is 1.4 V and can be considered as constant over the observed current range. In a similar manner, the forward conduction drop of the diode D_1 is equal to the voltage across the capacitor C_1 plus the voltage drop across the diode D_1 (approximately 1.4 V in the observed cases).

IV. Laboratory Setup

In order to experimentally verify the theoretical investigations, the laboratory setup of the half bridge inverter was built in the Laboratory of Power Electronics of the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture in Split. Specifications of the setup and the experimental results are presented in this section.

A. Electric Scheme and Specifications of the Inverter

The half bridge inverter consists of the power section and



a)

b)

Fig. 5. Power section a) and control unit b) of inverter.

the control unit. The power section includes: the IGBT module with freewheeling diodes (D_1 and D_2), made by Semikron, type SKM100GB125DN, the capacitor of 1000 μF and the diode type BY255 (D_3). The photo of the power section and the control unit is shown in Fig. 5.

The control unit controls the half bridge inverter by triggering pulses. The rectangular waveform of these pulses is determined by the amplitudes of +15 V and -7 V. The set requirements to the control unit are to ensure two outputs, as they are 180 degrees out of phase with variable frequency, for each transistor separately. In addition, the adjustment of the dead time between the two signals must be ensured. In this paper, the set requirements for the control unit are carried out by the integrated circuit type SG3525A, which presents a pulse width modulator, and the integrated circuit SKHI22B, which presents a dual IGBT driver. In order to ensure the proper

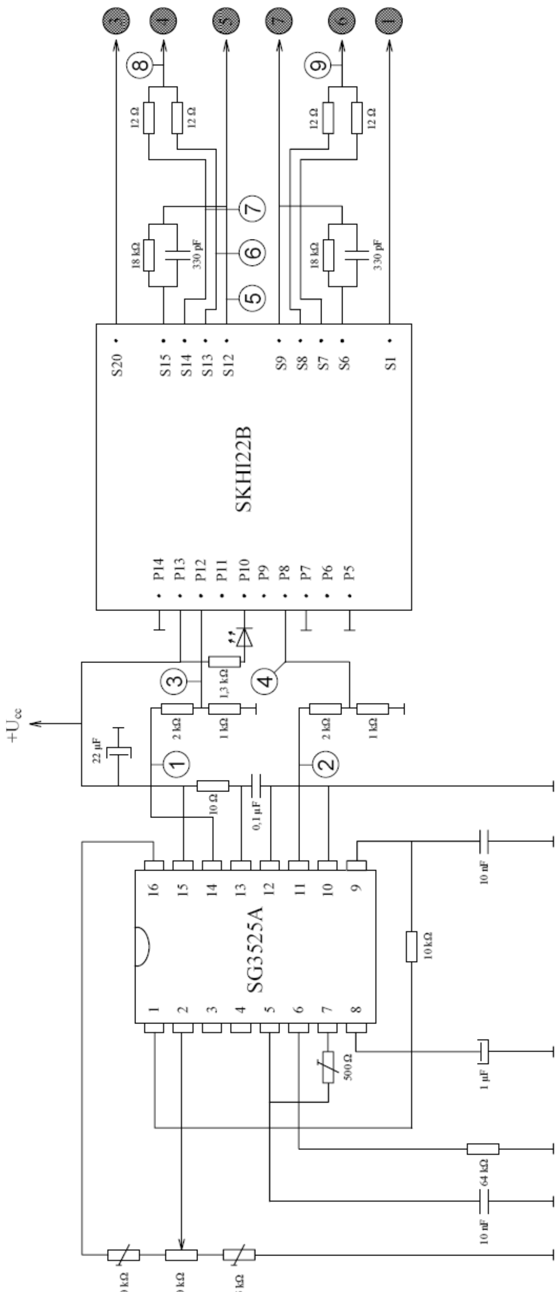


Fig. 6. Electric scheme of control unit.

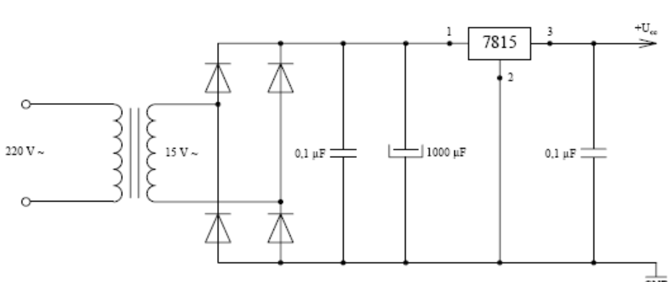


Fig. 7. Electric scheme of controlled voltage source.

operation of the control unit it must be supplied by the controlled voltage source. The electric scheme of the control unit is shown in Fig. 6, and the electric scheme of the controlled voltage source for the control unit supply is shown in Fig. 7.

The controlled voltage source shown in Fig. 7 includes: the power transformer, the diode bridge rectifier, the capacitor as an output filter and the series voltage regulator type 7815. Using this regulator on the output side of the rectifier, the stabilised voltage source of $U_{CC} = +15$ V is ensured.

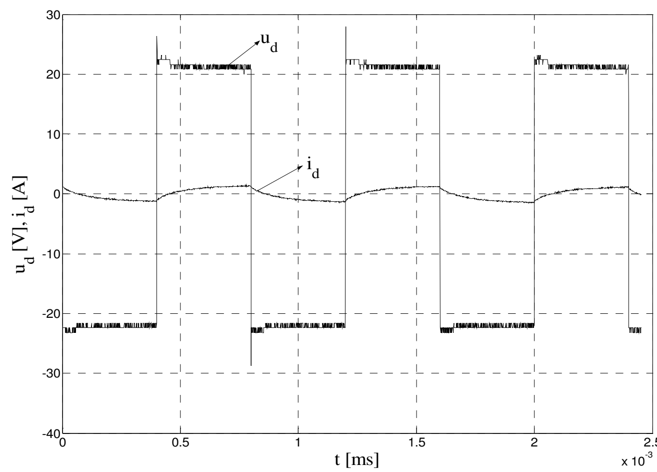


Fig. 8. Load voltage and current waveform; R_d =19 Ω, L_d = 1.5 mH.

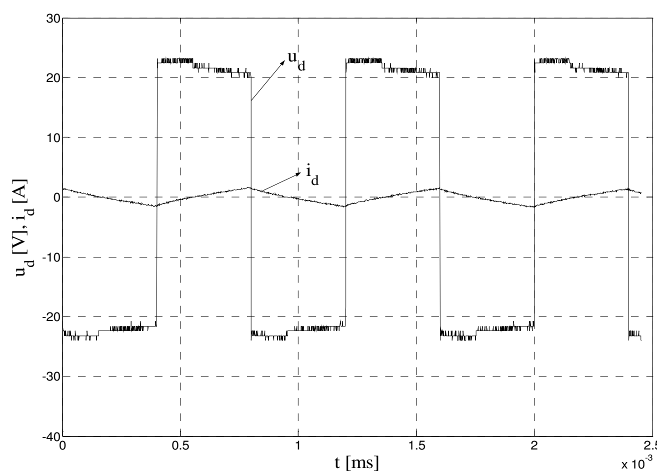


Fig. 9. Load voltage and current waveform; R_d =5 Ω, L_d = 3 mH.

B. *Experimental Results*

Figs. 8 and 9 show the experimental waveforms of the current and voltage of the resistive-inductive load obtained in the same operation modes and for the same parameters of the load as in Figs. 3 and 4.

The simulation results shown in Figs. 3 and 4 are in good agreement with the experimental results shown in Figs. 8 and 9, verifying the Simplorer model. The load voltage amplitude shown in Figs. 8 and 9 is approximately 2.5 V lower than the load voltage amplitude shown in Figs. 3 and 4 (during the forward conduction of both transistor and diode). The explanation for this effect is the following: the real battery used has an inner resistance, so the output voltage of the battery is the no-load voltage of the battery (48 V) minus the voltage drop across its inner resistance. This voltage drop is higher as the load current increases. Because of this, when a transistor forward conducts, the load voltage varies by a 0.4 V. Therefore, this voltage variability is caused by the non-ideal voltage source and not by the output characteristic of the transistor as in conformity with its data sheet ([4]). The variability of the load voltage during forward conduction of a transistor is not notable in Fig. 8, because the load current peak is approximately 30% lower than the load current peak shown in Fig. 9.

V. CONCLUSION

The laboratory setup of the half bridge inverter with the power section and control unit was developed in the Laboratory of Power Electronics. The power section presents the power module with two IGBTs and two freewheeling diodes. Inside of the control circuit there is the pulse-width modulator SG3525A, which outputs positive triggering pulses that are 180 degrees out of phase with an adjustable dead time interval between 0 and 22μs. These pulses are applied to the input of the hybrid dual driver SKHI22B, which outputs pulses for the transistor switching. The overall dead time in the output of the control unit is the sum of the dead times of the pulse-width modulator and hybrid dual driver. In this paper, the overall dead time is set to 3.9μs.

The validity of the simulation results was verified by the experiments carried out in the single-phase inverter laboratory setup. Through the analysed regimes and chosen parameters of the resistive-inductive load we noted very good agreement between the simulation and experiments results. The highest noted difference between the measured and simulated values of the load current is less than 5%, and the highest noted difference between the measured and simulated values of the load voltage is less than 10%. The difference in the load voltage was caused by the non-ideal DC voltage source.

In this paper, theoretical and experimental prerequisites for an analysis of voltage distortion effects caused by the dead time are made. This is one of the trends of modern research in the field of industrial electronics.

REFERENCES

[1] B. K. Bose, Modern Power Electronics and AC Drives,Oxford: Elsevier 2003.
[2] A. R. Muñoz, T. A. Lipo, "On-Line Dead-Time Compensation Technique for Open-Loop PWM-VSI Drives", IEEE Transactions on power electronics, Vol. 14, Issue 4, pp. 683-689, July 1999.
[3] M. Barnes, Practical Variable Speed Drives and Power Electronics, New York: Prentice Hall PTR 2007.
[4] Data sheets of electric components made by different manufacturers, http://alldatasheet.com, November 10th 2008.
[5] Ž. Jakopović, V. Šunde, Z. Benčić, Electrothermal Modeling and Simulation with Simplorer, IEEE International Conference on Industrial Technology, pp. 1141 -1145, 2003.
[6] D. Zovko, Half bridge inverter with IGBTs, graduate work (in Croatian), Croatia, Split: FESB, May 2008.

Benefits of Using OLAP versus RDBMS for Data Analyses in Health Care Information Systems

Srebrenko Pešić, Tatjana Stanković, and Dragan Janković

***Abstract*—In modern times, High qualities Information Systems are unconditional need in health care structure of developed countries. In our country, initiation process of such IS is at the very beginning. Despite that, we have tried to construct case ”what-if” and give some answers which could be useful in further implementation of most advanced techniques in statistical calculations, analysis and decision-making processes in health care. OLAP systems have capabilities for fast and readable insight, making strong basis for the top-class business decisions. This paper describes advantages of OLAP over classic queries in relational databases which are in use in health care. As an example, OLAP system was created based on a database of Clinic of Neurology in Nis, as well as statistical data from Yearbook of Administration for Economics, Sustainable Development and Environment Protection.**

***Index Terms*—OLAP in medicine, Health Care Information System.**

I. Introduction

INFORMATION systems are realized to help efficiency and consistency in business processes. After a while of IS existence a large amount of valid data is usually collected. In large companies or government institutes there is often a need for making some serious decisions, important for further business strategies, and based on large-scale data analysis. It is necessary to provide tools for efficient analysis on one side, and on the other, simplicity of those analyses, so people who are not IT experts could use them, like company management for example. One special segment of data processing is processing with hypothesis establishment cause, in another word – for studying. The area of medicine and medical information systems is one such area. Medical databases are very important part of every informatics society because they are directly related to the country health status, and so they affect all other society segments, so they need to be treated with special carefulness. Students, researchers, professionals and other people use medical databases to gain some data important for their activities. Those databases are further used for medicine improvement attendance, as given services quality marks, or like confirmation of some hypotheses (about certain trends and modern way of life negative aspects).

Statistic techniques and machine learning techniques are usually applied over medical data. Complexity of those techniques fluctuates from those extremely simple like histograms, to the most complex like prediction systems are. Statistical tests have wide appliance in medical researches, because they give them a simplicity, flexibility, and reliability. Basically, most experiments are performed to discover some important medical facts, which are confirmed through statistical calculations. Statistical tests are based on the hypothesis on the statistical characteristics of the analyzed medical data. The end cause is to prove correctness of the hypothesis with big trustiness. This problem becomes more complex when user has to analyze more than few data subgroups, with different combinations of risk attributes.

In the early eighties, some new methodologies for existing databases exploring have been developed. One of them is OLAP (online Analytical Processing [1]). Another definition that gives better description of the approach is Fast Analysis of Shared Multidimensional Information (FASMI) [2].

In OLAP system, users try to gain interesting but unexpected results by analyzing data subsets aggregated on different levels. OLAP techniques can have qualitative appliance in medical area, because they are intuitive, reasonable, and efficient, and on the other side again, they do not require advanced informatics knowledge from end-user. In OLAP systems, the biggest part of calculations is based on simple aggregations and counting, which are bases of the statistical tests.

Statistical methods have curtain benefits. They have simple assumptions about probability of distribution among datasets. There are no problems in such assumption when those methods are used in research with parameters which can easily be compared by querying over RDBMS. This is applicable until the moment of requirements for manipulation over data matrix. Also, statistical calculations give good results over small datasets just like over large ones. On the other side, statistical calculations have their imperfections. Basically they require great number of attempts and miss shots before they lead to some valid result. Every new attempt requires new choosing of parameters to accomplish under datasets partition.

S. Pešić is with Health Care Center, Niš, Serbia.
T.Stanković and D. Janković is with the Faculty of Electronic Engineering, University of Niš, Niš, Serbia.

One of the key benefits that OLAP system has comparing to statistical calculations is fast interactive querying through multidimensional and hierarchically organized data. Also, OLAP can be used for efficient reporting, quality control of given services, and for data integrity checking. The only deficiency of all is consumption of time needed for data warehousing.

This paper's goal is to present OLAP capabilities in Health Care systems. The contribution that OLAP systems can give to health services is not limited to one area - it refers to simplified decision-making (for management) or better tracking of medical parameters, such is frequency of some diagnosis referred to patient's age, gender, territorial partitioning, etc. Data structures and methods used in OLAP systems will be explained in next chapters. The results of applying OLAP over real medical data, and some prognosis related to larger Health Care systems will be presented after. At the end we will present summarized results and some directives for such system further developing.

II. Data and Methodology

Online Analytical Processing (OLAP) systems are very efficient tool used in complex Management Information Systems (MIS). These systems resolve next problems:

- Data analyzing according to parameters that user evaluates to be important.
- Reporting that requires exceptions or aggregations related to key indicators, trends, comparing by periods or territories, and other similar analyses.
- Business reports that require summing, exceptions and trends over different subjects.

The most important characteristic of OLAP system is multidimensional data which facilitates moving through data over "dimensions" and "measures". OLAP translates existing data from relational schemas by assigning key indicators (measures) to adequate contest (dimensions). When data is placed in multidimensional database (cube), all measures are easily and quickly available. The relation between dimensions and measures can be presented by star schema. The simplest schema presents tables with dimensions surrounding the table with the main data that comprises measures, called fact table. This is presented in Fig. 1. Table in which measures are comprised, comprise relations to dimensions also (foreign keys to outside tables).

Additional important characteristics of OLAP system are embedded and programmable analytic possibilities, and different options for data presenting and reporting. OLAP algorithms run over large datasets, and their result which overcame by using simple grouping and aggregating functions is unknown in advance.

To use OLAP system, previously we need to develop adequate multidimensional database. That process consists of standard steps [3] from which the most important and most demanded are data filtration and data importing to curtain dedicated OLAP tools. Data filtration (PREPARATION)

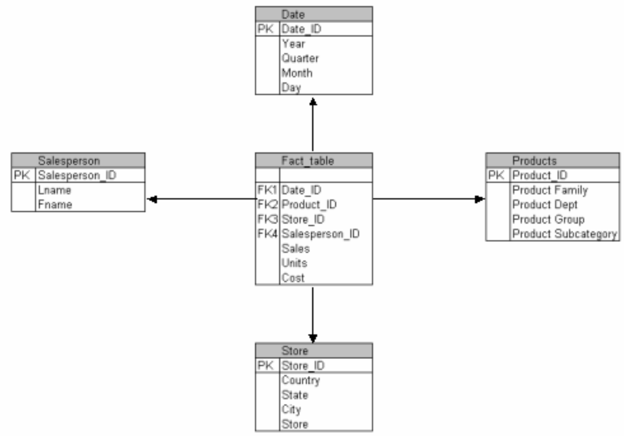


Fig. 1. Star schema example.

implies fault elimination (irregular inserts, duplicates, data inconsistency, and violation of referential integrity). This step can significantly slow down OLAP system developing. Data importing problem has technical nature, because multidimensional databases can very often overcome system hardware limits. The largeness of cube can be decreased by data aggregation before transporting to OLAP system. Besides, problem with available space can be determined by dividing cube to two or more overlapping multidimensional hyper cubes (analyses are divided to several sub-analysis).

Health Care has experienced great improvement lately by bringing computers in many clinical and administrative processes. Yet, there is no full potentiality of using medical data as management and diagnostic decision-making source. This paper describes OLAP capacity developing process from on-line transaction processing (OLTP) system (electronic health record [4]).

The Electronic Health Record (EHR) or the Electronic Patient Record (EPR) is on-line transaction processing (OLTP) system that enables on-line inserting and updating of given health care services and documentation, medical results tracking, and real-time deciding support. Because it contains information details about daily activities, such OLTP system has great OLAP capabilities in medical, financial, and administrative area. Health Care employers have understood lately the benefits of those systems, and have been beginning to show curtain interest in data analysis which would have helped them to easily achieve answers to number of every day situation questions. Unfortunately, the most part of classical OLTP EHR systems has not suitable support for OLAP systems.

Very often there is a situation that data in HER systems are not well structured. Also is not rare that there are well structured medical information systems, but analyses that would bring to useful conclusions are not system-integrated, or cannot be performed in a real-time.

According to all above, the challenging is to:

- Choose the right data that can be important.

- Choose the data that are relevant to analyzing in a contest of easiness for later using.
- Make the connections between seemingly separated data placed in different subgroups of the system.
- Analyze great deal of data through more different parameters.
- Come to conclusions in the analysis of data, which indicate the dependence of consequential and causal connection.

OLTP systems based on a traditional RDBMS without OLAP support are not convenient for performing those tasks because much more hardware resources are needed to accomplish same results, while response time for every day usage is too slow. Therefore, to gain multidimensional system suitable for easy manipulation above datasets, we need to pass curtain phases [5]. Those phases are:

- Multidimensional model creating (determining measures, dimensions and schemas),
- Extracting, transformation, and storing data to created schemas,
- Creating and manipulating with reporting by using relational or multidimensional sources, and
- Generating information from system by using created reports (algorithm).

In our paper we used star schema for OLAP multidimensional cube. Basic (fact) table is surrounded with dimension-tables, as is presented in Fig. 2. We have created OLAP system with several measures and more then several dimensions, based on the available database of Clinic for Neurology of Clinical Center Nis. Database contains patients records collected from the beginning of year 1996. until the end of year 2008. Database migration from MS Access 2000 to MS SQL Server 2005 platform was performed. After successfully finished first step (data transformation) we have got star-schema that was suitable for developing OLAP (Fig. 2).

In an effort to establish analytics related to the possibility, need, benefits of using OLAP in public health, as well as the existence of reasons for the necessity of OLAP in a close future, we selected all possible parameters that were able to represent the measures, and for the dimensions we chose different types of patient population: gender, occupation, age, marital status, as well as doctors and diagnoses as shown in Fig. 3.

III. Study and Actual Results

OLAP is built over Clinic of Neurology Nis database, in Mucrosoft Business Intelligent Developement Studio 2005. Analytics has been done in this software package, and in ProClarity Desktop Proffessionall 6.2 during september-november 2008 period of time. Star-shema fact table has been reduced to 27000 records after significant data transformation. Those record contains informations about patients health examinations, their hospitalisations, dehospitalisations, deaths, etc. The basic idea was to come to the conclusions related to

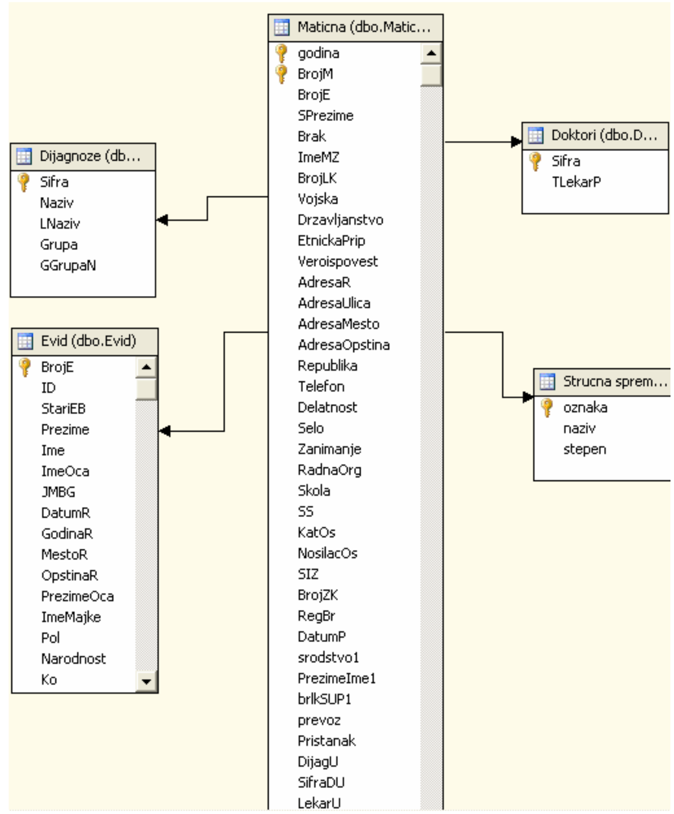


Fig. 2. Clinic of Neurology Nis star-schema.

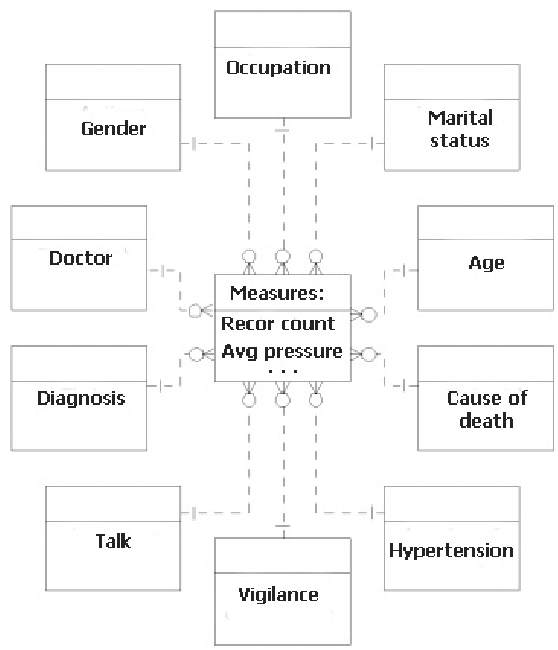


Fig. 3. Clinic of Neurology Nis hypercube model, developed for analysis on the necessity of establishing OLAP in public health.

the use of OLAP in public health in general, and over the data from all Residential Health Clinic in the region. There is a plan for using such system in Health Center Nis, after the introduction of information systems and data collection period of at least a year.

The results obtained in the built-in OLAP showed the following:

- Cube processing time is not of importance. Cube processing on database on server cofiguration (Intel Pentium Dual CPU E2160 1.80GHz, 3.00GB RAM) lasted from 10 to 15 seconds, depending on the number of dymensions included to cube;
- Data analytics related to diagnoses, such as for example, most diagnoses that emerged in more than 400 patient examinations in the period of 10 years, rare diagnoses and doctors that establish them, diagnosis in relation to age, degree, sex and marital status of patients and similar, proved to be very simple for the end users. With the help of tools that were used for analysis (Microsoft ProClarity Desktop Profesionall), and system users without training to work with computers could very easily acquire the analysis of OLAP, only if filed names (dymensions and measures) were concise and understandable for the end users. According to that, more in database design should take into account the naming of objects and attributes.
- During the report analyses, we came very quickly to expected, but what was even more important by us, to completely UNEXPECTED results. Example: analytics of the number of patient treathments by gender, marital status and diagnosis, unexpectedly showed that there were significantly more treatments of men who are married, but of all other population – as is shown in Fig. 4. Good material for the neurologist to do research on the topic: whether married men are the most endangered population in terms of neurological?
- The time needed for OLAP quering is significantly less then the time needed for quering relational database to get the same results. For executing query that gives results (15395 records) about number of patient threatments by gender, marital status and diagnosys for frequent diagnoses on relational database, server needed ~ 7 seconds. At the same server, the time required to obtain the same results on developed cube was ~ 0.2 seconds.

Based on statistical data [4], we may be able to make the assessment for this system implementation to quantitatively greater volume of data. To get started, as the test center will be taken Health Center Nis as one of the largest institutions of its kind in the Balkans. So it will be interesting to compare the time of execution of queries – comparing the results from The Clinic for Neurology and the valuation for Health Center Nis or all health centers in the country. For this, it is necessary first to ensure the introduction of information systems in public health and their use in a given period in order to collect relevant data for the full research. However, lets look some of the statistical data that our public health has collected for years, even wothout information system. These data are presented in the Statistical Yearbook for the city of Nis for the year 2007 [6], and they are related only to the General Medicine Service (Table I).

In this spreadsheet can be seen that the number of visits to ambulance (only for the service of general medicine) a year is between 600 000 and 800 000. Observed for all primary health

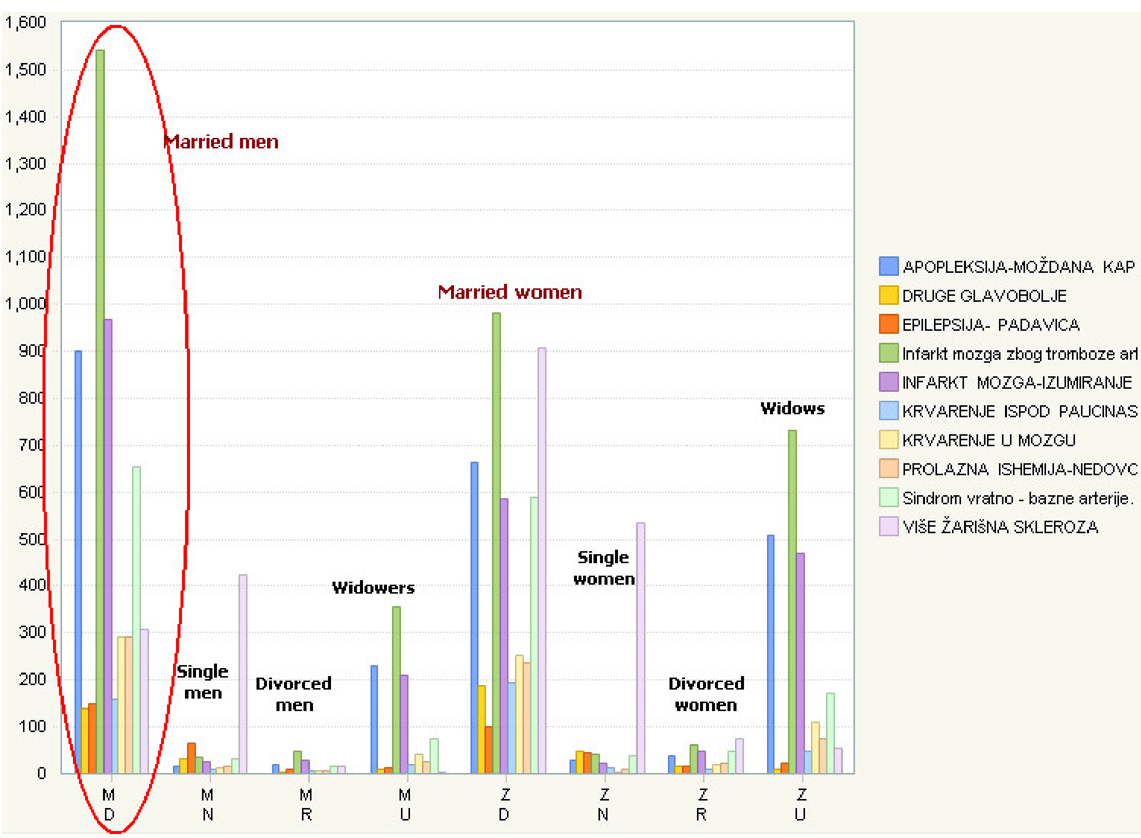


Fig. 4. Surprise factor - get unexpected results by analyzing OLAP for common diagnoses.

TABLE I
CLIPPING FROM THE TABLE 19.7. GENERAL MEDICINE SERVICE – SGN2007

Year	Threatments		**Total threatments**	Threat. per doctor	Home threatmen ts
	First thr.	Repeated threatment			
1998	220 551	385 475	**606 026**	7 390	17 715
1999	214 549	392 159	**606 708**	7 399	17 987
2000	261 378	465 199	**726 577**	8 146	18 429
2001	278 694	507 511	**786 205**	8 276	19 613
2002	288 092	454 697	**742 789**	7 902	19 811
2003	262 603	513 943	**776 546**	8 261	20 268
2004	287 352	486 403	**773 755**	7 661	12 138
2005	275 923	532 314	**808 237**	8 164	5 069
2006	268 735	536 795	**805 530**	7 897	7 662
2007	227 938	515 049	**742 987**	6 694	17 915

care in this area this number may be up to 7-8 million per year. As for the base at which we have built OLAP system for analysis, the number of visits included is not greater then 30000. Even in this case, we have received a significant difference by comparing the time needed for quering common relational database, and OLAP cube. The time required to execute the same queries over OLAP database is 35 to 100 times less then over RDBMS. Let us mention that this database still has a bunch of textual key columns. Considering all these facts, there is a logical thought: to put our public health in a situation to apply the modern pro-European health structures, OLAP technology for analysis and business decision-making will not only have advantages over the traditional report, but will in a way be necessary.

IV. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

The paper discussed the possibility of OLAP use in the analysis of medical data. OLAP system is implemented at the Clinic of Neurology in Nis, which has for 10 years had a total of 30,000 patient treatments. This system has enabled a quick overview of cumulative data and fast execution of complex queries. Else, such queries would not be possible over the classical relational base, or would be far slower over it. OLAP so now offers a new view of the data that have been collected for longer period of time. According to the data from the Statistical Yearbook for city of Nis, from which we have presented only few in this paper for the illustration, you may find the cost-effectiveness of this approach when applied to larger systems.

Cost-effectiveness of OLAP-for small and medium databases is questionable and must be considered for each case separately. The main question that is raised is: "Is the required time for the creation of OLAP systems worth the potential gains?“.

Today almost every RDBMS and statistical software packages include OLAP support (SQL Server, Oracle...), which is a sign that further development will go in the direction of more massive application of OLAP and appropriate techniques? In this way, access to data and the analysis of the data is provided to the experts of all different profiles that are not IT professionals. In order to use these techniques, it is necessary to pay attention to the way how databases are designed. Some aspects of this problem are presented in the paper.

REFERENCES

[1] E. F. Codd, S. B. Codd, C. T. Salley, Beyond decision support, Computerworld, 27, pp. 87-90, 1993.
[2] N. Pendse, What is OLAP, http://www.olapreport.com/ fasmi.htm.
[3] U. Fayyad, G. Piatetsky-Shapiro, P. Smith, Advances in Knowledge Discovery and Data mining, MIT Press, pp. 1-34, Cambridge, 1996.
[4] Rajković, P., Janković, D., "Electronic Patient Record as a Basis of Medical Information System", XXXIX International Scientific Conference on Information, Communication and Energy Systems and Technologies ICEST 2004, Bitola, Macedonia, June 2004.
[5] G. T. Monaco, An Introduction to OLAP in SQL Server 2005, http://www.devx.com/dbzone/Article/21410/.
[6] Uprava za privredu, održivi razvoj i zaštitu životne sredine, Statistički godišnjak grada Niša 2007., pp. 199-217, Niš, novembar 2008.

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

ELECTRONICS, VOL. 13, NO. 2, DECEMBER 2009

ISSN 1450 -5843