MACRo 2010

PROCEEDINGS OF THE 2nd INTERNATIONAL CONFERENCE ON RECENT ACHIEVEMENTS IN MECHATRONICS, AUTOMATION, COMPUTER SCIENCE AND ROBOTICS

SAPIENTIA HUNGARIAN UNIVERSITY OF TRANSYLVANIA

FACULTY OF TECHNICAL AND HUMAN SCIENCES

DEPARTMENT OF ELECTRICAL ENGINEERING

DEPARTMENT OF MECHANICAL ENGINEERING

MACRo 2010

PROCEEDINGS OF THE 2nd CONFERENCE
ON RECENT ACHIEVEMENTS
IN MECHATRONICS, AUTOMATION,
COMPUTER SCIENCE AND ROBOTICS

Edited by

József DOMOKOS

S C I E N T I A
Publishing House
Cluj-Napoca •2010

The book was published with the support of the:

SAPIENTIA FOUNDATION

The authors take all the professional responsibility for the present volume

Series cover:

Dénes Miklósi

First English edition: 2010

Descrierea CIP a Bibliotecii Naţionale a României

MACRo 2010. Proceedings of the 2nd Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics / ed.: József Domokos – [Cluj-Napoca] : Scientia, 2010

Bibliogr.

ISBN 978-973-1970-39-4

Contents

Foreword

The MACRo 2010 - International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics, organized by the Electrical Engineering Department and the Mechanical Engineering Department of Sapientia University is a young conference. It celebrates only the second edition and this is the first time when the accepted papers are published in a conference proceedings.

The conference aims to present original papers based on theoretical research, simulation, experimental work, design, development, testing and measurements with the financial support of "*Szülőföld Alap*" foundation.

The conference topics include, but are not limited to:

- Mechatronic systems
- Mechanical engineering
- Manufacturing
- Automation
- Robotics
- Computer science
- Telecommunication

Published papers were selected after a rigorous reviewing process. All papers have been reviewed by three specialists in the field and only the papers with two positive reviews have been accepted.

In this second edition we have five sections: the plenary section which includes the *"Bolyai János" Plenary lecture*, the *Industrial reality plenary session* and three Parallel sections: *Computer science, communication and signal processing*; *Automation and electrical drives*; *Mechatronics and robotics*, with a conference program which includes 27 papers, 4 invited lectures and a plenary lecture, bringing together authors from 7 countries.

In the end, I would like to express my gratitude to the members of the Program Committee for their effort in revising the papers, and to the Editorial board which has made the finalization of the Conference Proceedings possible.

Editor,
József DOMOKOS

Plenary session

„Bolyai János" plenary lecture:
Zsolt J. LACZIK
Department of Engineering Science, University of Oxford
3D Diffractive Beam Shaping

Invited lecture:
Attila SIPOS
Department of Telecommunications, Budapest University of Technology and Economics
On the Way Toward the NGN and NGOSS

Invited lecture:
Tibor TORÓ
Corresponding member of Hungarian Academy of Sciences
From Bolyai's Hidden Treasures to Einstein's Last Dream

Industrial reality session

Invited lecture:
Levente NAGY
CEE Business Development Manager, Fujitsu Microelectronics
Human - machine interfaces and applications

Invited lecture:
András OLTEÁN-PÉTER
Executive director, Nivelco Romania SRL
Nivelco News

3-D Diffractive Beam Shaping

Zsolt J. Laczik

Department of Engineering Science, University of Oxford,
Parks Road, Oxford OX1 3PJ, UK, e-mail: Laczik@eng.ox.ac.uk

Abstract: Laser beam shaping problems are conventionally defined by the required intensity and/or phase distribution in a single 2-D output plane. For a number of application areas it is beneficial to be able to work with beams that have a particular intensity distribution that is specified in a 3-D volume. Laser material processing, optical microscopy and laser trapping are a few examples for these. We will discuss how diffractive optical elements can be designed and used to generate beams with prescribed 3-D intensity profiles. Practical examples will be given to illustrate the application of these 3-D laser beams.

Keywords: Diffractive optics, 3-D beam shaping, direct binary search.

1. Introduction

By its 50th birthday the laser has really become ubiquitous. We can find it everywhere we look, in our DVD players at home, in complex experimental setups in research laboratories, and in many industrial environments which range from shipbuilding and car manufacturing to the precision manufacturing of microchips. In many of these applications the laser beam is collimated or focused, but essentially remains unmodified. There is, however, a number of applications where specific tailoring of the intensity distribution in the laser beam is desirable or even essential.

Diffraction gratings and lenses had been used since the early days of lasers to modify the properties of the laser beams. These arrangements provided solutions where a single laser beam was split into multiple angularly separated beams, or specific intensity distributions, for example spot or line patterns, were produced in the focal plane of a focusing lens.

There is a similar, but special class of laser beam shaping problems where the desired intensity distribution is specified in an extended 3-D volume. In the following I shall briefly describe how special diffraction gratings, so called diffractive optical elements (DOEs) can be designed and used to convert, or shape a conventional laser beam in order to satisfy such a 3-D intensity specification. Finally I shall provide a small selection of example applications to illustrate the type of effects that can be achieved.

2. Diffractive optics design

The main technique used for the design of DOEs is based on a direct binary search (DBS) algorithm, which operates as follows: in general it is assumed that a collimated input beam is focused by a lens and the 3-D volume where the light intensity distribution is specified is located at the focus of this lens. For a set of discrete target points the desired target intensities are specified. The pupil of the focusing system is divided into a large number of pixels. The pixel number is selected sufficiently large, so that within each pixel the complex transmittance can be approximated as being constant. Normally the aim of the optimization is to design a phase-only pupil filter, in which case the magnitude of the pixel transmittances is fixed at unity and only the pixel phases are manipulated.

At the beginning of the optimization a random phase is assigned to each pixel. During the next step the initial target intensities are calculated corresponding to the initial pixel phase values. From the target intensities (I_{Tk}) and predicted intensities (I_k) an error function (ε) is then calculated. The simplest and most frequently used form of the error function is the mean square error:

$$\varepsilon = \sum_k \left(I_{Tk} - I_k\right)^2 . \tag{1}$$

For most target specifications (1) is a good choice, however occasionally more complicated error functions are used. For example, the error contributions of the different target points may be weighted

$$\varepsilon = \sum_k w_k \left(I_{Tk} - I_k\right)^2 , \tag{2}$$

where the w_k weights are used to emphasise certain regions.

In the iterative optimization loop single pixels are selected at random, and the phase of the selected pixel is adjusted by the addition of a fixed phase step. The intensities at the target points are then recalculated and the new error value ε is determined. If the error improved the pixel phase change is kept, otherwise the pixel phase is reverted to its previous value. These steps are repeated until ε becomes smaller than a preset threshold, or until ε does not change for a preset number of iteration cycles.

The DBS technique is generally very flexible and robust and the only significant limitation is that the computational cost increases rapidly as the number of target point positions increases. This limitation means that although DBS is ideally suited to the optimization of the shape of one or more focused spots, for the design of extended intensity distributions (e.g. Gaussian to top hat converters) other techniques, such as the iterative FFT technique, may be better.

3. Application examples

A. Laser machining using dynamic DOEs

Laser marking of electronic components, integrated circuits in particular, is conventionally done sequentially by scanning a focussed laser beam along the text that is to be written. Alternatively, a series of DOE patterns can be designed to generate intensity distributions that match the character shapes and the surface profile of the component. These DOE patterns are then displayed on dynamic programmable phase modulators to inscribe the desired text on the component (*Fig. 1*).

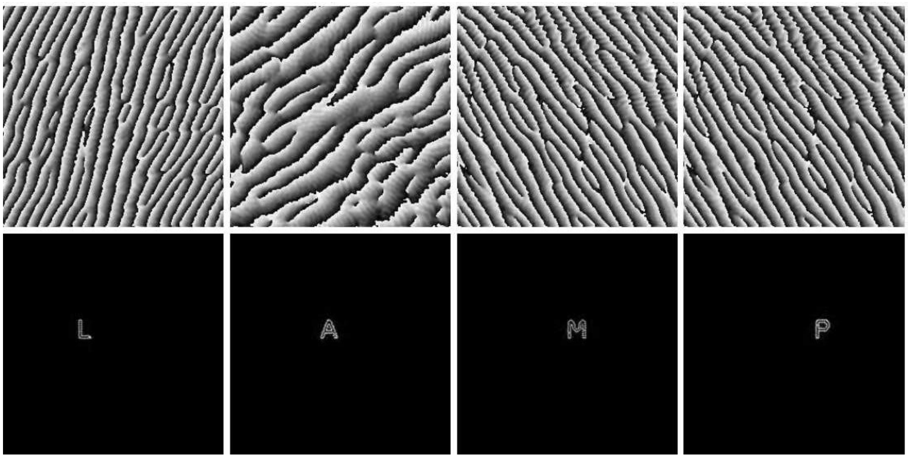


Figure 1: DOE phase patterns (top) and the corresponding intensity distributions (bottom) at the lens focus (component surface) in a laser marking system. The grey levels in the phase pattern images represent a phase shift range of 0 to 2π.

B. Confocal scanning optical microscopy

In a confocal microscope the specimen is illuminated by a laser beam focused to a single spot and light transmitted or reflected by the specimen is collected using a point detector. An image of the specimen is then built up by raster scanning the illumination over the specimen. As a first approximation, the resolution of the microscope is determined by the size of the focused spot. DOEs can be used to modify / reduce the size of the focused spot and to improve the lateral resolution and sectioning (axial resolution) of the microscope (*Fig. 2*) [1].

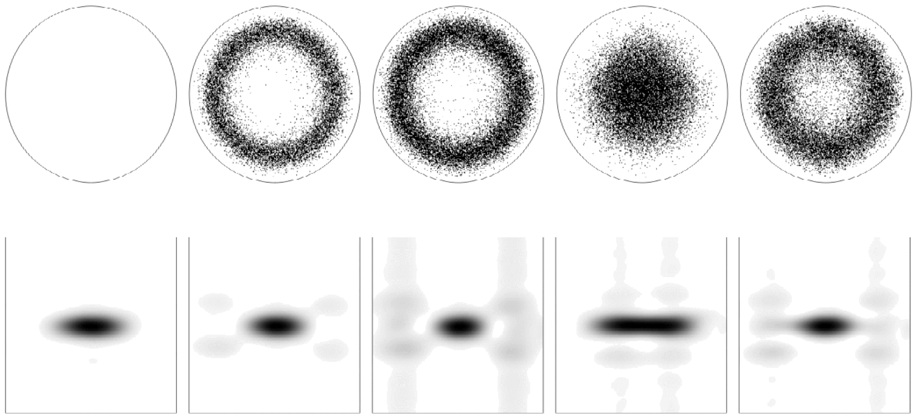


Figure 2: DOE phase patterns (top) and the corresponding intensity distributions (bottom) in a confocal scanning optical microscope. In the binary phase pattern images black and white represent a phase shift of 0 and π respectively. The intensity images represent axial-radial cross-sections through the focused spot in the microscope. The vertical and horizontal directions in the images correspond to radial and axial directions respectively in the focussed spot. The five image pairs show (left to right) the original microscope resolution, 15% axial, 25% axial, 15% lateral and 10% lateral plus 10% axial super-resolution.

C. Enhanced laser cutting

Focused laser beams are routinely used for the cutting of sheet material. One of the problems in these laser cutting systems is the limited axial size of the high intensity region in the cutting beam. This normally requires the mounting of the cutting head on a servo translation stage or the use of the robotic arm. A cost efficient alternative solution is to use a DOE to axially extend the high intensity region.

Fig. 3 shows the phase profile of a DOE designed for this purpose and the axial-radial cross-section intensity profile generated when the DOE is inserted in the system. For a corrugated Al foil, *Fig. 4* shows the incomplete cuts obtained using the original system with the servo stage turned off, and the improved cuts when using the DOE.

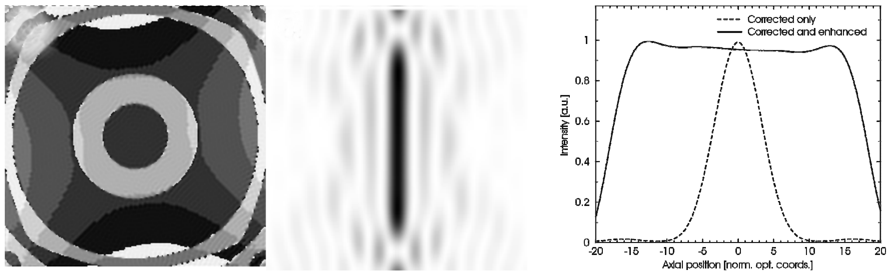


Figure 3: DOE phase pattern (left) and the corresponding intensity distribution (middle) in a laser cutting system. The grey levels in the phase pattern image represent a phase shift range of 0 to 2π. The intensity image is an axial-radial cross-section through the high intensity region of the cutting beam showing an approximately 4x axial extension when compared to the original system. This extension is also demonstrated by the axial intensity line traces from the same region as shown on the right.

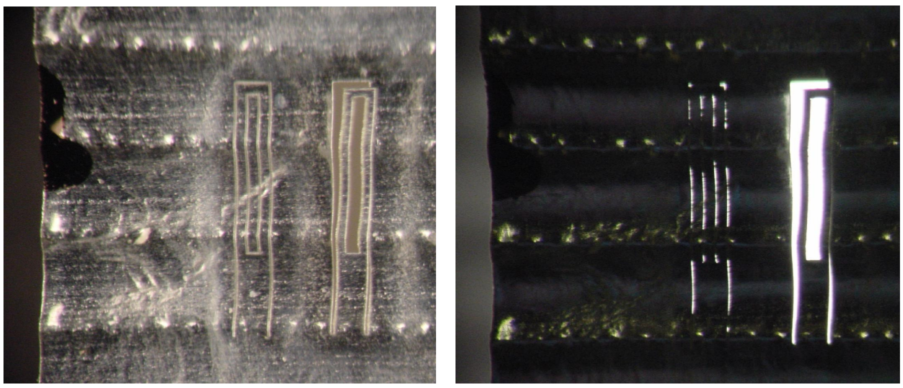


Figure 4: Reflected (left) and transmitted (right) light images showing imperfect and good quality cuts in a corrugated Al foil. In each image the cuts on the left were obtained using the original system, while the cuts on the right were obtained using the system incorporating the DOE.

D. Optical tweezers

High intensity focused light spots with strong intensity gradients polarise small dielectric objects and generate a net force that pushes the objects towards the highest intensity region in spot. Based on this effect focused laser beams can be used to trap and move / manipulate small objects, for example micron sized silica spheres, creating optical tweezers. Optical tweezers have been used for years to manipulate small specimens, but were limited either to a single trap or to time-multiplexing the trap to manipulate multiple objects. DOEs can be used to create efficient complicated multi-trap arrangements and if the phase shifts corresponding to the DOE patterns are implemented on dynamically programmable phase modulators, multiple objects can be trapped and manipulated in an extremely versatile manner. *Fig. 5* shows an image sequence where 18 silica spheres are initially trapped in a planar arrangement and then moved to create a zincblende unit cell [2]. The whole cell is then freely rotated and scaled demonstrating full 3-D control of the individual laser traps. The traps were created using a sequence of DOE patterns displayed on a programmable phase modulator.

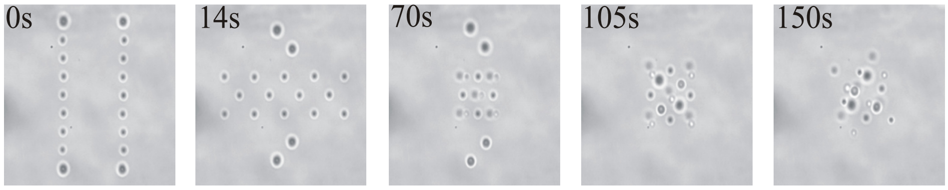


Figure 5: Sequence of video frames showing the morphing of a 15µm zincblende unit cell made from fourteen 1µm and four 2µm silica spheres.

4. Conclusion

The above examples can only give a flavour of the possibilities created by diffractive optical elements that are designed to generate laser beams to 3-D specifications and are implemented using programmable phase modulators. The development of these DOEs has become a new enabling technology and is now attracting significant interest and is finding many new applications.

Acknowledgements

The support of The Royal Society of London is gratefully acknowledged.

References

[1] Neil, M.A.A. et al., “Optimized pupil-plane filters for confocal microscope point-spread function engineering”, *Optics Letters,* vol. 25, no. 4, p. 245, 2000.

[2] Sinclair, G. et al., “Assembly of 3-dimensional structures using programmable holographic optical tweezers”, *Optics Express*, vol. 12, no. 22, p. 5475, 2004.

On the Way Toward the NGN and NGOSS

Attila SIPOS[1]

[1] Department of Telecommunications, Faculty of Electrical Engineering and Informatics,
Budapest University of Technology and Economics, Budapest

NETvisor Informatikai és Kommunikációs Zrt.

Abstract: The paper gives a global view about the architecture of Next Generation Network (NGN), the most important protocols, bandwidth demands of new services and main network technologies. This describes the main features of optical cables, broadband mobile and fixed access technologies, optical wavelength division systems and core routers. The presentation explains that the application of intelligent telecommunication equipments and the GIS based network inventory open a new perspective in network operation, construction, network failure elimination, customer assistance and management of network resources. The introduction of NGN gives the possibility and in the same time it demands the use of New Generation Operations Systems and Software (NGOSS).

Keywords: Next Generation Network, optical cable, broadband access technologies, Wave Length Multiplexers, routers, Geographical Information Systems, New Generation Operation System and Software.

From Bolyai's Hidden Treasures to Einstein's Last Dream

Tibor TORÓ

Corresponding member of the Hungarian Academy of Sciences
tt.neutrino@yahoo.com

Abstract: In this paper we attempt to obtain a connection between the two keywords, which are formulated in the title of our lecture: "Bolyai's hidden treasures" and "Einstein's last dream".

The expression of "hidden treasures" appeared in a János Bolyai's thesis in the following manner: "The space contains in its depths, such treasures which the man who is moving on the surface, never can see" (Paul Stackel: Bolyai Farkas és Bolyai János geometriai vizsgálatai, Bp. MTA, 1914, vol. II, p.293)

The notion of "Einstein's last dream" was formulated by the famous Nobel prize-winner physicist Abdus Salam, originated from Pakistan, in his Einstein's centenary memorial lecture, with following title: "The Einstein's last dream: the space-time unification of fundamental forces" (UNESCO, Paris 9 may, 1979). This is nothing else but the idea of geometrisation of fundamental physical forces in which was interested Bolyai himself. (T. Toro: Physica more geometrico, Bolyai's and Einstein's unfinished symphony, in Bolyai Memorial Volume, Ed Vince, Bp. 2004).

In our paper we will show which should be these hidden treasures of space-time (The Weyl's non-metricity, the torsion of Cartan, the Kaluza-Klein type extra-dimensions, the non-Abelian local gauge invariance of Yang-Mills, Utiyama, Kibble etc.) through which we will can arrive to realization of Einstein's last dream.

Computer science, communication and signal processing

Information Spreading in Self Organizing Mobile Network

Levente CSIKOR, Zoltán FEHÉR

Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, e-mail: levente.csikor@gmail.com, zolifeher@gmail.com

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: During the emergence of the information society social network based applications are becoming increasingly popular. Prominent examples are online acquaintance networks, chatting and blogging boards or virtual dating systems.

In the paper a novel information spreading system is proposed. The main idea is that data resided on mobile phones carried by persons are exchanged when devices get close to each other. Communication is carried out by the widely accepted Bluetooth protocol. Such a system can have many interesting applications, like infrastructure-less news propagation, direct marketing based on common interests or simply finding friends according to common daily routine like getting to work in the morning along the same transportation lines.

Keywords: Bluetooth, information spreading, self-organizing, social network, moving human groups

1. Introduction

As a result of the evolution of social networks most people have a virtual circle of friends. People like sharing information about themselves getting to know someone or joining to a group.

The basis of all social networks is that people can connect and communicate to each other. Acquaintances are "marked", i. e. the real contacts can be placed into a virtual world resulting in a traceable and transparent network where nodes are the persons and the edges indicate relationships. Online social connections today are formed though web based applications although similar networks may exist where dozens of people appear. Most interestingly a network can be formed also by people traveling on the same public transport or by visitors of a pleasure-ground. It is not a coincidence that people are in the same place at the same time. Similar purposes can form a link between attendees. These formations are often invisible for the superficial observer.

In this paper an information spreading system is proposed based on connecting isolated nodes represented by Bluetooth enabled mobile phones. The condition of spreading is based merely on spatial and temporal locality. Different measurement scenarios are presented to support the potential of the presented system.

A recently emerged related research carried out by NICTA and I^2R[1]. Their ultimate goal is to design a patent system that let people connecting each other in real-time using their mobile or smart phones, without the need for expensive communication infrastructure. The network is formed of devices that come within range of another. The main emphasis in the project is merely to create a working ad-hoc type networking environment using cheap prevailing communication technologies. In the related research program CRAWDAD[2] special-purpose hardware (iMote) is built to discover Bluetooth devices. Collected data includes a number of traces of Bluetooth sightings by groups of users carrying iMotes for a number of days. As of today, there is no further activity known in the project.

In the following section a measurement based analysis is outlined aiming to validate the idea of the possibility of a human motion based information spreading system. The measurements are performed by a *Bluetooth logger* application. Two kind of measuring scenario are discussed in short, one is for confined areas and one on public transport lines. Additional details and further results about the measurements are discussed in [5]. In Section 3 the proposed information spreading system is presented. After the general overview of the system the functional principles of the communication protocol is presented followed by a short discussion. In the last section before the conclusion possible extensions to the information spreading system are sketched up.

2. Connecting opportunities among people

Living in the information age enables that most people today have chance to use a computer, connect to Internet, send and receive emails or call someone at anytime from anywhere, but the most important thing a cell phone is always present in the pocket. As the result of the evolution of technology today's mobile devices are supported with short-range communication standards like Bluetooth and are programmable thanks to Java programming language. This language is ubiquitous and most cell phones are provided by a variant of versions of it. The key point of this article is providing a novel system where connecting people in real-time is permitted by an easy-to-use mobile application based on Bluetooth because of its incidence. As a result of the range of this technology (approx. 10m) only nearby people are reachable. Beginning to plan the specifics of the system, first, the important question emerges: How many

Bluetooth enabled devices can be discovered in our close environment and accordingly how many people can be connected in the same place at the same time by this type of communication method? To create comprehensive statistics and draw conclusions a special-purpose mobile application was developed felicitous called Bluetooth logger to answer these questions. As it is shown below and in [5] numerous devices were logged and many interesting relations discovered among moving groups of people supporting the idea of the Bluetooth-based system.

Wide range of mobile and smart phones with Java and Bluetooth capabilities meets the requirements to be part of the measurement. The application does auto inquiries in preset intervals[1], in case of finding a visible Bluetooth device a record is created with the physical addresses[2], time stamp, friendly name (if available) and device class. Records are stored in *.log* files in the memory of the phone or on memory card until uploaded to an SQL database, where location information is added.

A. Results of logging measurements

In current section according to [5] hidden relationships between people are demonstrated by discovering moving groups. Based on the results in Section 3 an information spreading system and its functional principles are shown. The measurements are carried out by about **10** people discovering more than **6600** different devices altogether. **368** (approx. 6%) devices were discovered by two different loggers at two different time. Measurements and empirical experiments show that every **20**$^{\mathbf{th}}$ person has a Bluetooth-enabled device, which

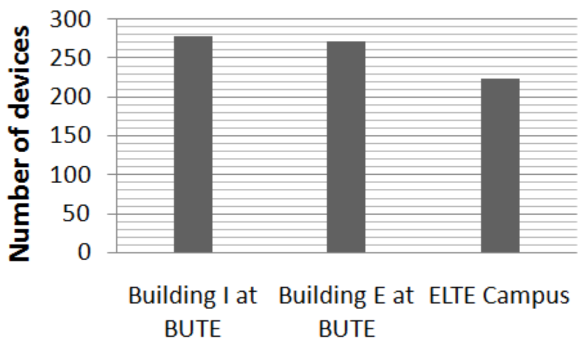


Figure 1: Number of discovered devices in campus areas.

1 Default period is 120 seconds
2 The logger's and the logged device's MAC address is stored

is significant, considering the fact that, as of today, besides headsets and synchronization, applications or services inspiring the usage of Bluetooth connections are practically non-existent.

One part of measurements was created in bounded areas covering two buildings of the *Campus of Budapest University of Technology and Economics* and one building at the near *Campus of Eötvös Lóránd University*. During the measurements numerous devices are discovered in the campus areas (See *Fig. 1*).

Few persons were discovered in both of the monitored places. **28%** of devices were logged more than once in *Building I* (77 of 278) and **25%** (67 of 272) of devices in *Building E*.

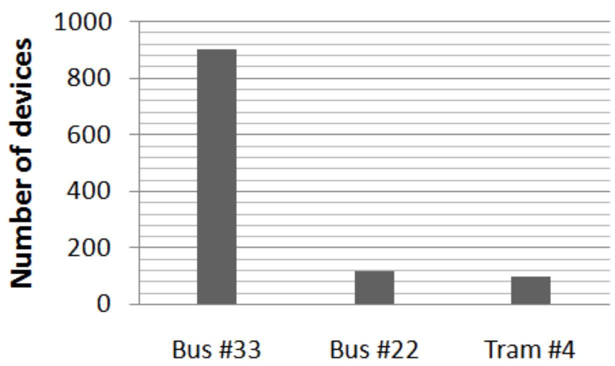


Figure 2: Distribution of devices on top public vehicles.

Different but also suitable locations for logging are public transportation vehicles (see *Fig. 2*). Most of the loggers live in a common dormitory, which is along the route of Bus #33, therefore most outdoor inquiries happened on this bus. More than **1000** distinct device were logged. Only one logger used Bus #22 daily where more than **100** devices are discovered meaning there is a strong relation between the number of discovered devices and the number of loggers while the transport line is less important. For further details about the measurements see [5].

3. Information spreading system

In the section a patent information spreading system and an integral protocol are presented that connects people to each other and exchange information merely based on spatial and temporal locality. The main idea is that data resided on mobile phones carried by persons are exchanged when devices get close to each other. Communication is carried out by the widely accepted Bluetooth protocol. The connections and information exchanges between people can be considered as links resulting a new way to connect people as one of many other opportunities of creating virtual relationships in social networks.

A. Communication among mobile devices

Bluetooth is a low-cost, short-range wireless technology to replace the wires and create *personal area networks (PANs)*. Each PAN is dynamically created of devices that come within range of another, enabling cellular phones to connect automatically and share data immediately. To support development of Bluetooth-enabled software on the Java platform, the *Java Community Process (JCP)* has defined *JSR*[3] *82*, the *Java APIs for Bluetooth Wireless Technology (JABWT)*.

B. Functional principles

During the connection building process two roles are defined, a server and a client role. A client consumes remote services. Nearby devices are being discovered and for each discovered device services are searched. Server registers services in the *Service Discovery Database (SDDB)* and advertise them. After this point connections are accepted as they come in. Finally, when the service is no longer needed it is removed from the *SDDB*.

In this system two types of nodes are defined, source node as server and destination node as client. The server advertise the service while other nodes (client) connect to them and write preset message into their memory. If both nodes have information to distribute an information exchange is done. Received messages also will be spread in the following connections[4]. The basic protocol messages are shown in *Fig.3*. Avoiding interference between advertised services an *UUID (Universal Unique Identifier)* is necessary for registering any supply. The promoted services can be searched and found in frequented areas without device discovery if remote devices' MAC address is known. Consequently, information spread/exchange accomplish faster on previously known devices also more battery power is saved. The service is described with a string which defines the communication profiles and contains a previously mentioned UUID and additional security settings (*authentication, authorization, encryption*). Three profiles are defined: *RFCOMM* (serial port emulation), *OBEX* (Object Exchange) and a lower-layer *L2CAP* (Logical Link Control and Adaptation Protocol). Using upper-layer profiles enables excluding L2CAP provided packet sizes, fragmentation, etc. Authentication and authorization requires access granting, hence every time when information exchange can be

3 **J**ava **S**pecification **R**equest
4 Accordingly an information can have more than one path

happened users have to grant incoming connections. Encryption cannot be used without authentication. To minimize user interaction, in our system all the security methods are set to false.

According to the logger application, every node of the system starts inquiries in every two seconds. If remote device founded and it is provided by the service connection will be established and information exchange will be made provided they reside in each other's close environment long enough.

The added messages are stored in *RMS (Record Management System)* because most Java provided mobile phones are supported with this type of persistent data structure[5]. Multitasking capable mobile operating systems, like Symbian, makes it possible the application to run in the background. When incoming message arrives a vibrating alert occurs denoting the user about it. The received information is stored in RMS as well, so all messages can be viewed at any time. All of them are stored in the same record store including the preset one, so during later connections more and more information can be spread and thus messages can have more than one variant paths.

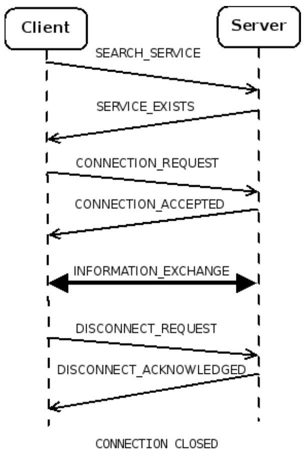


Figure.3: Basic protocol messages.

The procedure works fully automatically, as formulated messages *"infects"*[3] the nearby devices without any user interaction. Message types and user filtering can be defined to avoid flooding. A message could have many classification e. g. advertisement, interest, news, dating, etc. and users have ratings to denote confidence. Naturally scores are device specific because of the absence of a central server. Consequently, this type of social software is decentralized and self-organizing. Thanks to the specifications no expensive

5 Storing data on file system is complicated without signings from trusted certificate authorities

special-purpose hardware or built infrastructure is needed to become a member of the network The size of the network only[6] depends on the number of Bluetooth enabled devices in vicinity.

C. Technical capabilities

Usability depends on the state of the Bluetooth environment, the duration of the discovery process is affected by the number of devices in vicinity and computing assimilation of mobile phones. The interval length of exploration and the number of Bluetooth enabled cell phones has a linear relation. To find a remote device it takes 6-9 seconds while device discovery in a frequented area might takes up to several minutes, however, the applicability of the software is not narrowed by these facts. The key point here is to receive information irrespectively of the original poster, i. e. messages can have different paths meaning users can get information from many other participants.

4. Possible Extensions to the Information Spreading Application

The previously mentioned logger tool besides information spreading can have many useful extensions. People carrying cell phones with the Bluetooth switched on leave traces behind that can be detected by logger agents.

A. Following virtual footsteps

With careful planning of the number and movement patterns of the agents a human tracking system can be built. With these we can sketch up a moving track of the logged person. If we have records with special locations and time stamps e. g. being in some places or traveling somewhere, a daily routine of a person could be deducted.

B. Bluetooth message board

Stationary messages can be appropriate for information spreading as well as mobile ones. Information can be stored in a dedicated place (e.g. *Bluetooth hot-spot*) instead of distribution through moving devices in range. All the messages are filtered by recipient. When the specified device is in the vicinity left messages are sent. In this case the presence of an end-to-end connection is not necessary as it is in Delay Tolerant Networks (DTN) [4].

6 The software and its requirements are also important

C. Feature extension in Peer-to-Peer mobile networks

An increasing number of mobile phones are provided by GPS and mobile internet capabilities on the market, but the functions of older common cell phones can be expanded via Bluetooth by GPS and/or mobile internet provided mobile phones. If advanced phones are present in the same area common mobile phone can determine its position by sending a GPS request to it and after calculating the coordinates the smart phone answers. Running of the software in several cases is described In *Fig.4.* Five devices are in range and the cell phone in the middle wants to get its position. In case 1 device B is found but it is not running the software. In case 2 device C is located and the service is found but it lacks of GPS module and so no position information is sent. In case 3 GPS provided device D is discovered and it responses with its coordinates. Thanks to the range of Bluetooth (approx. 10 meters) the location is almost precise consequently the position can be determined on average cell phones.

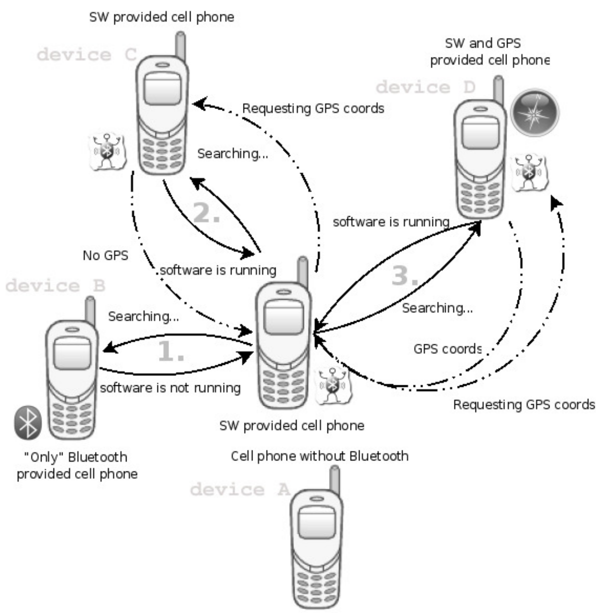


Figure 4: Expanding device capabilities.

D. Basic acquaintance network

As the result of the evolution of information technology and permeation of internet the consumers who want to buy a computer for stand-alone applications have almost fully disappeared. These days the aim of buying a personal computer is to use it as a communication device, which we can use to access the Internet, read e-mails and do social networking. Nowadays the same can be said about the mobile devices, which also went through a major evolution in the network era. Today's cell phones are provided by camera, GPS, mobile internet, relatively big screen for more comfortable usage and large computing capacity like personal computers had 5-10 years ago. In our days mobile phones can be more appropriate sharing information about people because position calculation, pictures, videos and messages can be created and posted anytime and anywhere.

Similarly, to social networking *Facebook* like profile pages can be created wherein beyond elementary data, tag-like entries are added. Two kind of tags are present. One is set by the user (interest, workplace, groups, etc.) and the other is created automatically by the phone (visited web pages, listened songs, number of sent messages, etc.). Users can create their own tags e. g. interests, favorite food, etc. On the other hand people can be tagged via hot-spots as well. These labels also contain place-specific information. This type of system can provides the people to get information about others residing in the same area without any explicit conversation.

5. Conclusion

In the paper a patent information spreading system were demonstrated where no expensive special-purpose hardware or built infrastructure is necessary. Only Bluetooth enabled common cell phones needed and messages can be sent by devices coming within range. As the results of storing and forwarding all incoming messages different routing paths can be drawn, consequently several users can be the poster (naturally not the original) of the same information. Thanks to its protocol messages can be spread fully automatically without any user interaction. Measurements were also presented proving the possibility of existence of the system. In the paper several kind of further extensions were sketched up as well. With the presented system ad-hoc and self-organizing social networks can be created the bases only on spatial and temporal locality. In a hypothetical world where all people carrying Bluetooth enabled cell phones, using such a system, the information can be spread faster than orally with no fixed infrastructure needed.

Acknowledgements

The authors would like to thank all the participating loggers for helping collect measurement data. They are also grateful to the Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics for their general support.

References

1. NICTA, I2R: "I2R to collaborate with NICTA in building innovative wireless network using mobile phones", Media release, Singapore, 2009 September 30.
2. Natarajan A., Motani M., Srinivasan V., "Understanding Urban Interactions from Bluetooth Phone Contact Traces", Lecture Notes in Computer Science, 2007,vol. 4427/2007, pp. 115-124
3. Wang P., González M. C., Hidalgo C.A., Barabási A-L., "Understanding the Spreading Patterns of Mobile Phone Viruses", Science, 2009, vol. 324, pp. 1071-1075.
4. Michael D., Eric B., Kevin F., Susant J., Melissa H., Rabin P., "Implementing delay tolerant networking", Technical Report IRB-TR-04-020, Intel Research Berkeley, 2004.
5. Csikor, L.,Fehér Z., "Exploring Hidden Relations in Moving Human Groups", To appear at POSTER2010, Prague, Czech Republic

A Platform for Implementing Secure Wireless Ad Hoc Networks

Gyula FARKAS[1], Béla GENGE[2], Piroska HALLER[3]

Department of Electrical Engineering, Faculty of Engineering,
"Petru Maior" University, Tg. Mureş, e-mail: [1]farkas_gyula@yahoo.com,
[2]bgenge@engineering.upm.ro, [3]phaller@engineering.upm.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: We propose a new platform for implementing secure wireless ad hoc networks. Our proposal is based on a modular architecture, with the software stack constructed directly on the Ethernet layer. Within our platform we use a new security protocol that we designed to ensure mutual authentication between nodes and a secure key exchange. The correctness of the proposed security protocol is ensured by Guttman's authentication tests.

Keywords: Wireless, ad hoc, security protocols.

1. Introduction

Unlike traditional wireless networks, wireless ad hoc networks do not rely on any fixed infrastructures. Nodes communicate with other nodes that are within their radio range and by using routed messages to communicate with nodes that are not directly reachable. This flexibility ensures that wireless networks can be set up in an ad hoc manner where there is no infrastructure available.

Ensuring a secure communication that also provides security properties such as authentication, secrecy, integrity and non-repudiation involves using security protocols [1]. There have been several security protocols and solutions proposed in the literature to ensure security goals [2, 3, 4]. However, these solutions are mainly focused on designing new security protocols and secure routing protocols and do not focus on providing a modular platform where security protocol implementations can be easily replaced with other implementations.

In this paper, we propose a solution for the authentication problem of nodes communicating directly and a solution to the secrecy, integrity and non-repudiation of messages exchanged between nodes. Instead of focusing on the security of routing protocols, we provide a platform that ensures security of directly linked nodes with a modular architecture that can be used to implement secure discovery, secure routing, and other applications for wireless ad hoc networks.

In order to ensure mutual authentication and a secure key exchange between directly communicating nodes, we design a new protocol, based on Guttman's authentication tests [5]. This protocol can be replaced with other implementations in order to satisfy the security requirements of applications. The platform requires a Linux OS, the OpenSSL [6] security library and a wireless network adapter for execution.

The paper is structured as follows.

2. Platform architecture

The software stack of the proposed platform is shown in *Fig. 1*. Our platform uses the *Ethernet* layer as a base package transfer layer. The advantages of using this layer are multiple. First of al, building ad hoc networks over well known network or transport layers is not possible because of the dynamic of ad hoc networks where nodes can change their location and there is no fixed infrastructure to provide IP addresses or to route packages based on IP addresses.

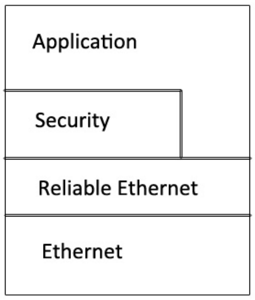


Figure 1: Platform software stack.

As another advantage, the overhead that would be generated by IP/TCP/UDP protocols is completely removed. However, this means that we must now deal with package loss/fragmenting. For this purpose we introduce a new layer on top of the simple Ethernet layer: *Reliable Ethernet*.

The *Security* layer is the one that provides an implementation of security goals for the platform. It ensures that nodes are properly authenticated, it ensures secrecy and integrity of data.

Based on the security layer, we can implement several applications on the *Application* layer. Applications can also bypass the security layer if security is not a requirement, or if it is handled at the application layer.

The block diagram of the proposed platform can be seen in *Fig. 2*. The platform consists of several modules. The "Connection Management" module is the core of the platform, it is responsible for creating message headers and maintaining connections. The "Reliable Network" is the implementation of the *Reliable Ethernet* layer. The "Container" implements the management of *devices* representing state machines corresponding to each node that has been discovered.

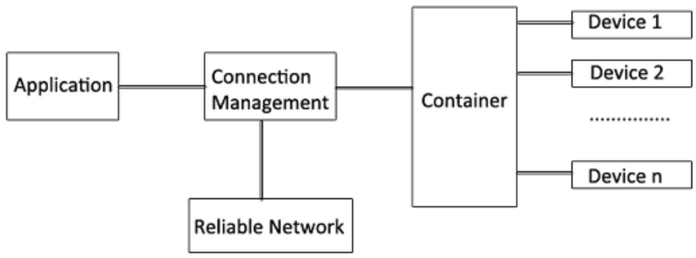


Figure 2: Internal block diagram.

3. Security considerations

The authentication can be based on password or certificate. We can distinguish three cases, where participants *A* and *B* both have certificates; only *A* has a certificate; no participant has a certificate. In this paper we present the protocol we used for mutual authentication where both parties are in the possession of certificates obtained from a Certificate Authority. The other cases are derived from this protocol, leading to the authentication of only one party, pre-defined key-based authentications and no authentication.

Guttman and Fabrega proposed a method for designing authentication protocols [5], based on authentication tests. In our proposed platform we used this method to design a new mutual authentication protocol that also provides a secure session key exchange. Session keys are used later to provide secrecy of data through symmetric block ciphers and integrity of data through message authentication codes.

The designed security protocol is shown in *Fig. 3*. The initiator of the protocol is *A*, who generated a random number N_1 and sends it to *B* along with his certificate C_A. *B* generates a random number N_2 and a session key *K* which is encrypted with its private key. *B* creates the hash $\{N_1, K, \{K\}_{Pk(A)}\}_h$ and signs it with his private key. Then, *B* sends the message $N_2, K, \{K\}_{Pk(A)}, C_B, \{ \{N_1, K, \{K\}_{Pk(A)}\}_h\}_{Pvk(B)}$. Finally *A* will confirm that he is running the same session by generating and sending to *B* the hash $\{N_1,N_2,K, \{K\}_{Pk(A)}\}$ signed with his private key.

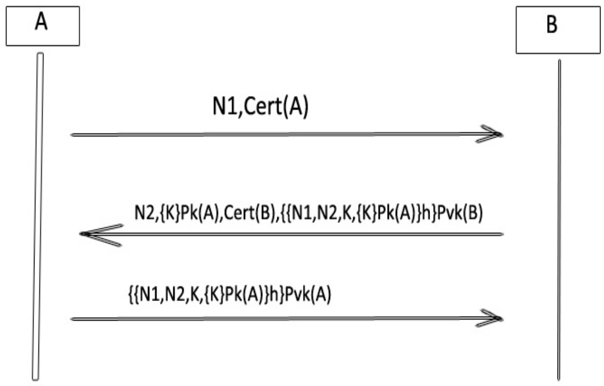


Figure 3: Proposed mutual authentication protocol.

4. Conclusion

We proposed a new platform for implementing wireless ad hoc networks. The platform has been designed for Linux OS, with a compiled OpenSSL library where a wireless network adapter is available. The proposed platform can be used to implement several application types such as: secure routing, secure data transfer, secure node discovery. Within the platform we use a new security protocol that we designed based on Gutman's authentication tests. The protocol ensures mutual node authentication and secure session key exchange.

References

[1] Cremers, C.J.F., and Mauw, S., "Checking secrecy by means of partial order reduction", In S. Leue and T. Systa, editors, Germany, september 7-12, 2003, revised selected papers LNCS, Vol. 3466, 2005.
[2] Zhou, L., Haas, J.Z., "Securing Ad Hoc Networks", Technical Report: TR99-1772, Cornell University, 1999.
[3] Park, J.C., and Kasera, S.K., "Securing Ad Hoc Wireless Networks Against Data Injection Attacks Using Firewalls", Wireless Communications and Networking Conference, pp. 2843-2848, 2007.
[4] Kumar, A., Reddy, L.C., and Hiremath, P.S., "Performance Comparison of Wireless Mobile Ad-Hoc Network Routing Protocols", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.6, pp. 338-343, 2008.
[5] Guttman, J.D., and Fabrega, F.J.T., "Authentication tests and the structure of bundles", Theoretical Computer Science, Vol. 283, No. 2, pp. 333-380, 2002.
[6] ***, OpenSSL Project, version 0.9.8h, available at http://www.openssl.org/, 2008.

ECG Tele-Monitoring Using Intelligent Network Services

Virgil CAZACU[1], Dan ROBU[2], Iuliu SZEKELY[3], Florin SANDU[4]

[1] BitDefender, Bucharest, Romania, e-mail: virgil.cazacu@gmail.com
[2] Siemens Program and System Engineering, Brasov, Romania, e-mail: dan.robu@siemens.com
[3] Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, Romania e-mail: gszekely@ms.sapientia.ro
[4] Faculty of Electrical Engineering & Computer Science, "Transilvania" University, Brasov, RO, e-mail: sandu@unitbv.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: The technology and the nature of the Telecom environment are changing rapidly. Network Operators and Service Subscribers have to differentiate themselves from the competitors to keep their market shares and to attack new market segments. One way of differentiating and of increasing revenues is to host IN (Intelligent Network) services that target market segments that usually were not in the operators' focus. In the current paper, such a service is presented, service that is related to the medical tele-monitoring. The paper is presenting a part of the technical aspects and market perspectives specific to such a service.

Keywords: Intelligent Networks, services, tele-monitoring, charging, Java, Distributed Object Technologies

1. General information

New Intelligent Networks (IN) services should be tailored to fulfill new market demands, thus extending the existing IN functionality towards support of SIP-based networks that provide the following functionality (seamlessly to both SS7 and IP Multimedia System domains): Freephone Service, Universal Access Number Service, Premium Rate Service, Televoting Service [1].

Intelligent Networks support service solutions in SS7 fixed and mobile networks as well as in these newly launched SIP networks, with a broad range of value-added service features such as user interaction, flexible routing, flexible screening, flexible pre-rating and others.

IN platforms and the installed IN services can serve multiple network types simultaneously. A typical example can be the VPN of a company that can consist of mobile phones, fixed telephones as well as VoIP (SIP-based) devices. Thus, seamless services for subscribers with mobile, fixed or SIP-based terminals can be provided. This feature is very interesting for Communication Service Providers (CSP) deciding to expand their business using the opportunities of the IMS domain to offer their customers new access-agnostic services.

Usualy, IN strengthens the customer loyalty by offering customized services targeted at the individual. By offering comprehensive world-class feature sets for all user groups, IN ensures maximized usage stimulation, subscriber attraction and efficient roll-out, resulting in increased network traffic and average revenue per user. Customer-specific modifications of the service will be considered with a time-to-market approach.

Increase in the efficiency of the company's internal and external communications or higher call completion attract high-ARPU (Average Revenue per User) customers. Therefore, CSPs are enabled to expand their customer base and to increase their revenues.

IN provides carrier-grade availability as well as highest scalability, capacity and performance.

Next Generation Intelligent Networks (ngIN) represent a well-defined evolution path from current IN platforms. Integrated interfaces enable provisioning of value-added services in SS7 and SIP networks. One system can be used for both domains without compromising customer care, Backup & Restore and OA&M systems [2].

Service Management is of significant importance for the CSPs and covers all management tasks related to intelligent services and to the corresponding service data. Efficient data entry, easy to use services and role specific interfaces enable easy and fast service administration, e.g. implementing customer specific service management applications, import and export of mass data, etc.

User interaction and subscriber self-administration via comfortable Web interfaces, Unstructured Supplementary Service Data (USSD) or Dual-tone multi-frequency (DTMF) interfaces also reduce service management costs.

Fig.1 gives an overview about the general layers used by IN platforms for the realization of IN services.

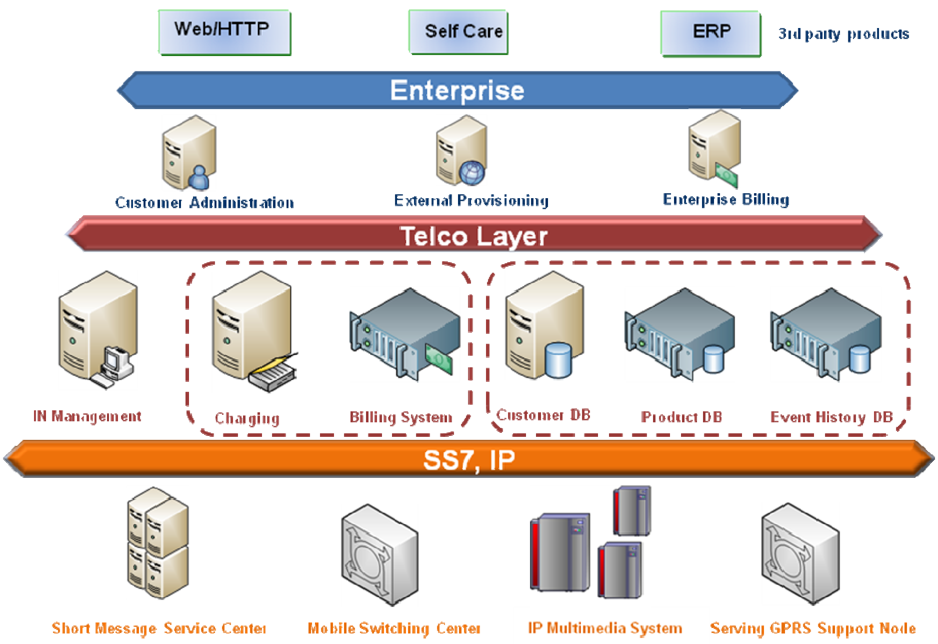


*Figure 1:*Overview of Intelligent Networks layers.

Understanding such tremendous opportunities offered by such service characteristics is the key of a successful business case. In this paper, such a service and solution will be presented, service that is acting in the tele-monitoring domain with direct appliances in medicine. Tele-observations of cardiac patients were and are a hot topic [9], different solutions being applied during the time [10]. The current solution description is a constellation of functionalities that cover users, physicians and operators or market areas.

From the user's point of view, one has to use a portable ECG device which is a non-invasive, handheld heart monitor used to record ECG tracing anywhere and anytime. The parameters are measured and transmitted in different ways depending on the device technology [9]- for example when the user places both thumbs on the electrodes, a wireless connection with the server is launched, for analysis and management of data.

From the physicians' point of view, they receive transmitted data in a common place, for example a database from where they can extract necessary data for patient health evaluation. They can be also alerted by SMS or even by MMS containing information from the users that are under certain observation and targeted for close monitoring.

From the operators' point of view, the solution that an IN service can offer, covers all communication areas, from patient to doctor, having the possibility to administrate even the submitted information. In this way, operators can present to their customers - which can be health institutions like hospitals - an integrated solution for tele-monitoring.

2. Technical information

In this paper section, the technical solution will be presented from all 3 perspectives: the user's, the doctor's and the mobile operator's, being in fact one integrated way out. In fig. 2, a general overview of the solution is presented, emphasizing the main interfaces used.

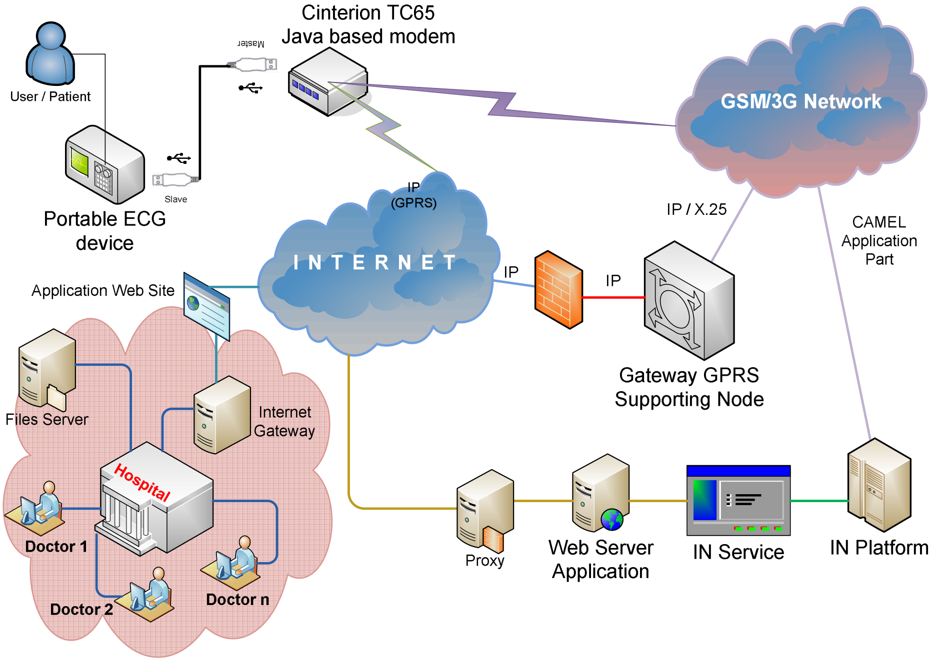


Figure 2: General Architecture of Solution.

A. Monitoring Solution

There are several ECG devices on the market which are cost effective in terms of affordability (from regular customer point of view) and offer standard communication interfaces like Universal Serial Bus (USB) or Bluetooth. For the solution presented in this paper, such a portable ECG device which offered an USB port for connection is used. Using this connection, data stored in the device (e.g. ECG waveforms) can be downloaded, usually on local PCs. Taking this into consideration, a Cinterion GSM/GPRS modem is used to access stored diagrams via USB port. Such a modem is chosen because, besides the GPRS [8], USB and TCP/IP connectivity, it has a build in Java Virtual Machine, having the possibility to develop and deploy fast solutions using Java MicroEdition.

The general steps to transmit the ECG diagrams remotely are:

1. Initialize the serial transfer between monitoring device and the modem
2. Retrieve ECG diagrams from monitoring device, in binary files format, to the modem memory
3. Initialize modem GPRS transfer
4. Send data to the IN service

The serial transfer of the 1st step is a well know serial initialization, the USB connection being reduced to a standard serial communication. Retrieving of data is done using the following Java code:

```
byte[] buffer = new byte[1024];
int lenbuffer = 0;
while ((data = in.read()) > -1) // Read the response...
    buffer[lenbuffer++] = (byte) data;
```

There are different possibilities to perform GPRS initialization, by using different AT commands [3] sequences from Java code [4]. In the next capture, one option is presented using the SJNET command, which allows creating a profile containing all the parameters needed for automated network access through a dialup connection.

```
ATCommand atc = new ATCommand(false);
String textAT = atc.send("AT+CREG?\r");//check if already is a
registration
String response="";
 if (textAT.indexOf("1,1")< 0) // if not a new profile is done
   {
   atc.send("AT+CREG=1");
   doPause(1);
   atc.send("AT^SJNET=\"gprs\",\"internet.vodafone.ro\",\"interne
   t.vodafone.ro\",\"vodafone\",\"193.230.161.3\",0\r");
   doPause(1);
   response=atc.send("ATD*99***1#\r");// establish communication
   if (response.indexOf("CONNECT") >= 0) {                ... }}
```

The last major step after retrieving ECGs data and establishing the GPRS connection is the sending of data to the IN service located on the mobile operator side. In this case, the HTTP POST procedure was adopted for sending the ECG data to the service, together with the additional user information, like International Mobile Subscriber Identity (IMSI) used for the GPRS connection. The next capture is extracted from the sending method deployed on the modem:

```
HttpConnection hcon = null;
DataOutputStream datastr = null;

// open http connection with web server

hcon = (HttpConnection)Connector.open("http://x.x.x.x:8080/
EcgServerCollector/Collector", Connector.READ_WRITE);

// setting the request method

hcon.setRequestMethod(HttpConnection.POST);

// obtain output stream for sending query

datastr = hcon.openDataOutputStream();
byte[] ECGDataByte = ECGDataString.getBytes();
for (int i = 0; i < ECGDataByte.length; i++)
{
// sending query with retrieved ECG data for web server
   datastr.writeByte(ECGDataByte [i]);
}
```

B. IN Service Interface and Functionalities from doctors' perspective

Besides collecting ECGs information from users, the IN service is also dealing with the filtering of received and stored information in order to offer different possibilities to interact with the physicians needs. In this section is presented how such interface is implemented.

The interface is developed as a client-side Java technology-based program using Java Network Launch Protocol (JNLP) and Java Web Start mechanism for program delivery through Web server [5].

Doctors can access a webpage from where they can follow a simple link and the application is downloaded / updated right away.

After the username and password are introduced by the doctor and checked by the application, based on these credentials, the application interface is customized according to the rights based on the roles that the user has in the application. In fig. 3 is presented the interface for a "power user", for example a chief doctor from a hospital. Besides his lists of patients and their ECGs, he can administrate the doctors from his team, offering different types of credentials.

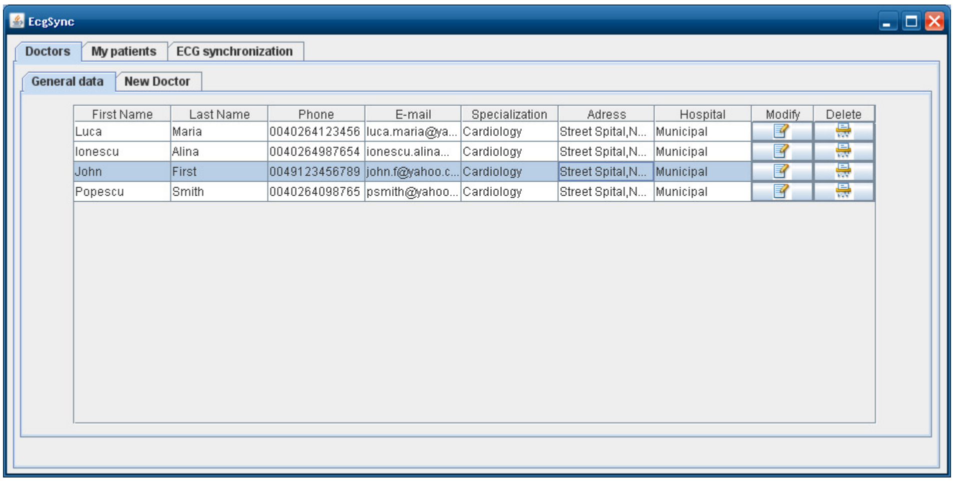


Figure 3: Doctors Team administration.

Entering new patients in the application and the ECGs administration is done using the same type of interface, as shown in the next figure.

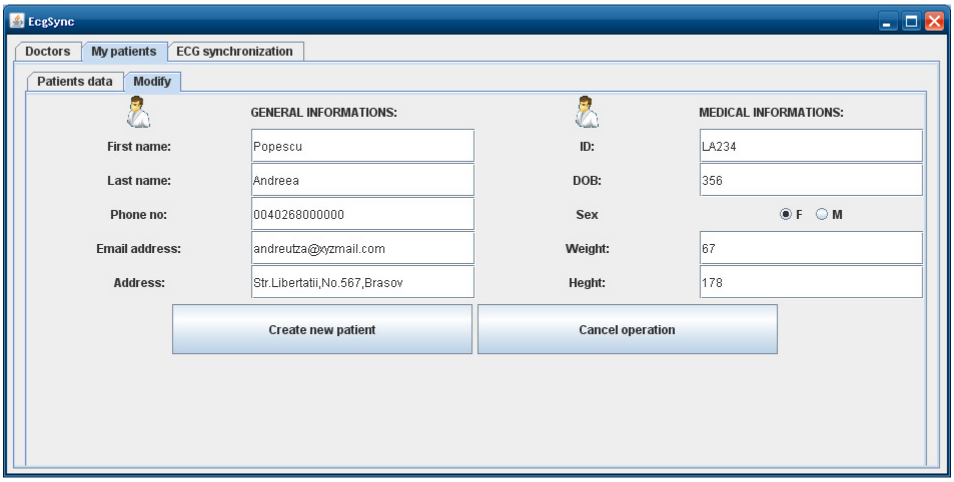


Figure 4: Patients administration.

The application is communicating with a database that was build using MySQL and that contains ECG data received from all the patients, patients related data and other information necessary in order to have the best synchronization with the application. The database can be located on the same server as the IN service itself.

C. IN Service from mobile operators' perspective

The IN service is developed in Java Enterprise Edition and from the functionality point of view, the service is split in 2 major parts [6]. One part is related to ECGs, patients and doctors management and the second part is related to specific IN functionalities such as charging and alarming. In this section an overview of both parts is presented.

In the management part of the service, the mechanisms for receiving ECGs, the specific database handlers and the Graphic User Interface generator are the major features.

The receiving mechanisms are handling the ECG data from the modem and convert them to files with the specific format in order to be viewed using dedicated software for ECG displaying. Receiving and conversion of data are not tied together in order to reuse the receiving part for other applications.

The database has a specific structure and in the next capture there is an example of table used in the case of ECG data.

```
CREATE TABLE `ecg`.`ecg_burst`
(
  `ecg_burst_id`   INTEGER UNSIGNED       NOT NULL AUTO_INCREMENT,
  `ecg_burst`      BLOB                   NOT NULL,
  `ecg_count`      INTEGER UNSIGNED       NOT NULL,
  `ecg_device_id`  VARCHAR(20)            NOT NULL,
  `ecg_receive_time`  TIMESTAMP           NOT NULL DEFAULT
  CURRENT_TIMESTAMP, PRIMARY KEY (`ecg_burst_id`)
)
```

Receiving of the information submitted by the modem is handled on the service side using *doPost* method from an extended HttpServlet class, where received data is parsed and introduced in the database tables. To avoid unnecessary traffic, the client PC user (doctors) will have a profile saved on the database that will contain the current list of ECG data and patient data. Because of this profile, only yet unavailable data (for client) will be transferred and thus updated. After the data transfer, the profile will be also updated. Patient specific data is synchronized with the one available on the server, as well. In this way more doctors can access the same data for the same patients.

The GUI was presented in the previous sections and it will not be detailed any more here.

From various dedicated functionalities of the IN service, just a part of the charging service strategy will be presented. The charging strategy for such a service, as it was described until now, is very complex but yet flexible because a service like the one presented, offers a number of possibilities to charge the customers in different ways. From the mobile operators' perspective, both the doctors and the patients must be charged, because they are using the service and the network infrastructure; on the other hand the doctors have to charge the patients because of the time spent to analyze their ECGs.

In the following part, the technical approach is presented for one charging scenario. When the doctor has a new patient who has to use a remote ECG for sending a diagram to be analyzed, the doctor has to enter the patient's coordinates in the application / service as presented in the previous sections. When he submits the form in the application backend, new items are entered in the management part of the service, but also new information is submitted to the charging part. In this way, when the patient will submit any ECG diagram he will be charged automatically for that with the amount of money established previously.

In order to build network operator-specific charging applications, a synchronous dialog-oriented interface is provided by the IN platforms in the form of the Service Management Agent / Access Function (SMAF) Common Object Request Broker Architecture (CORBA) interface [7]. Based on this, there are few steps that have to be followed in order to create users in the service with the purpose to be charged later on. In the next capture, there is an example of how this can be done based on the input information from the doctor, regarding the patients.

```
   // Initial credit

smafSDBeforeExecution.TheSubscriberData[0].OnPeakAccountID.Balanc
e = 10;
   //Creation of Subscriber

   ICorbaSubscribertheCorbaSubscriber =
                    ICorbaSubscriberHelper.narrow(
                    myOrb.string_to_object(ICorbaSubscriber_IOR));
   subscriberData = theCorbaSubscriber.newCustomer();
   //A new Subscriber type client is created.
   //There are mandatory values that have to be fill in.
   //Cgpa = CallingParty
   // MSISDN_String represents the modem phone number of the
patient based on which the charging is done

    subscriberData.TheSubscriberData[0].Cgpa = MSISDN_String;
    subscriberData.TheSubscriberData[0].CgpaOperationMode =
isDefined.value;
   //Each subscriber has to be part to a provider which in our
case it is already created.
   subscriberData.TheSubscriberData[0].ProviderID = "1";
   subscriberData.TheSubscriberData[0].ProviderIDOperationMode =
                                                  isDefined.value;
```

Based on this and additional settings that has to be done regarding the charging rates, the patients will start to be charged when they send ECG diagrams. Depending on the charging strategy, the verification of the patients' "accounts" can be done before sending any ECG data or after submitting data, but only before the doctor's analysis of the diagram. Part of the verification procedure of the service account balance is presented in the next capture.

```
ICorbaSubscriber ics = UtilityClass.getICorbaSubscriber();
Hashtable unusedSubscribers = SubscribersPool.getInstance().
                                            getUnusedSubscribers();
Object[] subscriberNumbers =
unusedSubscribers.keySet().toArray();
Arrays.sort(subscriberNumbers);

// Initiate Subscriber
_smafSubscriber = "1-" + smafProviderID + "-" + smafCC + smafNDC
+  subscriberNumbers[i];
// Get Subscriber by ID from the IN system
SubscriberDataData[] newSubscriberData = ics.
                 findSubscriberDataByCustomerId(_smafSubscriber);

// Calculate delta balance between the initial time and the
current one

long delta = smafSDBeforeExecution.TheSubscriberData[0].
OnPeakAccountID.Balance - newSubscriberData[0].OnPeakAccountID.
Balance-((Long)unusedSubscribers.
get(subscriberNumbers[i])).longValue();
System.out.println("Balance for subscriber " + _smafSubscriber +
": " + newSubscriberData[0].OnPeakAccountID.Balance +
". Expected charge: " + unusedSubscribers.get
(subscriberNumbers[i]) +". Delta from expected "+delta");
```

Using the balance calculated as above, different types of decisions can be taken in the service logic, all these having the possibility to be customized by the mobile operator.

As written in the first part of the paper, these services can have the possibility to offer also alarms based on already standard messaging SMS or MMS. What can be further developed are the triggering feature and the policy for such an approach.

3. Conclusions

In this paper was presented a new concept of Intelligent Network service. New because it comes hand in hand with the tele-monitoring area, thus creating new market possibilities in a mobile world, both technical and economical.

There were only few aspects covered, from technical- but also market- point of view, due to the complexity of the discussed topic and because of the service development difficulties. But, due to the global economical environment and the market demands, such a service can bring new ideas in terms of revenues strategy from the mobile operators' point of view.

As the solution is still under test, there are no final results but some measurements were done regarding the transfer speed of the ECG data. Thus, since GPRS is used for uploading, for five user ECG files at an average of 25-30 kbit/second upload rate the time required for a 100 kb total size is about 30 seconds. This test is very dependent on the information contained in the ECG files but it is relevant to determine how long a connection is active from service perspective in order to determine the accepted number of connections in a timeframe.

Since performance was not treated in this paper due to current status of the service development, further efforts will concentrate on the development of demo services in order to cover this issue and to bring this solution to a mature and consolidated stage.

Acknowledgements

The authors wish to thank to their colleagues who contributed with their effort to gain results presented in this paper, especially the colleagues located in Cluj-Napoca Siemens PSE site.

References

1. TMForum community, Information data model for the telecommunication business, http://www.tmforum.org/browse.aspx?catID=2008, October 2007
2. http://www.3gpp.org/
3. Cinterion Wireless Modules GmbH, "TC65i AT Command Set", TC65i_ATC_V01.000, pp. 317-360, Aug. 2008
4. Siemens AG," TC65 JAVA User's Guide", TC65 JAVA User's Guide_V05, 2005.
5. Application Server http://www.tomcat.apache.org
6. I.Venieris, F.Zizza, T.Magedan, „Object Oriented Software Technologies in Telecommunications: From Theory to Practice", pp.49-108, John Wiley&Sons, 2000
7. CORBA Specifications http://www.corba.org/
8. Y Jasemian, E Toft, L Arendt-Nielsen, "Real-time remote monitoring cardiac patients at distance", www.openecg.net/WS2_proceedings/Session07/S7.1_PA.pdf
9. C. Wena, M.-F.Yeha, K.-C. Changa, , R. Leeb ,"Real-time ECG tele-monitoring system design with mobile phone platform"
10. A.V.Nedelcu, F. Sandu, P.N. Borza, S.A. Moraru , "Mobile Data Acquisition and Tele-transmission by PDA" - Proceedings of the 19th INTERNATIONAL SYMPOSIUM OF DAAAM– 22-25th October 2008 - ISBN 3-901509

Localization of the Mobile Calls Based on SS7 Information and Using Web Mapping Service

Virgil CAZACU[1], Laura COBARZAN[2], Dan ROBU[3], Florin SANDU[4]

[1] BitDefender, Bucharest, Romania, e-mail: virgil.cazacu@gmail.com
[2] Softvision, Cluj-Napoca, Romania, e-mail: laura.cobarzan@gmail.com
[3] Siemens Program and System Engineering, Brasov, Romania,
e-mail: dan.robu@siemens.com
[4] Faculty of Electrical Engineering & Computer Science, "Transilvania" University, Brasov, RO,
e-mail: sandu@unitbv.ro

Manuscript received April 04, 2010; revised April 22, 2010.

Abstract: Localization of the calls is a topic that comes up even from the early time of the telephony. Calls made from mobile phones were even more interested to be localized due to their mobility. In this paper it is presented a localization solution that uses information from the mobile network. The technical solution ensures the acquisition of the localization information of the calls from the terminals in the mobile network and that delivers this data to a localization server. Due to the client-server architecture, the users of the system can access calls' locations using digital maps.

Keywords: Localization, mobile networks, service, client-server, integration

1. General information

A. The General Context

Localization of the calls is useful not only from the legal point of view but also in case of emergencies – e.g. in cases the short number 112 or 911 are called. In this case, the localization of the person which is in possible danger is more than vital. In the case of emergency call services, there are several regulations and recommendations from the European Union and from each European state government that recommends that public telephone network operators should provide location information "in a non-discriminatory way" [6]. Although it is suggested that open interface standard provided by European Telecommunications Standards Institute should be used [7], each country can adopt its own implementation strategy.

On the market, there are several emergency location solutions from big telecommunication vendors like Ericsson with the CoordCom product [8], Alcatel Lucent with the Alcatel-Lucent Enterprise Alert E911 software solution [9] or Huawei with the Police Command Center solution [10]. There are solutions that rely on the vendors' experience with large GSM networks.

Even if the same techniques can be used, there are still conceptual differences between Location Based Services (LBS) and location of the calls even if it's for emergency purpose. In the case of LBS, the solutions are oriented to the end-users, customers of the mobile operators. Thus, applications like receiving advertise notifications on the phone depending on the user location, requesting of services based on the user location like the nearest restaurant or other Points of Interest (POI) are some examples of LBS [11]. In the cases presented hereby, the end-users can be the state authorities or governmental agencies, thus different targeted customers.

B. Overview of the proposed solution

The localization solution presented in this paper has three major features: receiving calls information from mobile networks and obtaining the localization information from the Signaling System #7(SS7) frame; processing of data from the signaling frame and transmitting of this information via IP to a localization server; visualization of the call location on the map. The last feature is based on Google web mapping service and the main purpose of the presented solution is to demonstrate the possibility of usage of open technologies in this regard.

The SS7 localization approach brings a restriction which it is treated in the present solution: each mobile phone service provider supplies the localization information in its' own specific format, within the Initial Address Message (IAM) field of the ISDN User Part (ISUP) SS7 protocol [1]. Our solution offers the possibility to configure the necessary parameters, depending on where it is deployed.

From the design point of view, our solution ensures the seizure of the localization information for the terminals in the mobile networks and – for the calls selected to be localized – it delivers this information to a localization server. The processing of the calls – extracting the localization information and delivering this on the interface to the localization server – is made in real time (some delays appear only if the system is overloaded). The application could support between 50-100 requests per second, the response time for a request depending on several conditions. One of them, which it is very important and has a big impact in a real environment, is the network condition. As the current solution was not tested in a real environment, the response time was quite small, around 2-5 milliseconds – this time should increase insignificantly in a real environment. The scale of the database depends on the coverage area of this service and on the number of subscribers the operators may have. The solution is adaptable with minimal costs for future changes of the architecture, with the resize of the necessary input/output traffic and with the modification of the SS7 frame field where the localization information is transmitted by using parameterized components.

SS7 localization data is sent using “Cell ID” type from ISUP protocol, in the IAM message, “Location number” field and/or “Called number address” field” [1].

The next table presents an example of localization data format that is different, according to every mobile service provider.

Table 1: Localization data format.

Network code (e.g.72,74 or 4072, 4074)	Services bit (reserved)	Location area code	Cell ID
2-4 digits	1 digit	5 digits	5 digits

Based on these frames, each mobile operator maintains a database with geographic information that can help the find-out of caller’s position based on the positioning string. The database structure is different, according to the telephony provider and contains the equivalent geographical coordinates for the above data from SS7 ISUP frame as it is shown in the example of *Table 2*.

Table 2: Geo coordinates database structure.

Cell code	District	City	Street	Lat (Grade, Minutes, Seconds)	Long (Grade, Minutes, Seconds)	Azimuth	BSC	LAC
Specific Code	String	String	String	Int (6 digits)	Int (6 digits)	Int (3 digits)	Specific Code	Int (4 digits)

The location of such database is at each mobile operator, managed and maintained by employees of the mobile service providers and interfaced with a solution like the one presented in this paper.

The architecture of the solution is presented in *Fig.1*, where the Localization Server and the client side are introduced by our solution in a standard GSM network. Geographic Information Database Server is usually part of each GSM network, since it contains mobile operator data regarding the locations of different network entities.

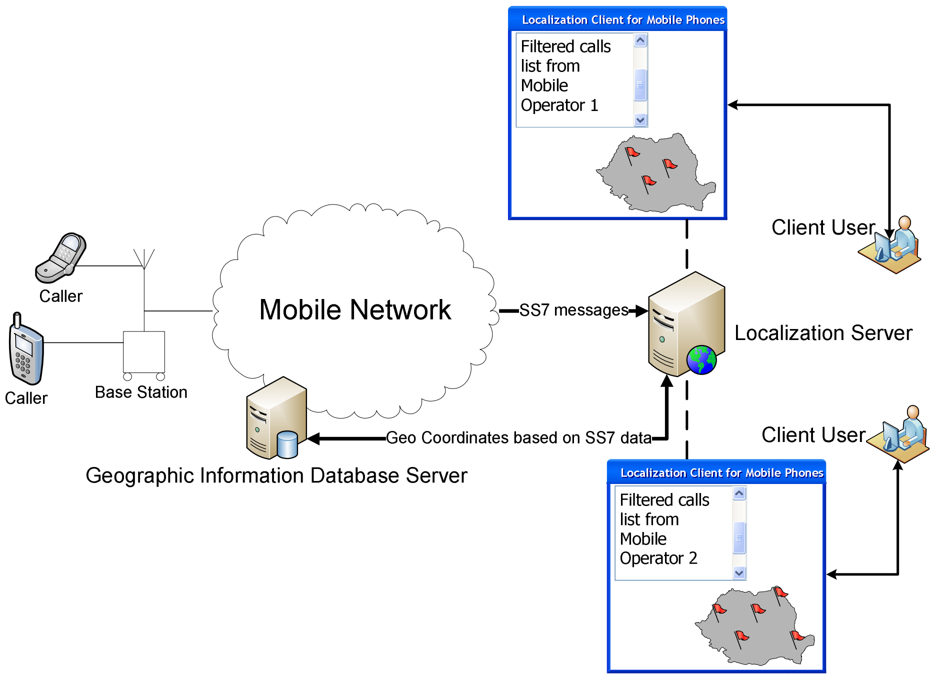


Figure 5: Overview of the solution architecture.

. The solution, as it can be seen in *Fig.1* above, is divided in four modules and is targeted to be used in operational centers that can coordinate emergency activities.

Extracting Localization String module from the Localization Server involve defining rules for parsing SS7 ISUP information – the main service of the server is to define interaction protocols as well as Geo Information Database Server interaction.

The Geographic Information Database Server is maintaining the geo-coordination of the radio cells.

The Client side module is handling the communication between the Graphical User Interface (GUI) module and the Localization Server.

The GUI module is handling specific interface functions and the digital maps – using web-mapping service available on the market at the solution's implementation time.

2. Technical information

In this section it is presented the technical implementation of the solution modules using as examples the case when the mobile users that are using the

Emergency Service 112 are customers of one of the Romanian mobile operators.

A. Localization Server

In this module, there are two major functionalities: SS7 frame parsing and communication protocols between the client and the Geo Database Server.

In the parsing module, there is the SS7 ISUP communication part as well as the parsing of the messages from the protocol. It is out of the paper's scope to detail the SS7 communication between our solution and the mobile network. As for the technical approach, using JAIN ISUP API gives the possibility to exchange ISUP messages in the form of Java Event Objects. [2], [4].

One rule for parsing SS7 information is the fact that – independent of the mobile operator – the SS7 frame is in a standard format and the parameters relevant for our solution can be found under the Initial Address heading. These relevant parameters are presented in *Table 3*.

Table 3: IAM parameters used for localization.

Parameter Name	Explanation
Calling party number	Nature of address: either National or International Calling address signal: the telephone number of the caller party (with a 2 digit prefix for international calls)
Called party number	Called address signals. For Emergency cases, the called telephone number is 112
Location number	The localization string, that contains all the localization information for the given provider
Cell ID	The mobile operator internal ID for the radio cell from where the call is done

The phone numbers are coming without the prefix digits, so in this implementation is taken into account the "Calling party number" parameter – for national calls, a 0 digit is inserted in the beginning of the caller's number and for international calls two 0 digits are inserted at the beginning of the caller's number. Also, to detect which mobile operator the call is performed from and knowing that every telephone number begins with 07XY, depending on the XY digits, the solution can extract the provider of the call based on a correspondence table.

In the implementation of the module, the extracted information is stored in an object called SS7Object with fields like String callingNumber, String calledNumber, String Nature, String LocalizationString, Date dateCreated, String Provider. The SS7Object is sent to the Localization Server module for further processing.

The communication protocol with the client is done using sockets and when new localization "objects" are received from the SS7 parsing module, the server will send the object to the client in order to use it on the GUI. After sending the localization object, the server is waiting for an answer from the client. If the client does not confirm receiving of the object in the previously defined time frame, the localization server will resend this information. The Localization objects that are not confirmed are maintained in a waiting list. When the Localization Server receives a message from GUI/Google Maps, it will delete the corresponding object from the "waiting list", meaning that it will not wait for the confirmation for that object.

The communication protocol between the Localization Server and the Geo Information Database Server is done by calling the getCoordinatesBy LocalizationString (localization_string) method which has as parameter a string, representing the localization string and as returned value, an object which contains the coordinates of the area from which the call was made. The coordinates are the latitude and the longitude, each one containing 3 fields: degrees, minutes and seconds. The calling of the class is done using Remote Method Invocation (RMI).

B. Geographical Information Database Server

As mention before, this server has to be located at each mobile operator, since it contains internal information about the place where radio cells are deployed – from the geographical point on view. For completeness of our solution description, it will be presented the RMI Database Server that has several classes aimed to extract the coordinates from the local database.

The Coordinate class is common with the RMI client, the CoordinateInterface class which contains the remote method and the CoordinateImplementation class, which implement the remote method as shown below:

```
public Coordinate getCoordinatesByLocalizationString(String
localization_string)
```

The Provider class has a static method String getProvider(String localization_str), which returns the provider, based on the localization string and the ExtractCoordinate class has a static method that returns the coordinates, based on the provider and the localization string, as it can be seen in the next capture.

```
public static Coordinate getCoordinatesByProvider(String provider, String localization_str)
```

The ExtractFromDatabase class contains one static method for each provider, to extract the coordinates from the local database, based on the localization string, as shown below:

```
public static Coordinate ExtractVodafone(String localization_string)
```

A database structure example for this server is presented in the next figure.

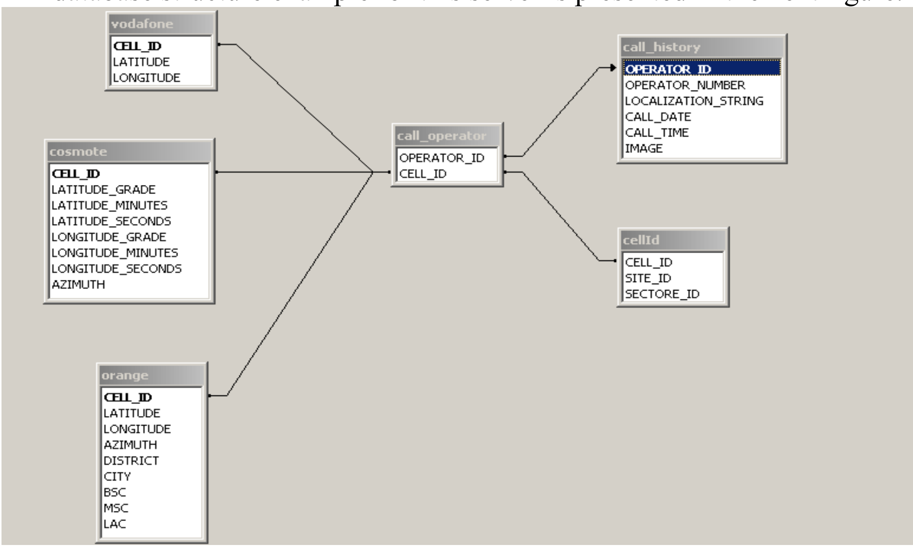


Figure 6: Example of structure for the geo coordinates database.

C. Client side – including GUI

As the communication between the client and the Localization Server is detailed in section B, this part is focused on the usage of web mapping service and of the user interface.

The graphical user interface presented in this paper has a proof-of-concept oriented design. The GUI of the solution is composed of two frames: one frame with 4 tabs: View Calls, View Archive, Options and Help. The other frame is displaying the digital map with all its options.

Only one tab is presented in this paper, View Calls tab which contains the recent calls information in a list. The call information is containing the exact time of the call, the caller's number, the provider and the location of the call as it is received from the Location Server. Each call is having its own checkbox, which will specify if the call was processed or not. When a call is selected in the list, the application is marking automatically the location of the call in the map frame. More calls can be selected simultaneously, so more locations can be marked on the map. Calls that are checked (processed) are deleted from the list and their marks from the map disappear.

For digital map integration into the solution, the Google Maps API was chosen due to several considerations [5].

Google Static Maps API embeds a Google Maps image without requiring JavaScript but the problem is that it returns the map as an image (GIF, PNG or JPEG) in response to a HTTP request via an URL. This way, the benefits of the zoom and navigation facilities disappear.

JXMapViewer embeds mapping abilities into Java application, but at the time of our solution's development, it was not possible to use it with Google Maps or Yahoo since there were some legal restrictions

Another strong reason why the Google Maps API was chosen to integrate the web mapping service was the possibility to control the zoom and navigation features from the application's back-end.

Since Google Maps API uses JavaScript, in the Java client application development was used class JWebBrowser from chrriis.dj.nativeswing.components package which offers the possibility to have a native web browser component in the application [3]. Because the client application has to be operating system-independent, the web browser component was created in development to use Mozilla engine.

```
NSOption opt = new NSOption(JWebBrowser.useXULRunnerRuntime());
web_browser = new JWebBrowser(opt);
JWebBrowser.useXULRunnerRuntime();
```

The digital map from Google is loaded using a code line like the next one [5].

```
web_browser.navigate(gmapfilelocation.getAbsolutePath());
```

The parameter gmapfilelocation points to the file containing the script which loads the map.

In *Fig. 3* and *Fig.4,* the client GUI is shown together with the calls' markers – each marker descriptor contains a string defining the location to place the marker and the visual attributes to use when displaying the mark.

Fig. 4 shows also the use case when one call is selected from the list and automatically the application zooms-in on the location where the caller is positioned. If another call is selected, so that 2 calls are on the map, the application automatically zooms-out exactly to the extent necessary to display the both callers on the map.

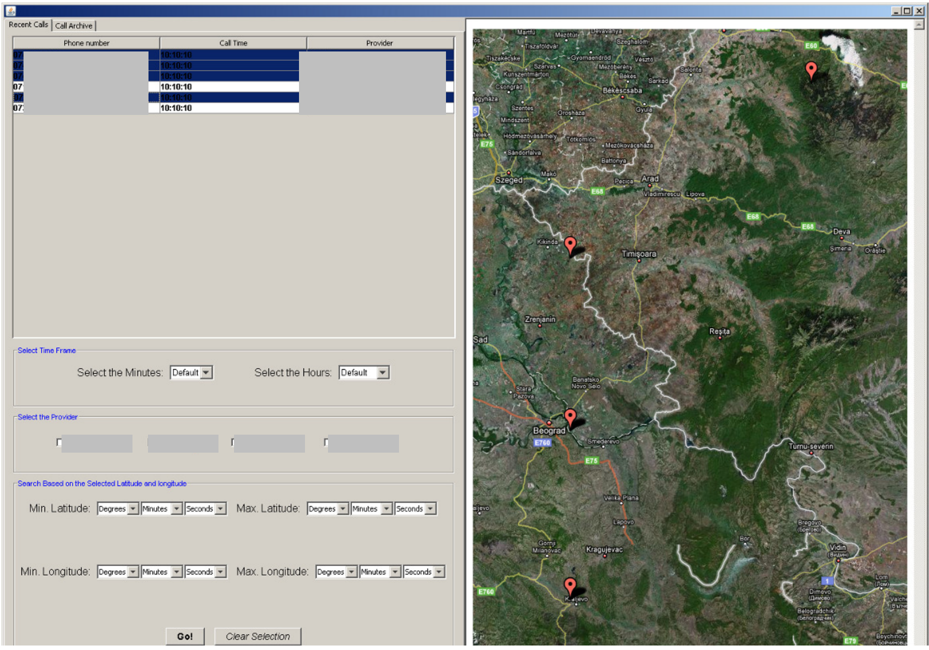


Figure 7: Calls markers on the map.

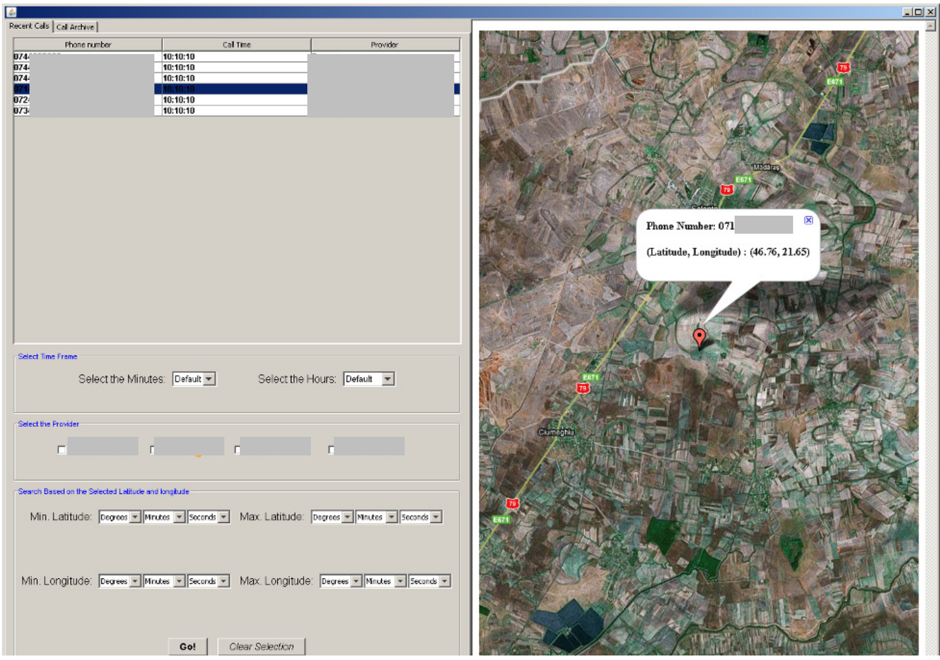


Figure 8: Zoom on caller location.

3. Conclusion

The solution of calls' localization presented in this paper is still under development, since topics like the high degree of availability or the ability to work in load-balanced and failed-over conditions between locations are not yet implemented. The application has the architecture accomplished by the authors of the paper. The solution allows further improvements, in order that features like accepted input traffic of a high number of simultaneous voice calls should be implemented as easy as possible.

The scope of our research was achieved in this phase, since the web mapping services usability for calls' localization was demonstrated by the solution described in this paper.

Acknowledgements

The authors wish to thank to their colleagues who contributed with their effort to the results presented in this paper, especially the colleagues located in the Siemens PSE site of Cluj-Napoca, Romania.

References

1. L. Dryburgh, J. Hewett.," Signaling System No. 7 (SS7/C7): protocol, architecture, and services", Cisco Press, 2005.
2. T. C. Jepsen, F. Anjum, "Java in telecommunications: solutions for next generation networks", John Wiley & Sons, 2001.
3. Framework DJ Native Swing http://djproject.sourceforge.net/ns/documentation/javadoc/ index.html
4. JAIN ISUP Specification http://jcp.org/en/jsr/summary?id=ISUP
5. Google Maps API Description http://code.google.com/apis/maps/documentation/ reference.html
6. European Union Commission Recommendation 2003/558/EC on the processing of caller location information in electronic communication networks for the purpose of location-enhanced emergency call services http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri= OJ:L:2003:189:0049:0051:EN:PDF
7. ETSI TS/102/164 Telecommunications and Internet converged Services and Protocols for Advanced Networking; Emergency Location Protocols http://docbox.etsi.org/STF/Archive/ STF321_TISPAN3_EC_Emergency_Call_Location/Public/Library/ETSI%20-%203GPP %20Documents/Published/TS102164_4e_EmrgncLocatProtcl.doc
8. Ericsson CoordCom http://www.ericsson.com/ourportfolio/products/coordcom-5.2?nav=fgb_101_390|fgb_101_212
9. Alcatel-Lucent Enterprise Alert E911 http://enterprise.alcatel-lucent.com/private/active_docs/EPG3310091214_E911_Emergency_Response_EN_Brochure.pdf
10. Huawei Police Command Center http://www.huawei.com/enterprise/solutions/ findSolution.action?id=3155
11. S. Steiniger, M.Neun, A. Edwardes , "Foundations of Location Based Services", http://www.geo.unizh.ch/publications/cartouche/lbs_lecturenotes_steinigeretal2006.pdf

EXAMBREV Integrated System for Patent Application Overview

Attila ASZALOS[1], József DOMOKOS[2], Tamás VAJDA[3], Sándor Tihamér BRASSAI[4], László DÁVID[5]

[1,2,3,4,5] Department of Electrical Engineering, Faculty of Technical and Human Sciences, Sapientia University, Tîrgu Mureş, e-mail: aszi.atti@hotmail.ro; domi@ms.sapientia.ro; vajdat@ms.sapientia.ro; tiha@ms.sapientia.ro; ldavid@ms.sapientia.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: In this paper we present the design and development of a patent application, examination and evaluation system based on the JEE platform. Our system reduces the time required for the application examination by automatic or semiautomatic verification of the formal aspect of a patent application. The system helps the inventors in the International Patent Code assignment and provides information to the domain experts to speed up the patent examination process. In this article we have focused on the search process in the International, European and Romanian patent databases and with the patent search engine provided by Google. This system will be implemented at the Romanian State Office for Inventions and Trademarks (OSIM).

Keywords: patent application, patent search, evaluation and examination system

1. Introduction

Worldwide patent applications are growing at an average rate of 4.7% per year, according to the 2007 edition of the World Intellectual Property Organization (WIPO)'s Patent Report [7]. Patent examination procedure has two stages: formal verification, which follows all the formal procedural steps and verify if applications are patentable and the evaluation stage which checks the grade of novelty and innovation of the patents [8, 9]. To reduce the patent examination time and increase the evaluation quality, despite that the number of the patent applications are growing, there are two possibilities: increase the number of employments of the State Office for Invention and Trademarks

(OSIM) or reduce the amount of work for registration, formal verification and evaluation by using an online integrated system.
In this article we present an integrated Patent Examination expert system which is a web application that helps the applicants to accomplish the entire registration procedure and verifies the formal correctness of the patent application. This system also manages the experts and presents the application for evaluation through a web interface. Our translator unit will give a helping hand for experts to search for similar technical solutions in a wide range of different language patent databases and also the Romanian Patent Database. The system also helps the management of the OSIM to see in which field to employ new experts in the future.

The outline of the paper is as follows. In section 2 we present our system architecture and describe the subsystems. In section 3 we present the results, conclusion and future work, and finally the acknowledgements and references.

2. EXAMBREV system architecture

A. System overview

Our system architecture is presented in *Fig. 1*.

Mainly our system has two different modules divided in multiple subsystems. First we have the *Interfaces and data preparation module* which manages the patent requests, common users (UCOM), expert users (UEX), civil servants (UFUNC), applicants (UAPP), administrators (UADM), civil servant managers (UFUNCM) and expert managers (UEXPM) and also prepares some initial data for the *Expert system module* (SIEXP). Second, we have the *Expert system module* (SIEXP) which gives the world wide novelty of a technical solution proposed by an inventor and contains the legal and procedural database.

In this paper we want to present the Interfaces and data preparation module and especially the search methods for similar technical solutions in the online patent databases.

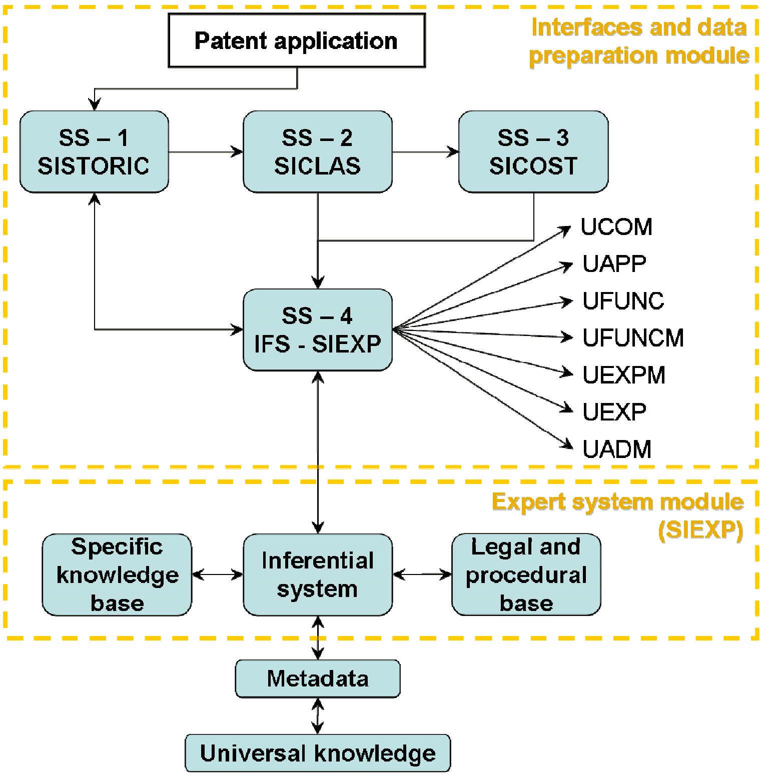


Figure 1: EXAMBREV System architecture.

The deployment diagram shown in *Fig. 2* illustrates the connections between the different subsystems of Interfaces and data preparation module and their deployment on the servers. As we can see, all the subsystems communicate with the system database through JPA (Java Persistence Application Programming Interface) which communicates with the database through the JDBC API (Java Database Connectivity).

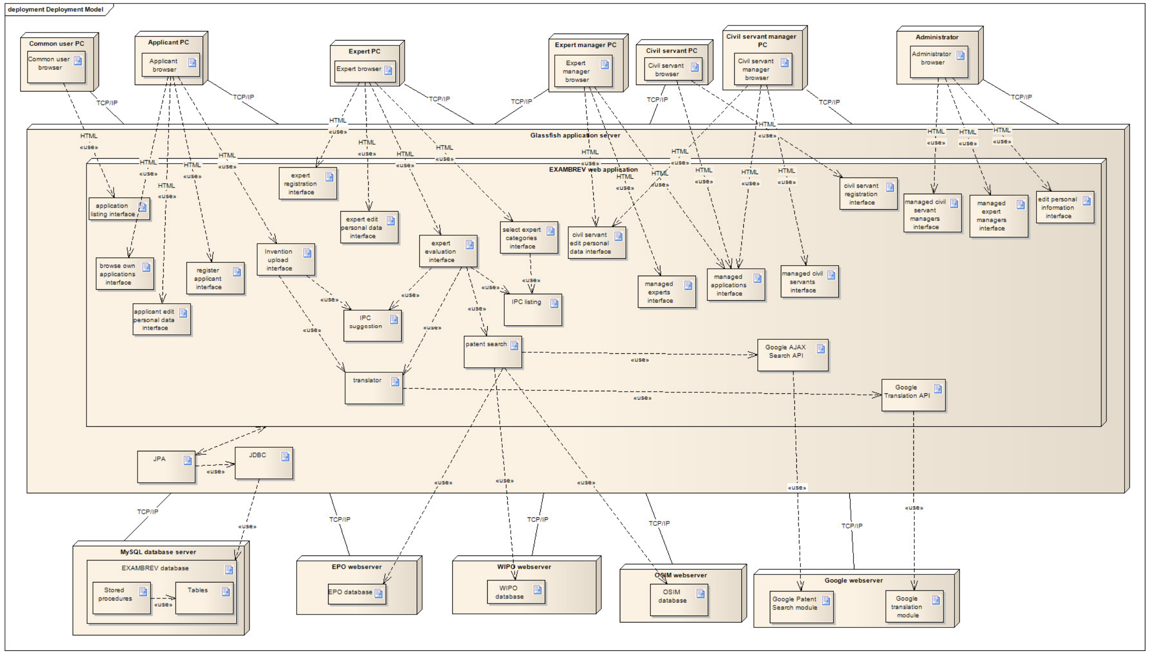


Figure 9: EXAMBREV deployment diagram.

B. The SISTORIC subsystem

The SISTORIC (SS-1) subsystem has three main functionalities [1]:

- to manage and store all the information regarding a technical solution proposed by an inventor;
- to add or modify the patent data according to the Expert system module decisions;

According to Romanian State office for Inventions and Trademarks we had to deal with three types of patents:

- EPC (European Patent Convention);
- PCT (Patent Cooperation Treaty);
- Romanian National Patent.

We have implemented since now the management system for the Romanian national patent type. The application is accessible from the Internet through a web browser and makes possible to deposit online a patent request. All the patent application data is stored in a MySQL relational database management system.

The software subsystem provides the following features [4]:

- Applicant user registration
- Civil servant user registration
- Expert user registration
- Account activation for the above users
- Online patent application;
- Editing patent application information;
- Editing account information;
- Patent application list;
- Semiautomatic IPC code assignment.

In the registration process the user starts the registration and fills in all the required personal information and specific user information. In the case of the expert users the specific information is the list of the IPC categories in which he is expert. After the data validation process the system sends an activation email to the registered person, containing a link to the activation page. The account of the user will be accessible after clicking the link in the received email and activating his account. In the case of the civil servant and expert users after the activation procedure their account must be confirmed. This can be done by the civil servant manager or the expert manager. These manager type users are promoted from the civil servant users by the administrator. If a civil servant is promoted to manager his account gets confirmed automatically. The user's personal information is saved in the database.

The login process is different for almost every user type, although there is a common part too. In the common part of the login process the system checks the username, password and the activation status of the account. If the username and password are matching and the account is activated the applicant and expert users can login. If the expert user's account isn't confirmed he has limited accessibility in the system and can only change the list of the categories he is expert at. The civil servant users can log in only if their account is confirmed.

The patent application process consist in filling of the online application form which is the same used at OSIM in present. This process is divided into 4 steps, because this way the required data on one page isn't too large and if there are validation errors, the user can correct them more easily.

C. The SICLAS subsystem

The SICLAS (SS-2) subsystem contains the IPC suggestion algorithm which can be used by the applicants or the experts.

After the applicant successfully submits an application he is asked to select the IPC code in which the invention can be categorized. This suggestion algorithm makes it easier for him to choose a correct IPC category.

When the expert inspects an application he has to enroll it into an IPC category. If he gives to this suggestion algorithm the keywords which describe the invention he gets a list of possible IPC codes for the invention.

This suggestion algorithm was presented in [10], and there were no modifications in the logic of the algorithm, only in the technical realization of it. It was necessary to rewrite the suggestion's algorithm implementation because it's execution time was very high. There were cases when for only 5 or 6 keywords the execution time was at least 30 seconds. In the case of the new implementation the execution time decreased by 25 seconds in the above cases.

D. The SICOST subsystem

The SICOST (SS-3) subsystem is responsible for Expert user (UEXP) data management [1, 2]. This subsystem also takes the IPC code given to a patent application by SS-2, and outputs it to the civil servant manager helping him to choose 3 expert users for this domain.

This subsystem also provides a web interface for the expert users to search the following online patent databases:

- WIPO database
- EPO database
- Romanian Patent Database

All of the above databases can be searched online from the esp@canet webpage. In our system we give the option to the expert to search these databases by keywords or by IPC code.

When the expert searches by keywords we build from them the URL with the required request parameters for the esp@canet webpage and execute it. The response given by esp@canet has a fixed structure containing an HTML table with the results of the search. This HTML table is parsed by us with a HTML parser and the results are organized in a list which contains the title of the invention, the link to the page presenting the invention in detail, the inventors, the applicants and the IPC code of the invention. The used HTML parser is HTML parser 2.0 SNAPSHOT, which is a public license library, submitted by Derrick Oswald to the http://sourceforge.net/projects/htmlparser webpage.

If the expert chooses to search in the patent databases by IPC code the process is similar to the search by keywords, except the fact that the URL is slightly different. It takes different request parameters with the value of the IPC code given by the expert.

Our system also provides the possibility for the experts to search with the Google Patent Search Engine. The results given by this search engine can be processed more efficiently, because it is a JSON (JavaScript Object Notation - http://www.json.org) object. It doesn't contain as much information about the patent as it contains the result from esp@canet. It only contains the title of the invention and the link to the patent application.

These results are presented to the expert immediately when the expert submits a search and help him to find similar technical solutions.

E. IFS-SIEXP subsystem

The IFS-SIEXP (SS-4) subsystem is the special interface for the SIEXP module [1, 2]. It makes data transfer between the Interfaces and data preparation module and Expert system module. It also communicates with UEXP, UCOM and UINV via a Web interface. This is the login point to the Web application for registered users.

3. Conclusions

We designed and developed a JEE based integrated system for patent examination. The system will help OSIM to offer online patent registration possibility for all three patent types discussed (EPC, PCT and Romanian national patent type). The system also helps OSIM patent evaluator experts management, employers management and patent management.

The main results obtained are the UCOM, UEXP, UINV, UFUNC and patent application registration interfaces. The interfaces were developed considering Java Server Faces technology and Primefaces 2.0 technology (http://www.primefaces.org).

We have developed an algorithm for semiautomatic IPC code assignment which was mentioned in this paper.

The patent database search and the patent search with the Google Patent Search engine were presented in more detail in this paper.

For the application development we have used NetBeans IDE.

As future work we can mention the SIEXP module development, and whole system integration.

Acknowledgements

This project is developed under Partnership in Anterior Domains Program of National Authority for Scientific Research in Romania, project code: 11-076/2007.

References

1. M., Radu, "Elaborarea strategiei de cercetare privind examinarea cererilor de brevet de invenţie şi studiu critic asupra procedurilor de examinare aflate în uz", Technical report EXAMBREV, stage I, PNII – Parteneriate, No. 11-076/2007, Centrul de Cercetări pentru Materiale Macromoleculare şi Membrane, Bucureşti, 2007.
2. S.,T. Brassai, L. Dávid, J. Domokos, T. Vajda, "Technical report EXAMBREV stage II", PNII – Parteneriate, No. 11-076/2007, Sapientia University, 2008.
3. J. Domokos, T. Vajda, S., T. Brassai, L. Dávid, "Integrated System for Patent Application Examination (EXAMBREV)", Proceedings of 17th International Conference on Control Systems and Computer Science – CSCS17, ISSN: 2066-4451, Bucureşti, 26 - 29 May 2009, p.p. 135-139.
4. J. Domokos, T. Vajda, S., T. Brassai, L. Dávid, "Realizarea, implementarea în faza de laborator şi testarea sistemului informatic de examinare a cererilor de brevet de invenţie*", Technical report stage II, EXAMBREV,* PNII – Parteneriate, No. 11-076/2007, Sapientia University, 2009.
5. EPO webpage: http://www.epo.org/
6. OSIM webpage: http://www.osim.ro/
7. WIPO webpage: http://www.wipo.int/
8. Implementing regulations to the patent law no. 64/1991, as republished in Official Gazette of Romania, Part I, No. 456/18 June 2008.
9. Patent law No. 64/1991, republished on Official Gazette of Romania, Part I, no. 541/ 8 august 2007.
10. J. Domokos, T. Vajda, S. T. Brassai, L. Dávid, A. Aszalos, "Development of EXAMBREV Integrated System for Patent Application", Inter – Eng 2009, "Petru Maior" University, 2009

Embedded System Based EEG Signal Processing

László BAKÓ[1], Sándor Tihamér BRASSAI[1], Lajos LOSONCZI[2],
László Ferenc MÁRTON[1]

[1] Department of Electrical Engineering, Faculty of Technical and Human Sciences,
Sapientia University, Tg. Mureş, e-mail: lbako@ms.sapientia.ro
[2] Lambda Communications Ltd., Tg. Mureş, e-mail: lajos@lambda.ro

Manuscript received April 11, 2010; revised April 25, 2010.

Abstract: Human EEG signals represent the global activity of millions of neurons. A brain–computer interface (BCI) is a communication system to translate brain activity into signals for a computer or other similar device. A BCI allows users to record signals upon which are possible to act on their environment by using only brain-activity, without using peripheral nerves and muscles. EEG signals are the recorded potentials of collective activity of synchronized cortical cell populations chained to an external system. Our basic tasks are to improve the signal-to-noise ratio and to resolve spatial and temporal impact of the measured signals on the external environment. For these reasons we propose a completely new concept of active electrodes, named 'Smart Active Electrodes (SAE)'.

Keywords: Electroencephalogram (EEG), BCI, signal processing, scalp electrodes, embedded implementation.

1. Introduction

An electroencephalogram-based brain-computer interface (BCI) should provide a control channel to individuals with severe motor impairments. Research areas include evaluation of invasive or noninvasive technologies to follow brain potential field activity. Electrical activity in the cerebral cortex contains important information to characterize brain function and is possible to study the role of individual cortical regions during event related processes, cognitive and motor tasks. Brain responses are, by definition cortical events translated into non-stationary EEG signals by active electrodes. We are involving special hardware and software procedures to handle the difficulties we met. Functional brain activity is associated with a variety of event-related changes in EEG spectra. Because the EEG signals reflect the summation on a small area of the synchronous cortical synaptic potentials, we are concentrated on positioning of the SAE on top of the scalp. In addition, it is important to study the activities often called event-related-synchronization. The functional correlations of brain activities depend on the frequency bands in which they occur. Signal energy may increase in some frequency bands while it decreases in others, and the energy within one frequency band may increase at one time and decrease at another depending on its temporal relationship to the functional brain activation. For the beginning we must emphasize that the EEG signals are ergodical, endogenous and nonstationary signals. For example, self-paced finger movements in humans are associated with power suppression in alpha (8–12 Hz) and beta (13–25 Hz) bands, while power in the gamma band (>30 Hz) is increased [1][8][9]. Usual EEG recording is performed with a cap equipped with integrated scalp electrodes. EEG potentials, referenced to the average of the left and right ear lobes, are recorded at the standard fronto-centro-parietal standard locations F3, F4, C3, Cz, C4, P3, Pz, and P4 (see *Fig. 1*. a general reference figure). At the beginning of this area of research, the minimal sampling rate was 128 Hz and it must be increased to get higher resolution of biosignals [3].

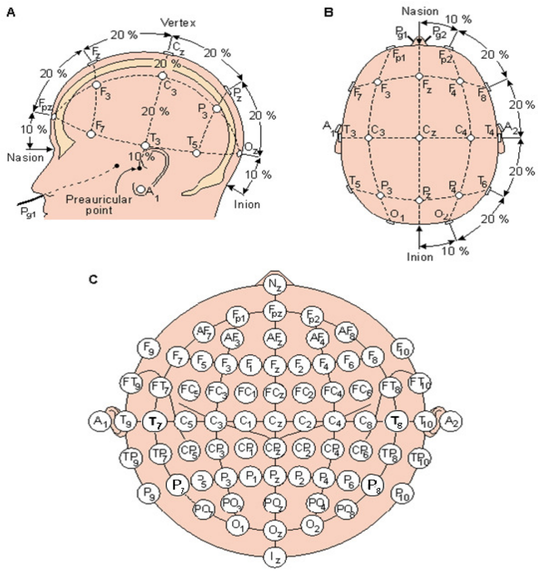


Figure 1: Electrode cap standard for 10-20, and enhanced structure.

In our SAE concept the recorded raw EEG potentials are transformed into pre-processed digital signals. In accordance with the literature, with a trial length around 2s, the information transfer rate should be between 5 to 25bit/min. A main goal of our BCI research was to overstep these transfer rate values. Active and passive electrode systems are designed for non-invasive electrophysiological derivations such as EEG, EMG, ECG, EOG and other biosignals [3][4][6]. Active electrodes have an additional ultra-low noise pre-amplifier located inside the electrode. The system is designed for use with biosignal amplifiers. Usually electrodes are made of silver/silver chloride electrodes (Ag/AgCl) and work in a frequency range of 0 - 100 kHz. The active electrodes can be used as an alternative to the passive standard electrodes. The system must avoid or reduce artifacts and signal noise resulting from high impedance between the electrode(s) and the skin. The biosignal amplifier must be used with or without an active electrode system [1][3][4]. Sometimes, conductive gel or saltwater is used to improve the conductivity between scalp and electrodes. We are avoiding this methods using special method to affix the electrodes to the scalp. EEG signals are always recorded with respect to reference electrodes, Scalp EEG signals are small potential differences (0.1 - 20 μV) between electrodes and reference electrodes placed at different positions on the scalp. Odd numbers indicate electrodes located on the left side of the head. Even numbers indicate electrodes located on the right side of the head. Capital letters are used to reference each cortical zone, namely frontal (*F*), central (*C*), parietal (*P*), temporal (*T*), and occipital (*O*). The *Fp* and *A* stands for frontal pole and auricular position. The percentage ratio of the inter-electrode distances with respect to the nasion-inion distance is important in complex measurement tasks and it has the significance of measured signal density on the scalp (*Fig.1.*).

Control tasks as results of EEG recording and processing, need a real-time bio-signal processing (high-speed online processing). For this reason, what is following is a summary of the new ideas about the above presented SAE ideas.

2. About electrodes

An accurate way to measure biosignals, because of anatomical problems, must be a noninvasive method. To find an efficient electrode is a big challenge. Our concept of how can electrode’s characteristics be enhanced in the measurement process of biological origin electrical signals, is one of the key elements of the whole research. The EEG measurement systems used in BCI applications meet much more difficult requirements than a usual recording procedure.

Because of its impedance, offset voltage and noise contribution, the electrode-skin interface is causing various problems, especially in multi-channel recording techniques. On the other hand, to ensure correct operation with typical electrode-skin impedances and typical interference sources, the amplifier should meet the following demands:

- Very high common mode input impedance (>100Mo),
- High differential mode input impedance (>10Mo),
- High common mode rejection ratio (>90 dB),
- High overall gain (10,000 – 100,000),
- Additional reduction of common mode voltage (ex. with a DRL circuit).

The ideal solution would be to perform signal amplification as close to the transducers as possible. As well, the filtering and digitization of biosignals measurements should be done in close proximity to the patient's skin. One of the main design goals is to keep the size and cost of the electrode comparable to a standard EEG electrode. We chose a good alternative (SAE, shown in *Fig. 2.*) to conventional transducers, which meet the above mentioned set of demands and can be achieved with a minimal number of electronic components. Use of SAE eliminates 50Hz interference picked up by the electrode wires and also eliminates the need for shielded wires. Because high electrode impedances are tolerated, the system can be used without the usual skin preparation to lower the impedance. This makes the application of the electrodes very fast and eliminates the risk that the electrodes should become a source of infection. To eliminate the poor Signal-to-Noise ratio problems in EEG recordings, we are considering the incorporation of amplification, filtering and A/D conversion electronics into each electrode.

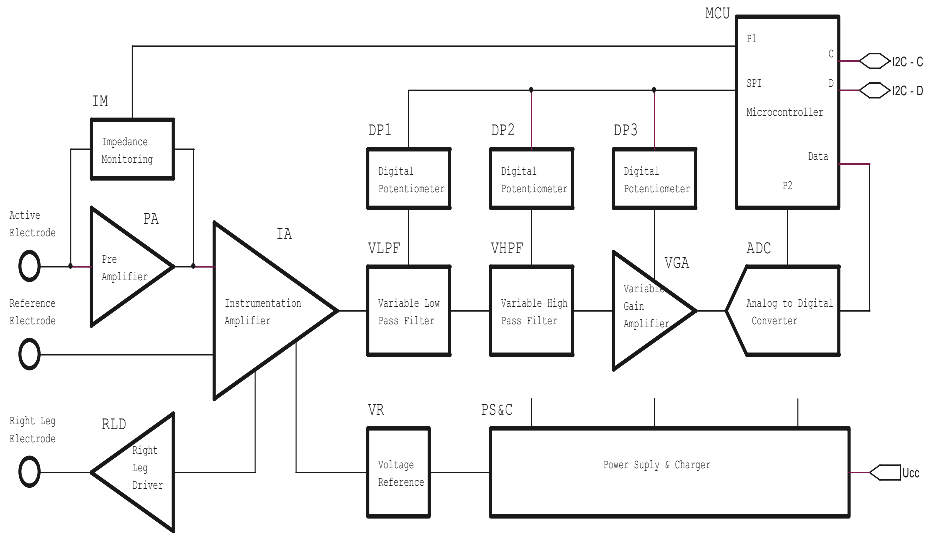


Figure 2. Block diagram of one of the Smart Active Electrodes (SAE).

Each electrode also contains a microcontroller, which coordinates the operation of the module and communicates with the next level, using the serial (I2C) bus. The measurement system uses 10 SAE embedded in a headset placed on the scalp of the patient. The *Fig. 3.* is the block diagram of measurement unit. The digitized signals are transmitted to the main microcontroller which communicates with the Signal Processing Unit through a two-way RF communications protocol.

Using this method of concentrating the analog circuitry on the electrode will be the right way of recording high-quality biosignals (using also ideas from [15][16][17]).

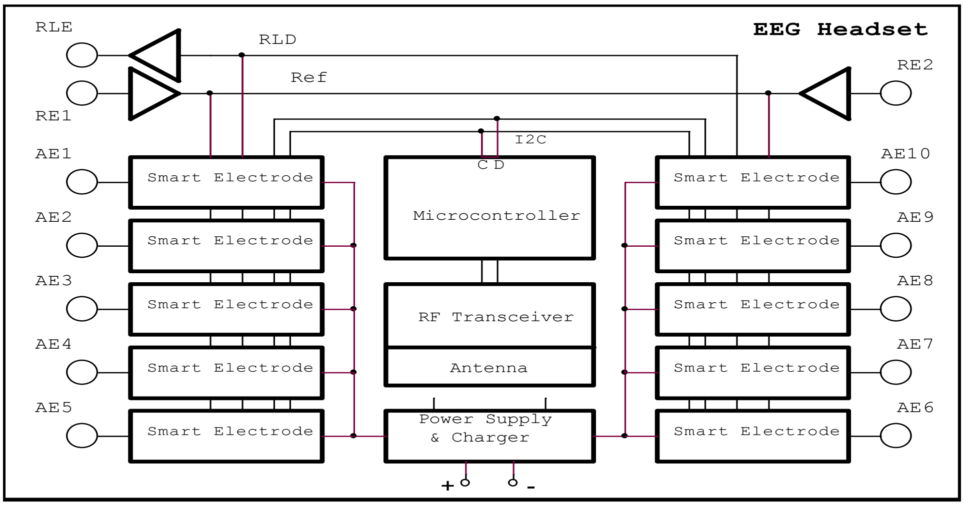


Figure 3. Block diagram of the biosignal measurement system

3. About hardware unit based signal processing

We have concentrated our effort to create hardware-based signal pre-processing units in parallel and based on our effort on software based processing modules. This is important because of real time considerations [7]. One of a well experimented procedure is based on using FPGA circuits as a hardware (embedded) platform. We are using this unit in creating filters, artificial neural networks for pattern recognition or control tasks. Traditionally, digital signal processing based applications were built using microprocessors. Decades before there were limited alternatives to implement special hardware devices with microprocessors with fixed structure and limited functionalities or very expensive systems with array of processors. The manipulation of the target algorithm logic at gate level allows us the design of an efficient processor dedication to the desired task. Like the microprocessors, the FPGA circuits can be reprogrammed easily. Design errors can be handled easily. Interfaces associated to microprocessors can be built in the reprogrammable logic [6]. We are using them because their high level of integration makes the system smaller, cheaper and with reduced power consumption. The above presented advantages indicate unambiguously to use the FPGA circuit for the implementation of algorithms necessary in the EEG signal processing, pre-filtering and other tasks needed in of our special noninvasive signal recording system. The block diagram of the design structure for the above mentioned EEG signal processing system implemented on FPGA circuit is presented on the *Fig. 4*. EEG signal measurement from SEA will be realized through a two-way RF interface. The pre-filtering unit has access to the EEG signal channel's signal through this RF interface. In the processing/filtering module, signals are distributed to channels.

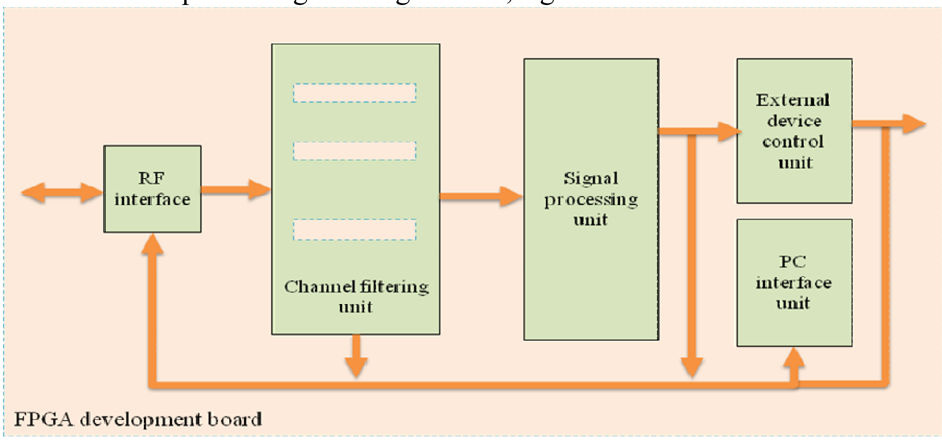


Figure 4. Block diagram of the FPGA based concept.

Filtering for all channels is made in parallel. More filtering algorithms are to be tested, and the best and less hardware consuming will be elected. VHDL encoding will be used for filter's hardware description. By using the filtered signals, the role of the processing module is to extract different characteristics of the recorded signals. After detecting any pattern and after the pattern's identification, the task of the control unit is to prepare the control signals for the executive unit. During the developmental phase, visualization of signals are made on a PC. Data is transmitted from the FPGA system to the PC. The PC interface is connected to all modules of the FPGA system in order that all signals in all step of processing can be visualized on a computer monitor. Communication is full duplex, parameters of the system (sampling time, selected channels, filter parameters, amplification factors) can be sent from the PC, and data (pre-filtered, filtered, extracted patterns) is sent from the FPGA system to the PC. At this phase, the control task of an external system is the PC's task [8][9].

Another technique involves the training of a spiking neural network (SNNs) to classify evoked potentials in continuous EEG recordings. EEG signals are to be pre-processed to remove the noise and to extract the relevant information. Analysis of the signals was performed on the ensemble EEG and the task of the neural network was to identify the prescribed components within the signal. The network employed leaky-integrate-and-fire (LIF) neurons as nodes in a multi-layered structure. Having a hardware implementation as aim (in order to maximize the pattern recognition speed), a choice has been made to use a spike-time encoding scheme of the input variables instead of spike-rate coding. This simplified the circuit needed for this encoding, while using the inherent sequential nature of these digital systems to encode the information [10][11][12]. Several versions of the encoding algorithm (of the input values into delayed spikes) have been experimented, considering the reconfigurable resources available in the FPGA chip of the utilized development board. In order to make the best use of the reconfigurable logic blocks, specifically, to free up as many logic cells as possible for the implementation of spiking neurons, different approaches were tested, to store the values of the functions of the input neurons in memory modules embedded into the reconfigurable FPGA circuit (BlockRAM) modules. The designer's dilemma is to decide how to use these resources, sacrificing – even if only partially – the pure parallel nature of the implementation or to exploit more BlockRAMs than it would actually be necessary to store the targeted data. The input space defined by the acquired EEG values was covered by input neurons with triangular activation functions. Each of these input neurons will fire (emit an output spike) a number of cycles after it has received the input value, depending on the position of the intersection point with its activation function. The inactive neurons will not fire for this input value. It is important to notice, that each input neuron will activate only for a reduced number of input values with a properly delayed spike. These values might be calculated in-circuit, but it would consume precious resources and computing power. Therefore, it has been decided to pre-compute these values using a Matlab program and then to store them in the BlockRAM modules of the FPGA circuit. Using this program, it is easy to vary the overlapping areas of the receptive fields as well as the number of input neurons used. On the other hand, the program generates the VHDL code for the BlockRAM initialization. These values will determine the moment when the input neurons will emit an output spike, relative to the moment of appearance of the new input value to be encoded (steps 0 to 15 of each timeframe). The values given as addresses to the memories are actually the input values of the SNN. A training data set was generated by manually marking durations of designated epochs. Training data was presented to SNNs with variable numbers of input, hidden and output cells. The performance of different types of networks was first examined with the training samples and then the best configuration was tested on novel sets of the EEG data. When testing the network with new patterns the manual and automatic evaluations were compared quantitatively. Three measures of performance are investigated: number of convergence epochs, computational efficiency, and classification accuracy. The network has a time-warp invariance property, which means that an input linearly compressed or elongated in time is still recognizable by the network. This enabled the network to train on one peak shape and generalize it to recognize similarly shaped peaks. These findings demonstrate the utility of SNNs in automatic recognition of EEG patterns.

4. Observation

Almost all elements of this BCI concept are completely new. We have the recording, preprocessing and signal transmission from electrodes to a computing unit in detail.

References

1. Krausz, G., Scherer, R., Korisek, G. and Pfurtscheller, G. "Critical Decision-Speed and Information Transfer in the 'Graz Brain–Computer Interface' "; Applied Psychophysiology and Biofeedback, Vol. 28, No. 3, September 2003 (C° 2003).
2. Leeb, R., Keinrath, C., Friedman, D., Neuper, C., Guger, C., Slater, M., "Walking from thought Gert Pfurtschellera", Research Report: Brain Research 1071 (2006) 145–152.
3. Piipponen, K.V.T., Sepponen, R. and Eskelinen, P., "A Biosignal Instrumentation System Using Capacitive Coupling for Power and Signal Isolation", IEEE transaction on biomedical, Vol. 54, No. 10, October 2007.
4. Harrison, T. C., Sigler, A., Murphy, T.H., "Simple and cost-effective hardware and software for functional brain mapping using intrinsic optical signal imaging", Journal of Neuroscience Methods 182 (2009) 211–218.
5. Yao, B., Salenius, S., Yue, G.H., Brownc, R.W., Liu, J.Z., "Effects of surface EMG rectification on power and coherence analyses: An EEG and MEG study"; Journal of Neuroscience Methods 159 (2007) 215–223.
6. Hampson, R.E., Collins, V., Deadwyler, S.A., "A wireless recording system that utilizes Bluetooth technology to transmit neural activity in freely moving animals", Journal of Neuroscience Methods 182 (2009) 195–204.
7. Patil, S. and Abel, E.W., "Real-time continuous wavelet transform implementation on a DSP processor", Journal of Medical Engineering & Technology, Vol. 33, No. 3, April 2009, 223–231.
8. Miller, L.C., Ruiz-Torres, R., Stienen, A.H.A. and Dewald, J.P.A. "A Wrist and Finger Force Sensor Module for Use During Movements of the Upper Limb in Chronic Hemiparetic Stroke", IEEE Transactions on Biomedical Engineering, Vol. 56, No. 9, September 2009.
9. Johansson, R.S. and Flanagan, J.R., "Coding and use of tactile signals from the fingertips in object manipulation tasks", Nature Reviews Neuroscience, Volume 10, May 2009, 345.
10. Shayani, H., Bentley, P.J. and Tyrrell, A.M., "An FPGA-based Model suitable for Evolution and Development of Spiking Neural Networks", ESANN'2008 proceedings,

European Symposium on Artificial neural Networks - Advances in computational Intelligence and Learning, Bruges (Belgium), ISBN 2-930307-08-0, pp. 197-202.

11. Kasabov, N., Benuskova, L., Wysoski, S.G., “Computational Neurogenetic Modeling: Integration of Spiking Neural Networks, Gene Networks, and Signal Processing Techniques”, ICANN 2005, LNCS 3697, pp. 509-514, 2005, Springer-Verlag Berlin Heidelber, pp.509-514.
12. Roggen, D., Federici, D. and Floreano, D., “Evolutionary morphogenesis for multicellular systems”, Genetic Programming and Evolvable Machines, 8(1):61–96, 2007.
13. Metting van Rijn, A.C., Peper, A., Grimbergen, C.A., “High-quality recording of bioelectric events. Low-noise, low-power multichannel amplifier design”, Medical & Biological Engineering & Computing, 1991, II. 29: 433-440.
14. Metting van Rijn, A.C., Peper, A., Grimbergen, C.A., “Instrumentation amplifiers for bioelectric events: a design with a minimal number of parts”, Medical & Biological Engineering & Computing, 1994, 32: 305-311.
15. Ferree, T.C. , Luu, P., Russell, G.S., Tucker, D.M., “Scalp electrode impedance, infection risk, and EEG data quality”, Clinical Neurophsiology, 2001, 112/3: 536-544.
16. Zaciu, C. “Metode electronice de masurare in neurobiofizica celulara”,1983,Cluj,Ed. Dacia
17. Teplan, M., “Fundamentals of EEG measurement”, Measurement Science Review, 2002, 2/2:17-28.

Analysis of Neuroelectric Oscillations of the Scalp EEG Signals

László F. MÁRTON[1], László SZABÓ[1], Margit ANTAL[1], Katalin GYÖRGY[1]

[1] Department of Electrical Engineering, Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, e-mail: martonlf@ms.sapientia.ro

Manuscript received April 11, 2010; revised April 25, 2010.

Abstract: Electroencephalography (EEG) or magneto-encephalography (MEG) are usual methods adopted in clinics and physiology to extract information relative to brain activity. EEG/MEG signals are a measure of the collective neural cell activity on restricted regions of the cortex. The brain activities ensue from the interaction of excitatory and inhibitory populations of neurons. Their kinetics vary depending on a particular task to be fulfilled, on the particular region involved in that task, and on any instants during the task. In order to improve our understanding of EEG/MEG signals, and to gain a deeper comprehension of the neurophysiological information contained, various mathematical models and methods have been proposed in past years. Based on these concepts we develop procedures for an optimal, online and hardware based EEG signal processing.

Keywords: brain-computer-interface (BCI), time frequency analysis, filters, EEG signals, stationarity of signals

1. Introduction

The description of neural network dynamics is difficult at the level of the individual cells (this technique is based on a so called invasive recording method, not the subject of our study). Because of our recording technique, a noninvasive method, the dynamics of a cortical neural population is summarized within an EEG signal. These potential types of biosignals are intrinsically very complex, including different rhythms and have a large frequency spectrum, which may vary rapidly in time, reflecting the nonstationarity of the underlying phenomena and the changes in the kinetics of the neural mechanisms [1]. A fundamental component of neural activity the action potentials, are very difficult to be detected in a noninvasive EEG recording. The main tasks of our study are grouped in the following diagram from *Fig. 1***.** There are three main streams in this signal processing work. One is the noise type recognition and reduction, the second is the signal visualization frame and the third is the pattern recognition. The visualization and noise type analysis have mainly offline importance. The results of these tasks are finally focused on signal's specifications recognition for the online noise reduction and a pattern recognition.

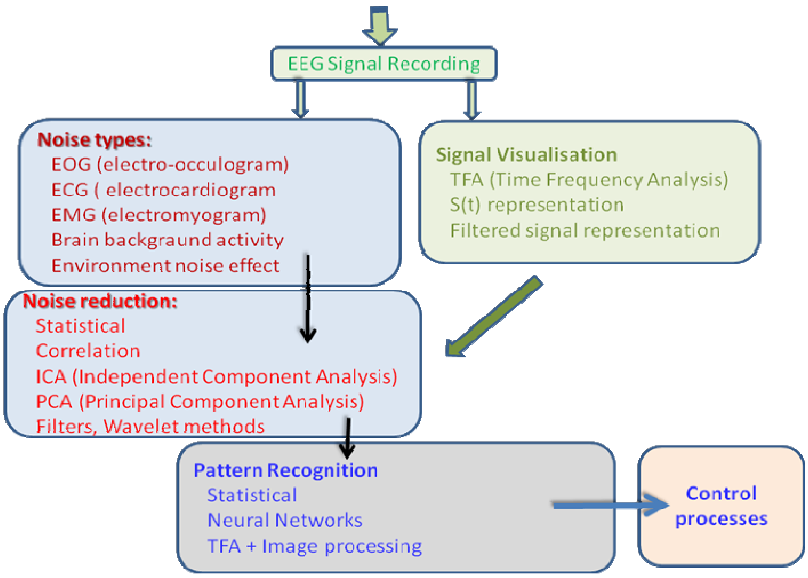


Figure 1:. Block diagram of the main EEG processing tasks.

A common goal in designing a BCI is to get the best electroencephalogram (EEG) signals, recorded during different mental tasks. The features from EEG signals such as power and ratios from different oscillatory bands (delta (0–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–20Hz) and gamma (20-70 Hz)) have been used in classifying different mental tasks. Asymmetry ratios in EEG are especially useful for recognizing mental tasks that elicit inter-hemispheric differences.

Human neuroelectric studies have provided strong evidence for a close link between the mind and the brain, so it should be, at least in principle, possible to decode what an individual is thinking extracted from their brain activity. However, this does not reveal whether such decoding of mental states can be practically achieved with current neuroelectric signal recording and decoding methods. The purpose of this paper is to improve information extracting procedures to get closer to understand the mapping of the environment into a brain task and to use the recognized information in different systems control procedures.

The basic question is how accurately and efficiently can a mental state be inferred? What is the maximum temporal resolution? Is it possible to provide a quasi-online estimate of an individual's current cognitive or perceptual state? Usually, EEG-signals are based on various phenomena like, for example, visual and P300 evoked potentials, slow cortical potentials, or sensory-motor cortical rhythms. The P300 shape is an event related potential (ERP) elicited by, usually stochastic task related stimuli. It is considered to be an endo-genous potential as its occurrence links not to the physical attributes of a stimuli it is a positive shape in voltage with a latency of roughly 300 to 600 ms. In a motion based BCI system, the variation in the EEG signal due to the movement or due to the idea of the movement of hands or feet must be recognized.

2. Noise reduction necessity

The main problem with most information extraction approaches is the low classification accuracy of extracted information, for which the main reason is the noisy nature of the EEG signals. One of the main sources of noise and artifact in the EEG signals is the interferences from other bio-potentials sources like the electro-occulogram (EOG), the electrocardiogram (ECG), the electromyogram (EMG), and most importantly the background activity of the brain itself[12][13]. As we have mentioned earlier we have classified the main tasks into two categories. The first category is a software package, important during the calibration signal recording and study procedures (offline bio-signal processing package, created for offline experiments to eliminate initial recording, filtering and feature extraction misinterpretations). The second category package is for online biosignal processing and analysis and it contains an extensive software package with specialized toolboxes for a wide range of scientific applications. All steps from raw data import to preprocessing, transformation, explorative analyses, parameter extraction are imperative for control tasks(our main goal)[7].

The result of our methods of background noise reduction can be seen in the following two figures (*Fig. 2.* and *Fig. 3.*). Noise type is asking for different de-noising algorithms. The first figure is showing a slow wave signal with background noise (first frame), and the steps to the biological signal using a de-noising procedure.

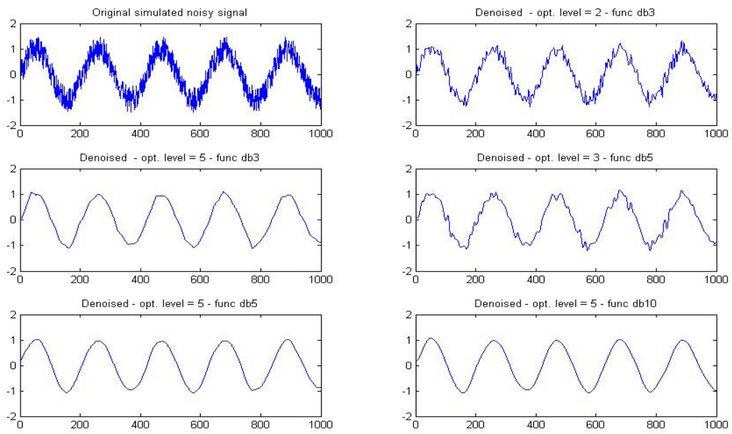


Figure 2: Different levels of de-noising of oscillatory EEG signal.

The next figure is an offline processing of a higher frequency band signal and is designed to remove the effects of volume conductance and other artifacts from our recording of a left hand moving recorded on right hemisphere of skullcap. This is only a detail of a longer EEG recording and this sequence is used further in this paper.

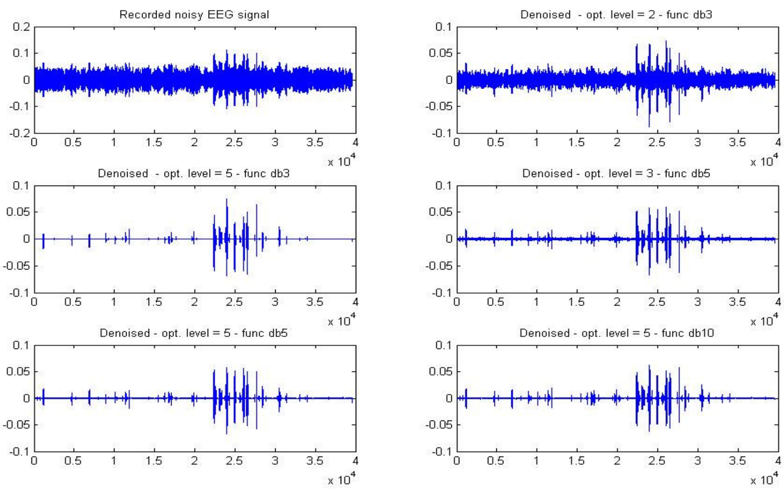


Figure 3: The left hand lift signature and filtered version of an EEG recording.

Artifacts can be due to physiological or non-physiological sources. Physiological sources for artifacts include eye movements and eye blinks, muscle activity, heart activity, and slow potential drifts due to transpiration.

Non-physiological sources for artifacts include power supply line noise (at 50 Hz), noise generated by the EEG amplifier, and noise generated by sudden changes in the properties of the electrode-scalp interface. Artifacts often have much larger amplitude than the signals of interest. Therefore, artifact removal and filtering procedures have to be applied before an analysis of EEG signals can be attempted. Each artifact has its characteristic. We are performing experiments to find out these characteristics and to provide the best way to identify and eliminate them. The above mentioned developed technique id named first category biosignal processing methods [14]. After the refinement of the techniques performed on a computer the next step must be to use them on a hardware based control unit (microcontroller, FPGA circuits, hardware filters [2]). Filter's design are based on frequency properties of the signals. To filter, we must know the basic characteristics of the artifacts and filters, especially the frequency band to be filtered out. In the references we can find several studies on artifact's descriptions [8][14]. Many methods have been proposed to remove eye movement and blink artifacts from EEG recordings [7][10].
Severe contamination of EEG activity is a serious problem for EEG interpretation and analysis. One method is rejecting contaminated EEG epochs with some loss of information, but this is difficult during online tasks.

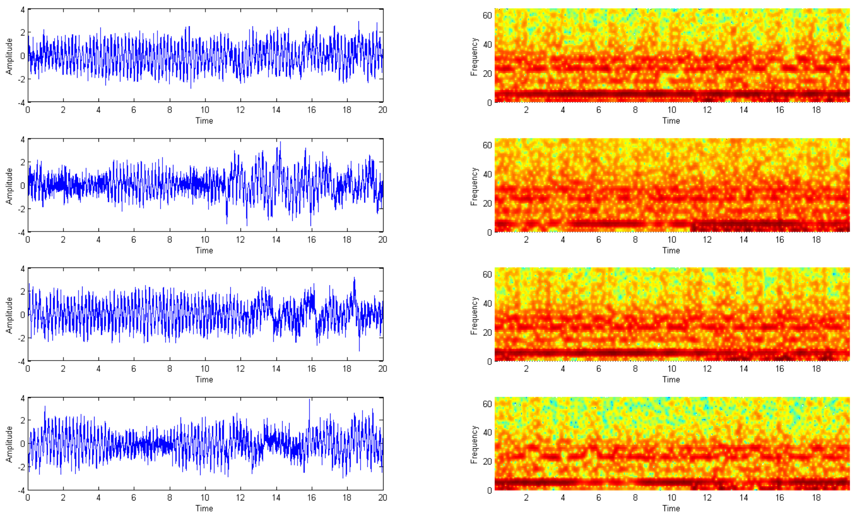


Figure 4: Four ICA component of the left hand lift of our EEG recording.

Another way is to use 'regression' based methods in time or frequency domain and are performed on simultaneous EEG and electrooculographic (EOG) time series. But in case of regression, not all the time, clear reference channels are available.

A new and often preferable alternative to remove a wide variety of artifacts is to apply Independent Component Analysis (ICA) to multichannel EEG recordings [8]. Literature shows that ICA can effectively detect, separate and remove wide variety of artifact activity. The independent components of the left hand EEG signal are visible on *Fig. 4.* The ICA method is based on the assumptions that the time series recorded on the scalp are spatially stable mixtures of the activities of temporally independent cerebral and artifact sources, that the summation of potentials arising from different parts of the brain, scalp, and body is linear at the electrodes, and that propagation delays from the sources to the electrodes are negligible and it is important that ICA does not require a reference channel for each artifact source [12].

3. Information extraction requirements

Event-Related Potentials **(**ERPs) are spatio-temporal patterns of brain activity occurring time-locked to an event, for example after presentation of a stimulus, before execution of a movement, or after the detection of a novel stimulus. Traditionally, ERPs are recorded with the EEG and have been used in neuroscience for studying the different stages of perception, cognition, and action. The above mentioned P300 is a positive deflection in the EEG, appearing approximately 300 ms after the presentation of rare or surprising, task-relevant stimuli[6][10]. New and emerging approaches must to be found for a direct assess how well a mental state can be reconstructed from non-invasive measurements of brain activity in humans. This is another goal of our research [1] [2] [3].

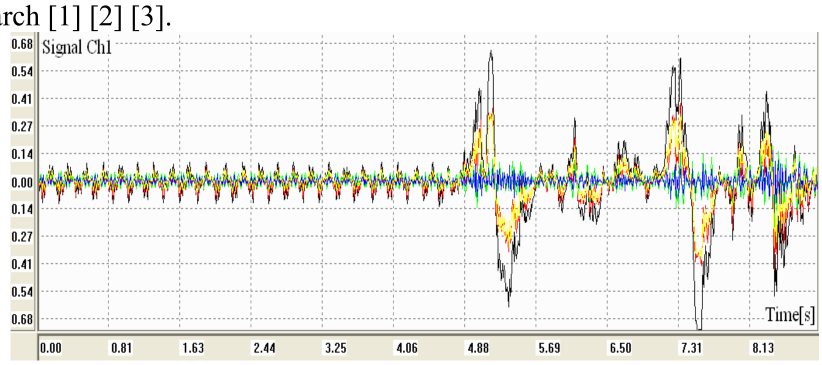


Figure 5: Time series of left hand lift EEG signal.

We consider now the EEG time series of the left hand lift signal, superimposed with different frequency band components (*see Fig. 5.*). The next figure is showing two different time-frequency representation (TFR) of the signal from previous figure (*Fig. 6.*).

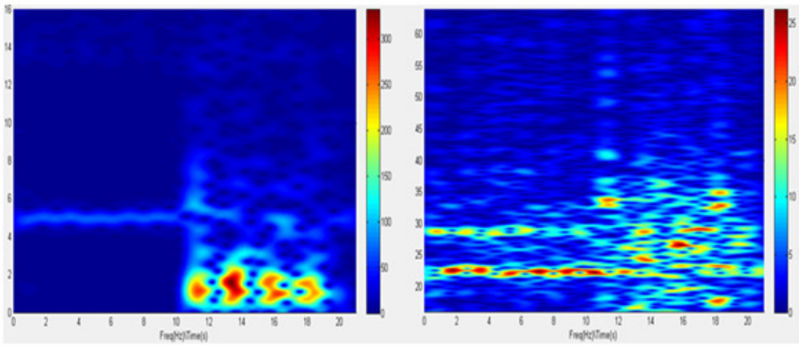


Figure 6: TFR of a time series signal.

It is visible that the information processing of the brain is reflected in dynamical changes in time, frequency, and cortical space of the electrical activity. Therefore, the concomitant studies required methods capable of describing the qualitative and quantitative variation of the signal in both time and frequency.

At this stage we have summarized the mathematically based procedures to perform the mentioned tasks relative to offline and online signal processing. As we have mentioned, the main goal is the pattern recognition of different external event related actions[9] reflected in our EEG recordings. After the pattern recognition of a followed event, the whole procedure supported by a software based tool, must provide a bitrate range between 10 and 25 bits/min of information to have a minimal capability to perform control tasks. The range of 10 to 25 bits/min are to be exceeded by new methods of signal processing.

The basic properties of biosignals (neuroelectric oscillations) are to be time varying and nonstationary, especially in some pathological conditions. Currently, the most widely used quantitative tool for such nonstationary signals is the above mentioned time–frequency representation which demonstrates the temporal evolution of different frequency components.

However, TFR does not provide a quantitative measure of nonstationarity level, e.g., how far the process deviates from stationarity. Techniques based on Fourier analysis (highly dependent on the stationarity of the measured signal) are not suitable for neuroelectric oscillations study [5] [6]. While the time frequency(TF) energy distribution of the averaged evoked potential allows us to characterize phase-locked activities, it is possible to account for both phase locked and non-phase-locked activities by averaging the TF energy distribution of each single trial. A way of our information recognition is the scalogram extraction procedure from a TF transformed signal. The black curve in the TF transforms from *Fig. 7.* is a scalogram and is representative for cortical events.

The signal used in this processing is a detail of the same left hand lift recording. One way of pattern recognition is based on the use of scalogram

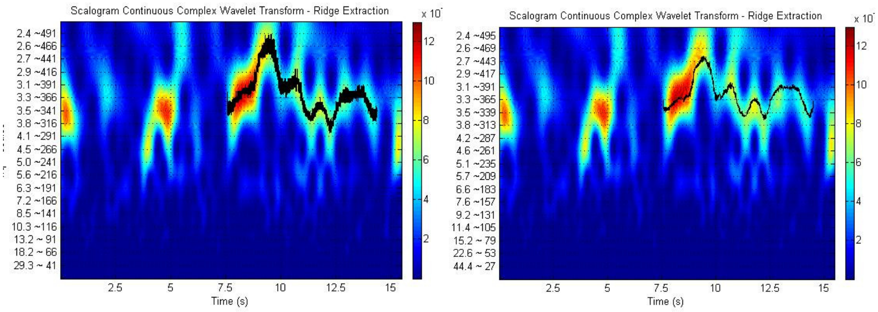


Figure 7: Scalogram extraction from TF transformed EEG event.

The use of time–frequency (TF) analysis is not sensitive to the information included in each single trial. The image in *Fig. 8*. is a general view of the same event from *Fig. 7*.

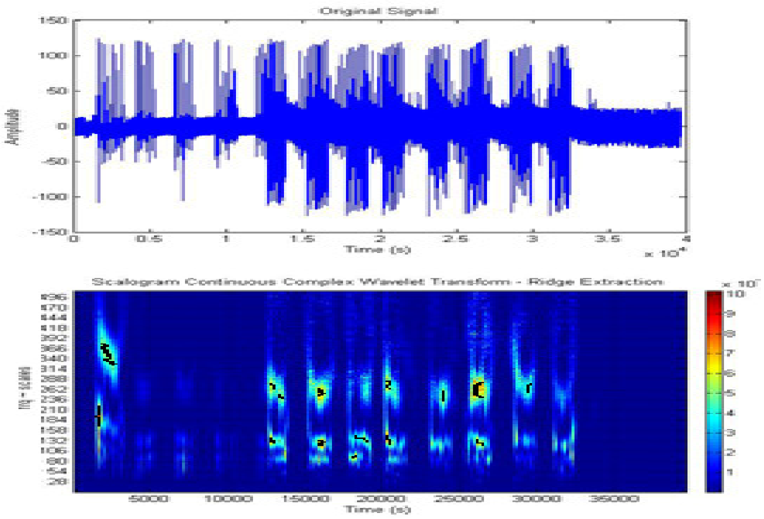


Figure 8: TF figure of one of our recording of left hand lift EEG.

Another way experimented are the discrete wavelet transform based neuroelectric signal analysis. In the presented processing, as noise suppression, wavelet methods have been experimented. At first, it is important to preprocess the EEG signal with the Hilbert transformer to yield an analytic signal. The transformed signal is then easily wavelet transformed using the linear phase complex scaling and wavelet filtering. Wavelet transform then can provide methods to find phase relations from scalogram and to design operative online filters and best methods of feature extraction. To enhance the above mentioned methods up to become online possibility, is the next step.

References

[1] Bard Ermentrout, Roberto F. Gala´n, and Nathaniel N. Urban “Reliability, synchrony and noise” G Review Trends in Neurosciences Vol.31 No.8

[2] Charles E. Schroeder and Peter Lakatos “Low-frequency neuronal oscillations as instruments of sensory selection” Opinion Trends in Neurosciences Vol.32 No.1

[3] Thomas C. Harrison, Albrecht Sigler, Timothy H. Murphy “Simple and cost-effective hardware and software for functional brain mapping using intrinsic optical signal imaging” Journal of Neuroscience Methods 182 (2009) 211–218

[4] K. C. Chua, V. Chandran, U. Rajendra Acharya, and C.M.Lim “Analysis of epileptic EEG signals using higher order spectra” Journal of Medical Engineering & Technology, Vol. 33, No. 1, January 2009, 42–50

[5] X. Guo, G. Yan and W. He “A novel method of three-dimensional localization based on a neural network algorithm” Journal of Medical Engineering & Technology, Vol. 33, No. 3, April 2009, 192 – 198

[6] Bor-Shyh Lin*, Member, IEEE, Bor-Shing Lin, Fok-Ching Chong, and Feipei Lai, Senior “Higher Order Statistics-Based Radial Basis Function Network for Evoked Potentials” Member IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 56, NO. 1, JANUARY 2009

[7] Luca Faes*, Member, IEEE*, Ki H. Chon*, Senior Member, IEEE*, and Giandomenico Nollo*, Member, IEEE* “A Method for the Time-Varying Nonlinear Prediction of Complex Nonstationary Biomedical Signals” IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 56, NO. 2, FEBRUARY 2009 205

[8] Jianhua Luo*, Yuemin Zhu, and Isabelle E. Magnin*, Member, IEEE* “Denoising by Averaging Reconstructed Images: Application to Magnetic Resonance Images” IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 56, NO. 3, MARCH 2009

[9] Arash Ajoudani and Abbas Erfanian∗*, Member, IEEE* “A Neuro-Sliding-Mode Control With Adaptive Modeling of Uncertainty for Control of Movement in Paralyzed Limbs Using Functional Electrical Stimulation” IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 56, NO. 7, JULY 2009 1771

[10] Koel Das, Daniel S. Rizzuto, and Zoran Nenadic∗*, Member, IEEE* “Mental State Estimation for Brain–Computer Interfaces” IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 56, NO. 8, AUGUST 2009

[11] *Miguel A. L. Nicolelis and Mikhail A. Lebedev “*Principles of neural ensemble physiology underlying the operation of brain–machine interfaces” 530 | July 2009 | Volume 10 www.nature.com/reviews/neuro

[12] Aihua Zhang and Weiping Li “Adaptive Noise Cancellation for Removing Cardiac and Respiratory Artifacts from EEG Recordings” Proceedings of the 5Ih World Congress on Intelligent Control and Automation, June 15-19, 2004, Hangzhou, P.R. China

[13] Nguyen T.K. Cuong, Vo Q. Ha1, Nguyen T. M. Huong, Truong Quang Dang Khoa , Nguyen Huynh Minh Tam, Huynh Q. Linh1, Vo Van Toi “Removing noise and artifacts from EEG using Adaptive noise cancelator and Blind”Source Separation Proceedings of the 3rd International Conference on the Development of BME in Vietnam, 11-14th Jan 2010

[14] T-P Junk, Scott Making, C. Humphries, T-W Lee, M. J. McKeown, V. Irugui and T. J. Sejnowski “Removing electroencephalographic artifacts by blind source separation” *Psychophysiology, 37* (2000), 163–178. Cambridge University Press. Printed in the USA.

Nonlinear Filtering of Electrocardiogram Signals

Zoltán GERMÁN-SALLÓ[1]

[1] Department of Electrical Engineering, Faculty of Engineering, "Petru Maior" University of Tg. Mureş, e-mail: zgerman@engineering.upm.ro

Manuscript received March 15, 2010; revised April 15, 2010

Abstract: This paper presents a non-linear filtering method based on the multiresolution analysis of the Discrete Wavelet Transform (DWT). The main idea is to use the time-frequency localization properties of the wavelet decomposition. The proposed procedure uses an extra decomposition of identified noise to reduce the correlation between the electrocardiogram signal and noise. The non-linear filter acts by thresholding the wavelet coefficients and by substracting the correlated noise The threshold used depends on the noise level in each of the frequency bands associated to the wavelet decomposition. Signals from the MIT-BIH database were used, gaussian white noise was added to perform the filtering procedure. The proposed procedure was compared with ordinary wavelet denoising and wavelet packet denoising procedures.

Keywords: wavelet decomposition, wavelet shrinkage, non-linear filtering, thresholding.

1. Introduction

The interest in using the wavelet transform to denoise the electrocardiogram (ECG) signal is increasing. The wavelet transform is a powerful tool in time–frequency analysis preferred for the interrogation of complex non-stationary signals. Its application to electrocardiogram signal processing has been found particularly useful due to its time-frequency localization capabilities. The discrete wavelet transform- based approach produces a dyadic decomposition structure. In this way, the wavelet packet approach is an adaptive method using an optimization of the best tree decomposition structure independently for every signal.

2. Methods

The wavelet transform (WT) of a signal $x(t)$ is defined as a combination of a set of base functions, obtained by means of scaling *(a)* and translation (*b)* of a mother wavelet function [1]:

$$W_a x(b) = \int_{-\infty}^{+\infty} x(t)\,\psi_{a,b}(t)dt = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty} x(t)\,\psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

where a and b are real ($a \neq 0$). A small value of *a* gives a contracted version of the mother wavelet function and then allows the analysis of high frequency components. A large value of the scaling factor stretches the basic function and provides the analysis of low-frequency components of the signal. The discrete wavelet transform (DWT) is defined as a discretized dilation and translation of a mother function, (or analyzing wavelet). In its most common form, the DWT applies a dyadic grid (integer power of two scalings with a and b) and orthonormal wavelet basis functions and it exhibits zero redundancy.

$$\psi_{(s,l)}(x) = 2^{-\frac{s}{2}}\psi\left(2^{-s}x - l\right) \tag{2}$$

The variables *s* and *l* are integers that scale and dilate the mother function ψ to generate wavelets (analyzing functions). The scale index *s* indicates the wavelet's width, and the location index *l* gives its position. The mother wavelets are rescaled, or "dilated" by powers of two, and translated by integer values. In this case we have a dyadic decomposition structure. This functions define an orthogonal basis, the so-called wavelet basis [3,5]. The Discrete Wavelet Transform decomposition of the signal into different frequency bands (according to Mallat's algorithm [2]) can be obtained by successive high-pass and low-pass filtering (digital FIR structures) of the time domain followed by down sampling to eliminate the redundancy, as shown in *Fig. 1.*

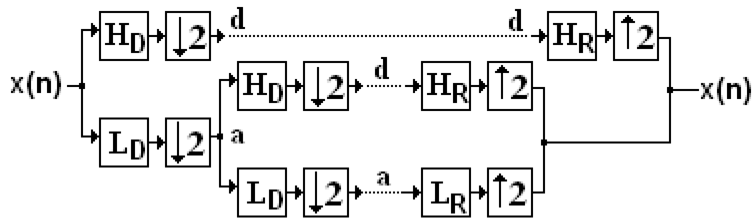


Figure 1. Second order dyadic scale decomposition and reconstruction.

In discrete wavelet analysis x(t) is decomposed on different scales, as follows:

$$x(t)=\sum_{j=1}^{K}\sum_{k=-\infty}^{\infty}d_j(k)\psi_{j,k}(t)+\sum_{k=-\infty}^{\infty}a_K(k)\varphi_{K,k}(t), \tag{3}$$

where $\psi_{j,k}(t)$ are discrete analysis wavelets and $\varphi_{K,k}(t)$ are discrete scaling functions, $d_j(k)$ are the detailed wavelet coefficients at scale 2^j and $a_K(k)$ are the approximated scaling coefficients at scale 2^K. The discrete wavelet transform can be implemented by the wavelet and scaling filters

$$L(n)=\frac{1}{\sqrt{2}}\langle\varphi(t),\varphi(2t-n)\rangle \tag{4}$$

$$H(n)=\frac{1}{\sqrt{2}}\langle\psi(t),\varphi(2t-n)\rangle=(-1)^n L(n) \tag{5}$$

being quadrature mirror filters [3]. The estimation of the detail signal at level j will be obtained by convolving the approximation signal at level j-1 with the coefficients G(n). Convolving the approximation coefficients at level j-1 with the coefficients H(n) gives an estimate for the approximation signal at level j. *Fig. 2.* shows the wavelet decomposition tree and the time-frequency blocks for a third order decomposition.

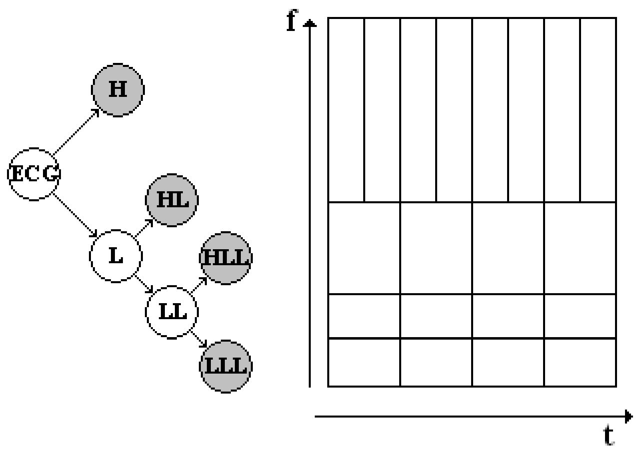


Figure 2. Time-frequency blocks for third order dyadic scale decomposition.

The wavelet packet analysis is a generalization of discrete wavelet analysis providing a redundant decomposition procedure, where both detail and approximation signals are split at each level into finer components. This produces a decomposition tree as shown in *Fig. 3*.

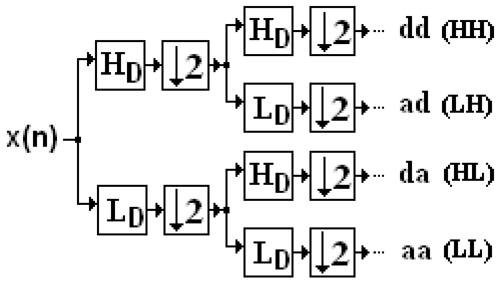


Figure 3. Wavelet packet decomposition tree.

The tree contains several admissible bases one of which is the wavelet basis itself. Having a large but finite library of bases it is possible to extract the best basis relatively to some criterion. The best basis algorithm finds a set of wavelet bases that provide the most desirable representation of the data relative to a particular cost function which may be chosen to fit a particular application. This basis can be any sub-tree of the initial entire tree.

Fig. 4 shows the time-frequency blocks for a second order wavelet packet decomposition, L and H are the low- and highpass filters with downsamplers.

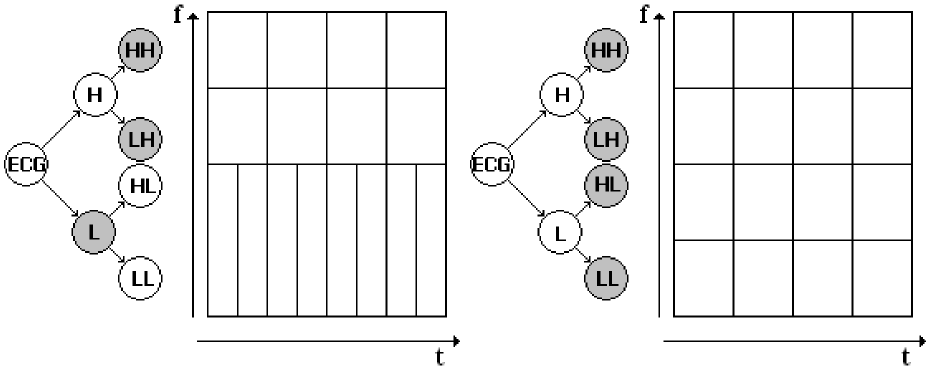


Figure 4. Time-frequency blocks for second order dyadic wavelet packet decomposition

3. The suggested procedure

The objective of any denoising procedure is to extract the useful signal from the noisy signal, by eliminating the identified noise. The model of the noisy signal is the superposition of the signal and a Gaussian type of noise. The main idea of non-linear filtering is to use the time-frequency localization properties of the discrete wavelet decomposition. The non-linear denoising approach assumes that every wavelet coefficient contains noise and the noise is distributed over all scales. The non-linear thresholding means discarding the detail coefficients exceeding a certain limit. There are two types of thersholding, the soft and the hard methods. With hard thresholding the coefficients which are lower than the threshold will be set to zero. In soft thresholding, the remaining non-zero coefficients are shrunk toward zero. In this paper we assume that the identified noise will be contained by the first order detail coefficients. As we can see the first order decomposition of an electrocardiogram signal shows some correlation between noise and signals characteristic points, as R peaks.

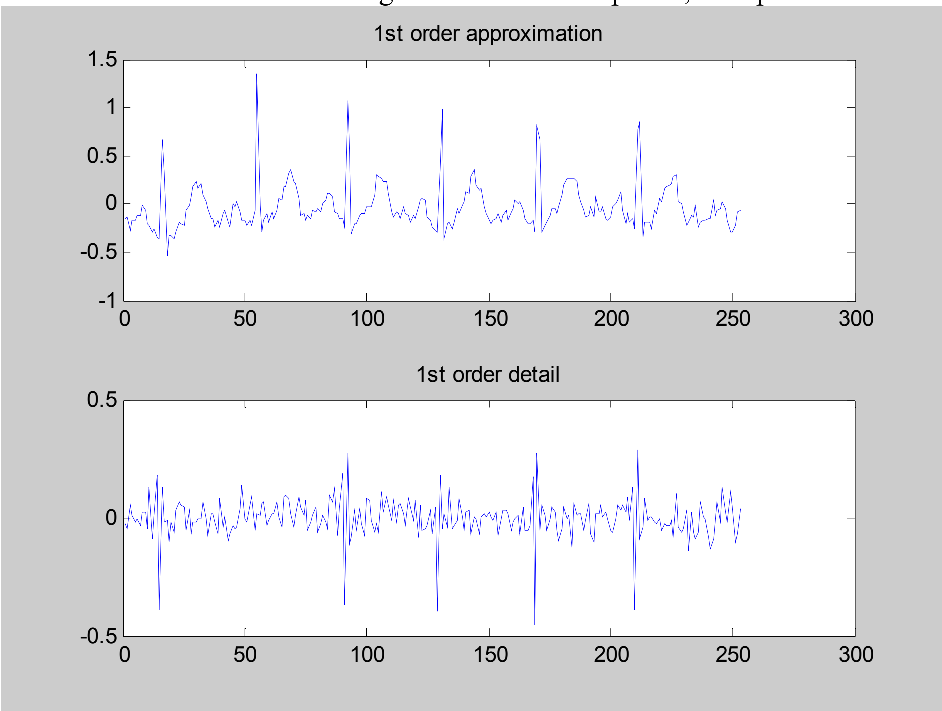


Figure 5. Correlation between first order detail and approximation coefficients.

An extra decomposition of identified noise can help to reduce the correlation between the electrocardiogram signal and noise. The non-linear filter acts by thresholding the detail coefficients. The estimated noise, n_e identified as the first order detail part is decomposed in a discrete wavelet structure, after that it is reconstructed only from the second order detail coefficients and is subtracted from the initial noise. The proposed smoothing procedure consists of:

1. one level DWT decomposition to estimate the noise from detail coefficients
2. third level DWT decomposition of the noisy ECG, resulting 3 detail coefficients (D1, D2, D3)

3. second level decomposition of the estimated noise (D1), resulting dD1, dD2
4. reconstruction of dD1' only from dD2 and subtracting from dD1,
5. thresholding the detail coefficients
6. reconstruction of the filtered signal

The procedure can be seen in *Fig. 6.*

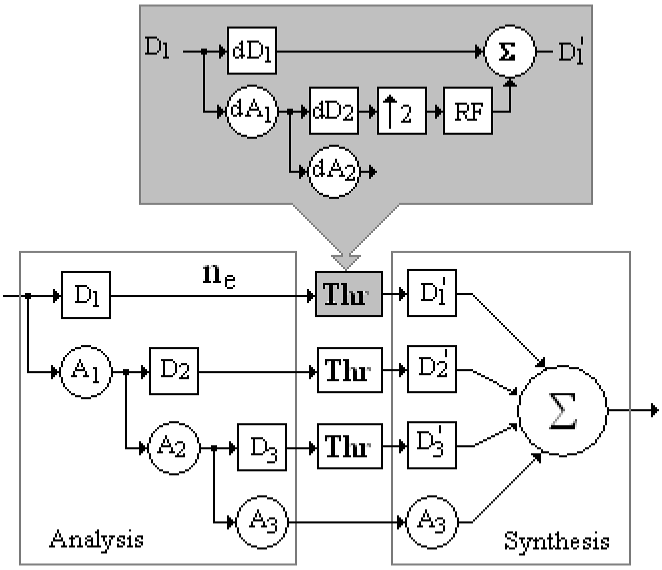


Figure 6. The proposed non-linear filtering procedure.

4. Results

The test signal used was from MIT-BIH database, gaussian white noise was added. To estimate the ability of this denoising procedure the parameters followed were the signal / noise ratio obtained, the absolute value of the error and the gain, defined as:

$$SNR_1[dB] = 10\lg\left(\frac{PoriginalECG}{PoriginalECG - PdenoisedECG}\right) \tag{5}$$

$$Error = abs(originalECG - denoisedECG) \tag{6}$$

$$Gain = SNR_1 - SNR \tag{7}$$

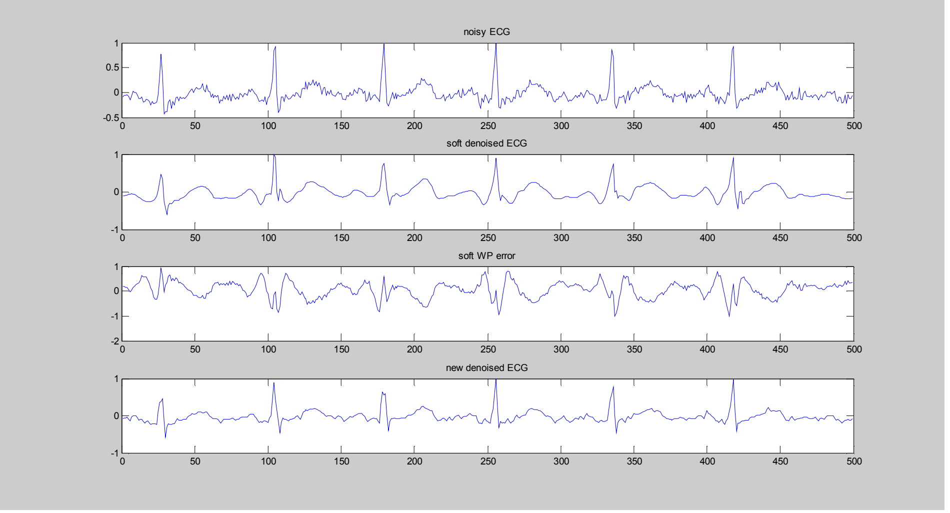


Figure 7. The obtained results.

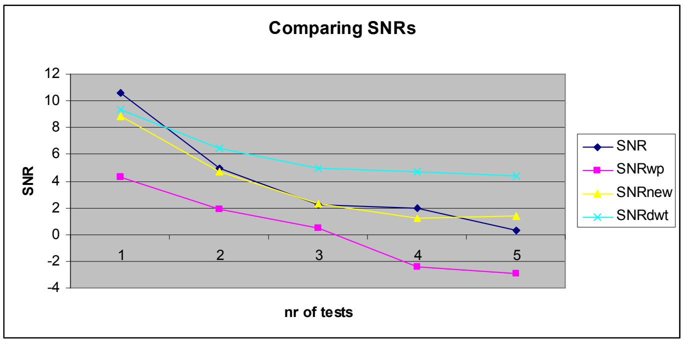


Figure 8 The approximation error.

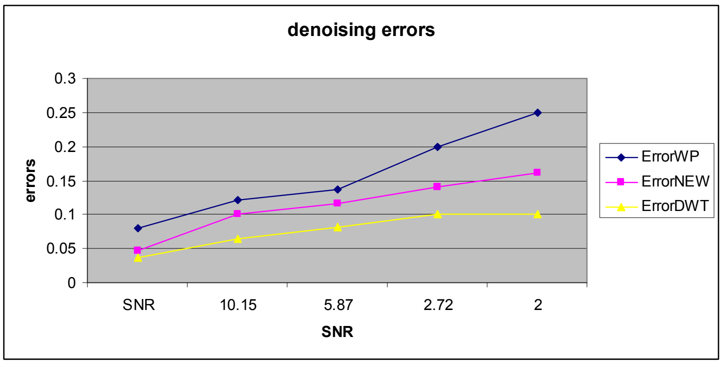


Figure 9 The tendency of denoising errors.

4. Conclusions

The main idea was to estimate the correlation between the noise and the signal. An extra decomposition of the noise tried to reduce this correlation. This method was compared with ordinary wavelet decomposition and wavelet packet decomposition based filtering techniques. Because of numerical errors the followed parameters do not necessarily inform about noise removal, sometimes, it is useful to check the denoised signals visually. The obtained error values cannot directly indicate the improvement for the ECG waveform detection. Wavelet and wavelet packets based denoising methods showed different results, due to the different division strategies of the signal decomposition structures.

References

1. Donoho, D.L. (1995), "De-noising by soft-thresholding," IEEE, Transaction on Information Theory, 41, 3, pp. 613-627.
2. Aldroubi, A., Unser, M.: Wavelets in Medicine and Biology. CRC Press New York 1996
3. Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J-M.: WaveletToolbox. For Use with Matlab. User's Guide. Version 2.The MathWorks Inc 2000
4. Coifman, R.R.: M.V Wickerhauser : Entropy-based algorithms for best basis selection IEEE Trans. on Inf. Theory, 1992, vol. 38, 2, pp. 713–718.
5. Mallat, S.A., 1989. A theory for multi-resolution signal decompositions: The wavelet representation. IEEE Trans. On Pattern Analysis and Machine Intelligence, 11(7):674-693. [doi:10.1109/34.192463]
6. Donoho, D.L.; I.M. Johnstone (1994), "Ideal spatial adaptation by wavelet shrinkage," Biometrika, vol 81, pp. 425-455 Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005
7. Chang, C.S., Jin, J., Kumar, S., Su, Q., Hoshino, T., Hanai, M., Kobayashi, N., 2005. Denoising of partial discharge signals in wavelet packets domain. IEE Proceedings of Science, Measurement and Technology, 152(3):129-140

Equivalence Relations for Modulated Signals and Implications on Astronomical Modulators

Vlad TURCU[1]

[1] Astronomical Observatory Cluj, Astronomical Institute of the Romanian Academy , Cluj-Napoca, e-mail: vladturcu@yahoo.com

Manuscript received April 13, 2010; revised April 25, 2010.

Abstract: An equivalence theorem between extremes of frequency modulated or phase modulated signals with pulse position modulated signals of natural type was formulated and demonstrated. General implications of this result in obtaining signals with pulse position modulation from frequency modulation or phase modulation were emphasized. Practical applications of the equivalence theorem for astronomical signals and astronomical modulators were also exemplified for X Tri binary system.

Keywords: Modulation-demodulation techniques, FM, ΦM, PPM, PWM, PAM, astronomical modulators time series, binary stars: individual X Tri.

1. Introduction

The present paper is dedicated to the application of modulation formalism and techniques to the astronomical signals and natural astronomical modulators.

Fundamental contributions in signal theory and in particular modulation theory and its applications can be found in [1], [2], [3], [4].

Modulation theory formalism was used in astronomy by Pop&Turcu [5] to describe stellar binary system with a pulsating component by means of frequency modulation. The principles of pulse position modulation were used by Turcu&Pop [6] in the study of period variability phenomena. The astronomical modulator concept was introduced by Turcu [7] for describing dynamical astronomical systems in terms of information transmission theory. The natural character of astronomical modulator systems was also emphasized comparing them with engineering modulators, designed and realized for optimization of transmission channels. A comprehensive study of the application of modulation theory and astronomical modulators was recently given in [8].

The methods developed in the present paper are both theoretical and practical. Firstly a theoretical background has been established for equivalence relations between different modulated signals. Secondly it has been developed a practical method for period variability investigation of variable stars using the previous equivalence theorem and the demodulation principles for pulse modulated signals.

The method was exemplified on X Trianguli, an eclipsing binary star system. It has been detected a periodicity of about 17 years, probably induced by an unseen third body. Meanwhile some residual signals have been identified from the usual heliocentric correction applied to every time series astronomical observations. Common points and differences between present method and classical “O-C” (observed-calculated) method were also emphasized.

2. Equivalence relations

Theorem: The time series representing moments of maxima for a frequency modulated signal (FM) or phase modulated signal (ΦM) is a pulse position modulated signal with natural sampling (PPMN) and carrier Dirac periodic function (δ_T).

Proof:

a. The case of frequency modulated signal.

We consider the following expressions for the modulating signal and the frequency modulated (FM) signal:

$$x_m(t) = A_m \cdot y_m(t), \quad \max|y_m(t)| = 1, \tag{1}$$

$$x_{FM}(t) = A_C \cdot \cos\left[\Omega_C \cdot t + \Delta\Omega \cdot \int_0^t y_m(\tau) d\tau + \Phi_C\right], \tag{2}$$

where $\Delta\Omega = k_\Omega A_m$ is the magnitude of the frequency deviation.

The extremes for $x_{FM}(t)$ are obtained by annulment of its first derivative:

$$\frac{dx_{FM}(t)}{dt} = -A_C \cdot [\Omega_C + \Delta\Omega \cdot y_m(t)] \cdot \sin\left\{\Omega_C \cdot t + \Delta\Omega \cdot \int_0^t y_m(\tau) d\tau + \Phi_C\right\} \tag{3}$$

The solutions for this equation are obtained either by cancellation of the square parenthesis expression or the solutions of by solving the trigonometric equation sin{z}=0.

The annulment of the square parenthesis gives:

$$y_m(t) = -\frac{\Omega_C}{\Delta\Omega}. \tag{4}$$

We also have $\Omega_C >> \Delta\Omega = k_\Omega A_m$, and obviously there is a contradiction with (1). Consequently the square parenthesis is always strictly positive.
Only the solutions for trigonometric equation remain:

$$\Omega_C \cdot t + \Delta\Omega \cdot \int_0^t y_m(\tau)d\tau + \Phi_C = k\pi, \quad k \in \mathbf{Z}. \tag{5}$$

If the condition is imposed that the first extreme is a maximum and this maximum is set at $t=0$, we obtain $\Phi_C=0$, for $k=0$. Equation (5) can be rewritten for times of maxima:

$$\Omega_C \cdot t + \Delta\Omega \cdot \int_0^t y_m(\tau)d\tau = 2n\pi, \quad n \in \mathbf{Z}. \tag{6}$$

The solutions for equation (6) can be given in the form:

$$t_n = n \cdot \frac{2\pi}{\Omega_C} - \frac{\Delta\Omega}{\Omega_C} \cdot \int_0^{t_n} y_m(\tau)d\tau, \quad n \in \mathbf{Z}. \tag{7}$$

With the substitution:

$$T_C = \frac{2\pi}{\Omega_C} \tag{8}$$

$$t_n - n \cdot T_C = -\frac{\Delta\Omega}{\Omega_C} \cdot \int_0^{t_n} y_m(\tau)d\tau, \quad n \in \mathbf{Z}. \tag{9}$$

The last equation gives the timing deviation of the FM signal maxima compared with uniform sampling timings nT_C, and it can be rewritten:

$$\Delta p(t_n) = -\Delta p_{FM} \cdot \int_0^{t_n} y_m(\tau)d\tau, \tag{10}$$

where the notations: $\Delta p(t_n) = t_n - n \cdot T_C$, and $\Delta p_{FM} = \Delta\Omega/\Omega_C$. has been used. Hence the time series representing moments of maxima for a frequency modulated signal (FM) has been obtained from a periodic signal δ_T, position modulated with modulating signal $\Delta p(t_n)$ given by relation (10):

$$x_{\max FM} = A_C \cdot \sum_{n=-\infty}^{\infty} \delta(t - t_n) = A_C \cdot \sum_{n=-\infty}^{\infty} \delta\left(t + \Delta p_{FM} \cdot \int_0^{t_n} y_m(\tau)d\tau - n \cdot T_C\right). \tag{11}$$

b. The case of phase modulated signal.

We consider the following expression for the phase modulated (ΦM) signal, and the modulating signal given by expression (1):

$$x_{\Phi M}(t) = A_C \cdot \cos[\Omega_C \cdot t + \Delta\Phi \cdot y_m(t) + \Phi_C], \tag{12}$$

where $\Delta\Phi = k_\Phi A_m$ is the phase deviation.

The extremes for $x_{\Phi M}(t)$ are obtained by annulment of its first derivative:

$$\frac{dx_{\Phi M}(t)}{dt} = -A_C \cdot \left[\Omega_C + \Delta\Phi \frac{dy_m(t)}{dt}\right] \cdot \sin\{\Omega_C \cdot t + \Delta\Phi \cdot y_m(t) + \Phi_C\}. \tag{13}$$

Similarly with the FM case the solutions for this equation are obtained either by cancellation of the square parenthesis expression or solutions of trigonometric equation sin{z}=0.

The annulment of the square parenthesis gives:

$$\frac{dy_m(t)}{dt} = -\frac{\Omega_C}{\Delta\Phi}. \tag{14}$$

We also have $\Omega_C >> \Delta\Phi = k_\Phi A_m$, and obviously there is a contradiction with (1). Consequently the square parenthesis is always strictly positive.

Only the solutions for the trigonometric equation remain:

$$\Omega_C \cdot t + \Delta\Phi \cdot y_m(t) + \Phi_C = k\pi, \quad k \in \mathbf{Z}. \tag{15}$$

If we impose the conditions for the first extreme to be a maximum and we set this maximum at $t=0$, we obtain $\Phi_C = -\Delta\Phi \cdot y_m(0)$, for $k=0$. Equation (15) can be rewritten for times of maxima:

$$\Omega_C \cdot t + \Delta\Phi \cdot (y_m(t) - y_m(0)) = 2n\pi, \quad n \in \mathbf{Z}. \tag{16}$$

The solutions for equation (16) can be given in the form:

$$t_n = n \cdot \frac{2\pi}{\Omega_C} - \frac{\Delta\Phi}{\Omega_C} \cdot (y_m(t_n) - y_m(0)), \quad n \in \mathbf{Z}. \tag{17}$$

With the same substitution (8) can be written:

$$t_n - n \cdot T_C = -\frac{\Delta\Phi}{\Omega_C} \cdot (y_m(t_n) - y_m(0)), \quad n \in \mathbf{Z}. \tag{18}$$

The last equation gives the deviations of the maxima timing of ΦM signal compared with uniform sampling timings nT_C, and it can be rewritten:

$$\Delta p(t_n) = -\Delta p_{\Phi M} \cdot (y_m(t_n) - y_m(0)), \tag{19}$$

where we used the notations: $\Delta p(t_n) = t_n - n \cdot T_C$, and $\Delta p_{\Phi M} = \Delta\Phi/\Omega_C$. Hence the time series representing moments of maxima for a phase modulated signal (ΦM) has been obtained from a periodic signal δ_T position modulated with modulating signal $\Delta p(t_n)$ given by relation (19):

$$x_{\max \Phi M} = A_C \cdot \sum_{n=-\infty}^{\infty} \delta(t - t_n) = A_C \cdot \sum_{n=-\infty}^{\infty} \delta(t + \Delta p_M \cdot (y_m(t_n) - y_m(0)) - n \cdot T_C). \tag{20}$$

Relations (11) and (20) represent the mathematical form of the theorem that has just been demonstrated.

If a harmonic modulating signal $x_m = A_m cos(\omega_m t)$ is considered and the relations (9) and (11) are particularized, the following results are obtained:

$$\begin{aligned} t_n - n \cdot T_C &= -\frac{k_\Omega A_m}{\Omega_C} \cdot \int_0^{t_n} \cos(\omega_m \tau) d\tau = -\frac{1}{\Omega_C} \cdot \frac{k_\Omega A_m}{\omega_m} \cdot \sin(\omega_m t_n) = \\ &= -\frac{\beta}{\Omega_C} \cdot \sin(\omega_m t_n), \quad n \in \mathbf{Z}, \end{aligned} \tag{21}$$

$$x_{\max FM}(t) = A_C \cdot \sum_{n=-\infty}^{\infty} \delta(t - t_n) = A_C \cdot \sum_{n=-\infty}^{\infty} \delta\left(t + \frac{\beta}{\Omega_C} \cdot \sin(\omega_m t_n) - n \cdot T_C\right). \tag{22}$$

Observations:

1. The modulating ΦM signal is directly represented in PPM, and for the modulating FM signal, its integral corresponds to the PPM signal.
2. The present theorem offers a method of obtaining PPM signals from FM or ΦM signals.

3. Results for astronomical modulators

Observing the timing for extremes of light curves for variable stars represent an important approach for exploration of physical and dynamical mechanisms existing in variable stars. The basic idea is to study the deviations of the observed period for light curves, comparing them with an average preliminary value, during an interval of time lasting as long as possible. In astronomy this kind of studies are usually realized by the classical method of “O-C”.

For a variable star with constant period it is possible to predict the timing for extremes (maxima for pulsating variables or minima for eclipsing variables) with a linear ephemeris:

$$t_n = t_0 + P_0 \cdot n, \quad n \in \mathbf{Z}, \tag{23}$$

where P_0 is the period, t_0 is an initial epoch and n is the number of cycles of variability. If the period is variable, the equation (21) must be rewritten in a more general form:

$$t_n = t_0 + P_0 \cdot n + \tau(n), \quad n \in \mathbf{Z}, \tag{24}$$

where $\tau(n)$ is a function which represents the deviation from linear ephemeris with a constant period to the actual variable period. The problem is to determine an adequate expression for $\tau(n)$, justified by the physical and dynamical model.

The equivalence theorem offers a signal theory background for describing time series for extremes for light curves of variable stars having variable periods with pulse position modulation signals. More precisely the time series of maxima or minima for a variable star in a binary system could be described by a PPM signal with natural sampling, and the modulating signal is associated with Doppler or “light-time” effect induced by the dynamical binary system.

The frequency spectrum for PPM signal is not appropriate for direct interpretations or reconstruction by filtration of the modulating signal due to harmonic distortions. However direct information for the main frequency of the variable star in the binary system can be obtained, which is associated with the carrier frequency for the PPM signal. This main frequency is not priory known in astronomy, unlike in engineering modulation-demodulation systems where a precise value for it is known.

In the classical “O-C” method the main approach is to ameliorate a preliminary value for period using statistical criteria (usually Least Square Method).Using frequency spectrum of PPM signal for the main frequency / period determination is more efficient and precise.

In the demodulation of a PPM signal usually PWM and subsequent PAM intermediate modulation are used [1], [9]. The PAM signal is demodulated by envelope detection.

The method using PPM for star period variability studies consists of three steps.

1. The first step consists of carrier frequency determination (f_0) from the frequency spectrum of the PPM signal associated with the time series of extremes for variable star system.
2. In the second step a PWM signal is calculated, from the original time series $d(t)$ and the periodic Dirac signal $\delta_{P0}(t-t_0)$, with the main period previously determined P_o and an initial epoch t_0. Subsequently the

PWM signal is transformed in a PAM signal with uniform sampling. The amplitudes of pulses of PAM signal correspond to the widths of the pulses in the PWM signal.

3. In the third step the PAM signal is demodulated and the corresponding frequencies in the original modulating signal are found.

A supplementary complication appears in the astronomical time series signals due to the non-uniform data sampling. Inherent to ground-based astronomical observations is the non-continuous data acquisition (daily, monthly and seasonal interval of time during which measurements are taken). In the frequency domain this "windowing" manifests as convolution with the spectrum of the original modulated signal. Consequently we must account for spurious frequencies in the computed spectrum.

The method was applied to the eclipsing binary system X Trianguli. The spectrum for the PPM signal is exemplified in *Fig.1*, from which the carrier frequency: $f_C = 1.029299934$ *cycles/day* is determined. The frequency spectrum for the PAM signal is presented in *Fig.2*. Two frequencies are subsequently subtracted. The PAM signal with two frequencies modeled signal is finally presented in the *Fig.3*.

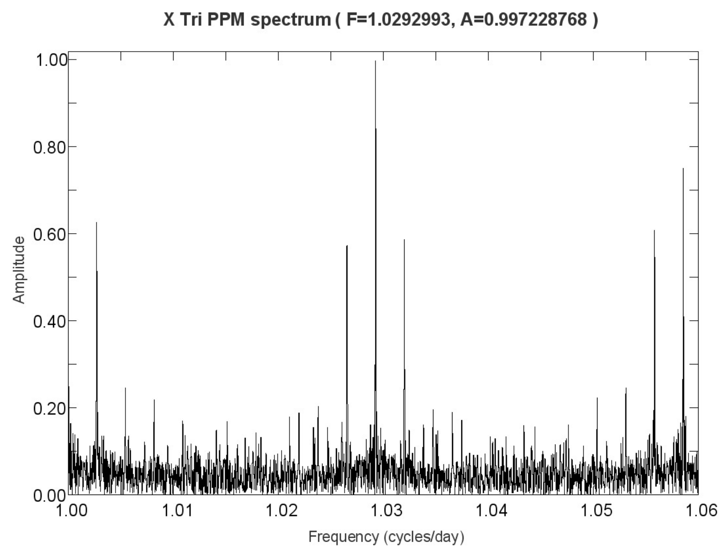


Figure 1: The PPM spectrum detail for X Trianguli system.

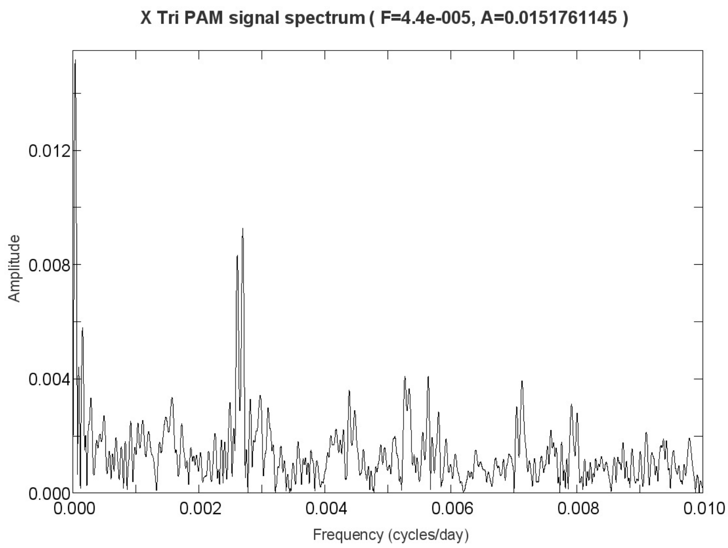


Figure 2: The PAM spectrum detail for X Trianguli system.

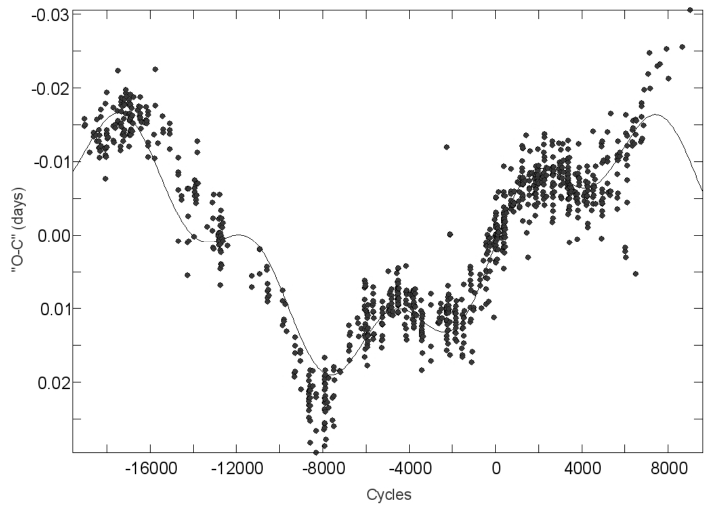


Figure 3: The PAM signal for X Trianguli system, adjusted with two frequencies.

The harmonic modulation for X Trianguli minima time series could be produced by a third body with the periodicity of about P_m = 6225 days = 17 years. Seasonal dispersion subsists in the PAM signal (*Fig.*3) despite of "heliocentric correction" which is applied to every astronomical time series. [10], [11]. Application of this correction determines an incomplete demodulation. The proposed method would operate better on raw astronomical time series (uncorrected data).

Spectral calculus and graphical representations were made using **Period'04** – version 1.0 beta 6.1 [12].

4. Conclusions

The equivalence theorem between extremes of frequency modulated or phase modulated signals and pulse position modulated signals of natural type which was formulated and demonstrated gives theoretical support for practical implementations of modulating and demodulating systems for pulse position modulation. Based on the theorem and on the astronomical modulator concept, a new method for astronomical time series for maxima or minima of variable stars has been given. The proposed method was successfully applied to the X Trianguli binary system. We found a possible periodicity of about 17 years which could imply an unseen third body. The application of this method emphasized the inaccuracies in the traditional "heliocentric correction", which is usually applied to astronomical time series for variable stars.

Acknowledgements

The author expresses his gratitude to the referees for their valuable observations which led to the improvement of the manuscript.

References

1. Spătaru, A., "Teoria transmisiunii informaţiei. Semnale şi perturbaţii", Ed.Tehnică, Bucureşti, 1965. (in romanian)
2. Cartianu, G., "Modulation de fréquence", Ed. Dunod, Paris, 1968.
3. Mateescu, A., "Semnale Circuite şi Sisteme", Ed. Didactică şi Pedagogică, Bucureşti, 1984. (in romanian)
4. Mateescu, A., Dumitriu, N, Stanciu, L., "Semnale şi sisteme. Aplicaţii în filtrarea semnalelor", Ed. Teora, Bucureşti, 2001. (in romanian)
5. Pop, A., Turcu, V., "The Study of Binary Systems with Pulsating Components using the Frequency-Modulation Theory", Astrophysics and Space Science, Kluwer Academic Publishers, vol.204, pp.233-241, 1993.

6. Turcu, V., Pop, A., “On the use of Pulse Position Modulation theory in the study of period variability phenomena”, Interplayof Periodic, Cyclic and Stochastic Variability in Selected Areas of the H-R Diagram, Edited by C.Sterken, ASP Conf.Ser San Francisco, Astron. Soc. of thePacific., vol.292, pp.373-376, 2003.
7. Turcu, V., ”Astronomical Modulators”, The Light-Time Effect in Astrophysics, Causes and cures of the O-C diagram, Edited by C.Sterken, ASP Conf.Ser San Francisco, Astron. Soc. of thePacific., vol.395, pp.69-74, 2005.
8. Turcu, V., „Contribuţii asupra metodelor de achiziţie şi prelucrare ale semnalelor in astronomie”, Teza de doctorat, Universitatea Tehnică din Cluj-Napoca, 2010. (in romanian)
9. Sato, M., Murata, M., Namekawa, T., “Pulse Interval and Width, Modulation for Video Transmission”, IEEE Transaction on Cable Television, 10/1978, pp.165-173, 1978.
10. Henden, A., Kaitchuck, R., “Astronomical Photometry”, Van Nostrand Reinhold Company Inc., New York, NY, 1982.
11. Landolt, A., Blondeau, K., “The Calculations of Heliocentric Corrections”, Pub. Astron.Soc Pacific, vol.84, pp. 784-809, 1972.
12. Lenz, P., “Period’04”, Institute of Astronomy, University of Vienna http://www.univie.ac.at/tops/Perod04, 2004-2006.

Human Activity Recognition for Intelligent Video Lecture Recording

Muhammad Haroon YOUSAF [1], Kanza HAROON [2],
Khubab AHMED [3], Hafiz Adnan HABIB [4]

[1,4] Department of Computer Engineering, Faculty of Telecom and Information Engg.,
University of Engg. & Tech. Taxila, Pakistan,
e-mail: haroon.yousaf@uettaxila.edu.pk , adnan.habib@uettaxila.edu.pk
[2] Department of Computer Science & Engineering
University of Engg. & Tech. Lahore, Pakistan e-mail: kanza.haroon@gmail.com
[3] Department of Electrical Engineering, Faculty of Elect. and Electronics Engg.,
University of Engg. & Tech. Taxila, Pakistan e-mail: khubab@hotmail.com

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: Human Activity Recognition for Intelligent Video Lecture Recording is a project of intelligent systems from e-learning perspective. Research work aims to develop an algorithm for recording video lectures in a classroom environment by analyzing class room scene and gesture of instructor. Such videos are recorded manually which involves considerable time and cost of human personnel. Low level features are used to identify the actions perform by the human and then Non parametric approach like template matching has been used for the human activity recognition. Two cameras are involved in this analysis procedure: Master camera will analyze class room scene and after activity recognition localization information and intended area of interest is passed to Slave camera mounted on a moving platform (two degree of freedom) which capture the video accordingly.

Keywords: Intelligent Systems, Digital Image Processing, Activity Recognition, Human Gestures, Video Recording.

1. General information

Gesture recognition is the technology of transforming human body parts movements into some meaningful information. Gesture recognition technology has been successfully implemented in various research oriented applications [1]: sign language understanding, human computer interaction, human activity recognition etc. Human body movements are, partially or wholly analyzed and interpreted into meaningful information. This meaningful information can be used in some type of decision making based upon the activities performed in that moments made by the humans. Gesture recognition is divided in two major parts: feature detection and recognition. Feature detection step extracts some meaningful information from the acquired gesture. These features are further used in the next stage for the recognition purpose. Various techniques are used for the gesture recognition phase, e.g. Template Matching, Statistical, neural network and hidden Markov models. Naturally gestures are of two types: Static and Dynamic. Static gestures are also called as postures. However dynamic gestures usually prolong in time. Information available in the dynamic gestures is in the form of video sequences. Another variable added in dynamic gestures is the time Therefore, time is also considered as a major factor in representation of dynamic gestures. This is the reason why we called dynamic gestures as temporal gestures.

2. Problem Formulation

Video lecture recording is a major activity now a day especially for the distance learning education. Currently this job is performed by a human being (*Fig. 1*), who analyzes the instructor activity and record the video of particular area. Major disadvantage is the labor cost involved in this job. Automated camera management system has already been discussed in past [2]. Research work aims to automate this activity by capturing the video lecture in a classroom by analyzing class room scene and gesture of instructor. Two cameras are involved in this analysis procedure: Master camera will analyze class room scene and gesture of the instructor to make a decision about area of interest i.e. area where instructor is focusing, students, white board or power point presentation area. Video captured by first camera localizes the area of interest and decides about the intended field of view (FOV). Localization information and intended FOV is passed to second camera which capture the video accordingly. *Fig. 2* illustrates the framework for the automated system for the video lecture recording in a classroom environment.

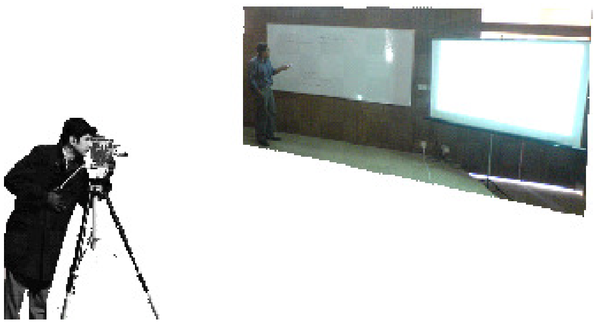


Figure 1: Human being analyzing the classroom scene.

A. Environment

The real time environment contains the following:

- Instructor,
- White Board,
- Presentation Area.

These areas will be registered in the registration phase. A graph representation based scene representation will be developed comprising instructor, white board and presentation area. Presentation area and white board are static nodes while instructor is dynamic in terms of location and relationship. This class room graph representation will be frequently utilized in scene analysis and decision making.

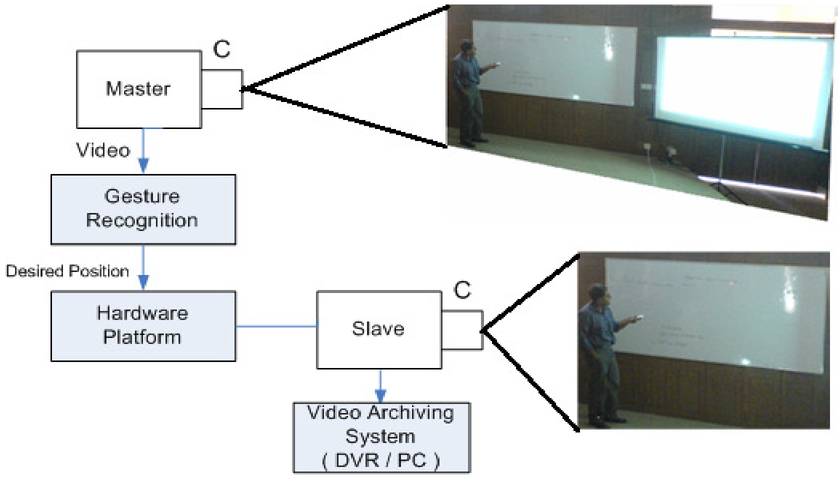


Figure 2: Gesture recognition based system.

Tables should be numbered. Table captions should be aligned to left and placed above the table.

3. Problem Solution

Fig. 3 describes the detailed algorithmic architecture for the above said problem. Knowledge base is used to save the information about the reference frame of the classroom scene.

A. Classroom Scene Analysis

Live video stream is processed for instructor activity recognition. Second task is to convert the video stream into frames. When the system is ON, the first frame will take as the reference frame for all the frames. First frame of master video is processed to locate the position of the white board and power point presentation area. First of all process is carried out to recognize white board and presentation area then storing their position in the knowledge base. This is further divided into two steps.

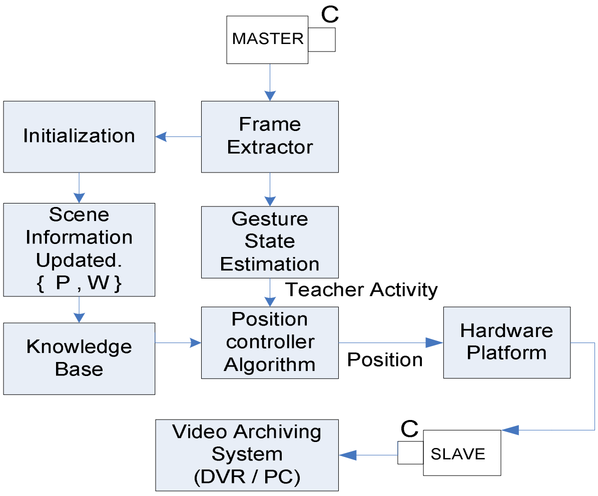


Figure 3: Algorithmic architecture.

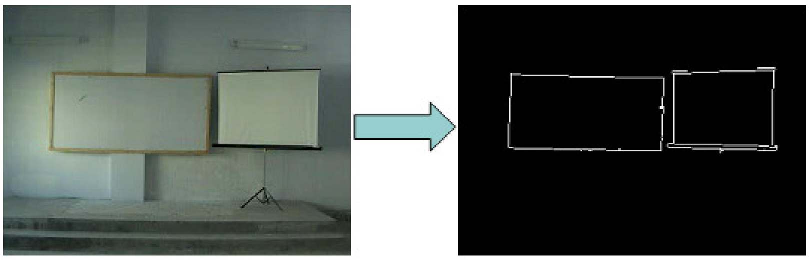


Figure 4: Scene segmentation.

Scene Segmentation: Scene segmentation is carried out by using edge detection using canny edge operator and some morphological operations, thus detecting the white board and presentation area in the scene. But it is not clear till now that which one of these two objects is white board and which one is presentation area. Experimental results of scene segmentation are shown in *Fig. 4*.

Region Labeling: In this step we recognize the both of the objects. Region having more aspect ratio is registered as white board and other is presentation area. We store the position of white board and presentation area for future use in camera coordinate mapping system.

B. Gesture Recognition

In gesture recognition stage, we have to detect the human being in the classroom scene and then recognize the activity. Three types of activities we have to recognize:

- Activity 1: Teacher is addressing to the students;
- Activity 2: Teacher is writing on the white board;
- Activity 3: Teacher is pointing on the board / presentation area.

Gesture recognition stage is further divided into two steps that are as follows.

Teacher Detection: To detect the teacher in the classroom scene, background subtraction scheme is used. In which absolute difference is taken b/w reference frame and the real time tested frame. The difference b/w two frames show the teacher in the scene. Example of teacher detection is shown in *Fig. 5*. After detecting teacher, we extract the area from the frame that is occupied by the teacher and then converting that area into binary for further processing.

Activity Recognition: Teacher activity recognition is carried out by u5sing template matching algorithm, for this purpose we have already stored some templates in the knowledge base. Templates used for activity recognition are shown in *Fig. 6*. For template matching we first used simple correlation process, if *C* is the teacher extracted from current frame and *T* is the template image, the correlation b/w these two will be carried by following equation

$$R = \sum_{I=1}^{r} \sum_{J=1}^{c} C(I,J)T(I,J) \tag{1}$$

where *r* and *c* is the number of rows and columns.

But to improve the efficiency of the system we have modified the correlation method. Correlation process computes the similarity b/w both, template and the teacher extracted from current frame. But in our system it didn't perform well, so to improve the efficiency we also compute the dissimilarity b/w template image and the current frame, this is carried as follows.

$$R = \sum_{I=1}^{r} \sum_{J=1}^{c} \left(C(I,J)T(I,J) - C(I,J)^{\wedge} T(I,J) \right) \tag{2}$$

where ^ shows.the XOR operation.

Now value of R will decide the teacher activity. Template that has greater R value will be the best match to the current frame and the activity shown in the template will be the activity of the teacher.

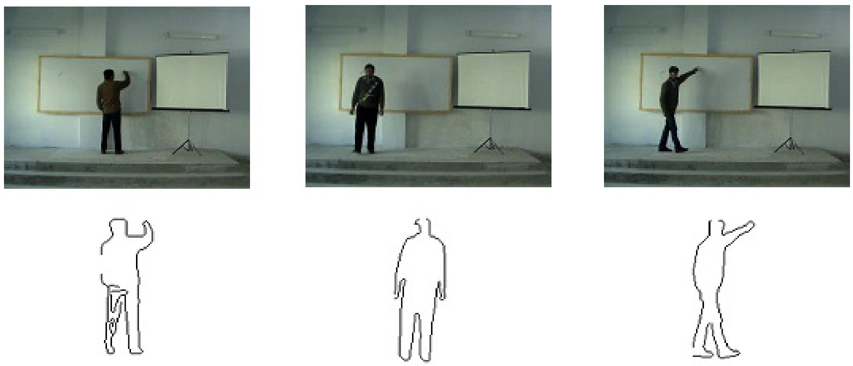


Figure 5: Instructor detection.

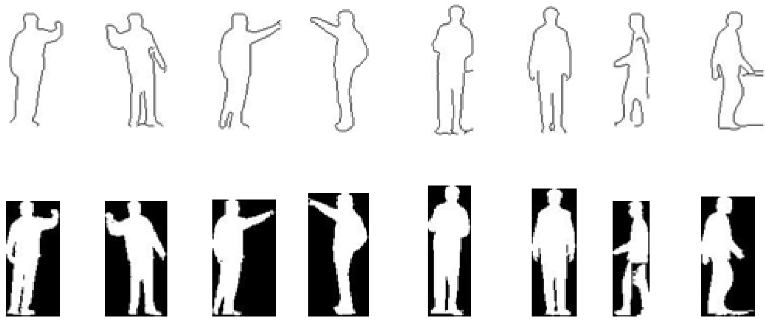


Figure 6: Template matching.

C. Experimental Setup

Dedicated hardware platform is designed to automate the lecture recording activity. Hardware platform has mainly two parts: First is the fixed part of the platform, on which the master camera (CCD Camera) is mounted. Master camera video is used for the classroom scene analysis. Second is moving part of the platform having two degree of freedom, on which slave camera (Sony Handycam) is mounted to capture the video of intended area of interest. Recorded video is then stored in any video archiving system. Video processing has been performed in Matlab using the Image Acquisition and Image

Processing Toolboxes. Localization information for the intended area of interest has been communicated to the driving circuitry of the hardware platform through parallel port. Driving circuitry for the hardware platform has been designed using Microcontroller 80C51.

D. Platform Coordinate Mapping

This intended area is mapped from master camera coordinates into slave camera coordinates assuming the following condition [4]:

$$S(x, y, t) \in M(x, y, t) \tag{3}$$

where $S(x, y, t)$ represents the video acquired by the slave camera and $M(x, y, t)$ represents the video acquired by the master camera. x and y represents the pixel coordinates and t represents the time axis. Above equation describes that pixel information content of slave video will always be a subset of pixel information of master video. Master camera video is analyzed and intended area of scene is selected based upon gesture recognition and scene analysis. Coordinates of selected area in the master video is recorded.

4. Results

Using the experimental setup and the above mentioned methodology for the human activity recognition, Performance of the sub-units of the system is analyzed by the percentage of successful execution of those units. Sub-units are as follows:

1. White board detection,
2. Presentation area detection,
3. Teacher detection.

By using the proposed methodology all of the sub-units work efficiently. System accurately detected the white board, presentation area and teacher. Algorithm is applied on variety of different video dataset and in each of these video, sub-units performance is approximately 100 %.

Performance analysis for the human activity recognition in the classroom environment is represented using the confusion matrix. We have already categorized three major activities of the instructor in the classroom.

In gesture recognition stage, we have to detect the human being in the classroom scene and then recognize the activity. Three types of activities we have to recognize:

1. Writing on the whiteboard,
2. Pointing on the board or presentation area,

3. Addressing to the students.

Tables 1-3 show the confusion matrices for the instructor activity recognition in three different videos.

Table 1: *Activity recognition confusion matrix for Video 1*

Activity / Result	Writing on the Board (100)	Pointing on the Board / Presentation (100)	Addressing to the Students (100)
Writing on the board	87	10	04
Pointing on Board / Presentation	09	83	02
Addressing to the Students	04	07	94

Table 2: *Activity recognition confusion matrix for Video 2*

Activity / Result	Writing on the Board (100)	Pointing on the Board / Presentation (100)	Addressing to the Students (100)
Writing on the board	84	14	05
Pointing on Board / Presentation	11	78	03
Addressing to the Students	05	08	92

Table 3: *Activity recognition confusion matrix for Video 3*

Activity / Result	Writing on the Board (100)	Pointing on the Board / Presentation (100)	Addressing to the Students (100)
Writing on the board	87	14	03
Pointing on Board / Presentation	07	81	04
Addressing to the Students	06	05	93

5. Conclusion and Future Work

Video lecture recording is performed intelligently. White board and presentation area are detected accurately. Teacher activities are also recognized; template matching algorithm works efficiently for human gesture recognition. The region of interest suggested by the algorithm is captured by slave camera mounted on the moving platform. By implementing this system in a practical environment human being or cameraman labor cost can be saved. Here the system is proposed for class room lecture recording. However, in future this concept may be tested for various situations in professional, industrial, and real time image / video acquisition.

Acknowledgements

This work has been funded by Video and Image Processing Laboratory (VIP Lab) and Advance Studies, Research and Technological Division (ASR & TD) UET Taxila.

References

1. Watson, R., "A survey of Gesture Recognition Techniques", Technical Report TCD-CS-93-11, July 1993.
2. Liu, Q., Rui, Y., Gupta, A., and Cadiz, J.J., "Automatic Camera Management for Lecture Room Environment", in Proc. International Conference on Human Factors in Computer Systems (SIGCHI 2001), pp. 1-8, 2001
3. Hsuan M, and Ahuja, N., "Face Detection and Gesture Recognition for Human-Computer Interaction" Springer, 2001.
4. Habib H.A., Yousaf, M.H., and Mufti, M. "Gesture and Scene Recognition based Autonomous Video Lecture Acquisition Framework for Distance Learning", Proc. International Conference on Distance Learning and Web Engineering, Lisbon, Portugal, pp. 97-100, 2006.
5. Ren, H., and Xu, G., "Human Action Recognition in Smart Classroom", Proc. 5th IEEE International conference on Automatic Face and Gesture Recognition (FGR'02), pp. 417-422, 2002.
6. Wang, F., Ngo, C.W., and Pong, T.C., "Gesture Tracking and Recognition for Lecture Video Editing", Proc. 17th International Conference on Pattern Recognition (ICPR'04), vol. 3. pp. 934-937, 2004.

Gujarati Handwriting Recognition Using 2-D Correlation and Concatenation Tree

Hemant JUNEJA[1], Mansi GULATI[2], Karan TANEJA[3]
Under the guidance of
Dr. T.V. PRASAD[4], Rina PATEL[5]

[1] Department of Computer Science, Lingaya's Institute of Management and Technology, India, e-mail: hemantjuneja4u@gmail.com
[2] Department of Computer Science, Lingaya's Institute of Management and Technology, India, e-mail: mansi_gulati2001@yahoo.com
[3]Department of Computer Science, Lingaya's Institute of Management and Technology, India, e-mail: kaayyy@gmail.com
[4]Department of Computer Science, Head of Department, Lingaya's Institute of Management and Technology, India.
[5]Department of Computer Science, Faculty, Lingaya's Institute of Management and Technology, India.

Manuscript received April 02, 2010; revised April 18, 2010.

Abstract: In this paper, we present a technique to detect and recognize gujarati character. Segmenting and recognizing gujarati characters (aksharas) is a challenging task because of the variations in the script and inter-class similarity. The OCR system extracts text zone and segments the text into symbols. The symbols are recognized with the help of 2-D correlation method. Since some characters in the script can be formed from two or more symbols, we are making use of concatenation tree to combine the ropes of recognized symbols to form a single character.

Keywords: Gujarati script; Segmentation; image processing; concatenation tree; 2-D correlation; Post-processing.

1. Introduction

Optical Character Recognition (OCR) is a process of converting scanned documents into machine editable text. Prior to Optical Character Recognition, if someone wanted to turn a book into a word processing file, each page would have to be typed word for word. But With the help of OCR, this process has become almost automatic. OCR is widely used to convert books and documents into electronic or word processing files making them fully searchable and editable documents and thus find applications in most of the fields like banking, legal, healthcare etc. It can also help a blind to read.

Any OCR consists of five major stages:

1. Pre-processing
2. Segmentation
3. Feature Extraction
4. Classification
5. Post processing

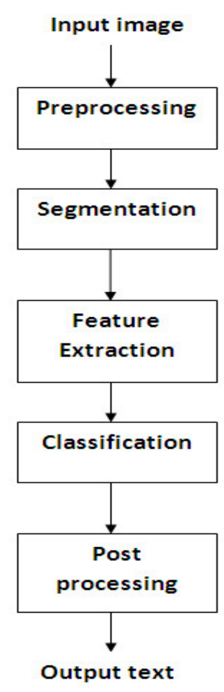


Figure 1: Block diagram of OCR.

The preprocessing phase (*Fig. 1*) includes binarization of the input document image, noise removal, normalization and smoothing which further includes thinning and filling. In the segmentation stage, characters or words are isolated at once or in different stages. In the feature extraction stage, some essential features of the symbols are extracted. The feature extraction techniques are often divided into three main groups [1], where the features are found from:

1. The distribution of points,
2. Transformations and series expansions and
3. Structural analysis

In the classification stage, each character is identified and is assigned to the correct character class. The post-processing phase includes grouping of the symbols into strings and detection and correction of errors made in recognition. Very less work has been done in developing an OCR for Gujarati script [2], [3], [4].

2. Gujarati script

Gujarati script is descended from Brahmi and is part of the Brahmic family. The Gujarati script was adapted from the Devanagari script and is used to write the Gujarati language spoken by about 50 million people in the western part of India and also in Bangladesh, Fiji, Kenya, Malawi, Mauritius, Oman, Pakistan, Singapore, South Africa, Tanzania, Uganda, United Kingdom, USA, Zambia and Zimbabwe.

Figure 2: Gujarati Script

In accordance with all the other Indic scripts, Gujarati is written from left to right, and is not case-sensitive. Gujarati script (*Fig. 2*) has 12 vowels and 34 Consonants. Vowel extensions shown in the figure called 'matras' denote the attachment of vowels to the consonants which can appear above, below, before or after the consonant. Gujarati script also uses consonant clusters called 'conjucts' and constant-vowel combinations. More detailed description on Gujarati script is given in the references [5], [6].

A. Challenges in the script

- Many characters in Gujarati and Devanagari scripts appear similar in shape but the major difference between the two scripts is the lack of the top horizontal bar (*Fig. 3*) in Gujarati, thus Gujarati requires a different treatment from the other Indo – Aryan scripts.

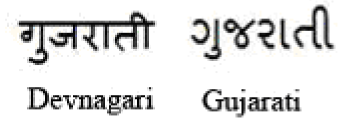


Figure 3: Gujarati vs. Devanagari.

- In Gujarati scripts, there is intrinsically high inter-class similarity between some pairs of symbols.

Figure 4: Similar Characters in Gujarati.

3. Approach for making Gujarati OCR

The Gujarati characters are recognized by the method proposed in the *Fig. 5*.

A. Preprocessing and Segmentation

In preprocessing, Skew detection & correction is done since sometimes pages are scanned crookedly. Thus there is a need to align the skewed text before further processing is done. After character segmentation and preprocessing of image, 2-D matrices of characters are classified into different classes (1 class for each character). Various samples of each character are taken and placed into their corresponding character classes. Thus, classes of all symbols which can be segmented and identified are formed and datasets for each class are created (about 50 samples for each class).

The input document image is first scanned. Segmentation of lines, words and characters is done resulting in the isolation of a character for recognition. The line segmentation is done by clipping the first row from the document. This is followed by word segmentation and then the character segmentation for the line clipped. The document then contains one less the number of lines present in the original document. The remaining lines are also clipped by repeating the above procedure till segmentation of all the lines containing words is achieved. This procedure can be shown as follows (*Fig. 6*):

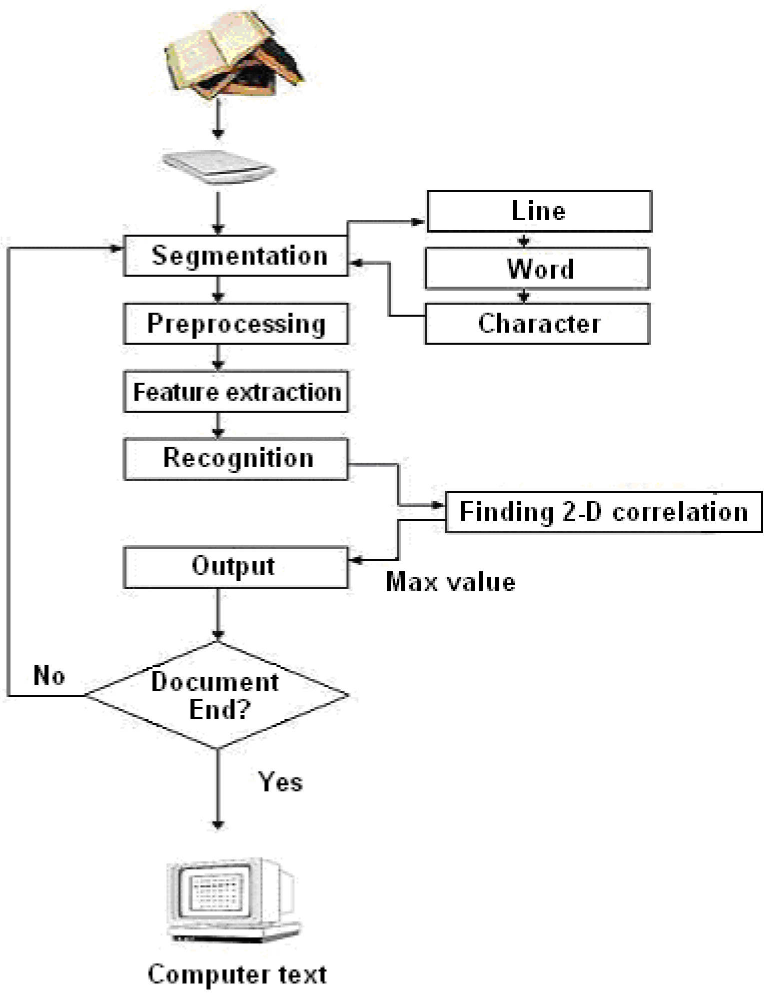


Figure. 5: Basic methodology.

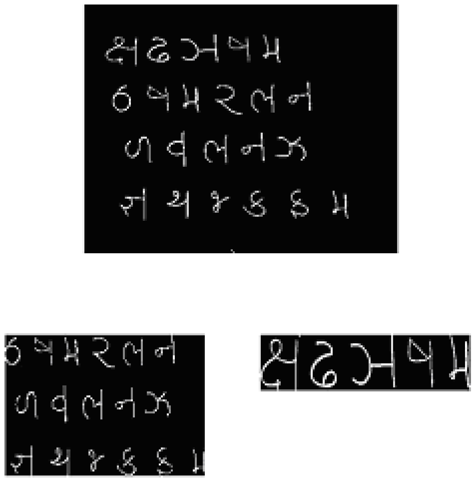


Figure 6: Preprocessing and Line Segmentation.

B. Recognition Using 2-D Correlation Coefficient

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. For example, height and weight are related; taller people tend to be heavier than shorter people. The relationship isn't perfect. People of the same height vary in weight, and we can easily think of two people we know where the shorter one is heavier than the taller one. Nonetheless, the average weight of people 5'5" is less than the average weight of people 5'6", and their average weight is less than that of people 5'7", etc. Hence, correlation can tell, just how much of the variation in people’s weights is related to their heights.

For recognizing the input image, the value of correlation coefficient (fig. 7) between the input image and each sample in each class is calculated. The maximum coefficient value from each class becomes the output of that class. The values are calculated by the following,

Formula used: A and B are 2-D matrices:

$$r = \frac{\sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \overline{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \overline{B})^2\right)}}$$

$\overline{A}$ = `mean2(A)`, $\overline{B}$ = `mean2(B)`, `r` is a scalar `double` vectors of the same size.

The main result of a correlation is called the correlation coefficient (or "r"), which ranges from 0 to +1.0. If r is close to 0, this implies that there is no relationship between the images being compared. But if the value of correlation coefficient is close to +1, then the two images are said to be related.

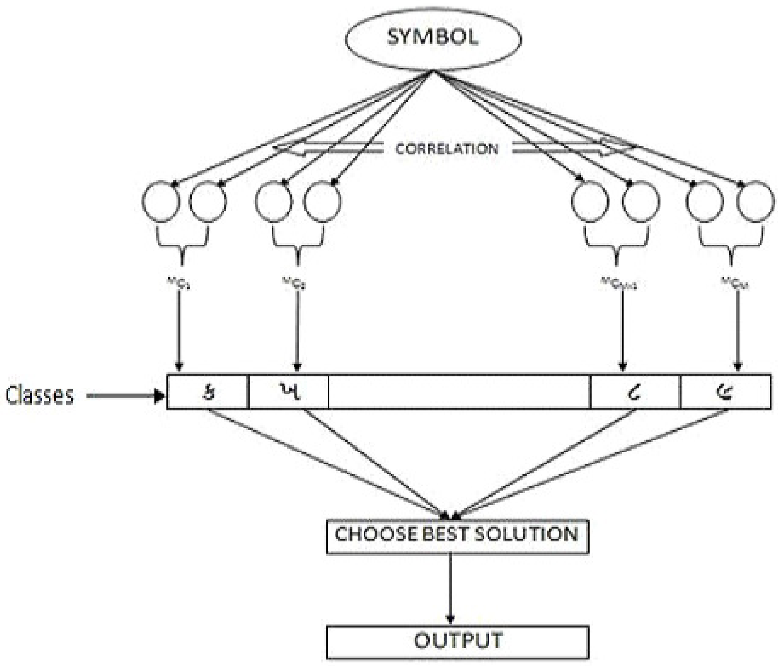


Figure 7: Calculation of correlation values

The class with maximum value of the correlation coefficient close to +1.0 is the final output for the given image.

C. Combining Symbols Using Concatenation Tree

A rope is a heavyweight string, involving the use of a concatenation tree representation [7]. It is essentially a tree whose leaves are arrays of characters. A node in the tree represents the recognized symbols. Concatenation of two ropes simply involves the creation of a new tree node with ropes as children. This is illustrated in *Fig. 8*. Here, the leaf nodes form the basis of building the concatenation tree for each character. The main advantages of ropes as compared to storing strings as character arrays are that they enable much faster concatenation than ordinary strings and don't require a large contiguous memory space for storing a large string. The main disadvantages are greater overall space usage and slower indexing, both of which becomes more severe as the tree structure becomes larger and deeper. However, many practical applications of indexing involve only iteration over the string, which remains fast as long as the leaf nodes are large enough to benefit from cache effects.

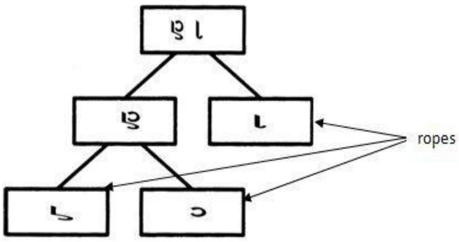
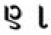


Figure 8:  character.

4. Conclusions

A novel technique for recognition of Gujarati characters using 2-D correlation has been described. This approach can be further extended to other scripts, since this is a script-independent technique. This paper has highlighted advantages and disadvantages of ropes, which will help developers to further explore ropes. In the entire process, the toughest challenges are faced in recognizing similar looking Gujarati characters. Also due to the variations in the handwriting of the people, problems occur in the recognition. However, this method produces an accuracy of 92.8% for characters but for similar characters it is 70.2%. There is still some work required to build a suitable post-processor to enhance the overall process. Use of Natural Language Processing (NLP) with dictionary could lead to an improvement in the performance of the system. Hence, it can be concluded that despite a delayed start in the research work of Gujarati OCR development, the area is gaining attention now.

References

1. Eikvil, L., “OCR Optical Character Recognition”, Dec. 1993, available at: http://www.nr.no /~eikvil/OCR.pdf
2. Antani, S, and Agnihotri, L., ” Gujarati character recognition”, in Proc. 5th International Conference on Document Analysis and Recognition, pp. 418–421, 1999.

3. Shah, S.K., and Sharma, A., "Design and Implementation of Optical Character Recognition System to Recognize Gujarati Script using Template Matching", IE(I) Journal-ET, vol. 86, pp. 44-49, Jan. 2006,
4. Dholakia, J., Negi, A., and Mohan S.R., "Progress in Gujarati Document Processing and Character Recognition", in: Govindaraju, V, and Setlur, S. (eds), "Guide to OCR for Indic Scripts. Document Recognition and Retrieval", Springer London, pp. 73-95, 2009.
5. http://tdil.mit.gov.in/GujaratiScriptDetailsApr02.pdf
6. http://www.statemaster.com/encyclopedia/Gujarati-script
7. Boehm, H.J., Atkinson, R., and Plass, M., "Ropes: An alternative to strings", Software: Practice and Experience, vol. 25, no. 12, pp. 1315-1330, Dec. 1995.
8. http://en.wikipedia.org/wiki/Rope_(computer_science)

Automation and electrical drives

Performance Curves of Starter Based on Mathematical Model

Csaba BLÁGA[1], Ernő KOVÁCS[2]

[1] Department of Electrical Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, e-mail: elkblaga@uni-miskolc.hu
[2] Department of Electrical Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, e-mail: elkke@uni-miskolc.hu

Manuscript received April 04, 2010; revised May 01, 2010.

Abstract: This paper presents a simulation model of a starter and its performance curves. At the terminal of every manufacturing line of starters there is a test stand to measure the performance curves of the starter. This provides important parameters for the validation of a starter. These curves cover the following functions: torque, rpm, voltage, output power, input power, and efficiency versus current. The paper introduces a mathematical model of the starter and its theoretical background. The model is validated through available characteristics known from literature and industrial datasheets.

Keywords: automotive electric and electronic systems, starter, simulation, performance curve.

1. Introduction

At the terminal of every manufacturing process line of starters there is test stand to measure the most characteristics performance curves of the starter. This gives important parameters into the validation process of starter. The parameters should be read from the performance curves of starter. These curves are the following functions versus current:

- torque, rpm, voltage, output power, input power, and efficiency.

These quantities are important for us. An internal combustion engine can not start itself, so it needs an outer starting system. The torque to rotate the main shaft of the internal combustion engine is quit big and depends on many parameters, such as: engine cycle type - Otto or Diesel -, volume of engine, number of cylinders, value of compression, type and quality of lubricant, temperature, etc.

On the other side the internal combustion engine needs a minimal rotation speed to be able to maintain its operation. The voltage is supplied by battery which current can be very high during the starting process, but the battery has to keep its terminal voltage over a certain level. In other case not only the starter will not work, but it might occur that the battery will be damaged. In addition the starter should have small size and small weight. This can be realized due its very short time of operation. It can be said that a starter is under dimensioned. It is not able to operate continuously or for a longer period because it will "burn down".

The battery supplies DC voltage so the easiest solution is to choose a DC motor as starter. There are different kinds of DC motors. The series DC motor has the most convenient mechanical characteristic (rpm versus torque) to satisfy the needs of smooth starting process. Nowadays in the case of smaller internal combustion engine we can find permanent magnet DC motors, too. Moreover there are attempts to combine the starter and the generator. Couple of years ago it operated on the base of shunt DC motor - dynamo, but nowadays there is a high advance given by AC synchronous motors - generators. We can meet starter-generators built in the so called micro-hybrid vehicles which main function is the start-stop system. Interesting solution is given by ISAD (Integrated Starter Alternator Damper) system. It can be mentioned as new solutions the total elimination of the starter that can be done in the case of direct fuel stratified injection (FSI) engines.

2. Mathematical model of series DC starter

Building up the mathematical model starts from the main two relations of DC motors:

$$M = k \cdot \phi \cdot I_a \tag{1}$$

$$V_i = k \cdot \phi \cdot \omega \tag{2}$$

where M = torque [Nm], k = constant relevant to DC motor,

ϕ = magnetic flux [Wb], I_a = current of armature [A],

V_i = induced voltage of armature [V], ω = speed of armature [rad/s].

The other relations result from the electrical circuit from *Fig.1*, where:

V = voltage [V], I_a = current of armature [A],

V_i = induced voltage of armature [V], I_e = excitation current [A],

R_a = resistance of armature [Ω], R_e = resistance of excitation [Ω].

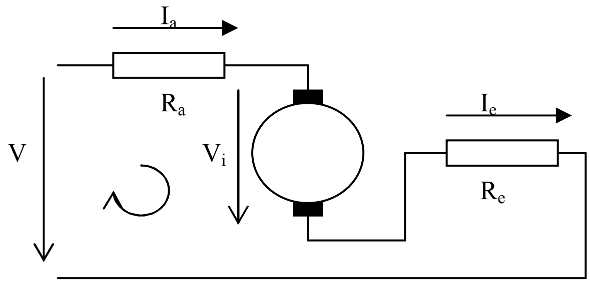


Figure 1: Electric circuit of series DC motor.

Due to the 2nd law of Kirchhoff it can be written:

$$-V + I_a R_a + V_i + I_e R_e = 0\,. \tag{3}$$

Due to the series circuit it can be written:

$$I_a = I_e = I\,. \tag{4}$$

It is known that the flux is given by the excitation current:

$$\phi = \chi \cdot I_e \tag{5}$$

where χ is a relation factor.

Arranging the system of equations given by relations (1)-(5) we can get the needed functions:

$$M = K \cdot I^2 \tag{6}$$

$$\omega = \frac{V}{K \cdot I} - \frac{R}{K} \tag{7}$$

$$P_{out} = V \cdot I - R \cdot I^2 \tag{8}$$

$$\eta = 1 - \frac{R \cdot I}{V}\,, \tag{9}$$

where $K = k \cdot \chi$, $I = I_a$ and $R = R_a + R_e$.

Instead of rotation speed ω it is more usual to use rpm noted n. The relation between them is very simple

$$\omega = \frac{\pi \cdot n}{30} \tag{10}$$

and

$$n = \frac{V}{C \cdot I} - \frac{R}{C}, \tag{11}$$

where

$$C = \frac{\pi \cdot K}{30}. \tag{12}$$

A starting idea was given on the base of what the main shape of these curves can be drawn up:

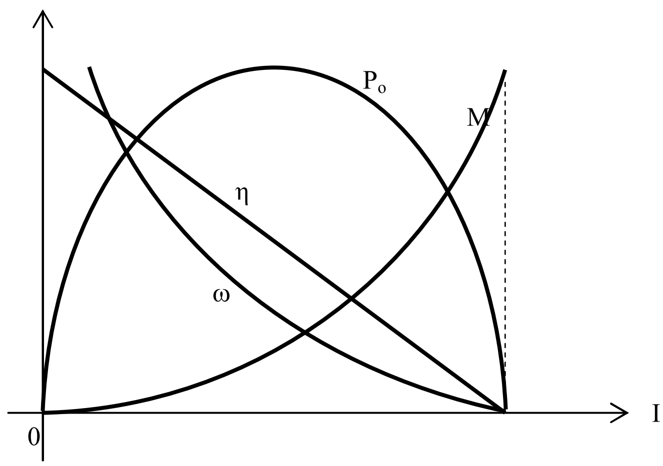


Figure 2: Theoretical performance curves of series DC starter.

3. Simulation model of series DC starter

The presented relations are still very theoretical relations, because there are some parameters neglected as the internal resistance of battery that supplies the electric power, the iron losses of armature, the friction and ventilation losses of rotor, saturation of magnetic flux, etc. If we would like to take into consideration all parameters that would complicate too much the mathematical relations. But if we would like to understand better the behaviour of starter and influence of parameters we should do some simulation. Such a model was built up as it is shown in *Fig. 3*.

The simulation is run in such a way that in the beginning the series DC starter will run without load and from a certain moment the load torque will start to increase. Several scopes are used to be able to draw the different functions versus current. The results of simulation are presented in *Fig.5* and *Fig.6*.

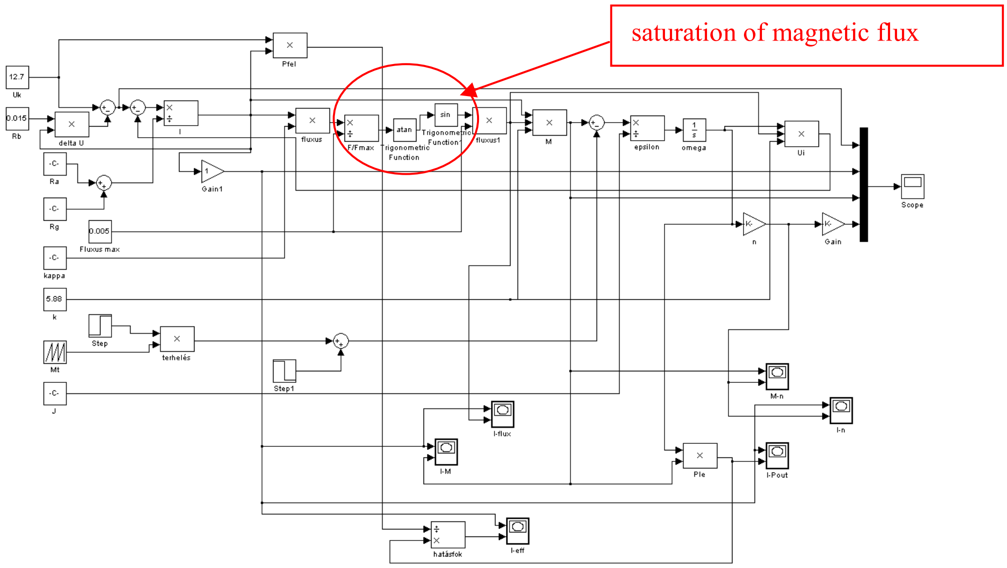


Figure 3: Simulation model of series DC starter.

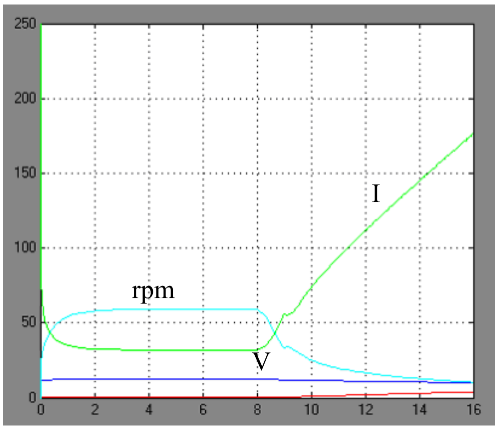


Figure 4: Result of time base simulation, the resulted rotation speed was divided by 100 in order to be presented in the same diagram.

The simulation model of series DC motor presented in *Fig. 3.* also contains the saturation of magnetic flux. So there is no more a linear relation between the flux and excitation current as it was written in relation (5), but a non-linear one (see *Fig.5*).

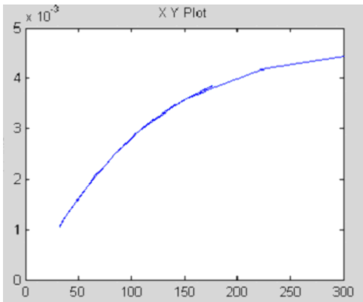


Figure 5: Flux versus current.

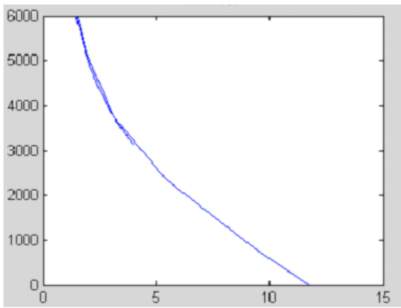


Figure 6: Rpm versus torque.

This means that the flux grows less and less as the excitation current increases and there is a certain value that can not be passed. The relation used for describing the relation between the magnetic flux and the excitation current is given in literature [3].

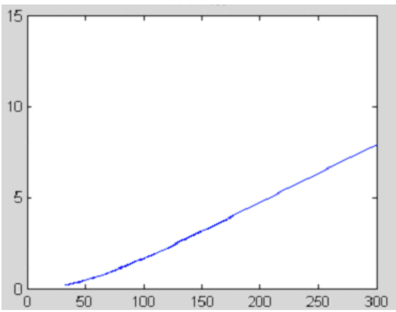


Figure 7: Torque versus current.

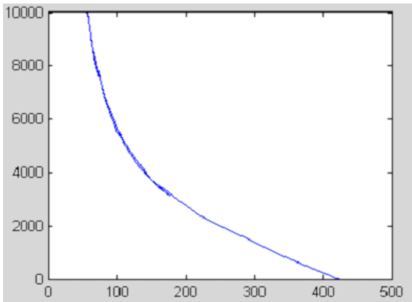


Figure 8: Rpm versus current.

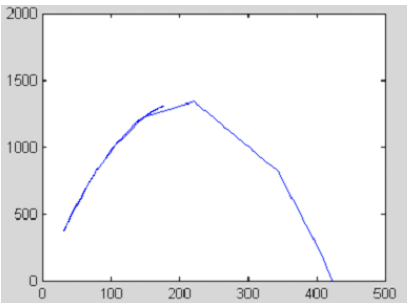


Figure 9: Output power versus current.

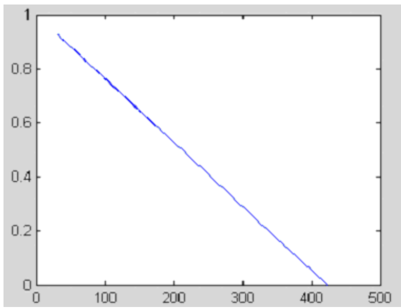


Figure 10: Efficiency versus current.

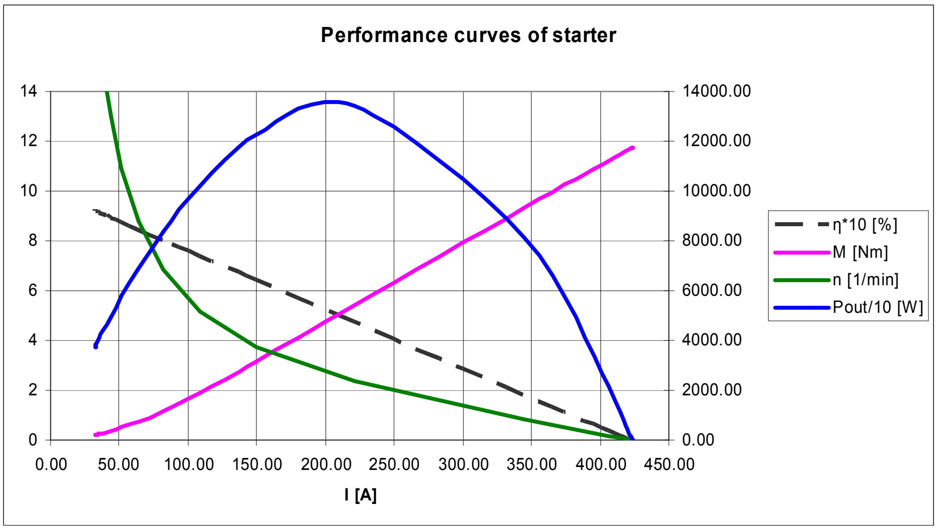


Figure 11: Performance curves of a starter by simulation efficiency (η) and torque (M) - gradation at left vertical axis rpm (n) and output power (P_{out}) - gradation at right vertical axis.

A next step should be the validation of the model. This paper will not present a detailed validation, only compares the behaviour of simulation model to those known from literature and industrial measurements presented in *Fig. 12-14*. It can be observed the characteristics provided by the simulation and the known characteristics show similar behaviour.

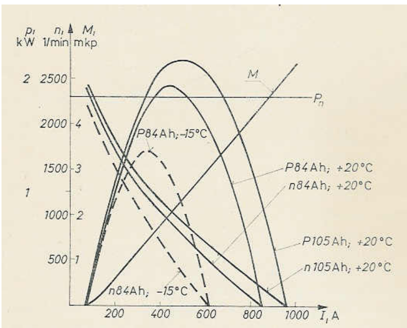


Figure 12: Performance curves in [1].

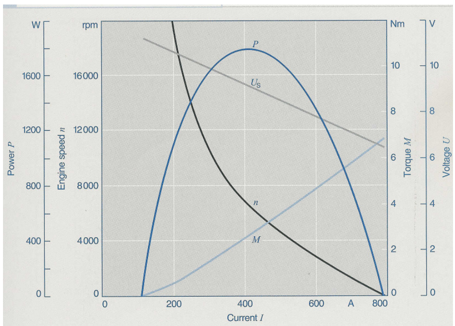


Figure 13: Performance curves in [2].

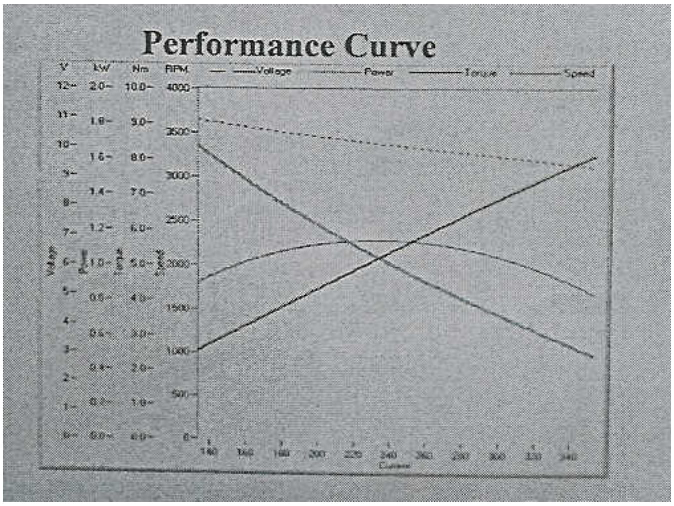


Figure 14: Performance curves at industrial test measurement.

4. Conclusions

The mathematical and simulation model of the series DC stater was built and the performance curves could be simulated. The model which contains some nonlinearity was fine tuned and validated by known and measured characteristics. It can be observed that the resulted performance curves from simulation (*Fig. 11.*) correspond to those written in theory (*Fig. 12.* and *Fig. 13.*) and practice (*Fig. 14.*).

Acknowledgements

The authors highly appreciate the generous support of the BAROSS_EM07-EM_ITN3_07-2008-0039 project of National Office of Research and Technology.

References

[1] Bakos, I., "Járművillamosság", Műszaki Könyvkiadó, Budapest, 1979.

[2] Bauer, H., "Alternators and starter motors", *Robert Bosch GmbH*, Stuttgart, 2003.

[3] Blága, Cs., "Munkásság tézisszerű összefoglalása", *a Miskolci Egyetem Doktori (PhD) Tézisfüzetei, Gépészmérnöki Kar Doktori Tanácsa*, Miskolc, 1997.

Measurement And Identification of a Starter Motor System

Viktor FÜVESI[1], Ernő KOVÁCS[2], Csaba BLÁGA[3]

[1] Department of Electrical and Electronic Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc, e-mail: elkfv@uni-miskolc.hu

[2] Department of Electrical and Electronic Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc, e-mail: elkke@uni-miskolc.hu

[3] Department of Electrical and Electronic Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc, e-mail: elkblaga@uni-miskolc.hu

Manuscript received April 14, 2010; revised May 01, 2010.

Abstract: In this paper the investigation of a starter motor is presented. First the mathematical model is set and analysed. A measurement test rig was designed to make some measurement on the motor. A validation process was used to validate the mathematical model of the system.

Keywords: Starter motor, mathematical model, validation.

1. Introduction

The main energy source of the vehicles on the roads is the internal combustion engine. The combustion engine is not able to start alone its operation. Starter motors are used to raise the rotation of the engine from zero to a certain velocity level, where the combustion engine can maintain its own operation. In the starting process the shaft of the starter is connected to the engine with a gear. A solenoid is used to drag the gears of the starter motor to gear of the engine. After the engine reached the necessary rpm level, the starter DC motor is finished its operation. After the short starting process the solenoid releases the connection between the motor and the engine is bake off.

The power source of the starter motor is the battery, which supplies starter motor with a constant voltage level. The armature current of the motor, meanwhile the starting operation, can be very high because the moment of inertia of the engine is high. The starter motor can burn down if the starting process takes longer than it was designed.

One of the easiest solutions is to choose a DC motor for starter. There are more different types of DC motors but in the vehicles serial wound or permanent magnet DC motors are used commonly, because of their good characteristics [1]. In the next section a series DC motor type will be mathematically investigated.

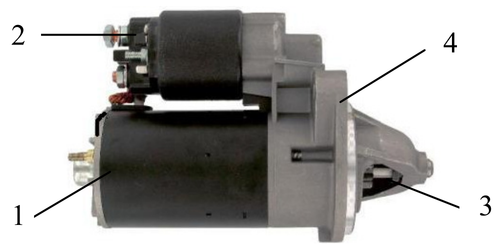


Figure 1: Main components of the starter motor
1 - starter motor, 2 -solenoid, 3 - gear, 4 - housing.

2. Mathematical model of the starter

The electric model of the system can be seen in Fig. 2.

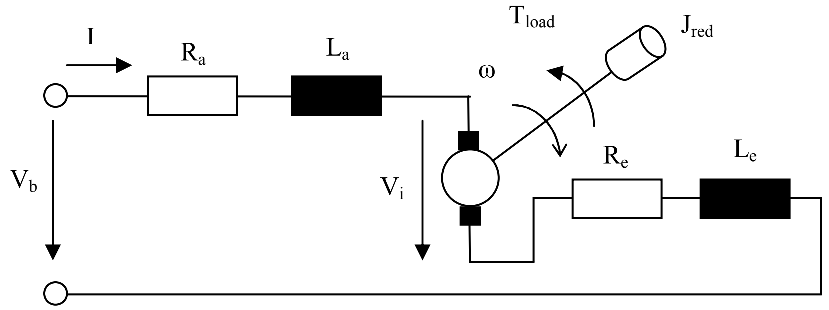


Figure 2: Schematic of the system.

It is a serial wound DC motor model, which is well known from the literature. Using Kirchhoff 2nd equation, the armature equation (1) can be written as

$$\frac{dI}{dt} = \left(V_b - (R_a + R_e)I - c\Phi\omega\right)/(L_a + L_e) \tag{1}$$

where V_b is the voltage of the battery, R_a and R_e are the resistance of the armature and the exciting winding, L_a and L_e are the inductivities, I is the current of the armature, c is the motor constant (the torque and electric motor constants are set equal in this model). The V_i induced voltage is a linear function of angular velocity of the output shaft of the starter (2).

$$V_i = c\Phi\omega \tag{2}$$

Here c is a motor constant, Φ is flux linkage, ω is angular velocity of the shaft. Motion equation of the system comes in the next form.

$$\frac{d\omega}{dt} = \left(c\Phi I - T_{load}\right) / J_{red} \; . \tag{3}$$

In this mechanical equation the difference between the electromagnetic torque of the motor and the load torque (T_{load}) which is created by the brake is equal with the shafts' angular acceleration multiplied with the reduced moment of inertia (J_{red}) of the rotor and of all loads.

At the first try the shaft of the motor and the clutch which connects the motor's terminal to the shaft of the powder brake can just be modeled as a rigid joint. In the equations the losses are enclosed, such as ventilation losses or bearing frictions.

The model was implemented in Scilab/Scicos [3].

3. Measurement pad and measurements

To validate the mathematical model of the system a measurement arrangement has been designed. (*Fig. 3*) The main motor parameters were measured such as the voltage of the battery (V_b), the armature current (I), the angular velocity of the motor (ω) and the load torque of the motor (T_{load}).

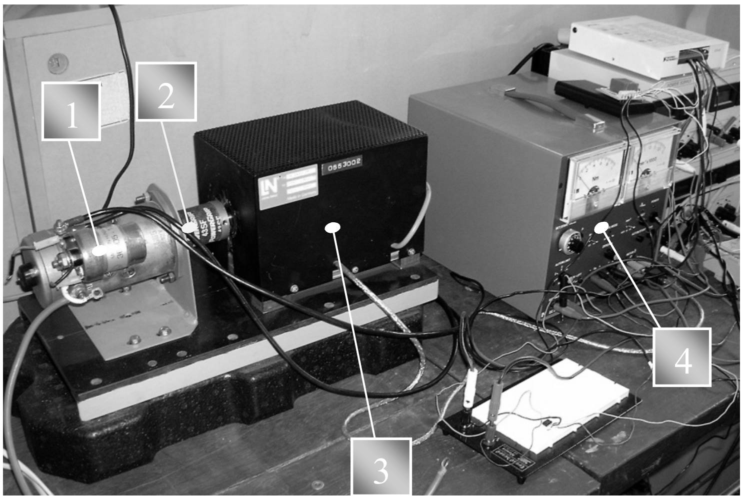


Figure 3: Measurement pad
1 - starter motor, 2 - clutch, 3 - powder brake, 4 – controller for the powder brake.

The load was dynamically generated with a powder brake. A computer based measurement system was applied on the base of NI technology. The sampling rate of the measurements was 800 Hz. A CVI measurement software was developed both for the measurement and control of the brake. Some result can be seen in Fig. 4.

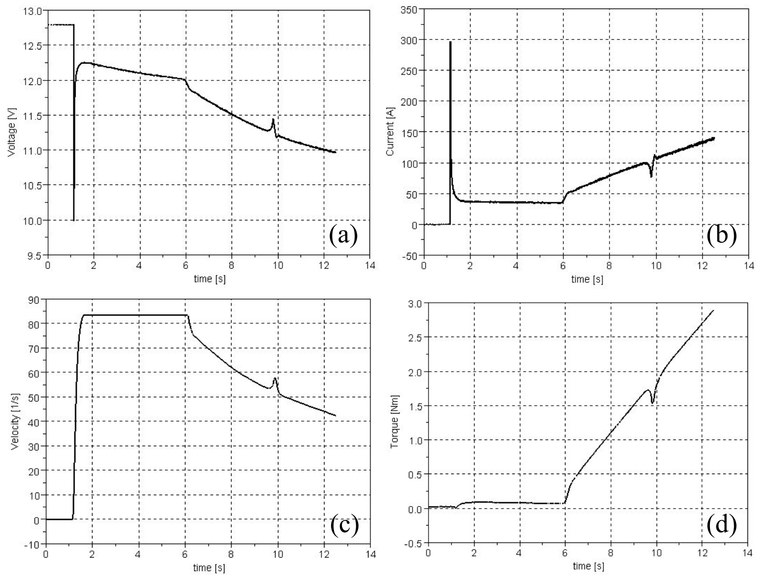


Figure 4: Measured parameters in the function of time
Voltage of the battery (a); Armature current (b);
Velocity of the starter (c); Load torque (d).

4. Identification of the system

To validate the system, least square approximation was used [2]. The main essential of the method is to minimize the squared error between the estimation of the parameters and the model parameters.

Using (1-3) equations, two linear equations can be written, where the unknown parameters are the parameters in the next form.

$$\underline{y} = \underline{\underline{W}}\,\underline{K} \tag{4}$$

Here $\underline{\underline{W}}$ is the regression matrix, $\underline{K}$ is the vector of the unknown parameters [R, L, $c\Phi$, J_{red}] in the system. R is the sum of the resistances in the electric circuit (Fig 1), L is the inductivity of the circuit.

Using the estimated parameters, the relative error (Fig. 5/a) of the modeled system results compared to the measured value (Fig. 5/b).

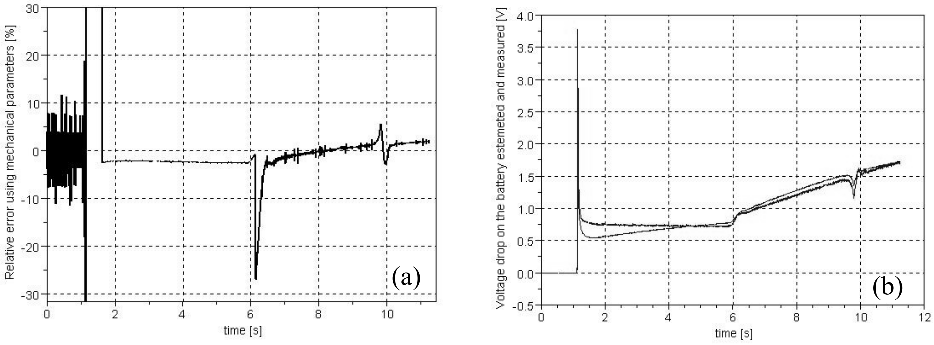


Figure 5: Validation results in the function of time
(a) Relative error; (b) Estimated and measured voltage drop on the battery.

5. Conclusion

The constructed model is proved to be adequate via validation. The mechanical and electric parameters of the system could be estimated well. The system of starter motor was modeled, implemented and validated. The estimated model of the motor has described well the behavior of the real system.

Acknowledgements

The authors highly appreciate the generous support of the BAROSS_EM07-EM_ITN3_07-2008-0039 project of National Office of Research and Technology.

References

[1] Blága, Cs, Kovács, E.: “Simulation of performancs curves of starter”, XXIV microCAD, International Scientific Conference, University of Miskolc, ISBN: 978-963-661-915-2, March, 2010, pp. 127-132

[2] Chiassion, J.:”Modeling and High-Performance Control of Electric Machines” IEEE Series on Power Enginnering, ISBN: 0-471-68449-X, John Wiley & Sons Inc, 2005

[3] Campbell, S.L., Chancelier, J.-P., Nikoukhah, R.: “Modeling and Simulation in Scilab/Scicos” ISBN 0-387-27802-8, Springer, 2006

[4] Vishniac-Bush: “Electromechanical sensors and actuators”, *Springer Verlag, 1999.*

[5] Fodor Gy.: “Elméleti elektrotechnika”, *Tankönyvkiadó, Budapest 1979.*

[6] Kovács E., Szalontai L., Füvesi V.: “Investigation of the Solenoid Switch of an Electric Starter Motor with Simulation”, *X. International Conference on Energetics-Electrical Engineering, Kolozsvár, Romania, 2009.*

Comparison between Centralized and Distributed Model Predictive Control

Katalin GYÖRGY[1] , László DÁVID[2]

[1]Department of Electrical Engineering, Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, e-mail: kgyorgy@ms.sapientia.ro
[2] Department of Electrical Engineering, Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, e-mail: ldavid@ms.sapientia.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: The implementation of advanced multivariable control techniques like Model Predictive Control (MPC) in industry is complex, time consuming and is therefore expensive. Nowdays it is a popular research area to reduce the complexity of the MPC algorithm while preserving the control performance. This paper presents a comparison between centralized and distributed model predictive algorithms and describes a possible implementation of distributed MPC algorithm using Simulink environment.

Keywords: model predictive control, distributed control, prediction horizon.

1. Introduction

The model predictive control (MPC), - also called receding horizon control (RHC)– is the most important advanced control technique which has been very successful in practical applications, where the control input can be obtained by solving a discrete-time optimal control problem over a finite horizon. The most important advantage of MPC algorithms is the fact that they have the unique ability to take into account constraints imposed on process inputs, process state variables and outputs, which usually determine the quality, efficiency and safety of production. Implementation and maintenance of centralized state space (SS) MPC algorithms is becoming an important issue for different multivariable industrial processes. The main idea of our work is to develop a multi-agent software that could be implemented in low cost embedded system, with parallel computational facilities. These software agents are valid for a default model and can be multiplied and customized according to the control horyzon. Each one would solve the problem of finding one of the control actions. This procedure would be repeated several times before the control action values would be delivered to the final control elements. An agent, as an executive, has to know general information about the system and some others which are specific of his own department. It is important to notice that the algorithm to be solved by each agent while computing its control action is much simpler than the one to be solved by the centralized solution.

Previous work on distributed MPC [2],[3],[4] use a wide variety of approaches, including multi-loop ideas, decentralized computation using standard coordination techniques, robustness to the actions of others, penalty functions, and partial grouping of computations. The key point is that, when decisions are made in a decentralized fashion, the actions of each subsystem must be consistent with those of the other subsystems, so that decisions taken independently do not lead to a violation of the coupling constraints. The decentralization of the control become more complex when disturbances act on the subsystems making the prediction of future behaviour uncertain.

We will analyse the influence to the overall performance of distributed system if one or more agent -except the coordinating agent-, fail or obviously underperforms from some reasons. The objective is to solve SS-MPC problems with locally relevant variables, costs, and constraints, but without solving a centralized SS-MPC problem. The coordinated distributed computations solve an equivalent centralized SS-MPC problem. This means that properties that can be proved for the equivalent centralized MPC problem (e.g., stability, robustness) are valid to the above distributed SS-MPC implementation. The significance of proposed distributed control scheme is to reduce the computational requirements in complex large-scale systems and to develop fault tolerant control systems.

2. Centralized State Space Model Predictive Control

All the MPC algorithms possess common elements and different options can be chosen for each one of these elements: prediction model, objective function and algorithms for obtaining the control law. In this paper the process model is a discrete input-state-output relationship:

$$\begin{aligned} \underline{x}_{k+1} &= \underline{\Phi} \cdot \underline{x}_k + \underline{\Gamma} \cdot \underline{u}_k \\ \underline{y}_k &= \underline{C} \cdot \underline{x}_k \end{aligned} , \tag{1}$$

where $\underline{x}_k$ is the state vector (n x 1), $\underline{u}_k$ is the input vector (m x 1), $\underline{y}_k$ is the output vector (p x 1), and $\underline{\Phi}_k$, $\underline{\Gamma}_k$ and $\underline{C}_k$ are the matrices of the system. If these matrices (parameters) are unknown, we have to implement a system identification module in control algorithm.

The centralized model predictive algorithm looks for the vector $\Delta\underline{U}_k$ that minimizes a cost function represented by the scalar J, defined as:

$$J\left(\Delta\underline{U}_k\right)=\left(\underline{Y}_k-\underline{Y}_k^{ref}\right)^T\cdot\underline{Q}\cdot\left(\underline{Y}-\underline{Y}_k^{ref}\right)+\Delta\underline{U}_k^T\cdot\underline{R}\cdot\Delta\underline{U}_k\,, \tag{2}$$

where $\underline{Y}_k^{ref}$ is the vector with the future references, $\underline{Y}_k$ is the vector with the predictions of the controlled variables (output signals), $\Delta\underline{U}_k$ is a vector with future input changes, $\underline{Q}$ is a diagonal matrix with weights for set-point following enforcement, $\underline{R}$ is a diagonal matrix with weights for control action changes. If the prediction horizon is N and the control horizon is N_c this vectors and matrixes are:

$$\underline{Y}_k=\begin{bmatrix}\underline{y}_{k+1/k}\\ \vdots\\ \underline{y}_{k+N/k}\end{bmatrix},\ \underline{Y}_k^{ref}=\begin{bmatrix}\underline{y}_{k+1/k}^{ref}\\ \vdots\\ \underline{y}_{k+N/k}^{ref}\end{bmatrix},\ \Delta\underline{U}_k=\begin{bmatrix}\Delta\underline{u}_{k/k}\\ \vdots\\ \Delta\underline{u}_{k+N_c-1/k}\end{bmatrix}, \tag{3}$$

$$\underline{Q}=\begin{bmatrix}\underline{Q}_1 & 0 & \dots & 0\\ 0 & \underline{Q}_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \underline{Q}_N\end{bmatrix},\ \underline{R}=\begin{bmatrix}\underline{R}_0 & 0 & \dots & 0\\ 0 & \underline{R}_1 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \underline{R}_{N_c-1}\end{bmatrix}. \tag{4}$$

An incremental state space model can be used if the model input is the control increment $\Delta\underline{u}_k=\underline{u}_k-\underline{u}_{k-1}$. The following representation is obtained for predictions:

$$\underline{Y_k}=\underline{\Phi}^*\cdot\underline{x}_k+\underline{\Gamma}^*\cdot\underline{u}_{k-1}+\underline{G}_y\cdot\Delta\underline{U}_k\,, \tag{5}$$

where

$$\underline{\Phi}^*=\begin{bmatrix}\underline{C}\cdot\underline{\Phi}\\ \vdots\\ \underline{C}\cdot\underline{\Phi}^{N_c}\\ \underline{C}\cdot\underline{\Phi}^{N_c+1}\\ \vdots\\ \underline{C}\cdot\underline{\Phi}^{N}\end{bmatrix}\ \underline{\Gamma}^*=\begin{bmatrix}\underline{C}\cdot\underline{\Gamma}\\ \vdots\\ \sum\limits_{i=0}^{N_c-1}\underline{C}\cdot\underline{\Phi}^i\cdot\underline{\Gamma}\\ \sum\limits_{i=0}^{N_c}\underline{C}\cdot\underline{\Phi}^i\cdot\underline{\Gamma}\\ \vdots\\ \sum\limits_{i=0}^{N-1}\underline{C}\cdot\underline{\Phi}^i\cdot\underline{\Gamma}\end{bmatrix}\ \underline{Gy}=\begin{bmatrix}\underline{C}\cdot\underline{\Gamma} & \dots & \underline{0}\\ \vdots & \dots & \vdots\\ \sum\limits_{i=0}^{N_c}\underline{C}\cdot\underline{\Phi}^i\cdot\underline{\Gamma} & \ddots & \underline{C}\cdot(\underline{\Phi}\cdot\underline{\Gamma}+\underline{\Gamma})\\ \vdots & \vdots & \vdots\\ \sum\limits_{i=0}^{N-1}\underline{C}\cdot\underline{\Phi}^i\cdot\underline{\Gamma} & \dots & \sum\limits_{i=0}^{N-N_c}\underline{C}\cdot\underline{\Phi}^i\cdot\underline{\Gamma}\end{bmatrix} \tag{6}$$

The cost function can be written as:

$$J\left(\Delta\underline{U}_k\right)=\frac{1}{2}\Delta\underline{U}_k{}^T\cdot\underline{H}\cdot\Delta\underline{U}_k+\underline{f}^T\cdot\Delta\underline{U}_k+const\,, \tag{7}$$

where

$$\begin{aligned} &\underline{H} = 2 \cdot \left[\underline{G}_y{}^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}\right], \\ &\underline{f} = -2 \cdot \underline{G}_y{}^T \cdot \underline{Q} \cdot \underline{E}_k \\ &\underline{E}_k = \underline{Y}_k^{ref} - \underline{\Phi}^* \cdot \underline{x}_k - \underline{\Gamma}^* \cdot \underline{u}_{k-1} \,. \end{aligned} \quad (8)$$

For problems without constraints the centralized model predictive control determines the vector $\Delta\underline{U}_k$ that makes

$$\frac{\partial J(\Delta\underline{U}_k)}{\partial(\Delta\underline{U}_k)} = 0 \Rightarrow \quad \Delta\underline{U}_k^{opt} = \frac{1}{2} \cdot \left(\underline{H} + \underline{H}^T\right)^{-1} \cdot \underline{f} = \left(\underline{G}_y^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}\right)^{-1} \cdot \underline{G}_y^T \cdot \underline{Q} \cdot \underline{E}_k \quad (9)$$

We remarks that only the first control actions are taken at each instant, and the procedure is repeated for the next control decision in a receding horizon fashion.

3. Distributed State Space Model Predictive Control

The implementation of distributed model predictive control needs to search for $\Delta\underline{u}_{k/k}, \Delta\underline{u}_{k+1/k}, \ldots, \Delta\underline{u}_{k+N_c-1/k}$ [5], that makes the

$$\frac{\partial J(\Delta\underline{U}_k)}{\partial(\Delta\underline{u}_{g/k})} = 0 \,. \quad (10)$$

for $k \le g \le k + N_c - 1$.

If the cost function is written as:

$$J\left(\Delta\underline{U}_k\right) = \Delta\underline{U}_k{}^T \cdot \left(\underline{G}_y^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}\right) \cdot \Delta\underline{U}_k - 2 \cdot \underline{E}_k^T \cdot \underline{Q} \cdot \underline{G}_y \cdot \Delta\underline{U}_k + \underline{E}_k^T \cdot \underline{Q} \cdot \underline{E}_k \quad (11)$$

then the first order optimality condition is determinable as:

$$\begin{aligned} \frac{\partial J(\Delta\underline{U}_k)}{\partial(\Delta\underline{u}_{g/k})} &= 2 \cdot \left[\underline{G}_y^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}\right]_{g-k+1, g-k+1} \cdot \Delta\underline{u}_{g,k} - 2 \cdot \sum_{i=1}^{N} \left(\left[\underline{Q} \cdot \underline{G}_y\right]^T_{i, g-k+1} \cdot \left[\underline{E}_k\right]_i \right) + \\ &+ \sum_{\substack{i=0 \\ i \ne g-k}}^{N_c-1} \left(\left([\underline{G}_y^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}]^T_{i+1, g-k+1} + [\underline{G}_y^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}]_{g, i+1} \right) \Delta\underline{u}_{k+i,k} \right) \end{aligned} , \quad (12)$$

and the variation of input signal is

$$\begin{aligned} \Delta\underline{u}_{g/k} &= \left(2 \cdot \left[G_y^T \cdot Q \cdot G_y + R\right]_{g-k+1, g-k+1}\right)^{-1} \cdot (2 \cdot \sum_{i=1}^{N} \left(\left[\underline{Q} \cdot \underline{G}_y\right]^T_{i, g-k+1} \cdot \left[\underline{E}_k\right]_i \right) + \\ &- \sum_{\substack{i=0 \\ i \ne g-k}}^{N_c-1} \left(\left([\underline{G}_y^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}]^T_{i+1, g-k+1} + [\underline{G}_y^T \cdot \underline{Q} \cdot \underline{G}_y + \underline{R}]_{g, i+1} \right) \Delta\underline{u}_{k+i,k} \right)) \end{aligned} . \quad (13)$$

The first value of every $\Delta\underline{u}_{g/k}$ is only an approximation since it depends on the other $\Delta\underline{u}_{i+k/k}$ values ($i \ne g - k$). It should be noticed that the computation burden to obtain $\Delta\underline{u}_{g/k}$ is much smaller than the one needed to compute the whole vector $\Delta\underline{U}_k$. As already discussed, in this distributed approach, the vector $\Delta\underline{U}_k$ is determined by software agents using a combination of repeated computation of $\Delta \underline{u}_{g/k}$ and exchange of information.

The equation (13) can be written in the following general form

$$\Delta \underline{u}^{n}_{k+j/k} = \sum_{\substack{i=0 \\ i\neq j}}^{N_c-1} \left(\underline{A}_{j+1,i+1} \cdot \Delta \underline{u}^{n-1}_{k+i/k} + \underline{B}_{j+1} \right), \tag{14}$$

where $0 \leq j \leq N_c - 1$, the matrices $\underline{A}_{i,j}$ will have dimension $m \times m$ and vector $\underline{B}_i$ dimension $m \times 1$, where m is the number of inputs. Matrix $\underline{A}_{i,i}$ is nil. Writing a centralized expression for $\Delta \underline{U}_k$ using equation (14)

$$\begin{bmatrix} \Delta \underline{u}_{k/k} \\ \Delta \underline{u}_{k+1/k} \\ \vdots \\ \Delta \underline{u}_{k+N_c-1/k} \end{bmatrix}^{n} = \begin{bmatrix} \underline{0} & \underline{A}_{1,2} & \cdots & \underline{A}_{1,Nc} \\ \underline{A}_{2,1} & \underline{0} & \cdots & \underline{A}_{2,Nc} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{A}_{Nc,1} & \underline{A}_{Nc,2} & \cdots & \underline{0} \end{bmatrix} \cdot \begin{bmatrix} \Delta \underline{u}_{k/k} \\ \Delta \underline{u}_{k+1/k} \\ \vdots \\ \Delta \underline{u}_{k+N_c-1/k} \end{bmatrix}^{n-1} + \begin{bmatrix} \underline{B}_1 \\ \underline{B}_2 \\ \vdots \\ \underline{B}_{Nc} \end{bmatrix} \tag{15}$$

which, in a compact form becomes

$$\Delta \underline{U}^{n}_{k} = \underline{A} \cdot \Delta \underline{U}^{n-1}_{k} + \underline{B}\,. \tag{16}$$

The convergence of the $\Delta \underline{U}_k$ vectors to their true values has to be assured for a reliable application. For unconstrained applications the developments made on the area of distributed computation can be used [2]. The Jacobi over relaxation approach is adopted here by recomputing $\Delta \underline{U}_k$ as a linear combination of the value computed by equation (16) and the value obtained in the previous iteration,

$$\Delta \underline{U}^{n}_{k,filtered} = (I - diag(\underline{\alpha})) \cdot \Delta \underline{U}^{n}_{k} + diag(\underline{\alpha}) \cdot \Delta \underline{U}^{n-1}_{k} \tag{17}$$

where $\underline{\alpha}$ is a vector with filter parameters. Applying the filter according to equation (16)

$$\begin{aligned} \Delta \underline{U}^{n}_{k} &= \left(\left(\underline{I} - diag(\underline{\alpha}) \right) \cdot \underline{A} + diag(\underline{\alpha}) \right) \cdot \Delta \underline{U}^{n-1}_{k} + \left(\underline{I} - diag(\underline{\alpha}) \right) \cdot \underline{B} = \\ &= \underline{A}(\underline{\alpha}) \cdot \Delta \underline{U}^{n-1}_{k} + \left(\underline{I} - diag(\underline{\alpha}) \right) \cdot \underline{B} \end{aligned} \tag{18}$$

A sufficient condition for convergence of the iterative process is to have $\left\| \underline{A}(\underline{\alpha}) \right\| < 1$ for $\alpha \in (0,1)$. The search for a filter vector $\underline{\alpha}$ minimizing $\left\| \underline{A}(\underline{\alpha}) \right\|$ can be reduced to a linear constrained optimal programming problem and is rather dependent on the matrix norm.

4. Numerical simulation

This section presents the application of the centralized and distributed model predictive algorithm to a multiple input and a multiple output theoretical system which characterized by following state space model:

$$\underline{x}_{k+1} = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & -0.5 \end{bmatrix} \cdot \underline{x}_k + \begin{bmatrix} 4 & 5 \\ 8 & 9 \\ 10 & 1 \\ 0 & 2 \end{bmatrix} \cdot \underline{u}_k \quad \underline{y}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \underline{x}_k \tag{19}$$

For both algorithms were realized one Simulink diagram and the following parameters were used for both simulations:

$$N = 4; \quad N_c = 3; \quad R = 0.1 \cdot I_2 \quad Q = 10 \cdot I_2 \tag{20}$$

The Simulink diagram of centralized predictive control is shown in *Fig. 1*, where the "Centralized_MPC_control" subsystem contain one complex S-function modul for centralized control algorithm. The Simulink model of distributed predictive algorithm is shown in *Fig. 2*, where the "Distributed_MPC_control" subsystem is presented separately in *Fig. 3*, where we can see the certain three dependent modules for calculating $\Delta\underline{u}_{1/k}, \Delta\underline{u}_{2/k}, \Delta\underline{u}_{3/k}$. The structure of these modules are one and the same, just the input signals and parameters are different.

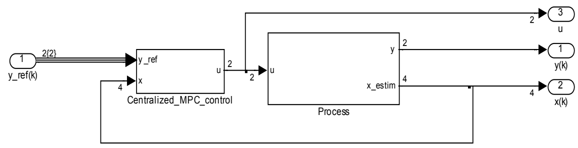


Figure 1. Simulink diagram for numerical simulation of centralized predictive control.

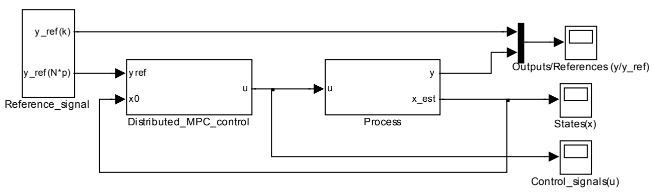


Figure 2. Simulink diagram for numerical simulation of distributed predictive control.

The choice of alpha provides all eigenvalues of matrix $A(\underline{\alpha})$ of equation (18) smaller than 1, which is sufficient to assure that the iterative method converges. These values were determinated before numerical simulation, and one optimal constrained problem was solved in Matlab environment. It seems that the parameters tuning for the distributed algorithm do not need to be exactly the same as the one used for the centralized version. For the same amount of information exchange among agents, a faster reference filter improves the response.

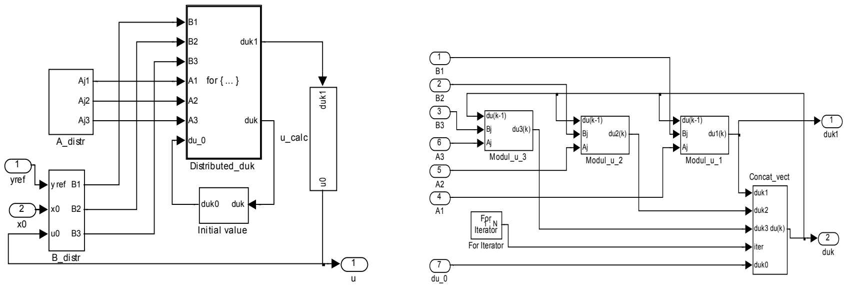
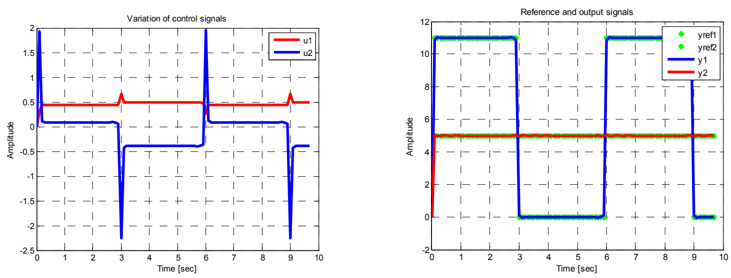


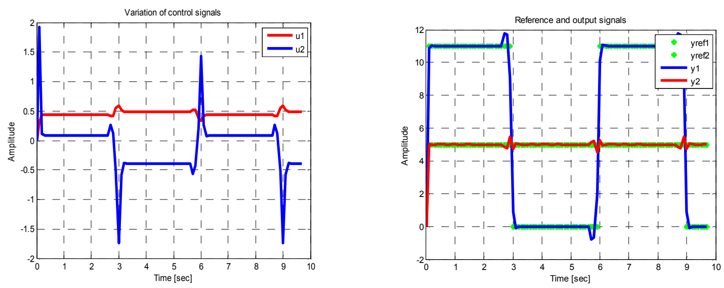
Figure 3. Subsystem diagram for distributed predictive algorithms (Nc=3).

The results obtained in numerical simulation using the centralized control algorithm using a variable reference signal are shown in *Fig. 4.* and results of numerical simulation of the distributed algorithm are shown in *Fig. 5.*



(a) (b)

Figure 4. Variation of control signals (a) outputs signals(b) with centralized algorithm



(a) (b)

Figure 5. Variation of control signals (a) outputs signals(b) with distributed algorithm

5. Conclusion

The performance of the distributed control applied on the example was comparable to that obtained with the centralized model predictive control. The computation power needed to solve the distributed problem is smaller than what is needed for the centralized case. This fact may allow the utilization of model predictive control executed in distributed hardware with low computational power. The size of the centralized problem grows considerable with the number of inputs/outputs while the distributed problem size remains the same for the same control horizon. One point to mention is that most of the multivariable problems do not have a complete interaction as the example presented. In the distributed algorithms the problem is to choose the convenient sample time and the correct filter parameters vector. The filter choice should be done off-line and the condition presented is enough to ensure the convergence of the algorithm. Future developments are needed to provide a best filter option (assuring the fastest convergence with robustness) and to introduce some constraints in model predictive applications. The main benefit expected with the distributed MPC control is the system maintenance and the 'apparent' simplicity to the user.

References

[1] E.F. Camacho, "Model Predictive Control", Ed. Springer, 2004;
[2] E. Camponogara, D. Jia, B.H. Krogh, and S. Talukdar, "Distributed model predictive control." IEEE Control Systems Magazine, 1:44–52, February 2002.
[3] Venkat, A. N., Rawlings, J. B. e Wright, S. J. (2006), "Implementable distributed model predictive control with guaranteed performance properties", American Control Conference 2006.
[4] Mercangoza, M. e Doyle III, F. (2007), "Distributed model predictive control of an experimental four tank system", Journal of Process Control 17: 297–308.
[5] A. Plucenio, D. J. Pagano, E. Camponogara, H.F. Sherer, M. Lima, "A simple distributed MPC algorithm", Rio de Janerio, Brasil.

Vector Control of the Current-Excited Synchronous Generators

Mária IMECS

Department of Electrical Drives and Robots, Faculty of Electrical Engineering,
Technical University of Cluj-Napoca, e-mail: imecs@edr.utcluj.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: The paper deals with the vector-control procedure of the synchronous generators with exciting winding, based on the field-orientation principle, disregarding in transient operation mode the effect of the damper bars. Two vector-control structures are treated for the wound-excited synchronous generator: connected to the 3-phase AC-grid, operating with leading current, and coupled to a DC-distribution line, working at unity power factor, respectively. There are presented the appropriate space-phasor diagrams showing the stator-field-oriented armature- and exciting-current components and their correspondence with the active and reactive powers. It is demonstrated that the stator voltage and frequency decoupled control loops are re-coupled by means of the exciting current field-oriented components.

Keywords: Field-orientation, decoupled control, leading current, unity power factor.

1. Introduction

The synchronous machine (SM) is the principal source of the electrical energy obtained by conversion from mechanical energy. It is capable to produce both active and reactive power. The applications of AC drives cannot dispense with the synchronous machines, not only for energy saving purposes, but also for their high control performance.

The classical control principle of the large power wound-excited synchronous generators (SG) is well-known, considering the frequency and voltage control by means of the active- (P) and reactive-power (Q) adjustment, respectively. The active energy is coming from the driving turbine and the reactive one is commanded with the excitation (field winding) rectifier. The two control loops usually are operating independently from each other.

Consequently, it may be considered as a scalar control (SC) procedure, which disregards some fundamental phenomena, i.e. the natural coupling effect inner the SG [1]–[3].

A vector control (VC) structure is based on the field-orientation (FO) principle. It represents the up-to-date control method not only for the AC motor drives, but also for proper generator running, due to its high dynamic performance and stability during different transient operation modes, approaching the performances of the DC machines. In the VC loops the natural behavior of the AC machine is taken into account by using its dynamic mathematical model in identification of the feedback variables and also in computation of the control ones [1], [4], [5].

The dynamic model of the SMs is based on the Park's two-reaction theory published in 1929 [6]. In 1959 it was treated by means of the space-phasors (SPh) [7], and later also in matrix form [8]. The mathematical background of the VC procedure is given properly by the SPh theory [1], [5].

The idea of the field-oriented VC applied for the large power SG was introduced in Romanian in 1989 [2]. In English first it was published in 1990 [3], and then described in detail in 1993 [1].

2. The stator-field orientation principle

The FO principle was initially proposed by *Blaschke* in 1971 in the *Siemens Zeitschrift* [9], and it referred to the decoupled control of the mechanical and magnetic phenomena of the cage (short-circuited) rotor induction motor drive by means of the stator-current rotor-field-oriented components.

In the same time the FO was extended also to the synchronous motor drives and it was published in 1971 in the same periodical [10].

Since the power factor control implies the stator quantities, in the most cases the stator flux is used for orientation [11], [12].

If the longitudinal reaction is cancelled in control schemes with orientation according to the air-gap flux, the motor will operate with lightly inductive stator current [10], [13]. This is analyzed and mentioned also in [1].

The FO principle leads to the analogy of the AC machines with the DC ones [9]. The rotor-position-oriented (or exciting-field-oriented) SM model (given by the Park's equations) is analogue to a non-compensated DC machine [1], [5]. It is suitable for the simulation of the SM [14], but the VC should be realized with the field-oriented model based on the resultant stator flux, which leads to the analogy with the compensated DC machine, with shifted or non-shifted brushes, depending on the value of the power factor [1], [4], [5], [15].

The FO-ed model allows the independent control of the two variables which produce the electromagnetic torque of the AC machine.

As procedure, stator-field-orientation (SFO) means that the complex plane real axis ($d\lambda_s$) is taken according to the direction of the SPh corresponding to the stator resultant flux [1], [5], [15]:

$$\Psi_{sd\lambda s} = \underline{\boldsymbol{\Psi}}_s = |\underline{\boldsymbol{\Psi}}_s| = \Psi_s \text{ and } \Psi_{sq\lambda s} = 0. \tag{1}$$

Angle λ_s is the angular position of the resultant stator flux $\underline{\Psi}_s$. The stator-field-oriented components of the stator- (armature-) (see *Fig. 1*) and the exciting-current (see *Fig. 3*) SPh-s are:

$$\underline{\mathbf{i}}_s = i_{sd\lambda s} + \mathbf{j}\, i_{sq\lambda s} \text{ and } \underline{\mathbf{i}}_e = i_{ed\lambda} + \mathbf{j}\, i_{eq\lambda} \tag{2}$$

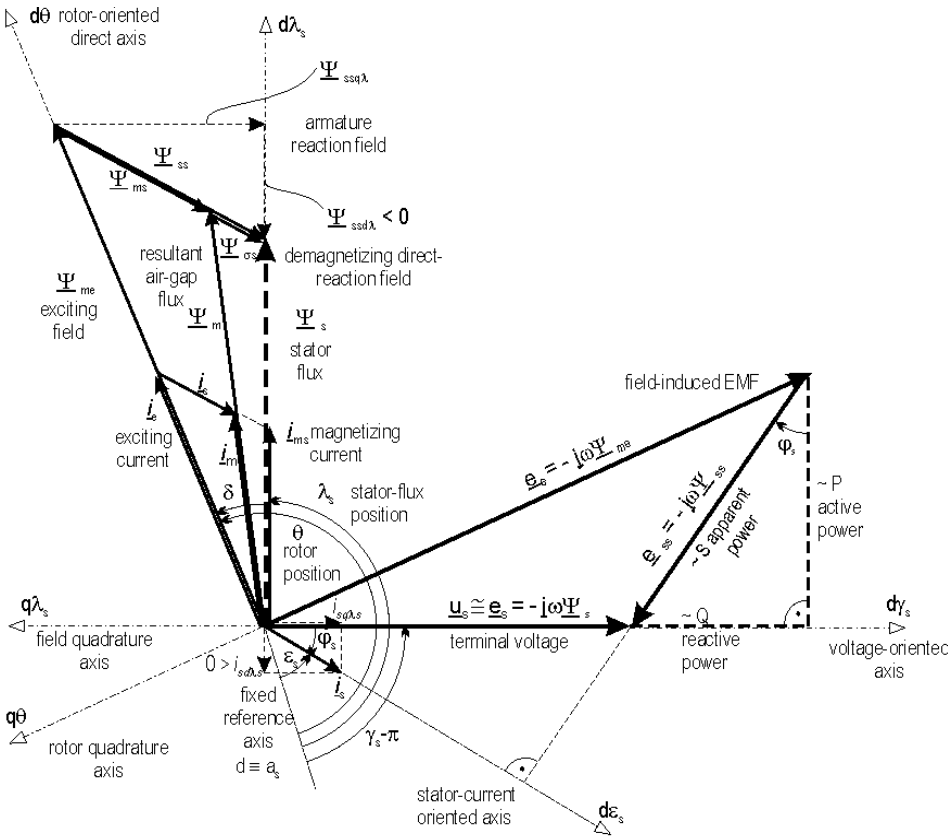


Figure 1: Phasor diagram of the synchronous generator operating with leading current.

In the stator-field-oriented (SFOed) axis frame the current direct component $i_{sd\lambda s}$ has reactive character because it determines the power factor (leading or lagging). The quadrature component is the active one, i.e. the torque-producing component, because it is perpendicular onto the stator flux:

$$m_e = K_M\, \Psi_s\, i_{sq\lambda s}\,. \tag{3}$$

If the SPh module is equal to the amplitude of the sine-wave variables, than the torque coefficient is $K_M = 3\, z_p/2$, where z_p is the number of the pole pairs [1]. In the SFOed axis frame the armature reaction (AR) flux is

$$\underline{\boldsymbol{\Psi}}_{ss} = \Psi_{ssd\lambda s} + \mathbf{j}\, \Psi_{ssq\lambda s} = L_{sd}\, i_{sd\lambda s} + \mathbf{j}\, L_{sq}\, i_{sq\lambda s}\,, \tag{4}$$

where the stator inductances are expressed by the useful and the leakage inductances $L_{sd} = L_{md} + L_{\sigma s}$ and $L_{sq} = L_{mq} + L_{\sigma s}$.

Disregarding in transient operation mode the damper effects of the rotor bars, the resultant stator flux may be written

$$\underline{\boldsymbol{\Psi}}_{s} = \underline{\boldsymbol{\Psi}}_{me} + \underline{\boldsymbol{\Psi}}_{ss}\,,\ \text{where}\ \underline{\boldsymbol{\Psi}}_{me} = L_{md}\, \underline{\boldsymbol{i}}_{e} \tag{5}$$

is the field produced by the exciting current i_e in the air gap, which is deviated with the load angle δ with respect to $\underline{\Psi}_s$ the resultant stator flux, as it is shown in *Fig. 1*. The flux $\underline{\Psi}_{me}$ exercises its magnetizing effect in the direction of the rotor longitudinal axis dθ. Considering $L_m = L_{md} \approx L_{mq}$ (i.e. a uniform air-gap SM), the AR flux may be expressed by means of the stator current $\underline{i}_s$, as follows

$$\underline{\boldsymbol{\Psi}}_{ss} = \underline{\boldsymbol{\Psi}}_{mss} + \underline{\boldsymbol{\Psi}}_{\sigma s} = (L_m + L_{\sigma s})\, \underline{\boldsymbol{i}}_{s} = L_s\, \underline{\boldsymbol{i}}_{s}\,, \tag{6}$$

where $\underline{\Psi}_{\sigma s}$ is the stator leakage flux. If $\sigma_s = L_{\sigma s} / L_m$ is the stator leakage coefficient, the stator-flux-based magnetizing current, which it is proportional to the resultant stator flux (including also the leakage one) is defined as

$$\underline{\boldsymbol{i}}_{ms} = \underline{\boldsymbol{\Psi}}_{s} / L_m = \underline{\boldsymbol{i}}_{m} + \sigma_s\, \underline{\boldsymbol{i}}_{s} .= \underline{\boldsymbol{i}}_{e} + (1 + \sigma_s)\, \underline{\boldsymbol{i}}_{s}. \tag{7}$$

Based on (5) and (7) the triangle of the fluxes will be similar to that of the corresponding currents. Furthermore it is similar also to that of the induced electro-motive forces (EMF).

The apparent power may be written in complex form as

$$\underline{\mathbf{S}} = P + \mathbf{j}\, Q \tag{8}$$

In the case of the uniform air-gap SyM, neglecting the stator resistance ($R_s \approx 0$), the right-angled triangle of the AR-flux SPh and its stator-field-oriented components may be put in evidence according to (2) and (4). This triangle looks similar to that one formed by the stator-currents and also to the power triangle.

Consequently the active (P) and the reactive (Q) power may be controlled by means of the stator-field-oriented armature current components, $i_{sq\lambda s}$ and $i_{sd\lambda s}$, respectively [1]–[3]. Substituting in (7) the stator-field-oriented two-phase components of the armature- and exciting-current, respectively, from (2), result

$$i_{ed\lambda} = i_{ms} - (1+\sigma_s)\, i_{sd\lambda} \text{ and } i_{eq\lambda} = -(1+\sigma_s)\, i_{sq\lambda}\,. \tag{9}$$

According to the above expression the double (magnetizing and torque producing) role of the exciting current, split into two stator-field-oriented components, may be considered, too [16], [17].

3. Stator-field oriented vector control structures

The actuators in a classical control structure of the wound-excited SGs are the automatic speed regulator (ASR), which controls the torque of the driving turbine and a thyristorized phase-controlled rectifier for the control of the exciting current, as it is shown in *Fig. 2*. Consequently, due to the two control variables (torque and exciting current), in the control structure it is possible to impose only two reference values, the stator voltage and its frequency [1]–[3], [17]–[21].

A. Synchronous generator connected to the 3-phase AC grid

In the generator operating mode the quadrature component of the AR flux $\Psi_{ssq\lambda s}$, (and also the stator current $i_{sq\lambda s}$), which determines the active power (*P*), will be negative due to the reversed active energy flow with respect to the motoring operation. The longitudinal component $\Psi_{ssd\lambda s}$ (and $i_{sd\lambda s}$, too) corresponding to the reactive power (*Q*) for leading current is also negative, due to its demagnetizing character.

According to (9) the exciting current may be also considered as containing an active component, which produces the torque in the air-gap. In the SM only the longitudinal component of the exciting-current vector contributes to the resultant magnetic field in the air-gap, consequently it may be considered with reactive character [1]–[5], [15]–[17], [19], [21].

In order to control the power flow in the SM the $\underline{i}_s$ SPh of the stator current has to be oriented according to the magnetizing direction of the resultant stator flux $\underline{\Psi}_s$. Mathematically the orientation is realized by means of a coordinate transformation block (CooT), as it is shown in *Fig. 2*. In this way the stator-current components may be taken into account in the decoupled control loops of the magnetic and mechanical quantities.

The re-coupling of the control loops is made by means of the reference value computation of the exciting current module in block i_eC:

$$i_e = [(i_{ed\lambda s})^2 + (i_{eq\lambda s})^2]^{0.5}, \tag{11}$$

where the SFO-ed components are given by (9).

The stator-flux-based magnetizing current i_{ms} is generated in the reactive control loop by the stator-flux controller [1]–[3], [17].

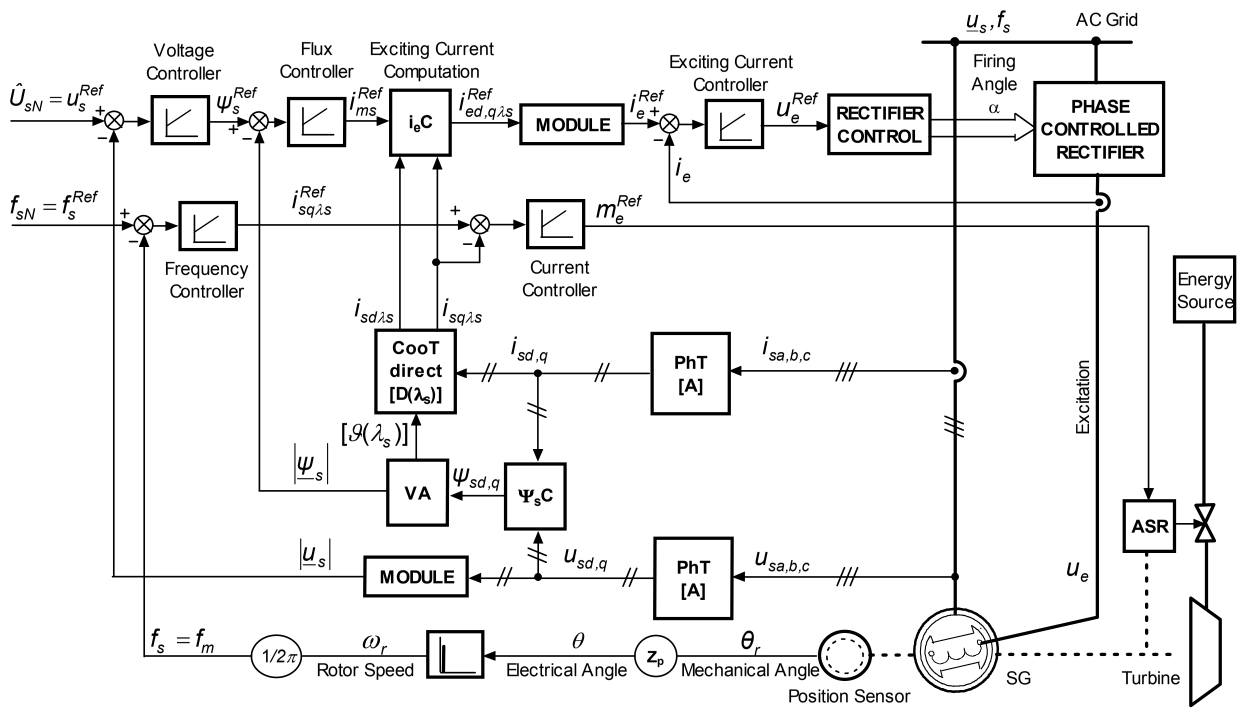


Figure 2: Vector control structure neglecting the damper effect of the synchronous generator with stator-field orientation connected to the three-phase AC grid.

Due to the fact the stator-current SPh is a three-phase vector, for the space vector (SV) of the exciting current corresponding to a single-phase winding an equivalent three-phase SPh must be deduced, i.e. the exciting winding has to be referred to the armature number of turns and number of phases [1], [5], [8].

B. Synchronous generator connected to a DC distribution line

The SG is connected to the DC distribution line by means of a rectifier. If the SG operates at UPF the longitudinal AR is zero, and the armature current is perpendicular onto the resultant stator flux, as it is represented in *Fig. 3*.

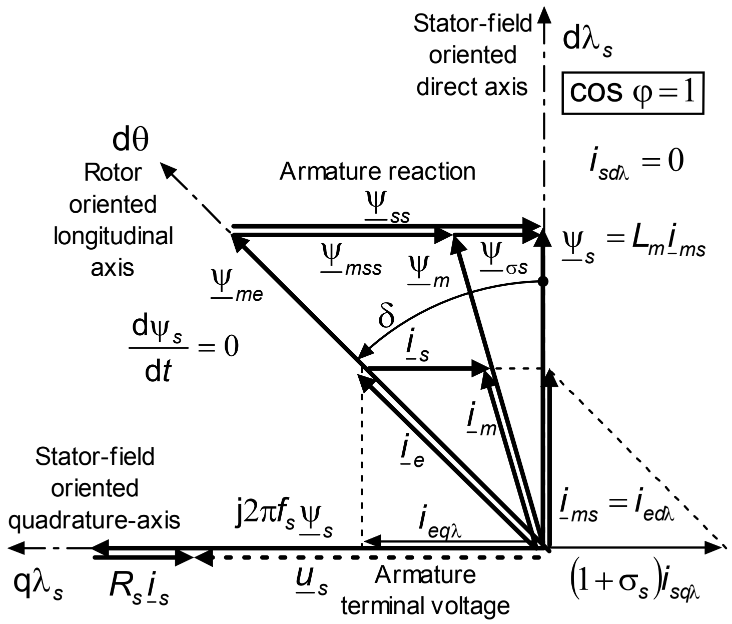


Figure 3: Stator-field oriented space-phasor diagram of the synchronous generator operating at unity power factor.

The exciting current is controlled by a DC-to-DC chopper, as it is shown in *Fig. 4*. [20], [21]. The SG is operating at unity power factor (UPF). The rectifier needs active current filtering (ACF), because it operates with 120° square waves. The ACF is realized by means of a PWM voltage-source filter (VSF), which is connected directly on the DC line in tandem arrangement, as it is shown in *Fig. 5*, similarly to a tandem converter [21]–[25].

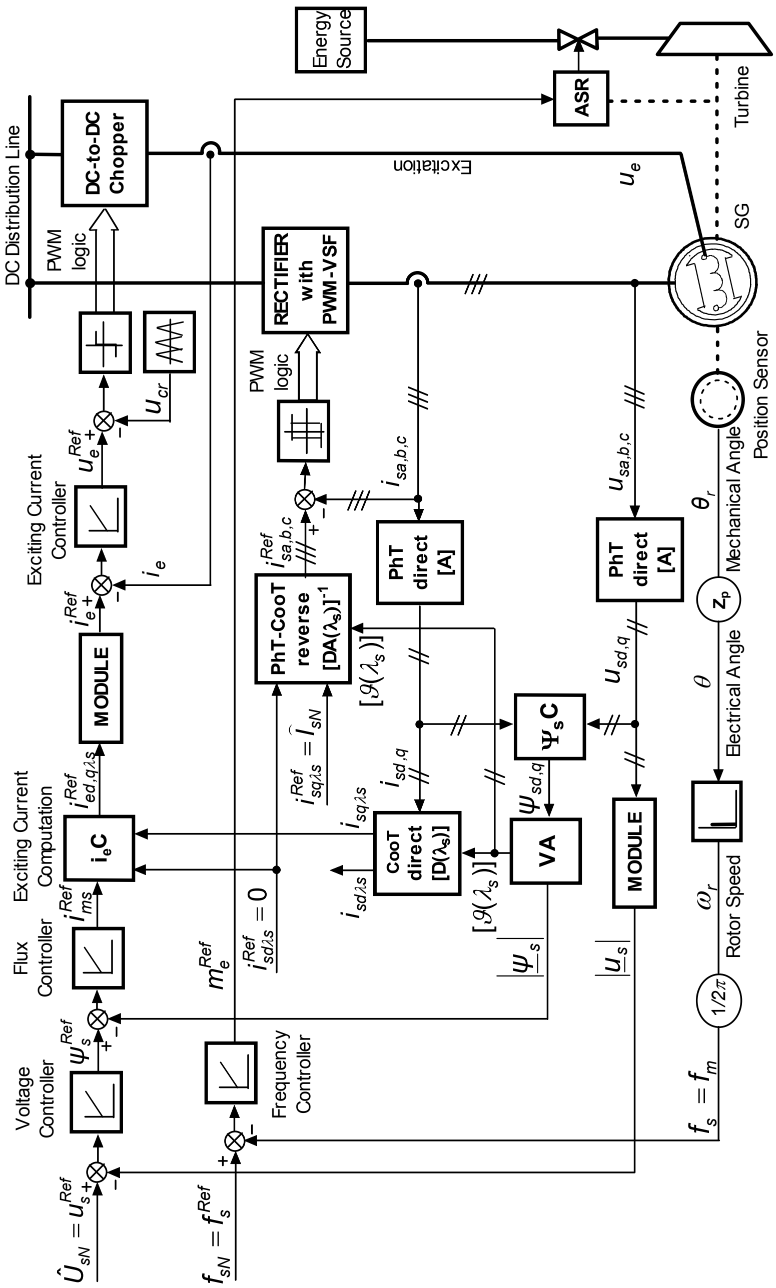


Figure 4: Vector controlled synchronous generator connected to a DC distribution -line with stator-field orientation operating at unity power factor.

The PWM-VSF operates synchronized with the square-wave rectifier, which needs pulse amplitude-modulation (PAM) of its current in order to prevent the energy transfer through the PWM-VSF.

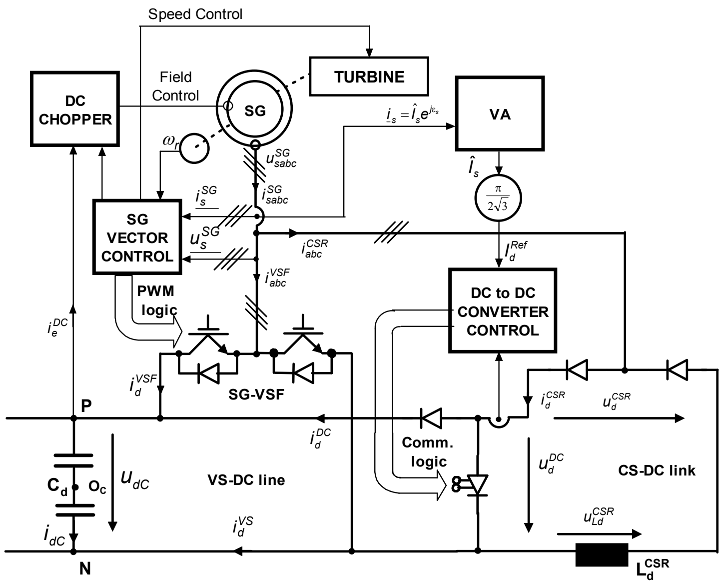


Figure 5: Synchronous generator with ACF connected to a DC-line.

If $d\Psi_s/dt = 0$, for $\cos\varphi = 1$ from (9) results $i_{ed\lambda s} = i_{ms}$, because $i_{sd\lambda s} = 0$. As a consequence, the direct component ($i_{ed\lambda}$) of the field-oriented exciting current in fact is the proper reactive current, which magnetizes the armature, because it is proportional to the resultant stator flux in the SG.

Based on (3), (7) and (12) the SM electromagnetic torque may be written also depending on the exciting-current components:

$$m_e = K_M L_m i_{ms} i_{sq\lambda} = - K_{Ms} L_m i_{ed\lambda} i_{eq\lambda}, \tag{12}$$

where the torque coefficient is $K_{Ms} = K_M (1+\sigma_s)^{-1}$.

The vectorial character of the control structure from *Fig. 4* is also given by the coordinate transformation block CooT, based on matrix $[D(\lambda_s)]$. It accomplishes the FO of the stator-current SPh by rotating the fixed axis frame with λ_s. The reverse Park's transformation block (PhT-CooT), marked with the matrix operator $[DA(\lambda_s)]^{-1}$, includes also the phase-transformation. In fact it is a vectorial sine-wave generator of the 3-phase stator-current references, corresponding to the $\underline{i}_s^{Ref}$ SPh with respect to the stator-field spatial position λ_s, which performs inherently a certain kind of self-commutation of the SG.

3. Conclusions

Generally the proper VC of the SGs (similarly to the SM drives) may be realized only by orientation of both the stator- ($\underline{i}_s$) and exciting- ($\underline{i}_e$) current SPh-s, splitting each of them into two ($d\lambda_s$-$q\lambda_s$) stator-field-oriented components. The orientation field $\underline{\Psi}_s$ includes the exciting and armature reaction fluxes.

In simulation results of the VC structure the SG presents a good dynamic at starting and a good dynamic response for a perturbation consisting of a sudden change of the electrical load. While the armature voltage is kept at constant level, the exciting voltage varies in accordance with the modified load conditions [17], [18], [20], [21].

The transient performance of the VC structure from *Fig. 2* and *Fig. 4* may be improved by taken into account in the exciting current computation, not only the FO-ed stator-current components, but also the damping circuit currents. These may be identified by on-line computation from the feedback variables based on the rotor-position-oriented (dθ-qθ) model of the damping windings [1], [2], [26].

The proper two-phase damping bars must be also transformed, like the exciting circuit, into another kind of two-phase equivalent winding, corresponding to a fictitious three-phase one [1], [5], [8].

For the vector controlled SM drives supplied from a static frequency converter, or inverter, which operate with sine-wave currents, the damper windings become less important for the motor starting, as well as in the rotor oscillation damping [1], [11]. In the case of the vector controlled SG the control performance may achieve such a level, that the damper bars lose their role and may be disregarded.

The VC structures of the IM and SM drives with stator-flux identification are already implemented [27]–[29]. Future works will be focused on the implementation of the PWM-active filters and SG control scheme.

The unified electrical machine theory treated with the space phasors [30] is the basis of the field-orientation principle, which leads to the unified control theory of the electrical machines. The mathematical model of the AC machines is transformed in order to become analogues to that of the DC compensated ones, and then they may be controlled with the same procedures.

Acknowledgements

In memoriam *Professor Árpád KELEMEN*, who raised first the idea to apply the field-orientation principle also for the vector control of the large power wound-excited synchronous generators.

References

[1] Kelemen, Á., Imecs, M., "Vector Control of AC Drives", Vol. 2: "Vector Control of Synchronous Machine Drives", Ecriture Publisher, Budapest, Hungary, 1993, ISBN 963 593 140 9.

[2] Kelemen, Á., Imecs, M., "Procedure and Driving Device for Unitary Automatic Vector Control of the Active and Reactive Power of Synchronous Generators by Means of Frequency and Voltage Regulation" (in Romanian), Patent of Invention Nr. 104278/30.10.1989, Romania.

[3] Kelemen, Á., Imecs, M., "Vector control system for frequency and voltage of synchronous generators", *Proceedings of International Conference on Electrical Machines, ICEM '90, Cambridge, Massachusetts, USA*, 1990.

[4] Imecs, M., "Vector Control Systems for Positioning of Induction Motors Fed by Static Converters" (in Romanian), PhD Thesis, Supervisor Prof. Á. Kelemen, Polytechnical Institute of Cluj-Napoca, Romania, 1989.

[5] Kelemen, Á., Imecs, M., Sisteme de reglare cu orientare după câmp ale maşinilor de curent alternativ (in Romanian), IPCN Press, Cluj-Napoca, 1987 and Romanian Academy Press, Bucharest, Romania, 1989.

[6] Park, R. H., „Two-reaction theory of synchronous machines . Generalized method of analysis", Part I and Part II, *Transactions of AIEE*, Vol. 48, July 1929 and Vol. 52, June 1933.

[7] Kovács, K. P., Rácz, I., "Transiente Vorgänge in Wechselstrommaschinen" (in German), Verlag der Ungarischen Academic der Wissenschaften, Budapest, Hungary, 1959.

[8] Kovács, K. P., "Analysis of the Dynamic Processes of the Electrical Machines" (in Romanian), translated and completed by Á. Kelemen and M. Imecs, Technical Press, Bucharest, Romania, 1980.

[9] Blaschke, F., "Das Prinzip der Feldorientierung, die Grundlage für die Transvector-Regelung von Drehfeldmaschinen" (in German), *Siemens-Zeitschrift* 45, Heft 10, 1971.

[10] Bayer, K. H., Waldmann, H., Weibelzahl, M., „Die Transvector-Regelung für feldorientierten Betrieb einer Synchronmaschine" (in German), *Siemens-Zeitschrift,* 45, Heft 10, 1971.

[11] Leonhard, W., "Control of AC machines with help of microelectronics", *3rd IFAC Symposium on Control in Power Electronics and Electrical Drives, Survey Paper, Lausanne, Switzerland*, 1983.

[12] Späth, H.: Steuerverfahren für Drehstrommaschinen, Springer Verlag, Berlin, Heidelberg, New-York, Tokyo, 1983.

[13] Vauhkonen, V., "A cycloconverter-fed synchronous motor drive having isolated output phases", *Proceedings of International Conference on Electrical Machines, ICEM '84, Lausanne, Switzerland,* 1984.

[14] Kelemen, Á., Imecs, M., "Mathematical model and digital simulation of a current-converter fed synchronous motor", *Proceedings of International Conference on Electrical Machines, ICEM '84,* Lausanne, Switzerland, 1984.

[15] Kelemen, Á., Imecs, M., "Analogy between DC- and synchronous machine for field-oriented control", *Proceedings of Beijing International Conference on Electrical Machines BICEM '87, Beijing, China,* China Academic Publishers, 1987.

[16] Imecs, M., Szabó, Cs., Incze, I. I., "Stator-field-oriented control of the variable-excited synchronous motor: numerical simulation", *The 7th International Symposium of Hungarian Researchers on Computational Intelligence HUCI '06, Budapest, Hungary*, 2006, ISBN 963 7154 54 X, pp. 95-106.

[17] Imecs, M., Incze, I. I., Szabó, Cs., "Stator-field oriented control of the synchronous generator: numerical simulation", *Proceeding of the 12th IEEE INES 2008, Miami, Florida, USA*, CD-ROM, ISBN: 978-1-4244-2083-4, IEEE Catalog Number: CFP08IES-CDR, Library of Congress: 2008900450.

[18] Imecs, M., Incze, I. I., Szabó, Cs., "Ecological power system with local distribution DC-line for AC-drive consumers", *IFAC Workshop ICPS'07*, *Cluj-Napoca, Romania,* 2007, ISBN: 978-973-713-180-5, pp. 71-76.

[19] Imecs, M., Szabó, Cs., Incze, I. I., "Stator-field-oriented vector control for VSI-fed wound-excited synchronous motor", *International Aegean Conference on Electrical Machines and Power Electronics, ACEMP-ELECTROMOTION Joint Conference, Bodrum, Turkey*, 2007, ISBN 978-975-93410-2-2, pp. 303-308.

[20] Imecs, M., Incze, I. I., Szabó, Cs., "Control of the energy flow in a dc distribution line, autonomous synchronous generator and ac grid by means of power electronic converters: modeling and simulation", *ICCC 2008, 9th International Carpathian Control Conference ICCC 2008, Sinaia, Romania*, May 26-28, 2008, Volume ISBN 978-973-746-897-0, pp. 255-258.

[21] Imecs, M., Szabó, Cs., Incze, I. I., „Modelling and Simulation of a Vector Controlled Synchronous Generator Supplying a DC Energy Distribution Line Coupled to the AC Grid", *19th International Symposium on Power Electronics, Electrical Drives, Automation and Motion, SPEEDAM 2008*, *Ischia, Italy*, June 11-13, 2008, CD-ROM, pp. 538-543, IEEE Catalog Number CFP 0848 A-CDR, ISBN: 978-1-4244-1664-6, Library of Congress: 2007936381.

[22] Imecs, M., Szabó, Cs., Incze, I. I., "*Active power filtering of line and motor currents for AC drives, modelling and simulation"*, *Workshop TEQREP, Bucharest, Romania*, Ed. R. Măgureanu, University "Politehnica" Bucharest, Printech Press (Romania), 2004, pp. 41- 48

[23] Imecs M., Incze, I. I., Szabó, Cs.:, "New topology for sine-wave current filtering of induction motor drives fed by tandem frequency converter*"*, *Acta Electrotehnica*, Vol. 44, No 3, 2004, Cluj, Romania, ISSN 1224-297, pp. 145-150.

[24] Imecs, M., Trzynadlowski, A. M., Incz, I. I., Szabó, Cs., "Vector control structures of the tandem converter fed induction motor*"*, *IEEE Transactions on Power Electronics*, March 2005, Vol. 20, Nr. 2, ITPEE8, pp. 493-501.

[25] Imecs, M., Szabó, Cs., Incze, I. I., "Modelling and simulation of controlled bi-directional power electronic converters in a DC energy distribution line with AC grid- and motor-side active filtering", *Proceedings of EPE 2007, Aalborg, Denmark*, CD-ROM, ISBN: 9789075815108/ IEEE Catalog No. 07EX1656C.

[26] Naunin, D., „Space-phasor representation of damper currents in synchronous machines at different wave-forms", Electric Machines and Power Systems, Vol. 10, Nr. 1, 1985.

[27] Incze, I. I., "Implementation of Scalar and Vector Control Structures for Induction Motors*"* (in Romanian)*,* PhD Thesis, Supervisor Prof. Maria Imecs*,* Technical University of Cluj-Napoca, Romania, 2004.

[28] Szabó Cs., Implementation of Scalar and Vector Control Structures for Synchronous Motors (in Romanian)*,* PhD Thesis, Supervisor Prof. Maria Imecs, Technical University of Cluj-Napoca, Romania, 2006.

[29] Szabó, Cs., M., Imecs, I. I., Incze, "Synchronous motor drive with controlled stator-field-oriented longitudinal armature reaction", *The 33rd International Conference of the IEEE Industrial Electronics Society, IECON 2007, Taipei, Taiwan*, CD-ROM, IEEE Cat. No. 07CH37855D, ISBN: 1-4244-0783-4, ISSN: 1553-572X and Conference Digest p. 105.

[30] Retter, Gy., "Matrix and Space Phasor Theory of Electrical Machines", Academic Press, Budapest, Hungary, 1987.

Simulation of Position Identification in Electrical Drives Equipped with Incremental Encoder

Ioan Iov INCZE

Department of Electrical Drives and Robots, Faculty of Electrical Engineering,
Technical University of Cluj-Napoca, e-mail: ioan.incze@edr.utcluj.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: Electrical drives use frequently incremental encoders as position sensor. The paper deals with modeling and simulation of an incremental encoder and associated units for processing the information provided by the encoder. Matlab-Simulink® simulation structures were realized for the position computation and identification of the direction of the rotation based on the encoder signals.

Keywords: Angle transducer, position sensor, incremental encoder, simulation, electrical drive.

1. Introduction

The incremental encoder is a device which provides electrical pulses if its shaft is rotating [1], [2], [4]. The number of the generated pulses is proportional to the angular position of the shaft. The incremental encoder is one of the most frequently used position transducers. The principle of an optical incremental encoder is presented in *Fig. 1*. Together with the shaft there is rotating a transparent (usually glass) rotor disc with a circular graduation-track realized as a periodic sequence of transparent and non-transparent radial zones which modulates the light beams emitted by a light source placed on one side of the disc on the fix part (stator) of the encoder. On the opposite side the modulated light beams are sensed by two groups of optical sensors and processed by electronic circuits. Each of the two outputs of the encoder (noted *A* and *B*) will generate one pulse when the shaft rotates an angle equal to the angular step of graduation θ_p, i.e. the angle according to one successive transparent and non-transparent zone. The number of pulses (counted usually by external electronic counters) is proportional to the angular position of the shaft. Due to the fact that the light beams are placed shifted to each other with an angle equal to the quarter of angular step of graduation $\theta_p/4$, the pulses of the two outputs will be also shifted, making possible the determination of the rotation sense. A third light beam is modulated by another track with a single graduation. The output signal (named *Z*) associated to this third beam provides a single pulse in the course of a complete (360°) rotation. The shaft position corresponding to this pulse may be considered as reference position. *Fig. 2* shows the output pulses of the encoder.

Usually for counter-clockwise (CCW) direction θ is considered as positive, for clockwise (CW) negative.

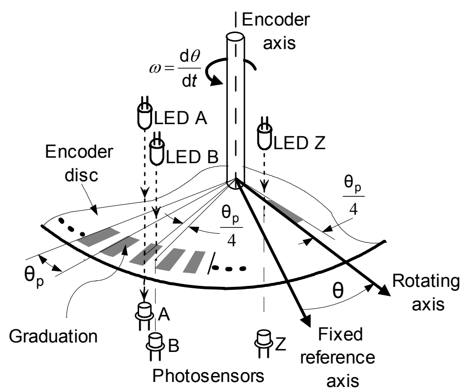


Figure 1: Construction principle of the incremental encoder: the gray surfaces are optically transparent.

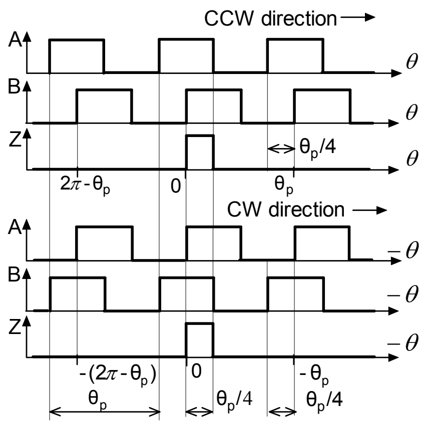


Figure 2: Diagram of the output signals for counter-clockwise (CCW) and clockwise (CW) rotation.

2. Incremental encoder modeling

The input signal of the incremental encoder is the angular position θ of its shaft with respect to the fixed reference axis. The output signals are the two pulses shifted by a quarter angular step $A(\theta)$ and $B(\theta)$, respectively the marker signal $Z(\theta)$. If θ_p is the angular step of the encoder, the outputs may be described by the following equations [2]:

$$A(\theta)=\begin{cases}1 & \text{if} \quad 0 \le (\theta \bmod \theta_p) \le \theta_p/2; \\ 0 & \text{if} \quad \theta_p/2 < (\theta \bmod \theta_p) \le \theta_p; \end{cases}$$
$$B(\theta)=\begin{cases}1 & \text{if} \quad 0 \le ((\theta-\theta_p/4) \bmod \theta_p) \le \theta_p/2; \\ 0 & \text{if} \quad \theta_p/2 < ((\theta-\theta_p/4) \bmod \theta_p) \le \theta_p; \end{cases} \tag{1}$$
$$Z(\theta)=\begin{cases}1 & \text{if} \quad \theta \bmod (2\pi) = 0; \\ 0 & \text{if} \quad \theta \bmod (2\pi) \ne 0. \end{cases}$$

During a rotation angle of the shaft, equal to the angular step of graduation θ_p, there are four switching events in the output pulses; therefore the minimal rotation-angle-increment detectable by the encoder is $\theta_p/4$ [3]. The number of pulses, generated by the encoder in the course of a rotation, is equal with the number of angular steps of the graduation on the circular track on the rotor.

$$N_r = \frac{2\pi}{\theta_p} \tag{2}$$

Based on (1) a Matlab/Simulink® a simulation structure shown in *Fig. 3* was built. The outputs A, B and Z are computed by Simulink® function blocks. θ_p is defined by a constant block. The structure is saved as a subsystem. The simulation structure of the incremental encoder may be integrated in any other Simulink® structure.

2. Encoder based position identification

In an incremental-encoder-based system the angular position θ is measured with respect to a fixed reference axis (θ=0 rad.) and it is obtained by algebraic counting of the number of the generated encoder pulses according to the CCW (ΣN_i pulses) and CW direction (ΣN_j pulses) and multiplying it with the angular step θ_p of the encoder [2]. Mathematically:

$$\theta = \theta_p \left(\sum_i^{CCW} N_i - \sum_j^{CW} N_j \right) = \theta_p N, \tag{3}$$

In order to compute the algebraic number of pulses it is necessary to know the direction of the rotation.

A. Identification of the rotation direction

The two $\theta_p/4$ shifted output signals of the encoder contain implicitly also the direction information which may be obtained in different ways.

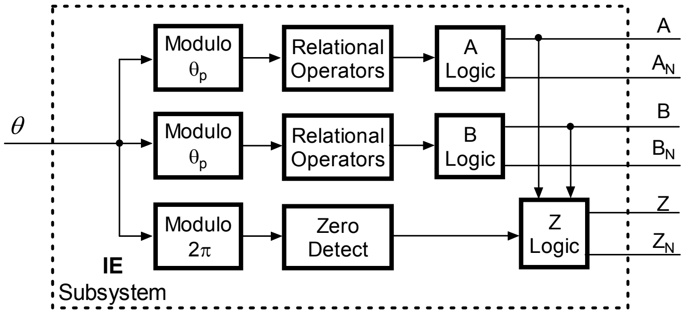


Figure 3: Simulation structure of the incremental encoder.
(Note: The subscript "N" denotes the negated logical variable.)

A trivial solution of the problem may be sampling at every rising edge of the *B* output pulses of the *A* output logic value. The resulted logic value will be 1 for counterclockwise (CCW) direction of rotation, and 0 for the clockwise (CW) direction. The method detects the direction changing only after a time interval according to a rotation of 3 $\theta_p/4$ – 5 $\theta_p/4$.

Taking into account the all four possible combinations of *A* and *B* signals for the reversals, it is possible to detect the direction changing in all cases during a rotation of the minimal detectable rotation-angle-increment $\theta_p/4$. *Table 1* shows the all combinations of signals which detect the reversal of rotation sense.

Table 1. Combination of signals which detect the reversal of rotation

From CCW to CW Q=1 to Q=0 (Triggered by R)		From CW to CCW Q=1 to Q=0 (Triggered by S)	
Occurs if		Occurs if	
A	*B*	*A*	*B*
0	0→1	0→1	0
0→1	1	0	1→0
1→0	0	1	0→1
1	1→0	1→0	1

Note: The 0→1 denote the raising-edge and the 1→0 the falling-edge of associated logic variable.

B. The position-identification structure

Based on (3) and the above presented direction identification method, a Simulink® subsystem was built for position computation. Its structure is presented in *Fig. 4*.

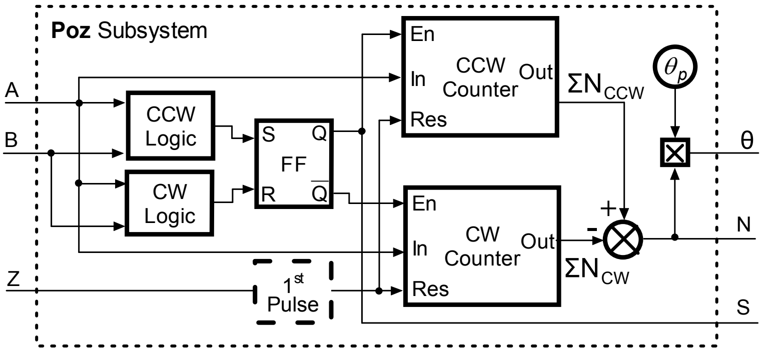


Figure 4. The simulation structure of the position computing block.

The left side of the structure identifies the direction of the rotation. The direction signal *S* enables the appropriate (CCW or CW) counter. The structure has to be provided with the "1st Pulse" block (broken line in *Fig. 5*) in order to extend the position measurement to more rotations.

3. Simulation results

The structure of the interconnected functional units for simulation of position computing is shown in *Fig. 5*. The reference angular position θ_{ref} (the input signal of the encoder block "IE") is generated by a user programmable "Function generator" block. The encoder generates the *A*, *B* and *Z* signals. Based on these, the position computing block "Poz" computes the position θ.

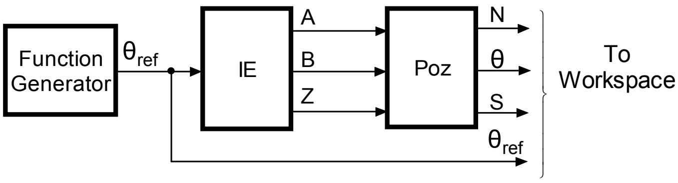
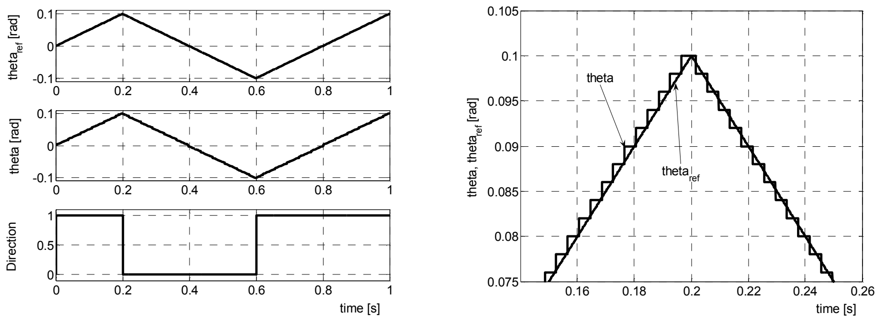


Figure 5: The structure of the interconnected functional units for simulation of position computing (IE – incremental encoder, Poz – position computing block).

In order to test the structure, the function generator was programmed in order to start the simulation generating a positive ramp-reference angle, which is the input signal for the "IE" encoder block. At 0.2 s the ramp is switched to negative (equivalent to a reversal from CCW to CW), decreasing in time until

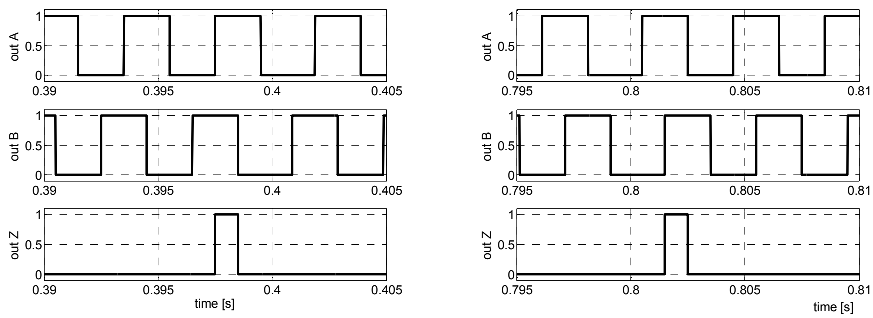
0.8 s, when it is again switched to positive. The time profile of the generated reference angle is presented in *Fig. 6 a)* (top trace). The block "Poz", using the encoder output signals, determines the direction of the rotation (in *Fig. 6 a)* down trace) and computes the position (shown in *Fig. 6 a)* middle trace). The computed position follows very well the reference one. *Fig. 6 b)* presents an enlarged detail of superposed reference and computed angle before and after the reversal at 0.2 s. The incremental character of the computed position is evident.



a) *b)*

Figure 6:. The simulation results representing the reference angle and computed angle during reversal.
a) from top to down: reference angle θ_{ref}, computed angle θ, direction signal *S*.
b) detail of the reference angle θ_{ref} and computed angle θ versus time.

Fig. 7 presents the *A*, *B* and *Z* signals at crossing the reference position ($\theta = 0$) in CW and CCW direction.



a) *b)*

Figure 7. The *A, B* and *Z* signals of the encoder at crossing the reference position.
a) in CW direction, *b)* in CCW direction.

The reversal process was analyzed. The simulated results are presented in *Fig. 8. a)–d)*.

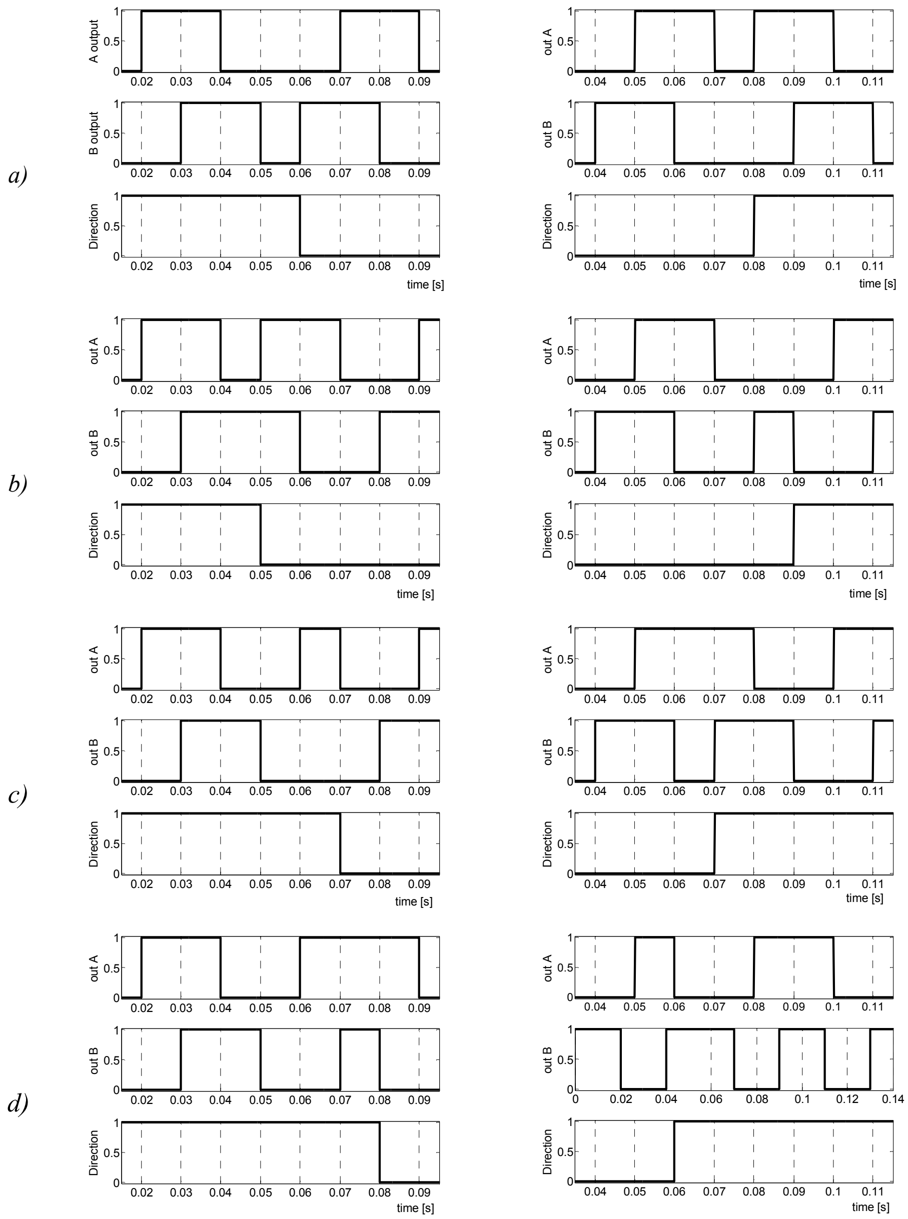


Figure 8: The simulation results showing all combinations of the reversal process.
Left column: Reversal from CCW to CW, *Right column*: Reversal from CW to CCW,
Top trace: output *A*, *Middle Trac*e: output *B*, *Down trace*: direction signal.
Reversal occurs at: a) *A*=0, *B*=0; b) *A*=0, *B*=1; c) *A*=1, *B*=0; d) *A*=1, *B*=1.

The parameters of the function generator were selected in such a manner, that all possible combinations of signals *A* and *B* at reversal (presented in *Table 1*) were captured. As the *Fig. 6* shows, in all cases the sensing of the reversal is done in a quarter of angular step.

The structure presented in *Fig. 5* may be integrated in the simulation structures of electrical drives [2], [5]. In this case the input signal of the encoder – i.e. the angular position – will be provided by the mathematical model of the electrical machine. The position computed by the “Poz” block is used as position feed-back signal by the control system of the drive.

4. Conclusion

The information provided by the incremental encoders is inherently digital.

The angular position of the encoder shaft is obtained by algebraic summing of the number of pulses provided by the encoder according to CCW and CW rotation.

The direction of the rotation may be determined by a digital decoding scheme using the two quadrature signals. The direction changes are detected in an angular interval equal to a quarter of the angular step of the graduation.

The presented simulation structure of the incremental encoder and position computing may be integrated in any Simulink® structure.

References

[1] Incze, J. J., Szabó, Cs., Imecs, Maria, “Modeling and simulation of an incremental encoder used in electrical drives”, in *Proc of 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics CINTI 2009*, *Budapest, Hungary*, 2009, pp. 97-109.

[2] Incze, I. I., Szabó, Cs., Imecs, Maria, “Incremental Encoder in Electrical Drives: Modeling and Simulation” in *Studies in Computational Intelligence* Editors: I. J. Rudas, J. Fodor, J. Kacprzyk, Springer Verlag, Germany, under press.

[3] Koci, P., Tuma, J., “Incremental rotary encoders accuracy”, in *Proc. of International Carpathian Control Conference ICCC 2006, Roznov pod Rashosten, Czech Republic*, 2006, pp. 257-260.

[4] Lehoczky, J., Márkus, M., Mucsi, S., „Szervorendszerek, követő szabályozások”, Műszaki Kiadó, Budapest, Hungary, 1977.

[5] Petrella, R., Tursini, M., Peretti, L., Zigliotto, M., “Speed measurement algorithms for low resolution incremental encoder equipped drives: comparative analysis”, in *Proc. Of International Aegean Conference on Electrical Machines and Power Electronics, ACEMP-ELECTROMOTION Joint Conference, Bodrum, Turkey*, 2007, ISBN 978-975-93410-2-2, pp. 780-787.

MACRo'2010- International Conference on Recent Achievements in
Mechatronics, Automation, Computer Science and Robotics

Synchronous Motor Drive with Double Field-Orientation Working at Maximum Power Factor

Csaba SZABÓ

Department of Electrical Drives and Robots, Faculty of Electrical Engineering,
Technical University of Cluj-Napoca, e-mail: csaba.szabo@edr.utcluj.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: The paper presents a vector control structure for a wound-excited salient-pole synchronous motor, fed by a voltage-source converter, working at unity power factor. Two types of orientation procedure are used: stator-field-orientation for power factor control, and rotor-orientation for computation of the voltage-control variables and self-commutation. There is also presented a speed-computation procedure used in practical implementation that is based on the incremental encoder position signal processing. Simulations were carried out in Matlab/Simulink® environment.

Keywords: synchronous motor, unity power-factor control, double field-orientation.

1. Introduction

For high performance dynamic applications the most suitable solution is the vector controlled (VC) AC drive fed by a static frequency converter (SFC). The wound-excited synchronous motor (Ex-SyM) is the only machine capable to operate at unity or leading power factor (PF). The structure of the vector control system is determined by the combination between the type of the SFC used including the pulse width modulation (PWM) procedure, the orientation field and its identification [2], [8], [9].

The rigorous control of the PF can be made only with the resultant stator-field orientation. If the PF is maximum, there is no reactive energy transfer between the armature and the three-phase power source.

In variable speed drive applications with Ex-SyM a shaft position sensor is usually required for natural commutation and for synchronization of the converter triggering pulses with respect to the rotor position θ [11].

In the proposed control structure both types of orientation are used: stator-field orientation for the control of the unity power factor and stator-flux, and also for the generation of the current control variables; rotor position (i.e. exciting-field) orientation for self-commutation and generation of the voltage control variables for the inverter control. The transition between the two orientations is performed by using a coordinate transformation block (CooT), which rotates the stator-field oriented reference frame with the value of the load angle ($\delta=\lambda_s-\theta$).

2. Angular speed computation

For the self-commutation of the Ex-SyM it is important an accurate information about the rotor position. This was realized using an incremental encoder, mounted on the motor shaft. The mounting is realized in a manner, that the encoder index signal is synchronized with respect to the rotor position. The encoder generates a number of pulses proportional to the angular position of the shaft. It gives information also about the sense of the rotating motion: positive values for direct and negative ones for reverse running. The counter resets it to zero at every full rotation. The amplitude of this saw-tooth signal will be equal to the encoder resolution. In order to obtain a continuous signal, useful in the present application, where the coordinate transformation blocks (CooT) use the sine and cosine of the angular positions, this θ_{enc} encoder position signal has to be processed accordingly, in order to obtain a position signal θ between $[0, 2\pi]$ for direct, and $[0, -2\pi]$ for reverse rotation respectively, based on the following expression:

$$\theta = \frac{\theta_{enc}}{N_r} 2\pi \tag{1}$$

where N_r is the number of increments per revolution (the encoder resolution).

Applying the sine and cosine trigonometrical functions to the above position signal, continuous sine wave signals are obtained which period is equal to 2π mechanical angle. On the other hand, the computation of the angular speed, as a derivative of the position θ, is not accurate because of the zero crossing, where the obtained speed signal is infinite, and it cannot be processed.

In the following, a simple procedure is presented (see *Fig. 1*) that avoids the zero crossings, and gives an accurate result.

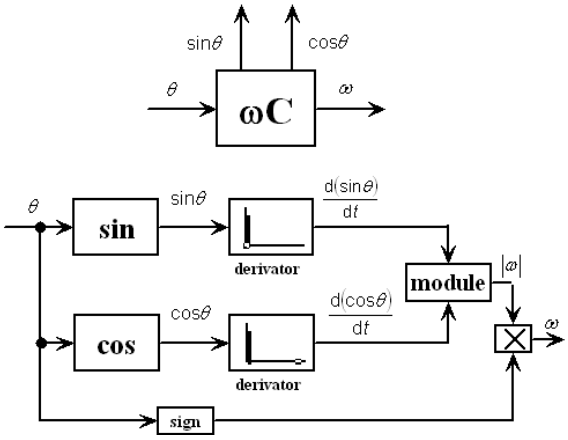


Figure 1: Block symbol and structure for computation of the rotor angular speed based on the encoder position signals.

The sign of the position signal θ will give the direction of the rotation. The computation of the angular speed is based on:

$$\omega = |\omega| sign(\theta) = \sqrt{\left(\frac{d(\sin\theta)}{dt}\right)^2 + \left(\frac{d(\cos\theta)}{dt}\right)^2} sign(\theta), \tag{2}$$

and its computation may be processed with the structure presented in *Fig. 1* [4].

4. Double field-oriented control of the Ex-SyM

In the proposed control structure, presented in *Fig. 2,* the salient pole Ex-SyM is fed by a voltage-source inverter (VSI), working with feed-forward voltage-PWM procedure. The exciting current is controlled, and its winding is fed by a DC chopper, that works also with voltage PWM. Three control loops are formed: two magnetical (for flux and PF control) and a mechanical one (for speed). The flux and the speed are directly controlled by PI controllers, and the PF is indirectly controlled. The PF is maximum, if the stator voltage and stator current are in phase. Consequently, the stator-flux vector $\underline{\Psi}_s$ results perpendicular onto the stator-current space phasor $\underline{i}_s$.[2], [8].

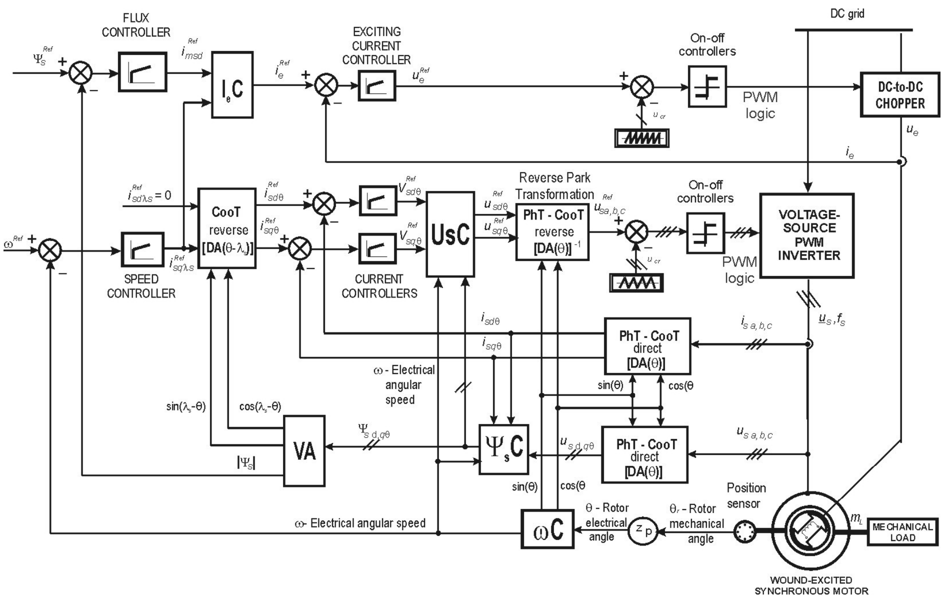


Figure 2: Vector control system of the adjustable excited synchronous motor fed by a static frequency converter with feed-forward voltage-PWM and double field orientation, operating with controlled stator flux and imposed unity power factor.

The perpendicularity can be achieved by canceling the stator-field-oriented longitudinal armature reaction ($i_{sd\lambda s}=0$). The reference armature-current components are oriented according to the resultant stator-flux phasor. The longitudinal component $i_{sd\lambda s}^{Ref}$ is an imposed value, and it is cancelled, while the quadrature component $i_{sq\lambda s}^{Ref}$ results at the output of the speed controller. But the stator-current control is recommended to be made in exciting field- (i.e. rotor-position-) oriented ($d\theta$–$q\theta$) rotating reference frame, because the self-commutation of the motor is made inherently by means of the reverse Park transformation block. The re-orientation (from the resultant stator field to the exciting one) is made by means of a reverse CooT block, which rotates the stator-field-oriented reference frame with the value of the load angle $\delta=\lambda_s-\theta$ [2], [8], [9].

The voltage-control variables are generated by the two current controllers in the active and the reactive control loops. In the voltage-computation block UsC the electromagnetic cross-effect is taken into account, realizing the re-coupling of the two decoupled control loops, the active and the reactive ones, by means of the rotating EMF components [2], [8].

The inverter control is made by means of a carrier-wave based voltage-PWM method, with simple on-off controllers [5].

In the third control loop the resultant stator-flux is directly controlled with a PI controller, which outputs the i_{ms} magnetizing current, necessary for the computation of the excitation current (IeC block) [2], [8].

The exciting winding is supplied from a DC-to-DC chopper working controlled by a feed-forward carrier-wave voltage PWM. With the before presented control structure simulations were performed for validation purpose.

4. Simulation results

Based on structure from *Fig. 2* simulations were performed in Matlab-Simulink® environment. The salient pole Ex-SyM rated data are: U_{sN} = 380 V, I_{sN} = 1.52 A, P_N = 800 W, f_N=50 Hz, n_N = 1500 [rpm], cosφ= 0.8 (capacitive).

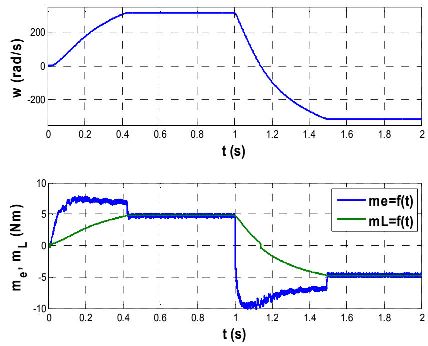


Figure 3: Electrical angular speed (***w***), electromagnetic (***m****e*) and load torque (***m****L*).

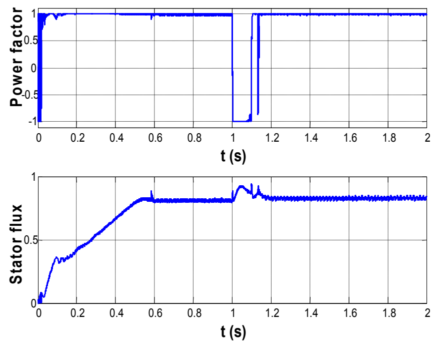


Figure 4: The power factor and the stator-flux amplitude versus time.

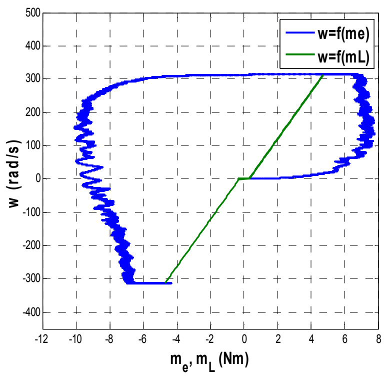


Figure 5: Mechanical characteristics*:* speed versus torque of the motor w=f(m_e) and of the mechanical load w=f(m_L).

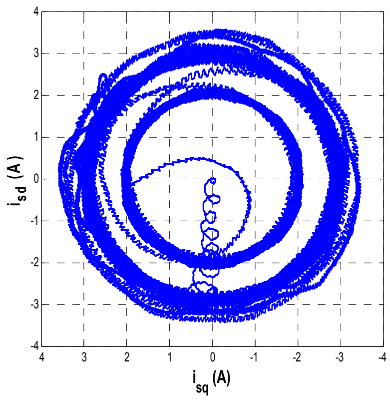


Figure 6: The trajectory of the armature-current in natural stator-fixed coordinate frame.

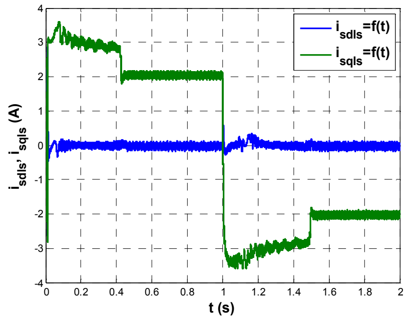


Figure 7. The armature-current two-phase components ($i_{sd\lambda s}$ and $i_{sq\lambda s}$) in stator-flux-oriented coordinate frame.

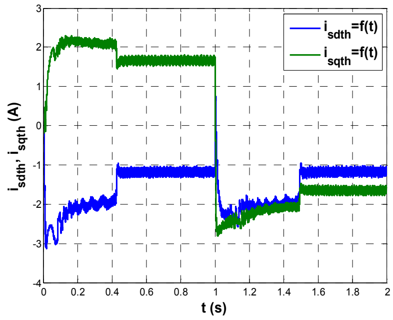


Figure 8. Armature-current two-phase components ($i_{sd\theta}$ and $i_{sq\theta}$) in the rotor-oriented reference frame.

After the starting process the motor will run at the rated speed value corresponding to a frequency of 50 Hz. At t= 1 s a speed reversal is applied, under the full rated load. The mechanical load has reactive character and it is linearly speed-dependent.

The simulation results show that the proposed control structure from *Fig. 2* is a viable one with improved performances with respect to the conventional VC systems [2].

The results show a good performance of the drive also in transient operation at starting, and also at speed reversal (*Fig. 3*). The power factor is maximum also during the speed reversal, when the drive is in regenerative running for a short period of time, as is shown in *Fig. 4*. Unity power factor is realized by canceling the stator-field-oriented longitudinal armature reaction, as in *Fig. 7*.

6. Conclusion

The presented control structure uses two types of orientations: resultant stator-field and exciting-field, i.e. rotor-position orientation.

For a rigorous control of the power factor, stator-field orientation was applied, and in order to achieve unity power factor operation on the reactive loop the stator-flux oriented longitudinal armature reaction was cancelled.

In order to obtain improved control performances, the computation of the control variables were made in rotor-oriented reference frame, so the self-commutation of the motor and the synchronization of the inverter trigger signal are made based on a directly measured quantity, namely the rotor position.

The presented control structure was validated by simulation in Matlab/Simulink®, and the obtained result shows the reliability of this method.

The practical implementation was realized on an experimental rig based on the dSpace DS1104 controller board. The results were published in [8].

References

[1] Kelemen, Árpád, Imecs, Mária, "Vector Control of AC Drives", Volume 1: *Vector* "Control of Induction Machine Drives", OMIKK-Publisher, Budapest, Hungary, 1991, ISBN 963 593 140 9.

[2] Kelemen, Á., Imecs, Mária, "Vector Control of AC Drives", Volume 2: "Vector Control of Synchronous Machine *Drives*" Ecriture Budapest, Hungary, ISBN963 593 140 9

[3] W. Leonhard, "Control of Electrical Drives", Springler Verlag. Berlin, Heidelberg, New York, Tokyo, 1985

[4] Szabó, Cs., "Implementation of Scalar and Vector Control Structures for Synchronous Motors "(in Romanian), PhD Thesis, Supervisor Prof. Maria Imecs, Technical University of Cluj-Napoca, Romania, 2006

[5] Kazmierkowski, M. P., Tunia, H., "Automatic Control of Converter-Fed Drives", Elsevier, Amsterdam, 1993.

[6] Vauhkonen, V.:, "A cycloconverter-fed synchronous motor drive having isolated output phases", *Proceedings of International Conference on Electrical Machines, ICEM '84, Lausanne, Switzerland*, 1984.

[7] Imecs, Mária, Szabó, Cs., Incze, I. I., "Stator-field-oriented control of the variable-excited synchronous motor: numerical simulation", 7th *International Symposium of Hungarian Researchers on Computational Intelligence HUCI 2006*, *Budapest, Hungary*, Nov. 24-25, 2006, ISBN 963 7154 54 X, pp. 95-106.

[8] Szabo, C., Imecs, M., Incze, I.I., "Vector control of the synchronous motor operating at unity power factor", *11th International Conference on Optimization of Electrical and Electronic Equipment, OPTIM 2008*, *Brasov, Romania*, 22-24 May, CD-ROM, ISBN 978-973-131-028-2, ISSN 1842-0133; Volume II-A: "Power Electronics, Electrical Machines & Drives" pp. 15-20, ISBN 978-973-131-030-5.

[9] Szabo, C., Imecs, M., Incze, I.I., "Synchronous motor drive with controlled stator-field-oriented longitudinal armature reaction", *The 33rd International Conference of the IEEE Industrial Electronics Society, IECON 2007*, *Taipei, Taiwan*, CD-ROM, IEEE Cat. No. 07CH37855D, ISBN: 1-4244-0783-4, ISSN: 1553-572X, and Conference Digest p. 105.

[10] Davoine, J., Perret, R., Le-Huy H., "Operation of a self-controlled synchronous motor without a shaft position sensor", *Trans. Ind Applications*, IA-19, No. 2, March/April 1983.

[11] Shinnaka, S., Sagawa, T., "New optimal current control methods for energy-efficient and wide speed-range operation of hybrid-field synchronous motor," *Trans. Ind Electronics,* Vol. 54, no. 5, Oct. 2007, pp. 2443-2450.

[12] Imecs, M., Incze, I. I., Szabó, Cs., "Double field orientated vector control structure for cage induction motor drive", *Scientific Bulletin of the „Politehnica" University of Timisoara, Romania, Transaction of Power Engineering,* Tom 53(67), Special Issue, 2008, ISSN 1582-7194, pp. 135-140.

[13] Imecs, M., Incze, I. I., Szabó, Cs., "Dual field orientation for vector controlled cage induction motors", *Proc. of the 11th IEEE Internat. Conference on Intelligent Engineering Systems, INES 2009, Barbados*, CD-ROM, ISBN: 978-1-4244-4111-2, pp 143-148.

[14] Imecs, M., Szabó, Cs., Incze, I. I., "Stator-field-oriented vector control for VSI-fed wound-excited synchronous motor", I*nternational Aegean Conference on Electrical Machines and Power Electronics, ACEMP-ELECTROMOTION Joint Conference, Bodrum, Turkey*, 2007, ISBN 978-975-93410-2-2, pp. 303-308.

Control System for Supercapacitor based ICE Catalyst Preheating

Cornel Aurel STANCA[1],Veneţia SANDU[2], Iuliu SZÉKELY[3]

[1] Faculty of Electrical Engineering & Computer Science, "Transilvania" University, Brasov, Romania, e-mail: cornelstanca@yahoo.com
[2] Faculty of Mechanical Engineering, "Transilvania" University, Brasov, Romania, e-mail: sandu@unitbv.ro
[3] Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, Romania e-mail: gszekely@ms.sapientia.ro

Manuscript received April 12, 2010; revised April 18, 2010.

Abstract: The environmental concern within Internal Combustion Engine leads nowadays towards updated management strategies implying benefits of new technologies and materials. The paper presents the implementation of a supercapacitor for the preheating of three way catalyst mounted on a spark ignited engine which lowers up to 5 times the specific emission during transient modes. Besides the thermal calculation of heat transfer, the paper stresses on calculation of parameters for preheating resistance, supercapacitor, converters and sensors. The phases of catalyst preheating are related to supercapacitor charging, charging preservation and discharging as well as with engine starting phases. An original control algorithm is defined between engine starting controls and catalyst preheating demands. Supercapacitor discharge time is interpreted according to variables: voltage and catalyst temperature.

Keywords: preheated ICE catalyst, supercapacitor, control system.

1. Introduction

In the operation modes of Internal Combustion Engines (*ICE*), the control of anti-pollutant devices has an important role. For spark ignition engines the catalyst efficiency is significantly affected by the air–fuel mixture and catalyst temperature. The control of air-fuel ratio is done by the Engine Control Unit (*ECU*) which keeps its value as close as possible to 14.7:1, the stoechiometric value for gasoline.

Fig. 1. illustrates a closed-loop control system for air-fuel mixture. The oxygen content in exhaust gas is measured with a lambda sensor upstream the catalytic converter.

Data as oxygen content, combustion air flow and temperature, combustion chamber pressure and acceleration pedal position allow the appropriate injection of fuel for any operation mode.

A key element in exhaust system is the catalytic converter or, for spark engines, the three way catalyst (*TWC*) which transforms carbon oxide (CO), gaseous hydrocarbons (HC) and nitrogen oxides (NO_x) in non-toxic products such as carbon dioxide (CO_2), water (H_2O) and nitrogen (N_2) [1].

The efficiency of *TWC* depends on the catalyst temperature, so it is vital to rapidly preheat the catalyst at engine starting in order to reduce the pollutants. When there is no catalyst preheating, the regulated exhaust gas pollutants (CO, HC, NO_x) are 2…5 times greater than in case of a heated catalyst [2].

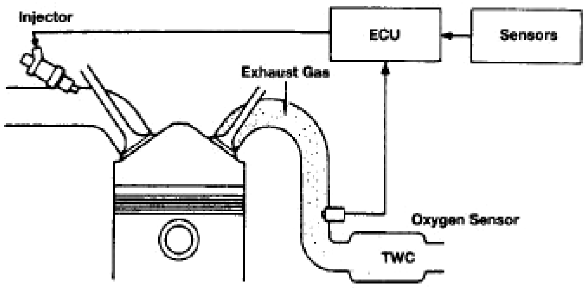


Figure 1: Schematic of the closed-loop air-fuel ratio control.

2. Control system of catalyst preheating

The control system of catalyst preheating contains a supercapacitor which is able to rapidly heat the *TWC* up to minimum operation temperature. So at engine cold starting with a preheated catalyst the exhaust gas pollutants reduction is important, as the *TWC* will work close to rated operation mode even from engine starting.

The electric supply for the supercapacitor is delivered by vehicle alternator during the engine operation. During acceleration, the supercapacitor is disconnected from the alternator and during stopping (before starting) the supercapacitor is kept charged by the starter battery. In the latter situation, the battery is not overloaded, as it delivers low currents of several mA (the supercapacitor leakage current).

The *TWC* preheating up to light-off temperature (temperature at which the catalyst efficiency is 50%) of around 250°C is done using an electric resistor on which the energy accumulated in a special dedicated supercapacitor is discharged.

The block schematic is presented in *Fig. 2*, the meaning of each block being as follows:

- R_{PH}, *TWC* preheating resistance;
- SC_{PH}, *TWC* preheating supercapacitor;
- *TS*, temperature sensor;
- K_{PH}, static switch which connects/disconnects the resistance to the supercapacitor;
- K_E, vehicle starting switch, engine function position;
- *DCDC-PH*, DC to DC converter from *12.06V* (the voltage of the battery charged to 50% of its capacity) to 48.6V (rated supercapacitor voltage);
- *SLPH* – (LED) signal light for TWC preheating;
- *μC*, microcontroller of management system.

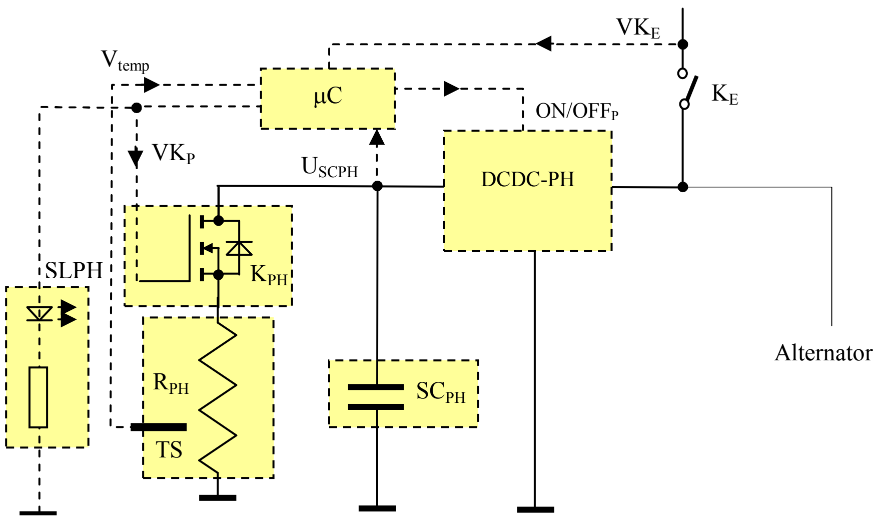


Figure 2: Block schematic of controlling TWC preheating system.

According to the initial catalyst temperature measured by the *TS* sensor the supercapacitor SC_{PH} is completely or partially discharged on the R_{PH} resistor, when the system processor controls it (before engine starting, at closing of the K_E switch). Discharge time is calculated according to the initial temperature, considering a *250°C* final temperature. The discharge is signaled by lighting the lamp *SLPH* and it triggers a dedicated switch in the vehicle starting system.

The supercapacitor SC_{PH} charging is done during a post-start phase, from a *DC-DC* converter *DCDC-PH*, when the active source is the alternator. During the Prestart phase SC_{PH} is maintained at the nominal voltage of *48.6V*, through the same *DCDC-PH* converter.

3. Sizing the main system components

A. Supercapacitor selection

In order to select the supercapacitor, it is necessary to know the energy required to heat the catalyst. For a stainless steel catalyst of 1 liter volume, the energy required to heat up to *250°C* from ambient temperature is as follows:

$$W_{\Delta t=250^\circ C} = \rho \cdot V \cdot c \cdot \Delta t = 0.8 kg/dm^3 \cdot 1 dm^3 \cdot 500 J/kg^\circ K \cdot 250^\circ K = 100 kJ \quad (1)$$

in which ρ is material density, V – catalyst volume, c specific heat and Δt – temperature increase.

For a ceramic catalyst the energy required is:

$$W_{\Delta t=250^\circ C} = \rho \cdot V \cdot (\frac{1}{8} c_{cer} + \frac{7}{8} c_{aer}) \cdot \Delta t = 0.5 kg/dm^3 \cdot 1 dm^3 \cdot (\frac{1}{8} \cdot 1465 + \frac{7}{8} \cdot 1010) J/kg^\circ K \cdot 250^\circ K = 133.36 kJ \quad (2)$$

in which c_{cer} is the specific heat of the ceramic material and c_{aer} is the specific heat of the air in the temperature interval of *20-250°C*.

The energy for TWC preheating can be obtained using a BMOD0165 P048 Maxwell supercapacitor module [3] shown in *Fig. 3*, with the characteristics in Table 1.

Figure 3: Supercapacitor module BMOD0165 P048.

Table 1: Supercapacitor characteristics.

C (F)	U (V)	ESR,dc (mΩ)	t_o (°C)	P (W/kg)	W (Wh/kg)	I_{leak} (mA)	I_{scc} (A)	$I_{max,cc}$ (A)
165	48.6	7,1	-40÷65	3200	3.81	5.2	4800	150

It can deliver *194 kJ* energy when discharged from *48.6V to 0V*, sufficient for both types of catalysts.

B. DC-DC Converter components

In order to manufacture the voltage-rising DC-DC converter, the circuit LT3751 is used [4] as controller of the capacitor charging (*Fig. 4*). The selection of the operation mode is performed applying a voltage on the FB terminal, in the *0-1.16V* range.

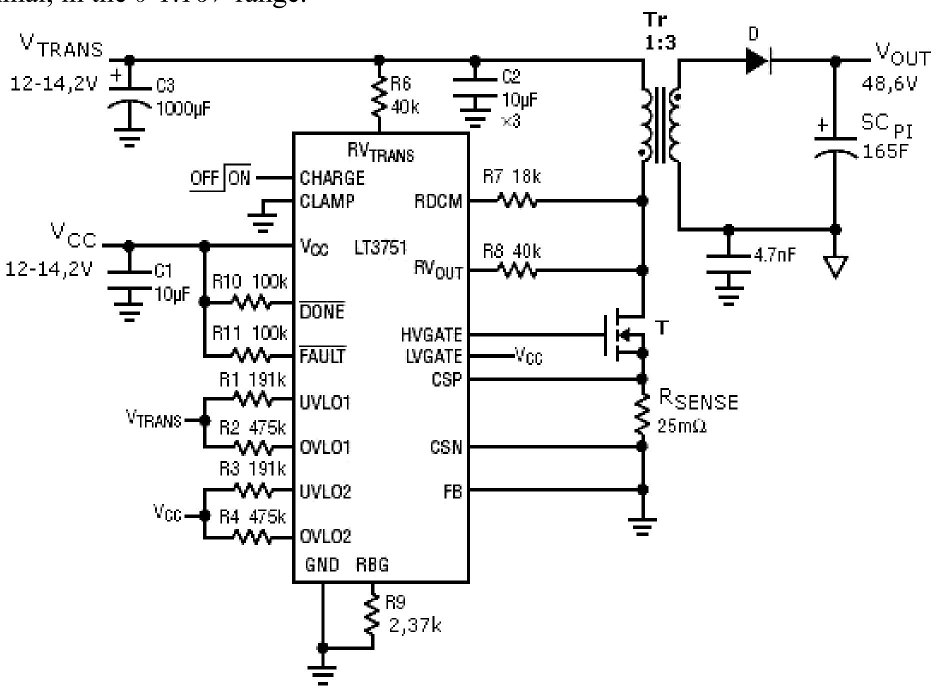


Figure 4: Electronic schematic of DC-DC 12-14.2V/48.6V converter for supercapacitor loading at 165F/48.6V.

For input parameters: $V_{TRANS}=12\div14.2V$; $V_{OUT}=48.6V$; $C_{OUT}=165F$, the transformer's ratio (transformation factor N) results as follows:

$$N \le \frac{V_{OUT}}{V_{TRANS}} = \frac{48.6V}{14.2V} = 3.42\,. \tag{3}$$

A transformer with the following characteristics has been chosen: *N=3;* $I_{PRI(MAX)}=10A$, $L_{PRI}=10\mu H$.

The peak current in the primary winding is:

$$I_{PK} = \frac{3\mu s \cdot V_{OUT}}{L_{PRI} \cdot N} = \frac{3\mu s \cdot 48.6V}{10\mu H \cdot 3} = 4.86A \tag{4}$$

The charging time t_{CHARGE} for charging from *0V* to the rated value of *48.6V* is:

$$t_{CHARGE} = \frac{(2 \cdot N \cdot V_{TRANS} + V_{OUT}) \cdot C_{OUT} \cdot V_{OUT}}{\eta \cdot V_{TRANS} \cdot I_{PK}} = \frac{[2 \cdot 3 \cdot (12 \div 14.2) + 48.6] \cdot 165 \cdot 48.6}{0.7 \cdot (12 \div 14.2) \cdot 4.86} = 23689 \div 22210s \tag{5}$$

in which η is the transformer efficiency, with a typical value of *70%.*

As from the charged energy of the supercapacitor SC_{PH} only a part is used, the supercapacitor voltage does not drop to *0V* at the end of preheating time. The final voltage U_f is calculated with the formula:

$$U_f = \sqrt{U_{SC0}^{\;2} - \frac{2 \cdot Wc}{C}}\,, \tag{6}$$

in which U_{SC0} is the initial open-circuit voltage of the supercapacitor (*48.6V*), Wc – consumed energy for preheating (for metallic catalysts 100 kJ and for ceramic catalysts *133 kJ*), C – supercapacitor capacity (*165F*). It results a final voltage on supercapacitor of *33.9V* for metallic catalysts, respectively of *27.4V*, for ceramic catalysts.

The variation of the supercapacitor voltage at charging is approximately linear, especially at voltage values close to final voltage, like it can be observed in the example of *Fig. 5*, in the case of a *1.2mF* capacitor, charged from *0* to *500V* [4]. Thus, the charging time can be estimated with the formula:

$$t_i = \frac{U_{SC0} - U_f}{U_{SC0}} t_0\,, \tag{7}$$

in which t_0 is charging time from *0V* to final voltage U_{SC0}. There result the charging times of *1h: 52min – 2h* in case of metal in catalysts, respectively, *2h: 43min – 2h:53min* in ceramic catalysts.

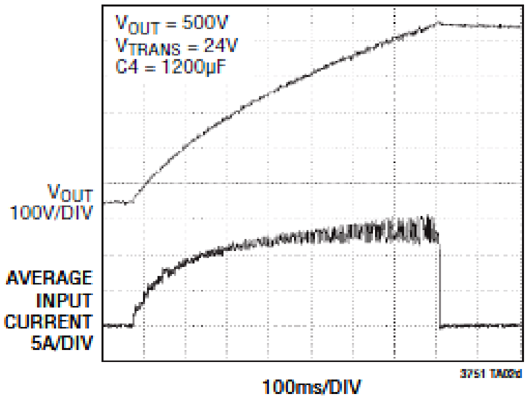


Figure 5: Waveforms of input current and output voltage during charging of a capacitor with 1.2mF capacity and 500V final voltage, with the LT3571 circuit.

The supercapacitor SC_{PH} charging is done during the *Poststarting* phase, being kept charged during *Prestarting*. As charging current is low, maximum *4.86A*, the alternator is not very loaded. During *Acceleration* mode the charger disconnects from alternator to make power available.
The resistance R_{BG} has the value:

$$R_{BG} = \frac{0.98V \cdot RV_{OUT}}{\dfrac{V_{OUT,TRIP} + V_{DIODE}}{N} - V_{TRANS}\left(\dfrac{RV_{OUT}}{RV_{TRANS}} - 1\right)} = \frac{0.98V \cdot 40k}{\dfrac{49.6V}{3}} = 2.37k \tag{8}$$

The drain-source transistor voltage $V_{DS(MAX)}$ is calculated with the formula:

$$V_{DS(MAX)} \geq V_{TRANS} + V_{OUT}/N = 14.2V + 48.6V/3 = 30.4V, \tag{9}$$

and the average drain current $I_{D(AV)}$ is calculated with the formula:

$$I_{D(AV)} \geq I_{AVG,M} = \frac{I_{PK} \cdot V_{OUT(PK)}}{2 \cdot (V_{OUT(PK)} + N \cdot V_{TRANS})} = \frac{4.86A \cdot 48.6V}{2 \cdot (48.6V + 3 \cdot 12V)} = 1.4A \tag{10}$$

A FDS2582 transistor is suitable, having I_D=4.1A and $V_{DS(MAX)}$=150V.

The maximum reverse (repetitive) diode voltage V_{RRM} is given by the formula:

$$V_{RRM} \geq V_{OUT} + N \cdot V_{TRANS} = 48.6V + 3 \cdot 14.2V = 91.2V \qquad (11)$$

and the average diode current $I_{F(AV)}$ is given by the formula:

$$I_{F(AV)} \geq I_{PK}/2N = 4.86A/6 = 0.81A. \qquad (12)$$

An ES1G diode is suitable, with $I_{F(AV)=}$1A and V_{RRM}=400V.

C. Selection of the preheating resistor

The dimensioning of TWC preheating resistor is based on the formula:

$$R_{PI} = -\frac{t_d}{C \cdot \ln \frac{U_{SCf}}{U_{SC0}}}, \qquad (13)$$

in which U_{SC0} is the supercapacitor initial voltage U_{SCf} is the supercapacitor final voltage and t_d is the discharge time. For a reasonable discharge time of 20s, sufficient for a heat transfer from preheater to the catalyst (with metal catalyst support), with $U_{SC0}=48.6V$, $U_{SCf}=33.9V$ and $C=165F$ the resistance is *376mΩ*. For ceramic catalyst support with $U_{SC0}=48.6V$, $U_{SCf}=27.4V$ and $C=165F$ the resistance is *212mΩ*.

The power dissipated on the series equivalent resistances of the supercapacitor and of the power switch can be neglected as having very low values (7.1mΩ, respectively 2.5 mΩ) compared to R_{PH} resistance.

D. Selection of the power switch

Considering the high value of discharge current required by the supercapacitor, the circuit BTS555 was chosen [5]. This circuit has the following parameters - important for our implementation:

- Working voltage: 5-34V;
- Conductive resistance 2.5mΩ;
- Load current: 165A
- Limit of short-circuit current: 520A.

E. Selection of the temperature sensor

As TWC temperature can reach 900°C, a K type thermocouple is the most suitable for temperature measurement, due to its 0÷1250°C [6] operating range.

In order to avoid special cables and sockets, a K type thermocouple connected to THAK 1250 amplifier can be used, with the following characteristics: supply voltage 5-16V, output signal 0-5V for temperatures within 0 and 1250°C.

F. Selection of the microcontroller

The 2 analog inputs, 1 digital input and 2 digital outputs can be controlled by an algorithm hosted by a PIC 18F4480/4580 microcontroller dedicated to auto appliances. This microcontroller offers capabilities like hardware 8 x 8 bit multiplier, serial in-circuit programming and debugging with only 2 pins, a serial port supporting SPI (as Master) and I2C (as Master or Slave), an USART supporting RS-485, RS-232 and LIN/ J6202, a CAN module supporting the CAN 1.2, CAN 2.0A, CAN 2.0B (passive and active) protocols up to 1Mbps, 5 bidirectional I/O ports, 11 analog channels with 100ksps and 10 bit ADC, 4 timers (3 of them on 16 bits) etc. All these allow this microcontroller to host also the algorithm for the control of the starting system [7].

4. Control system operation

The control system of TWC preheating is represented in *Fig. 6*. In the prestarting phase, the normal position of static contactors implied in the system is: K_{PH} – open, K_S – open (switch for controlling the starting time, connected in series with the K_{start} switch), K_M – open, Kstart – opens and converter control signal DCDC-PH , CHARGE ON/OFF activated (ON), in the charge preservation mode. Switching off K_M at t_{ppk} time (prestarting time when engine starting switch is ON, before starting), the TWC preheating is initiated stopping the supercapacitor charging / charging preservation mode (CHARGE ON/OFF dezactivated OFF) and closing K_{PH} switch which allows supercapacitor discharge on preheating resistor.

Until the elapse of the time required for heat transfer, necessary to reach the 250°C TWC temperature, any attempt of engine starting by switching the key from K_M in K_{start} (t_{pk} – key starting time) is disabled by the K_S switch which controls the effective starting moment.

At the t_{start} moment when the energy transfer is over, K_S is closed (“on” - validating the engine starting). In the same moment K_{PH} opens. If K_{start} is closed (“on”), the starting phase begins in the moment of switching on (closing) K_S If not, the starting phase begins only when K_{start} is closed (“on”).

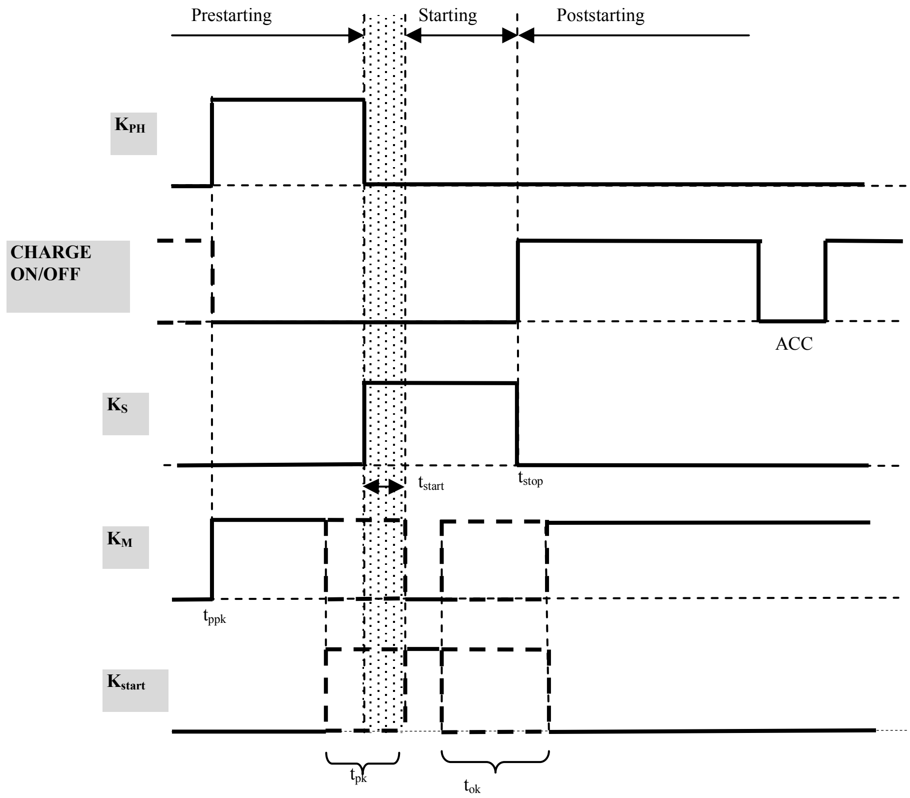


Figure 6: Time diagram in catalyst preheating.

The starting phase ends with the starter reaching the “free-run” mode (no more mechanical load as the engine just started), by stopping its supply opening K_S (“off”) at the t_{stop} moment.

Beginning of the post-starting phase does not depend anymore on the contact-key position (then no more with the t_{ok} time, the stopping time of K_{start}, at the moment of switching the contact key from starting position to normal engine operation position).

In the acceleration mode, the SC_{PH} supercapacitor charging is stopped to allow the engine to use its whole power.

In order to estimate the moment when SC_{PH} supercapacitor discharge can be stopped, the following procedure is performed:

- The energy required to heat the catalyst up to 250°C is computed based on the initially measured catalyst temperature T_0, with the formula:

$$W_{nec} = \rho \cdot V \cdot c \cdot (250 - T_0) = K \cdot (250 - T_0), \tag{14}$$

where ρ is material's density, V – catalyst volume, c – material's specific heat. For metal catalysts K=400 J/°K and for ceramic ones K=533.44 J/°K;

- The total energy stored in the supercapacitor is computed -considering the open-circuit voltage at terminals before discharging U_{SCPH0} - with the formula

$$W_{tot} = \frac{C}{2} \cdot U_{SCPH0}{}^2 \tag{15}$$

- The consumed energy W_c is calculated iteratively, after every sampling, based on supercapacitor voltage, U_{SCPH}:

$$W_c = W_{tot} - \frac{C}{2} \cdot U_{SCPH}{}^2 \tag{16}$$

and it is compared to the required calculated energy W_{nec}. When values become equal, SC_{PH} is disconnected.

5. Conclusions

The catalyst preheating system improves catalyst efficiency and extends functionality during transient operation modes.

The key element in catalyst preheating is the supercapacitor which is charged from the alternator (with current values under 5A) and then rapidly discharges (with current values of tens of hundreds of Amps) on a resistor. The heat released by resistor is transferred to the catalyst in order to rapidly reach the optimum temperature.

Catalyst temperature and supercapacitor voltage are controlled in prestarting phase, because they determine the discharge time. Higher catalyst temperature shortens discharge time in order to avoid catalyst overheating; lower supercapacitor voltage increases discharge time in order to reach light-off temperature, around 250°C.

The system makes starting dependent on the end of catalyst preheating and informs the driver about preheating by on-board signaling.

The most expensive component of the system is the supercapacitor; nevertheless, on the extent of maturing manufacturing technologies and considering the importance of environmental protection, the costs matter will become collateral and the system will be increasingly implemented.

Acknowledgements

The authors wish to thank to Romanian Research and Education Ministry, the financer of the project "Energy-Optimized Electrical Systems for Land Transport Using Batteries and Supercapacitors: TRANS-SUPERCAP" in the frame of the Program Partnerships, PNII – P4, no. 21-018/2007, which supported some facilities used in the experimental tests.

References

[1] Burch, S.D., Potter, T.F., Keyser, M.A. and Benson, D.IS., *Progress in Emission Control Technologies*, SAE Technical Paper Series SP-1053, Paper #941998, pp. 137-142.

[2] Joumard R.et al - Emissions unitaires de polluants des vehicules legeres, Rapport INRETS 116, 1990

[3] Maxwell Technologies, Top 10 Reasons for Using Ultracapacitors in Your System Design, White Paper, 2008.

[4] Datasheet of LT3751 - High Voltage Capacitor Charger Controller with Regulation, Linear Technology, http://www.linear.com/pc/productDetail.jsp.

[5] Smart Highside High Current Power Switch Reversave, PROFET® Data Sheet BTS555, Infineon Technologies AG 2003, www.datasheetcatalog.com.

[6] http://www.atis.com.cn/pdf/Texys/4_thermocouple_amplifier.pdf, Texense® Thermocouple Amplifier K type.

[7] Stanca, A.C., Sandu V., Szekely I., "Embedded System for Vehicles Start-of-ICE with Hybrid Battery-Supercapacitor Source", Proceedings of the 10th International Conference on Development and Application Systems – „DAS 2010", Suceava.

Investigation of the Solenoid Switch of an Electric Starter Motor Using FEM Model

Levente SZALONTAI[1], Ernő KOVÁCS[2], Csaba BLÁGA[3]

[1]Department of Electrical and Electronic Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc, Hungary, e-mail: elkszl@uni-miskolc.hu
[2]Department of Electrical and Electronic Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc, Hungary, e-mail: elkke@uni-miskolc.hu
[3]Department of Electrical and Electronic Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc, Hungary, e-mail: elkblaga@uni-miskolc.hu

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: This paper introduces the investigation and a simulation of the solenoid switch of an electric starter motor that is used in automobile cars. The structure of the electric starter motor is examined focusing mainly on the architecture of the solenoid switch. In a new approach the simulation has been carried out using in the MATLAB programming environment the finite element model of the solenoid switch built with the FEMM program. The applied equations are presented together with the results.

Keywords: starter motor, solenoid switch, electromagnet, simulation, FEMM and MATLAB model

1. Introduction

Starter motors are widely used in automobiles from the twenties of the last century. The pistons of the Otto type and also of the Diesel type of engines should be pre-moved before the engine can be self-sustaining. This kind of piston pre-movement is ensured by the starter motor generally through a releasable gear joint. The connecting of the gears is commonly done by a linear solenoid switch, which is also connecting the poles of the starter motor to the battery - when the solenoid is energized [4] - , so the starter motor gets under voltage and can drive the piston's shaft.

The aim of the project is to learn more about the solenoid switch behaviour in time, also about the internal parameters, forces, etc. With the better understanding of the functionality of the solenoid switch we are going to be able to develop higher efficiency solenoid switches, and also cost effective ones.

In this article some selected simulation results are presented, which have been obtained with a method which is different from those used in [1,5], i.e.

using Finite Element Method Magnetics- a free FEM software- together with the Matlab programming environment.

2. Starter motor with the solenoid switch

The starter motor is shown in the *Fig. 1*. The five main parts are: driving gear (1), clutch mechanism (2), solenoid switch (3), IC (4), DC motor (5).
The parts of the solenoid switch are shown in the *Fig. 1.* Connection housing (6), clip (7), IC (4), solenoid (8), iron core with spring (9), metal housing (10).
After ignition the solenoid is excited and it magnetizes the moving iron core, which accelerates towards the solenoid. As the iron core reaches the maximum displacement, it connects the gears and closes the circuit of the DC motor. It was stated formerly that the IC has the role of protecting the DC motor and the coil, including thermal protection [2].

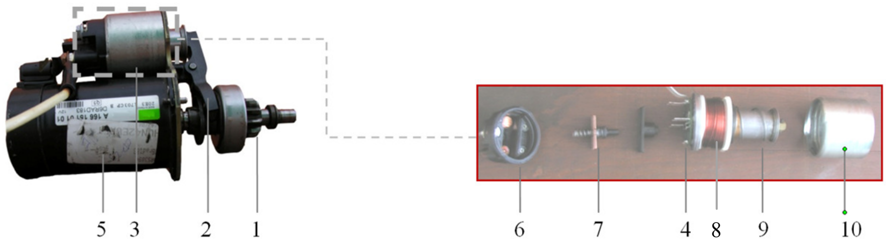


Figure 1: Starter motor and the solenoid switch.

3. Matlab and FEMM model

To model and simulate the pull-in process we need the equations which describe the system. The system consists of a mechanical part, and of an electrical part. Considering these we have the mechanical equation of the accelerating iron core (1), and the voltage equation of the solenoid winding (2) [4].

$$m \cdot \frac{d^2x}{dt^2} = F_{mag} - F_{friction} - F_{spring} \tag{1}$$

$$V = i \cdot R + \frac{d\Psi}{dt} \tag{2}$$

In (1) the $F_{friction}$ force is calculated as (3) and the force of the springs can be calculated after (4,5,6):

$$F_{friction} = c \cdot \frac{dx}{dt} \tag{3}$$

where (c) is the friction constant, and (x) is the displacement, (m) is the mass of the iron core. The (F_{spring}) force can be calculated with the appropriate spring constants *(k_i, $i=1,2,3$)* as follows:

- if $0 \leq x \leq 11[\mathrm{mm}]$:

$$F_{spring} = k_1 \cdot x \tag{4}$$

- if $11 < x \leq 14[\mathrm{mm}]$:

$$F_{spring} = (k_1 + k_2) \cdot x - 0{,}011 \cdot k_2 \tag{5}$$

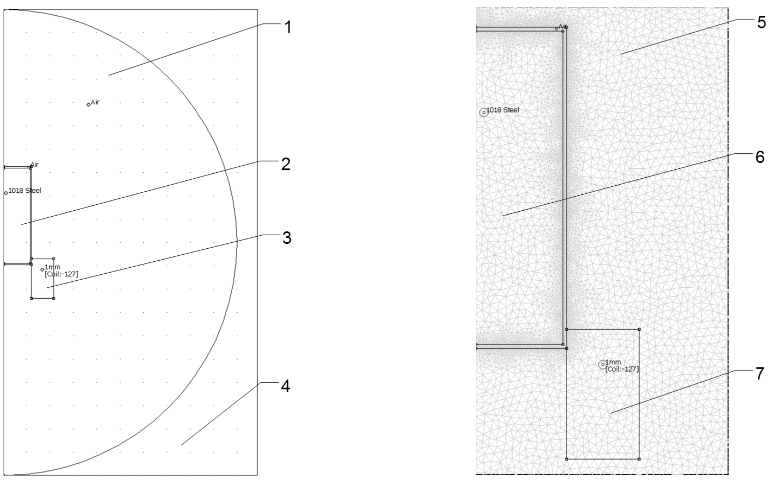
- if $14 < x \leq 14{,}9[\mathrm{mm}]$:

$$F_{spring} = (k_1 + k_2 + k_3) \cdot x - 0{,}011 \cdot k_2 - 0{,}014 \cdot k_3 \tag{6}$$

The magnetic attractive force (F_{mag}) - which acts on the iron core – will be calculated in the FEMM model. In the eq. (2) (V) is the terminal voltage of the solenoid, (i) is the current flowing through the winding, (R) is the ohmic resistance of the winding, (Ψ) is the flux of the solenoid which also can be written as in (7) with the help of the inductance (L) of the solenoid switch and with the current of the winding (i):

$$\Psi(t) = L(t) \cdot i(t) \tag{7}$$

The result of (1) and (2) is that a second order differential equation must be solved in Matlab. After solving the equation we get the value of the current and the value of the displacement, which can be transferred into the FEMM model.



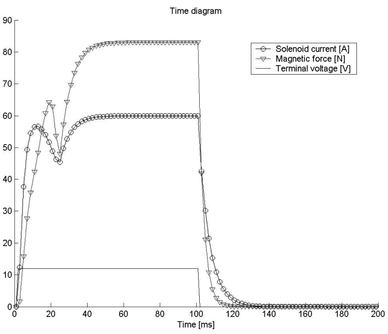
a) *b)*

Figure 2: FEMM model *a)* with boundary; *b)* close-up with mesh.

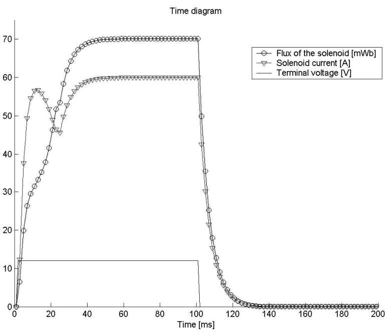
The FEM model is created as an axisymmetric problem (*Fig. 2.a*). The boundary region has a radius of 100 mm, on *Fig. 2.a.* (4) denotes the region that is outside of our interest. Inside the region as the environment we defined air (1,5) *Fig.2.* As material for the iron core (2,6) we used 1018 low carbon steel. In *Fig.2.*, numbers (3,7) indicate the coil, which is a 127 turns one, with 1 mm diameter copper wiring.

4. Simulation results

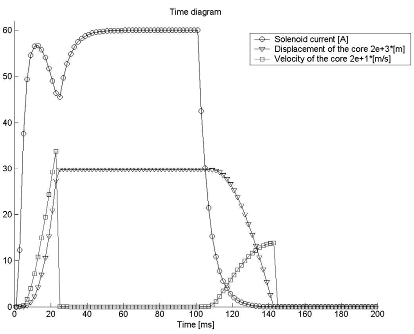
The following constants were considered: simulated timeframe 0,2s, time step size 0,001s, m=0,45kg, R=0,2Ω, k_1=1000N/m, k_2=13000N/m, k_3=5200N/m, max displacement 14,9mm, V=12V. Simulating a pull-in and release cycle of the solenoid switch we get the results shown in *Fig.3*.



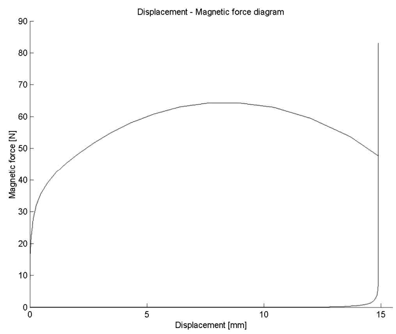
a)



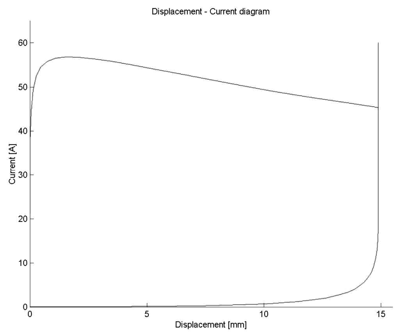
b)



c)



d)



e)

Figure 3: Simulation results.

Fig.3.a. shows the variation of the current of the winding (i), and the magnetic force acting on the iron core (F_{mag}) in function of time. It also shows the terminal voltage (V) of the solenoid which generates the current. In *Fig.3.b.* the flux of the solenoid (Ψ), the current and the terminal voltage are shown side-by-side also in function of time. *Fig.3.c.* shows the simulation results for the displacement (x) together with the current and the calculated velocity of the iron core. *Fig.3.d.* and *Fig.3.e.* show the current of the winding (i) and the magnetic force (F_{mag}) acting on the iron core in function of displacement (x).

5. Conclusion

The results of the simulation using a FEM model together with Matlab are promising, in fact in the literature the same current-time diagrams are mentioned [1, 3, 5].
With our simulations presented here it is possible to get more information, data on the solenoid and its behaviour. Thus it will open novel ways for us to redesign, simulate and for further develop solenoid switches.
In this article we presented our new simulation results on solenoid switches, made with FEMM and Matlab programs. Nevertheless, the simulation results need to be approved by measurements.

Acknowledgements

The authors highly appreciate the generous support of the BAROSS_EM07-EM_ITN3_07-2008-0039 project of National Office of Research and Technology.

References

[1] Blága, Cs.: "Lineáris elektromágnesek működésének számítógépes szimulációja", *Elektrotechnika, Magyar Elektrotechnikai Egyesület, 2005. 98.évf. 3.szám 73-76.*

[2] Blága, Cs.: "Hány ohmos az indítómotor behúzótekercsének ellenállása?", *Autótechnika, 2009. 7. szám 16-18 HU-ISSN 1588-9558.*

[3] Vishniac-Bush: "Electromechanical sensors and actuators", *Springer Verlag, 1999.*

[4] Fodor, Gy.: "Elméleti elektrotechnika", *Tankönyvkiadó, Budapest 1979.*

[5] Kovács, E., Szalontai, L., Füvesi, V.: "Investigation of the Solenoid Switch of an Electric Starter Motor with Simulation", *X. International Conference on Energetics-Electrical Engineering, Kolozsvár, Romania, 2009.*

Performance Curve of an Alternator Based on Mathematical Model

Ernő KOVÁCS [1], Csaba BLÁGA [2]

[1] Department of Electrical Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, e-mail: elkke@uni-miskolc.hu
[2] Department of Electrical Engineering, Faculty of Mechanical Engineering and Informatics, University of Miskolc, e-mail: elkblaga@uni-miskolc.hu

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: This paper presents a simulation model of an alternator in order to get its performance characteristic. Studying the construction and operation of alternators, a mathematical model was created and simulation can be carried out. It can be observed that the alternator is a synchronous generator and its amplitude and frequency of the induced alternating voltage changes as a function of the shaft revolution speed. The saturation of magnetic flux and frequency dependence of the impedance have great influence on the performance characteristic. The paper discusses briefly the influence of different parameters on the design process.

Keywords: automotive electric and electronic systems, alternator, simulation, performance curve.

1. Introduction

One of the most important characteristics of an alternator is the performance curve that distinguishes it from the DC generator called dynamo. This curve has a special shape that gives an advantageous behaviour to the alternator and this is the main reason why the alternator has totally substituted the dynamo in the current generating process at a car. The role of the battery is indisputable; it gives electric power when the internal combustion engine does not work. The battery has to be charged, so a generator is needed. As the battery needs DC current it would be easier to use DC generator, but the performance curve of it needs a lot of control and also does not supply enough charge at low rpm, so in urban traffic especially in winter time could result some inconvenient surprises: the battery remains uncharged and the car stops.

We can observe that the alternator starts to supply current at lower rpm and does not need current control because of the special shape of its performance curve. So in spite of the fact that alternator generates alternating current, it became more advantageous than dynamo. The low price and the better behaviour of diodes gives the possibility of rectifying the AC current and the charging of battery has been solved.

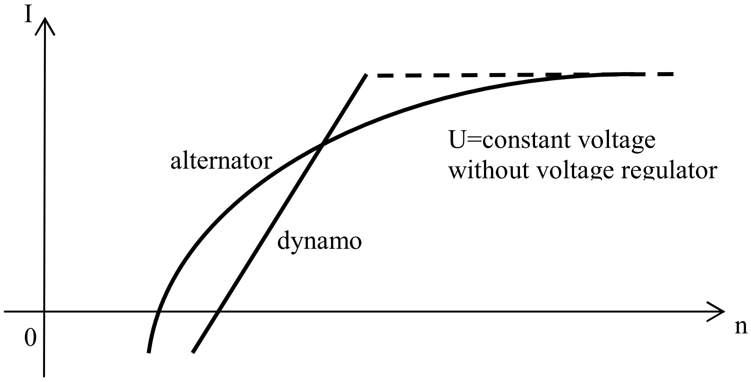


Figure 1: Performance curves [1].

Application of diodes eliminated the current relay, so one reason more to use alternator. Some other facts on the side of alternator are the followings: the current is induced in the stator, so it can be cooled better, and it can produce higher power; it has better efficiency, there are no commutators, the sense of rpm has no effect (only the wings of the ventilator should be taken into consideration), smaller size, less weight, less noise.

2. Improved model

In order to be able to simulate the performance curve of an alternator we should know the construction and the operation of it. The alternator basically is a synchronous generator as it is described in many books, for example [2].

A simulation model is presented in paper [3]. An improvement of this model has been carried out by introducing:

-the saturation of magnetic flux [4] and

-the effect of the revolution speed on the internal impedance of the synchronous generator.

The induced voltage depends on the magnetic flux. The saturation of magnetic flux has a great influence on the operation of an alternator. Moreover the frequency depends on the rpm of the shaft. The internal impedance also depends on the rpm.

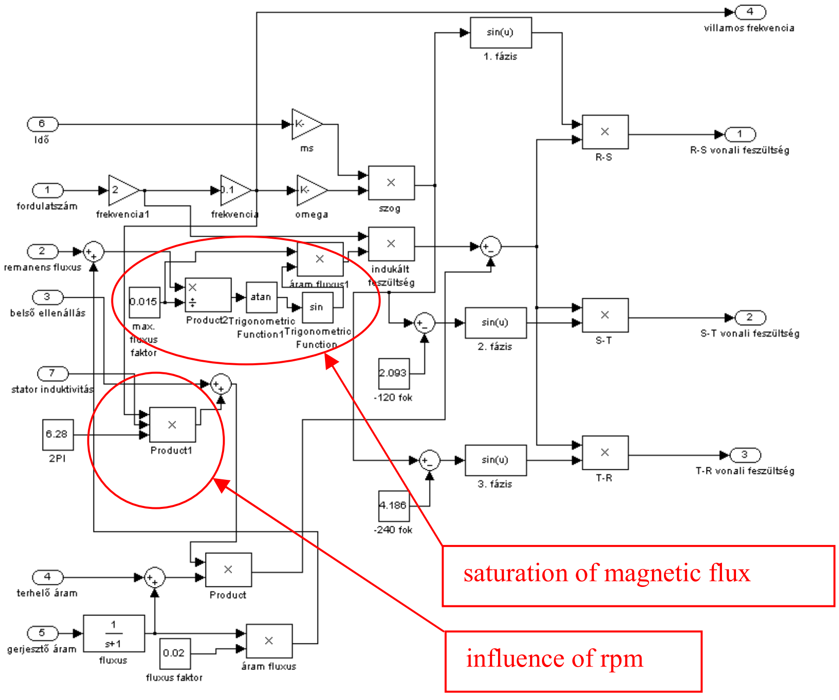


Figure 2: Improved model of synchronous generator.
Inputs: 1- rpm, 2 - remanent flux, 3 - internal resistance, 4 - load current, 5 - excitation current, 6 - time, 7 - inductance of stator
Outputs: 1 - line voltage R-S, 2 - line voltage S-T, 3 - line voltage T-R, 4 - frequency.

The rectifier model is the same as presented in [3].

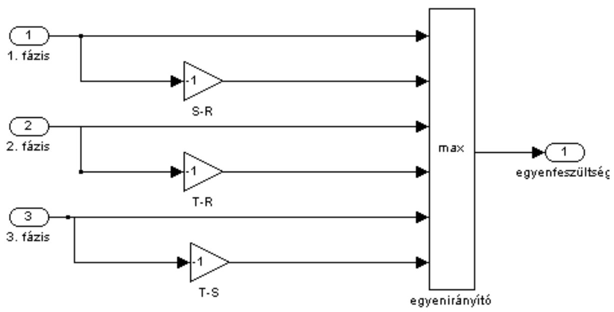


Figure 3: Rectifier.
Inputs: 1 - phase 1, 2 - phase 2, 3 - phase 3. Output: DC voltage.

The voltage regulator also has changed by introducing the effect of inductance influenced by rpm as it is presented in *Fig. 4*.

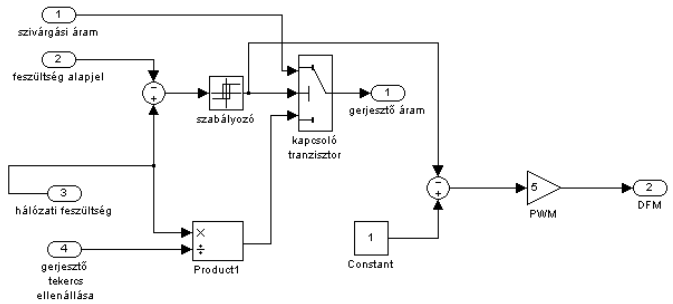


Figure 4: Voltage regulator.
Inputs: 1 - leakage current, 2 - voltage reference, 3 - network voltage, 4 - resistance of the excitation circuit, 5 - frequency, 6 - inductance of the excitation circuit, Outputs: 1 - excitation current, 2 - DFM.

The model of the battery was also modified.

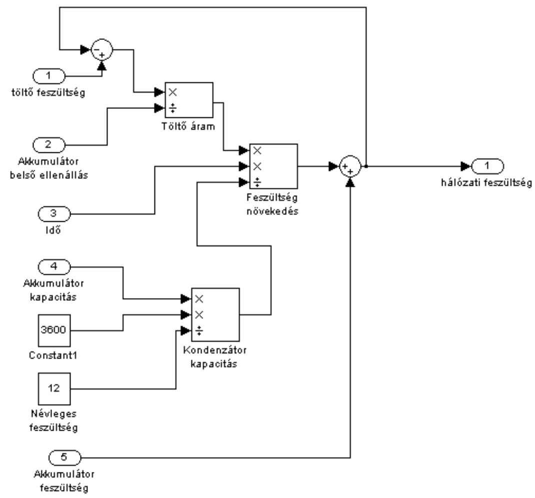


Figure 5: Model of the battery.
Inputs: 1 - charging voltage, 2 - internal resistance of battery, 3 - time, 4 - capacity of the battery, 6 - internal voltage of battery
Output: 1 - network voltage.

In such a way we got all the most important elements that build up the model of an alternator charging a battery. Setting the parameters of the inputs we could start the simulation. The rpm of the internal combustion engine is set to change from 1000 to 6000 rpm in every 50 s, and the load resistance changes from 5.5 to 0.5 Ω in every 10 s.

The result of the simulation is presented in *Fig. 6*. It can be observed that the induced voltage (1) has ripples due to the rectification and due to the voltage regulator. The network voltage is highly influenced by the level of charging of the battery and the state of the load current (3). The DFM signal gives information about the operation mode of the voltage regulator and so about the excitation current. The higher the duty cycle is, the higher excitation is.

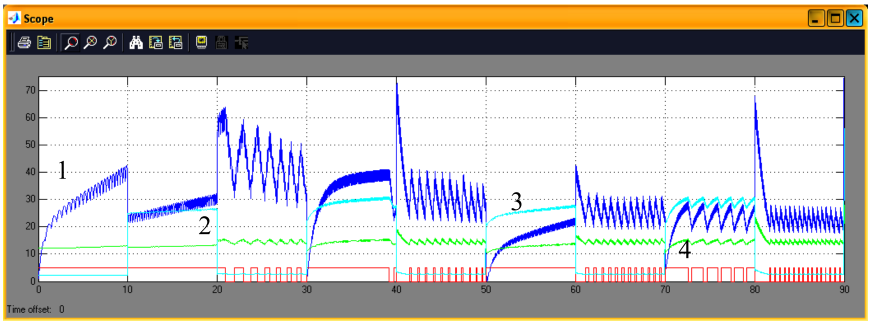


Figure 6: The result of the simulation.
1 - the induced voltage in the stator winding, 2 - voltage of the network,
3 - load current, 4 - DFM signal

3. Performance curve

In order to get the performance curve, the alternator has to operate in such a condition that the regulator has no influence on its work. This characteristic represents the variation of the load current in function of the alternator speed at constant voltage. In real cases the alternator has a built-in regulator, thus it is difficult to separate its influence. Similar happens on a test bench where in the first step the level of voltage regulator is found. The test voltage is set below this value, meaning that the voltage regulator will not act. So we have an easier situation because we can eliminate the voltage regulator from our model. Also the battery is taken out because its internal voltage and charging current would modify the behaviour of the alternator.

In *Fig. 7* can be observed that the output voltage (2) of the alternator is kept constant, the load current (1) increases and there is no any act in the excitation current (3). The performance curve is presented in the next figure.

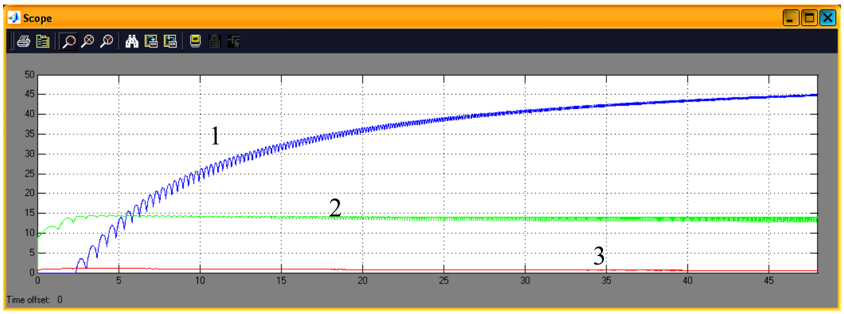


Figure 7: Time domain analysis of the rectifier output voltage.
1 - the load current, 2 - output voltage, 3 - excitation current.

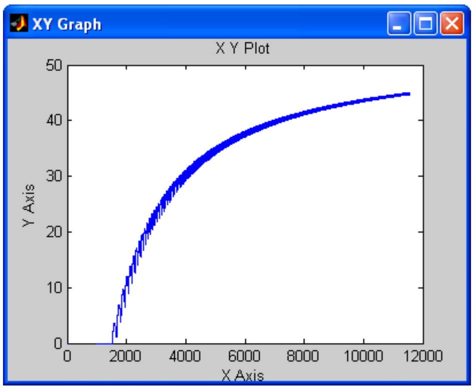


Figure 8: The performance curve I(n) determined by simulation.

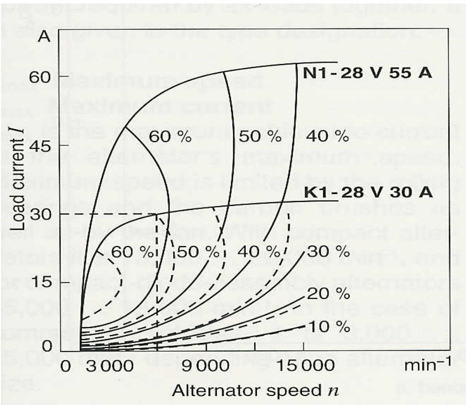


Figure 9: The performance curve I(n) in literature [2].

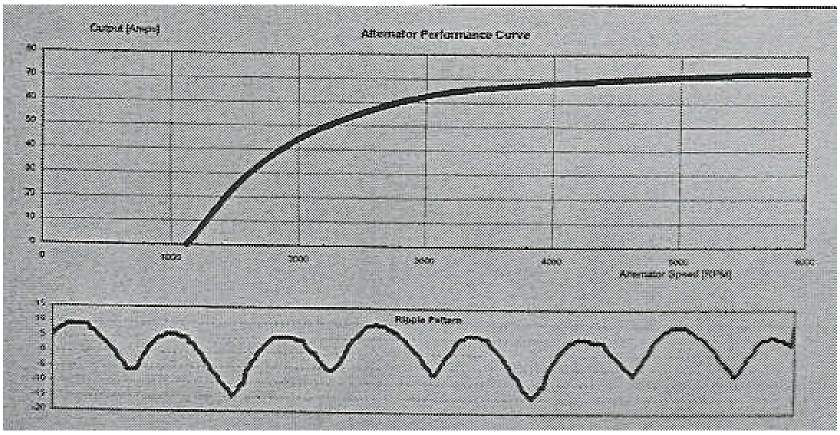


Figure 10: The performance curve I(n) from industrial computerised test report ALT-7.

4. Conclusions

The model of the alternator, partly developed earlier, was improved with some nonlinearities and the performance curve could be simulated. It can be observed that the resulted performance curve highly corresponds to that written in theory and measured in industrial environment.

Acknowledgements

The authors highly appreciate the generous support of the BAROSS_EM07-EM_ITN3_07-2008-0039 project of National Office of Research and Technology.

References

[1] Tömösy M. J., Frank, Gy., "Autóvillamosság", Műszaki Könyvkiadó, Budapest, 1975.

[2] Bauer, H., "Alternators and starter motors", Robert Bosch GmbH, Stuttgart, 2003.

[3] Kovács, E., Blága, Cs., "Személygépkocsi generátorának szimulációja", *X. ENELKO*, EMMTT, Marosvásárhely, 2009. október 8-11., (pp. 78-83).

[4] Blága, Cs., "Munkásság tézisszerű összefoglalása", *a Miskolci Egyetem Doktori (PhD) Tézisfüzetei*, Gépészmérnöki Kar Doktori Tanácsa, Miskolc, 1997.

Mechatronics and robotics

Microstructural Modification of $(Ti_{1-x}Al_xSi_y)N$ Thin Film Coatings as a Function of Nitrogen Concentration

Dominic BIRÓ[1], Sándor PAPP[2], László JAKAB-FARKAS[2]

[1] Department of Mechanical Engineering, Faculty of Technical and Human Sciences,
Sapientia University, Tg. Mureş, e-mail: dbiro@ms.sapientia.ro
[2] Department of Electrical Engineering, Faculty of Technical and Human Sciences,
Sapientia University, Tg. Mureş, e-mail: spapp@rdslink.ro, jflaci@ms.sapientia.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: Nanostructured (Al, Ti, Si)N thin film coatings were synthesized by DC reactive magnetron sputtering of a planar rectangular Al:Ti:Si=50:25:25 alloyed target, performed in Ar/N_2 gas mixture. Cross-sectional transmission electron microscopy investigation (XTEM) of as-deposited films revealed distinct microstructure evolution for different samples. The metallic AlTiSi film exhibited strong columnar growth with a textured crystalline structure. Addition of a small amount of nitrogen to the Ar process gas leads to a grain refinement. Further increase of N concentration resulted in fine lamellae growth morphology consisting of very fine grains in close crystallographic orientation showing up clusters of the chain-like pearls in a dendrite form evolution.

Keywords: Nanostructured (Ti,Al,Si)N thin films, cross-sectional XTEM microstructure, grain refinement, lamellae growth morphology.

1. Introduction

In the last decade intense research activity was devoted to investigate nanocomposite coating materials, consisting of a nanocrystalline transition metal nitride and an amorphous tissue phase. These coating materials are characterized by high hardness [1], enhanced elasticity [2] and high thermal stability [3], which define their unusual mechanical and tribological properties. Various studies revealed that in multiphase nanocomposite materials the microstructure and the ratio between hardness and elastic modulus H/E are important in the coating performance [4]. Recently the most studied material is the quaternary (Ti, Al, Si)N nitride system revealing the most promising results.

As was suggested by Weprek [5], in nanocomposite materials the structure and size of nanocrystalline grains embedded in the amorphous tissue phase altogether with the high cohesive strength on their interfaces, are the main parameters which controls the mechanical behaviours of the coatings. The reported results revealed that adatom mobility may control the microstructure evolution in multi-elemental coating systems, where the substrate temperature and the low energy ion/atom arrival ratio has significant effect on growth of nanocrystalline grains.

Microstructure and growth mechanism of arc plasma deposited TiAlSiN (35 at.% Ti, 42 at.% Al, 6.5at.% Si) thin films were investigated by Parlinska at al. [6, 7]. It was shown that compositionally graded TiAlSiN thin films with Ti-rich zone close to the substrate exhibited crystalline structure with pronounced columnar growth. Addition of Al+Si leads to a grain refinement of the coatings, and a further increase of the Al+Si concentration resulted in the formation of nanocomposites, consisting of equiaxial, crystalline nanograins surrounded by a disordered, amorphous SiN_x phase.

In our study (Al, Ti, Si)N single layer thin film coatings were deposited on Si(100) and high-speed steel substrates produced by DC reactive magnetron sputtering. We investigated the microstructural modification of $(Ti_{1-x}Al_xSi_y)N$ thin film coatings as a function of nitrogen concentration by conventional transmission electron microscopy.

2. Experimental details and characterization technique

Deposition experiments of (Al, Ti, Si)N quaternary nitride coatings were carried out in a laboratory scale equipment by DC driven magnetron sputtering, whose details are reported elsewhere [8]. The three independently operated sputter sources were closely arranged side by side on the neighbouring vertical walls of a 75 *l* octagonal all-metal high vacuum chamber. The closely disposed UM magnetrons arranged on an arc segment were highly interacting by their magnetic fields, leading to a far extended active plasma volume. In the presented deposition experiments only the central magnetron source was active, while the adjacent two magnetron sources contributed only in the closed magnetic field. A high purity planar rectangular target material of alloyed AlTiSi was used. Elemental composition of the PLANSEE GmbH. alloyed target was 50 at.% Al, 25 at.% Ti, and 25 at.% Si, with 165x85x12 mm^3 in size, which was partially covered on the erosion zone with a high purity 99.98% Ti sheet. Prior to deposition in vacuum chamber a base pressure of $2 \cdot 10^{-4}$ Pa was established by operating a 540 l/s turbomolecular pump.

Polycrystalline high-speed steel (HSS) substrates were used for tribological measurements, and monocrystalline <100> Si wafers were also used for structure characterization covered by thermally grown SiO_2 substrates. The target-to-substrate distance was kept constant at 110 mm in all runs. The substrates were positioned in static mode on a molybdenum sheet substrate holder, which allowed application of U_s= –75 V bias voltage. The Mo sheet was externally heated to a controllable substrate temperature of T_s= 400 ºC.

Prior to the starting of the deposition process, the surface of the substrates was plasma-etched by a DC glow discharge in argon for 10 min at 0.8 Pa, while the bias voltage was limited up to 350 V. During the ion etching of substrates, the target surfaces were also sputter cleaned by operating the magnetron unit at limited discharge power (pre-sputtering power 150 W). The substrate surfaces were shielded during the pre-sputtering. After about 10 min sputter cleaning of the targets, the discharge power at the target was raised to 500 W and the development of coating started with the deposition of a 50 nm thick AlTiSi seed-layer in pure Ar atmosphere. In the next step a gradient (Al, Ti, Si)(N) interlayer reactively deposited with the same sputtering power of the target, while PC controlled N_2 flow rate was increased slowly up to the pre-selected value. The argon gas flow was kept constant at 6.0 sccm. The typical thickness of the coatings was approximately 2 μm.

The reactive sputtering process was performed in a mixture of Ar and N_2 atmosphere at 0.28 Pa dynamic pressure. During the reactive sputtering process the nitrogen mass flow rate was controlled with an Aalborg DFC 26 flow controller, which contains a solenoid valve. The argon gas throughput (q_{Ar}=6.0 sccm, measured by GFM 17 Aalborg mass flow meter) was adjusted by a servo motor driven mass flow rate controller (MFC-Granville Phillips S 216). During deposition in the target a constant sputtering power with a current density of 10 $mA\cdot cm^{-2}$ was selected. The experimental conditions for preparation of (Al, Ti, Si)N coatings are listed in *Table 1*.

The microstructure of the as-deposited coatings, such as size and morphology of the crystalline grains was examined by a 100 kV operated JEOL 100U transmission electron microscope. In order to prepare cross-sectional XTEM samples for transverse observations, the samples were subjected to ion-milling in view of thinning up to electron beam transparency. Thin specimens for XTEM investigations were prepared in Technoorg-Linda model 4IV/H/L ion beam thinning unit. High energy ion beam thinning was completed with a low angle and low energy (200 eV) ion beam process in order to eliminate the amorphous by-products and etching defects induced by the high energy ions.

Table 1: Summary of deposition parameters used for preparation of (Al, Ti, Si)N coatings: P_d- DC magnetron discharge power, q_{N2}- nitrogen mass flow rate, T_s- substrate temperature, U_s- substrate bias voltage.

Samples	P_d [W]	q_{N2} [sccm]	T_s [°C]	U_s [V]
TiS_01	500	-	400	-75
TiS_07 TiS_08	500 500	1.0 1.0	400 400	-75 -75
TiS_04 TiS_09	500 500	2.0 2.0	400 400	-75 -75
TiS_10	500	2.0	100	-75
TiS-06	500	3.0	400	-75
TiS-03	500	4.0	400	-75
TiS-05	500	5.0	400	-75
TiS-02	500	6.0	400	-75

Bright-field (BF) and dark-field (DF) transmission imaging techniques were used for microstructure investigation of the as prepared samples. The identification of the crystallographic phases and the crystal orientation were also performed by evaluation of selected area electron diffraction (SAED) patterns. The SAED patterns were processed with the ´Process-Diffraction´ software tool developed by Labar [9].

3. Results and discussion

In this XTEM study of our prepared (Al, Ti, Si)N coatings, we combined direct imaging and selected area electron diffraction modes (SAED), which facilitate obtaining information on the microstructure morphology, grain size and crystallographic preferred orientation.

Cross sectional image of the polycrystalline AlTiSi coating's morphology is given in *Fig. 1*. The micrograph shows that crystallite in a conical shape evolution starts close to the substrate and grows in a competitive mode. The crystalline grains grow through the entire film thickness up to the top surface of the coating. The columnar grains are normally oriented to the substrate's surface. The large AlTi(Si) crystallites of approximately 80 nm in width are separated by the more electron transparent $TiSi_2$ phase segregated to the grain boundaries. The SAED patterns taken from the bulk region of the film exhibit well defined spotted diffraction rings (not to be seen here). The SAED pattern taken from the near substrate region of the coating proved crystalline character of the Si doped TiAl film (inset of *Fig. 1.*). The phases that can be derived from the diffraction pattern are mainly fcc-B1 NaCl-type of TiAlSi solid solution crystallites. The simulation of the diffraction rings was performed taking into account an fcc-type structure. It can be clearly seen the <200> preferential growth direction, indicated by significant brightness increase due to reflections from (200) crystallite planes that are oriented in parallel to the growing surface.

The chemical composition of the as deposited thin film was evaluated from EDS spectra analysis, and found to be of 34 at.% Ti, 46 at.% Al and 20 at.% Si.

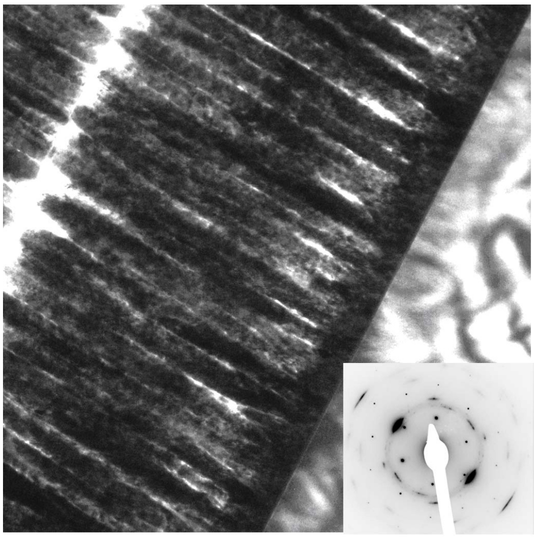


Figure 1: XTEM micrograph showing a cross sectional view of the columnar structured polycrystalline TiAlSi thin film coating (TiS_01 sample). The inset of SAED electron diffraction pattern indicates an fcc-structured TiAlSi solid solution phase, showing (200) texture evolution in the growing direction.

By adding a small amount of nitrogen as reactive gas in argon process gas, the growth morphology of the film dramatically changed (*Fig. 2*).

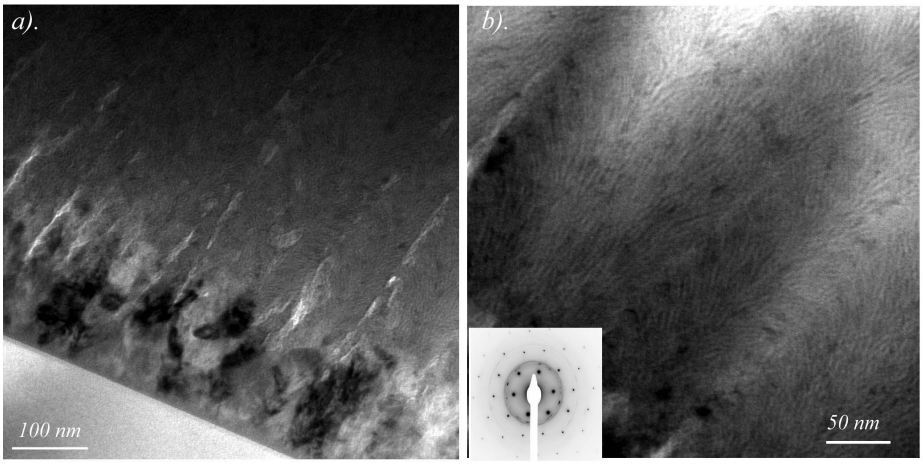


Figure 2: XTEM micrograph and SAED electron diffraction pattern of the (AlTiSi)N coating grown by nitrogen flow rate of q_{N2}=2 sccm (sample TiS_09): a). Bright field (BF) image indicates a weakly columnar structure evolution in close vicinity of transition zone from the ternary TiAlSi sub-layer to the quaternary (AlTiSi)N overgrown layer, b). On the enlarged micrograph slightly curved fine lamellae growth morphology could be identified inside the individual columns.

For a nitrogen flow rate of q_{N2}=2 sccm the microstructure indicates a weak columnar evolution (*Fig. 2a*). A slightly curved fine lamellae growth morphology could be identified inside of the individual columns (see on the enlarged micrograph, *Fig. 2b*).

Selected area electron diffraction pattern (SAED) performed in close vicinity of transition zone –including also the Si(100) bulk–, claims for two-phase mixture of fcc-TiAlN nanocrystals embedded in an amorphous tissue phase (inset of *Fig. 2b*). Furthermore, (200) preferential growth in close vicinity of transition zone from the ternary TiAlSi sub-layer to the quaternary (AlTiSi)N overgrown layer was slightly maintained. The presence of continuous reflection rings suggest a grain refinement of the coating with a strong tendency for evolution from the textured polycrystalline phase to a mixture of nanocrystalline AlTi(Si)N phase and possible formation of silicon nitride amorphous tissue phase.

Chemical composition of the as deposited thin films was evaluated from EDS spectra, and found to be 23 at.% Ti, 46 at.% Al and 30 at.%. Si.

With further increase of nitrogen amount in the coating deposition process (TiS_05 sample performed by q_{N2}=5 sccm nitrogen flow rate, which determined in our reactive magnetron sputtering process a p_N=0.0016 mbar nitrogen partial pressure) the crystalline character of the coating disappeared and formed an isotropic nanocomposite structure, possibly consisting of nanocrystalline $Ti_3AlN/TiSi_2$ grains of 2…3 nm in size surrounded by Si_xN_y and/or AlN amorphous matrix phase (*Fig. 3a*). The SAED diffraction pattern revealed that an increased nitrogen amount leads to a nanocomposite structure, consisting of equiaxially distributed $Ti_3AlN/TiSi_2$ nanocrystalline grains surrounded by a very thin amorphous Si_3N_4 phase.

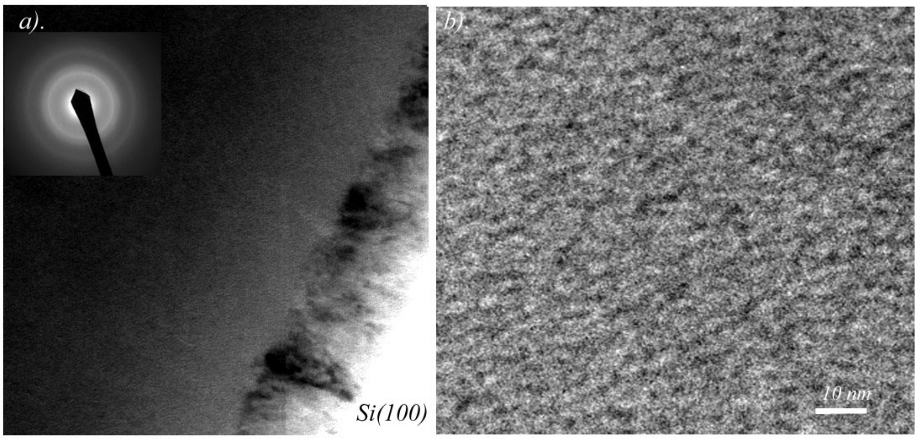


Figure 3: Bright field XTEM micrograph of (AlTiSi)N thin film deposited with an increased nitrogen flow rate (TiS_05 sample, q_{N2}=5 sccm): a). The coating's microstructure indicates the development of a competitive columnar evolution of the ternary TiAlSi sub-layer followed by the growth of the quaternary (AlTiSi)N overgrown layer developed in an isotropic morphology. b). The enlarged micrograph clearly shows randomly distribution of the very fine *nc*-$(Al_{1-x}Ti_x)N$ grains, having an average size of ~ 3 nm, with disordered grain limiting boundaries.

The chemical composition of the as deposited thin film was evaluated from EDS spectra analysis, and found to be 12 at.% Ti, 19 at.% Al, 23 at.% Si and 46 at.% N. The oxygen impurity content decreased to about 0.2 % witch was related to a prolonged out gassing process of the vacuum chamber and thermal degassing of the substrate by heating to 600 °C prior to the deposition process.

As Veprek et. al emphasized in their recent review paper [10-11] quaternary (Ti, Al, Si)N nitride coatings deposited under appropriate conditions assure the formation of stable *nc*-$(Al_{1-x}Ti_x)N$/*a*-Si_3N_4 nanocomposite structure by self-organization upon spinodal decomposition and phase segregation of few nanometer small $(Al_{1-x}Ti_x)N$ crystallites "glued" together by about one monolayer thin Si_3N_4 as tissue phase. The strong immiscibility of stoichiometric TiAlN and Si_3N_4 phase and the absence of Ostwald ripening, due to the low diffusion rate at the deposition temperature, are appropriate conditions for the development of nanocomposite coatings.

Our experimental results on fine lamellae growth morphology of (Ti, Al, Si)N nitride coatings, consisting of chain-like pearls in a dendrite evolution with very fine grains in close crystallographic orientation, may be explained in accordance with Veprek's theory by partial spinodal decomposition and phase segregation during the film's growth while percolation threshold composition is attained by an increased nitrogen activity.

The increase of deposition rate induces a decrease of the surface mobility related to the decrease of the ion-to-atom arrival rate ratio. These particular deposition conditions explain the columnar structure of TiAlSi solid solution crystallites, which can be clearly observed in the XTEM image of TiS_01 sample. Addition of minor amounts of nitrogen leads to an encapsulation of the growing TiAl(Si)N crystallites by process segregated $TiSi_2$ and Si_xN_y phase.

From the detailed observation of the SAED diffuse diffraction pattern of sample TiS_05 obtained with an increased nitrogen flow rate, the presence of an amorphous phase surrounding the Ti_3AlN nanocrystallites can be attributed to Si_3N_4 matrix phase (inset of *Fig. 3a*). The formation of AlN phase due to the partial segregation of Al atoms should be also considered due to the effect of enhanced ion bombardment provided by the focused plasma beam that is characteristic to the present experimental conditions [8].

When atomic surface mobility in the growing film is adequate, the segregated atoms can nucleate and develop the new phases due to the relative high deposition temperature and/or due to an energy transfer from an increased incident ion-to-atom arrival rate ratio, [12-14].

Further experiments are in progress to investigate the influence of deposition temperature on structure evolution of coatings and explain their high temperature stability.

4. Conclusions

In the present work it was shown that:

a). Columnar structured polycrystalline AlTiSi thin film coating evolved by non-reactive DC magnetron sputtering (performed in pure Ar atmosphere) applied to Al:Ti:Si=50:25:25 alloyed target (where the discharge power 500 W, T_s= 400 ºC substrate temperature, U_s=–75 V bias voltage were held constants).

b). Addition of a small amount of nitrogen to the process gas leads to a grain refinement of polycrystalline (Ti, Al, Si)N thin films. Increase of N concentration (q_{N2}=2 sccm flow rate) resulted in fine lamellae growth morphology of coatings, showing chain-like pearls in a dendrite evolution, consisting of clusters of very fine grains in close crystallographic orientation.

c). Further increase in nitrogen amount (qN2=5 sccm) leads to evolution of a nanocomposite coatings consisting of crystalline Ti_3AlN nanograins in 2…3 nm size surrounded by an amorphous Si_xN_y covalent nitride and/or AlN matrix phase.

Acknowledgements

The authors are thankful for the financial support of this project granted by Sapientia Foundation -Institute for Scientific Research, Sapientia University. The EDS analyses of the investigated coatings were performed in a CM 20 Philips 200kV TEM electron microscope by Professor P.B. Barna from RITPMS, Budapest. Professor P.B. Barna's contribution to investigating elemental composition and the valuable discussions with him are highly appreciated.

References

[1] Yoon, J.S., Lee, H.Y., Han, J., Yang, S.H., Musil, J., "The effect of Al composition on the microstructure and mechanical properties of WC–TiAlN superhard composite coating", *Surface and Coatings Technology*, 142-144 pp. 596-602, 2001.

[2] Duran-Drouhin, O., Santana, A.E., Karimi, A., "Mechanical properties and failure modes of TiAl(Si)N single and multilayer thin films", *Surface and Coatings Technology,* 163-164 pp. 260-266, 2000.

[3] Musil, J. and Hruby, H., "Superhard Nanocomposite $Ti_{1-x}Al_xN$ Films Prepared by Magnetron Sputtering" *Thin Solid Films*, 365 pp. 104-109, 2000.

[4] Ribeiro, E., Malczyk, A., Carvalho, S., Rebouta, L., Fernandes, J.V., Alves, E., Miranda, A.S., "Effect of ion bombardment on properties of d.c. sputtered superhard (Ti,Si,Al)N nanocomposite coatings", *Surface and Coatings Technology,* 151-152 pp. 515-520, 2002.

[5] Veprek, S., "New development in superhard coatings: the superhard nanocrystalline-amorphous compositics", *Thin Solid Films* 317, pp. 449-454, 1998.

[6] Parlinska-Wojtan, M., Karimi, A., Cselle, T., Morstein, M., "Conventional and high resolution TEM investigation of the microstructure of compositionally graded TiAlSiN thin films", *Surface and Coatings Technology*, 177-178 pp. 376-381, 2004.

[7] Parlinska-Wojtan, M., Karimi, A., Coddet, O., Cselle, T., Morstein, M., "Characterization of thermally treated TiAlSiN coatings by TEM and nanoindentation", *Surface and Coatings Technology*, 188-189 pp. 344-350, 2004.

[8] Biro, D., Barna, P.B., Szekely, L., Geszti, O., Hattori, T., Devenyi, A., "Preparation of multilayered nanocrystalline thin films with composition-modulated interfaces",*Nuclear Instruments and Methods in Physics Research* A 590 pp. 99-106, 2008.

[9] Lábár, J.L., "ProcessDiffraction: A computer program to process electron diffraction patterns from polycrystalline or amorphous samples", *Proceedings of the XII EUREM*, Brno (L. Frank and F. Ciampor, Eds.) Vol III pp. I 379-380, 2000.

[10] Veprek, S., Veprek-Heijman, M.G.J., Karvankova, P., Prochazka, J., "Different approaches to superhard coatings and nanocomposites*", Thin Solid Films* 476, 1-29*,* 2005.
[11] Veprek, S., Zhang, R.F., Veprek-Heijman, M.G.J., Sheng, S.H., Argon, A.S., *"*Superhard nanocomposites: Origin of hardness enhancement, properties and applications*",* in *Surf. And Coat. Technol.* 204, 1898-1096, 2009.
[12] Carvalho, S., Rebouta, L., Ribeiro, E., Vaz, F., Dennnot, M.F., Pacaud, J., Riviere, J.P., Paumier, F., Gaboriaud, R. J., Alves, E., "Microstructure of (Ti,Si,Al)N nanocomposite coatings", *Surfaec and Coatings Technology*, 177-178 pp. 369-375, 2004.
[13] Carvalho, S., Rebouta, L., Cavaleiro, A., Rocha, L.A., Gomes, J., Alves, E., "Microstructure and mechanical properties of nanocomposite (Ti,Si,Al)N coatings", in *Thin Solid Films* 398-399 pp. 391-396, 2001.
[14] Vaz, F., Rebouta, L., Goudeau, P., Pacaud, J., Garem, H., Riviere, J.P., Cavaleiro, A., Alves, E., "Characterisation of TiSiN nanocomposite films" *Surface and Coatings Technology* 133-134 pp. 307-313, 2000.

Creative Technique in the Domain of Bevel Gears

Erzsébet FALUVÉGI[1], Luciana CRISTEA[2]

[1] Department of Precision Mechanics and Mechatronics, Faculty of Mechanical Engineering, Transilvania University, Braşov, e-mail: faluvegi.erzsebet@gmail.com
[2] Department of Precision Mechanics and Mechatronics, Faculty of Mechanical Engineering, Transilvania University, Braşov, e-mail: lcristea@unitbv.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: This paper presents a creative technique in the domain of the Bevel gears. This creativity technique is called advanced multi-criteria analyses, which compares three methods for measuring the teeth's direction precision.

Keywords: bevel gears, precision, advanced multi-criteria analyses

1. Introduction

The gears are machine organs made of rotating bodies, equipped with internal or external teeth. They are used for transmitting the rotary motion. If a rack is used instead of one of the gears, the mechanism assures the transformation of rotary motion into linear motion. The bevel gears, which form the subject of this article, will transmit rotary motion between two intersecting axes.

2. The state of art in the field of measuring the precision of the teeth's flank

The tooth directional control is a relatively difficult operation and is done indirectly, by the contact pattern control or by the complex control of one or both flanks, [1].

In the case of bevel gears, the area and the position of contact path are of a great importance for it's characterization of the contact criteria between the teeth. It should be remarked that the contact path does not appear as a result of an instantaneous contact, but the effect is common to all lines of contact instantly.

The highlighting of the contact pattern is done by using control paint on the gears, be it on a special device, or installed inside it's own hood, [1]. The flanks of a tooth, which takes part of the gear, are painted with a mixture of vaseline, oil and, generally, a blue dye. The painted gear is rotated on both sides (clockwise and anticlockwise) and thus on each flank of an unpainted tooth of the driving gear remains a contact pattern which is copied with a filter paper.

There is an apparatus which is a letter patent that measures the direction of the tooth of the bevel gears. This device functions on the bases of a coordinate measuring system, [2]. The apparatus gives the idea for the advanced multi-criteria analyses, which can be found in the third section of the article.

The measuring instruments used to identify the errors of direction of the teeth in case of cylindrical gears can be modified to be used for those of bevel gears.

3. Directional type of the tooth

With the industrial development and for fulfilling the needs of the users, besides the bevel gears with straight and angular teeth also appeared those with circular teeth, which replace the angular teeth.

The teeth of the gears are identified by the help of the imaginary face gears (*Fig. 1*).

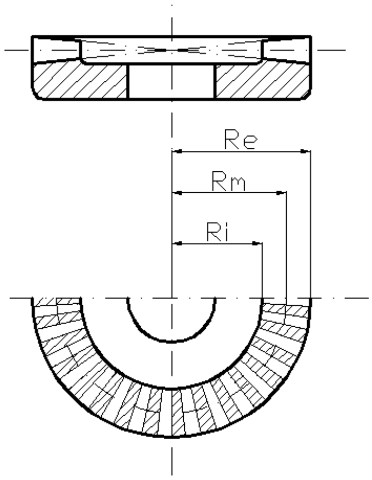


Figure 1: Face gear [3].

The face gear is like a rack which has a finite number of teeth.

The following table represents the types of bevel gears, [3]:

Table 1: Types of Bevel gears

Types of Bevel gears	The flank's direction on the face gear
Bevel gears with straight teeth	The flank's direction is straight line which passes through the center of the imaginary face gear.
Bevel gears with arched teeth shaped	The flank's direction is an arc of a circle, being on another concentric circle with the face gear.
Bevel gears with paloid teeth	The flank's direction is that of an involute arc, being described by a point of a line, which rolls without sliding on the basic circle
Bevel gears with eloid teeth	The flank's direction is that of a curly epicycloid arc described by the exterior points of the rotating circle, which rolls without sliding on the basic circle.

3. Advanced multi-criteria analyses

Creativity is a mental and a social process which generates some ideas or new concepts. There are many creative techniques which are heuristic methods for facilitating and stimulating the creativity of a person or a group of people.

The advanced multi-criteria analyses are an individual creative technique, which consists of five steps, [4]:

A. Stabilizing criteria

- the cost of construction – C;
- the speed of measuring – SM;
- the precision of measuring – P;
- the complexity of construction – CC;
- the energy source – ES;
- reliability – R;
- sensibility– S;
- sampling time – ST;

B. Determining the coefficients of each criteria

This determination is finalized by calculating the coefficient values for each criterion with a squared table:

Table 2: Coefficient values.

	C	SM	P	CC	ES	R	S	ST
C	½	½	½	1	½	½	½	0
SM	½	½	½	1	1	½	½	½
P	½	½	½	1	½	½	½	½
CC	0	0	0	½	0	0	0	0
ES	½	0	½	1	½	0	0	0
R	½	½	½	1	1	½	½	0
S	½	½	½	1	1	½	½	1
ST	1	½	½	1	1	1	0	½

Determining the coefficients:

Table 3: Coefficient values.

	Points	Level	γ_i
C	4	6	1,82
SM	5	3	3,33
P	4.5	4,5	2,4
CC	0.5	8	0,11
ES	1.5	7	0,5
R	4.5	4,5	2,4
S	5.5	1,2	4,25
ST	5.5	1,2	4,25

Coefficients are calculated through the formula, [4]:

$$\gamma_i = \frac{p + m + 0{,}5 + \Delta p}{-\Delta p' + \frac{N}{2}} \quad , \tag{1}$$

where:
- p is the sum of the obtained points of the considered element;
- m is the number of the surpassed criteria by the criteria taken into account;

- *Δp* is the difference between the score of the considered element and the score of the element at the last level;
- *Δp'* is the difference between the score of the considered element and the score of the element at the first level;
- N is the number of the considered criteria.

Substituting the values from the table above, the following results are:

$$\gamma_C = (4 + 2 + 0{,}5 + (4 - 0{,}5))/(-(4 - 5{,}5) + 8/2) = 1{,}82;$$
$$\gamma_{SM} = (5 + 5 + 0{,}5 + (5 - 0{,}5))/(-(5 - 5{,}5) + 8/2) = 3{,}33;$$

and so on.

C. Identifying all the variants

- Variant A: measuring systems with contact – a coordinate measuring system
- Variant B: a scanning system with a laser sensor with isolation for eliminating external light
- Version C: Ultrasonic scanning system.

D. Granting a note N

	Variant A	Variant B	Variant C
Criteria	N_i	N_i	N_i
C	9	7	10
SM	7	10	8
P	9	9	9
CC	8	9	9
ES	8	8	8
R	9	9	8
S	9	10	8
ST	7	10	10

E. Calculating the products between the notes N and the coefficients

		Version A		Version B		Version C	
Criteria	γ_i	N_i	N_i x γ_i	N_i	N_i x γ_i	N_i	N_i x γ_i
C	1,82	9	16,38	7	12,74	10	18,2
SM	3,33	7	23,31	10	33,3	8	26,64
P	2,4	9	21,6	9	21,6	9	21,6
CC	0,11	8	0,88	9	0,99	9	0,99
ES	0,5	8	4	8	4	8	4
R	2,4	9	21,6	9	21,6	8	19,2
S	4,25	9	38,25	10	42,5	8	34
ST	4,25	7	29,75	10	42,5	10	42,5
Final classification			**155,77**		**179,23**		**167,13**

3. Conclusion

By resolving the table above, it can be observed that the greatest value (179,23), is obtained in variant B, which is a measuring system with laser sensor with isolation against external light. In the second position the variant C is found, the ultrasonic measuring system, with 167,13 points, and in the third position variant A is found , the 3D coordinate measuring system, with 155,77 points.

By eliminating the errors in bevel gears, the machines which are using them become less noisy and reduce vibrations.

Acknowledgements

This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), ID59321 financed from the European Social Fund and by the Romanian Government References.

References

[1] Minciu, C. (1984) *Precizia şi controlul angrenajelor* Bucureşti: Editura Tehnică.

[2] Hollanda, D. (2004) *Aparat pentru controlul profilului danturii roţilor dinţate conice* nr. RO 118485 B1

[3] Hollanda, D. (1996) *Bazele aşchierii şi generarea suprafeţelor* Târgu Mureş: Universitatea Petru Maior

[4] Bobancu, Ş. (2005) *Creativitate şi inventica* Braşov: Universitatea Transilvania

Design of a 12-Dof Biped Robot

Mesut ACAR[1], Alper GERCEK[2], Burak TASASIZ,[3]
Zeki Y. BAYRAKTAROGLU[4], Hikmet KOCABAS[5]

[1]Department of Mechanical Engineering, Technical University, Istanbul,
e-mail: acarmes@itu.edu.tr
[2]Department of Mechanical Engineering, Technical University, Istanbul,
e-mail: alpergercek@itu.edu.tr
Department of Mechanical Engineering, Technical University, Istanbul,
e-mail: tasasiz@itu.edu.tr
Department of Mechanical Engineering, Technical University, Istanbul,
e-mail: zeki@itu.edu.tr
Department of Mechanical Engineering, Technical University, Istanbul,
e-mail: kocabash@itu.edu.tr

Abstract: This paper presents the mechanical and electrical designs of a 12-DoF biped walking robot which is under development at the Istanbul Technical University (ITU). The kinematic structure of the mechanism is similar to that of human legs. Electrical motors, harmonic reducers and ball-screw mechanisms are used for actuation of the joints and power transmission. The prototype design includes 6-axes force-torque sensor on each foot and absolute encoders on each joint. Static and dynamic walking simulations of the biped robot model have been realized by using Msc.Adams® and Simulink® softwares interactively.

Keywords: Biped walking, biped robot design

1. Introduction

In the last decades, great technological developments have occurred in sensor, actuator, and computer technologies. Humanoid robotics is one of the research fields which exploited these technological developments. Although the construction of human-like machines has been always a desire for mankind, it has only become possible recently.

In Japan, remarkable developments in humanoid prototypes have been observed. ASIMO developed by Honda is one of the most sophisticated prototypes with a running speed of 6[km/h] [1-3]. The HRP robots developed by AIST in Japan are also remarkable prototypes and have reached a maximum walking speed of 2.5[km/h] [4-7]. In Germany, the Technical University of Munich (TUM) has developed a humanoid prototype called Johnnie which has reached a walking speed of 2[km/h] [8-11]. A new prototype Lola is under development at the TUM [12].

In the USA, the MIT Leg Laboratory has developed a lot of walking and running legged robots. One legged hoping robot and quadruped running robots are good examples of these prototypes [13]. Moreover, a quadruped prototype BigDog and a biped prototype Petman have been developed by Boston Dynamics in collaboration with the MIT Leg Lab. [13]. Mechanical and electrical design and component selection of the biped robot at ITU has been completed recently. The design process is achieved by an iterative process between simulation and redesign (*Figure 1*). A screenshot from the CAD model of the robot can be seen on the *Figure 2*.

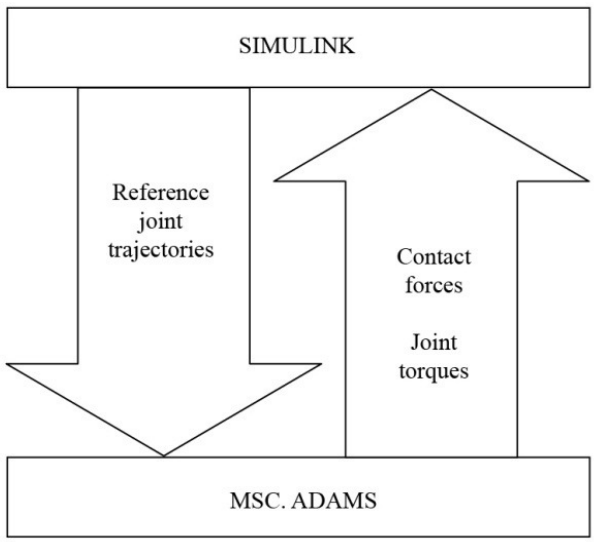


Figure 1: Simulation environment.

Although all of the prototypes mentioned here have achieved dynamically balanced walking, only ASIMO has achieved running motion. The first objective of our project is dynamically balanced walking. Dynamically balanced walking has been achieved in simulations and sizing and component selections are made due to these simulations.

2. Mechanical Design

The prototype designed in this project is a 12- DoF biped robot. Each leg consists of 6-DoF, 3 at the hip, 1 at the knee and 2 at the ankle joints (*Fig. 3*). All DoFs are driven by DC motors.

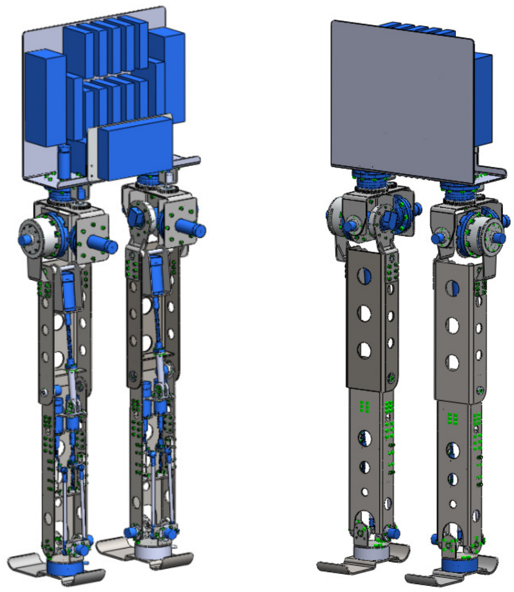


Figure 2: CAD model of the robot.

Three rotating DoF at the hip joint are driven by compact harmonic drivers (*Fig. 4*). Knee and ankle joints are driven by vertically placed actuators through ball screw mechanisms (*Fig. 5* and *Fig. 6*). In this design, actuators and transmissions driving the knee and ankle joints are situated vertically higher than the driven joints. In such a kinematic configuration, the centre of mass of the robot moves also higher in vertical with respect to that of robots with direct driven joints. This is an important design criterion for biped robots to be controlled by the inverted pendulum approach [14]. In this method, the biped robot is modelled as an inverted pendulum with a point mass situated as higher as the hip joints.

Lower and upper legs consist of U-profiles containing all actuation and transmission mechanisms of the knee and ankle joints. An industrial PC with real-time OS, all motor drivers and sensor interfaces and a battery are situated on an upper body upon the hip joints.

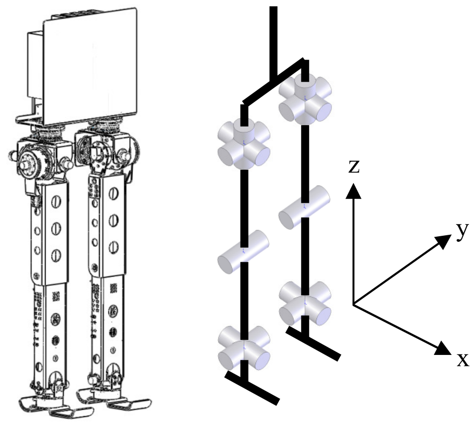


Figure 3: CAD model of the robot.

The entire mechanism has 130 cm tall and 37 kg weight. Dimensions are given in *Fig. 7*.

A. Hip Joint

Rotation axes of the hip joints intersect at one point. Such a configuration allows easier manipulation of the inverse kinematics.

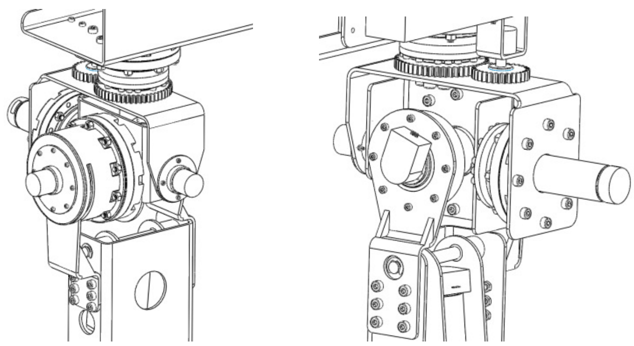


Figure 4: Hip joint

Each Dof is driven by a compact actuation unit with Maxon® and Harmonic Drive® components (*Table 1*).

Table 1: Selected components

Joint	Actuator	Gear
Hip Joint-x	Maxon RE-50	H.D. CSG-25-160
Hip Joint-y	Maxon RE-50	H.D. CSG-25-160
Hip Joint-z	Maxon RE-30	H.D. CSG-20-100
Knee	Maxon RE-50	B.S. Rexroth 1531-2
Ankle-1	Maxon RE-40	B.S. Rexroth 1531-1
Ankle-2	Maxon RE-40	B.S. Rexroth 1531-1

B. Knee Joint

The knee joint has 90 degrees motion range, and it is driven by a ball screw mechanism which is actuated by a DC actuator placed in the upper leg (*Fig. 5*). Instead of harmonic driver, a ball screw mechanism is used as velocity reducer to provide the required conversion ratio.

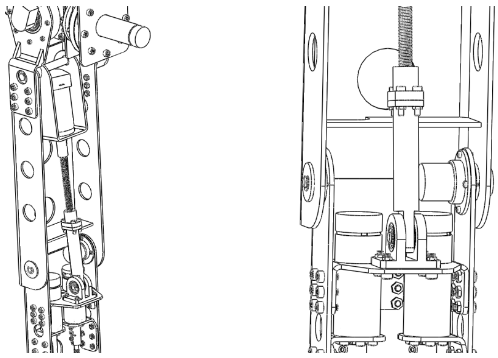


Figure 5: Knee joint.

In a similar way, the incremental encoder is connected to the DC motor shaft of the knee joint. The absolute encoder is placed on the knee joint shaft to compensate the measurement errors which may occur because of mechanically imperfect transmission mechanisms.

C. Ankle Joint

The actuators situated in the lower legs drive the 2 DoF of the ankle joints. Driving torque of the actuators is transmitted to the ankle joints by ball screw mechanisms similar to that of the knee joints (*Fig. 6*).

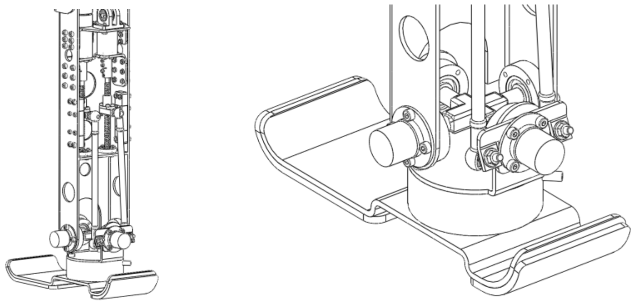


Figure 6: Ankle joint.

The nuts on the ball screw mechanism are guided on a polyamide part which is mounted on lower leg, therefore only linear displacements are allowed to the mechanism. Absolute encoders are also mounted to each ankle joints.

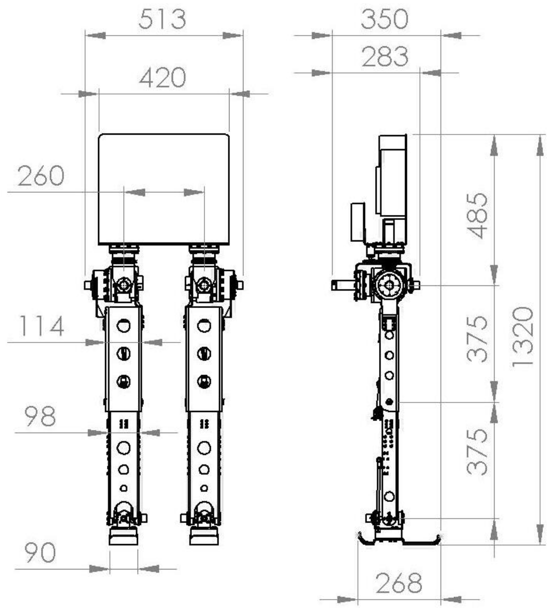


Figure 7: Dimensions of the robot

2 DoF of the ankle joints are provided by universal joints which are placed between the lower legs and feet. Besides, 6 axes force-torque sensors are placed on feet to measure the generalized contact forces which are required for the closed-loop control.

3. Electrical Design

A. Sensors

The sensory system of the biped robot includes incremental and absolute encoders at all 12 joints and two force-moment sensors at each ankle.

The incremental encoders (Maxon® MR L 500 series) attached to the motor shafts measure the angular displacements relative to initial values of each joint. Resolution of the incremental encoder which is 500 increments per revolution satisfies qualified measurement with highly stiff Harmonic-Drive gears.

Since an incremental encoder measures differentiation of angular positions, initial angular position of all joints must be measured by absolute encoders. The absolute encoders (Scancon® SCH24AB series) which are integrated to shafts of joints measure the initial values of joints before the biped robot starts walking. Measurement errors of the absolute encoders are negligible thanks to its 12-bit resolution per revolution.

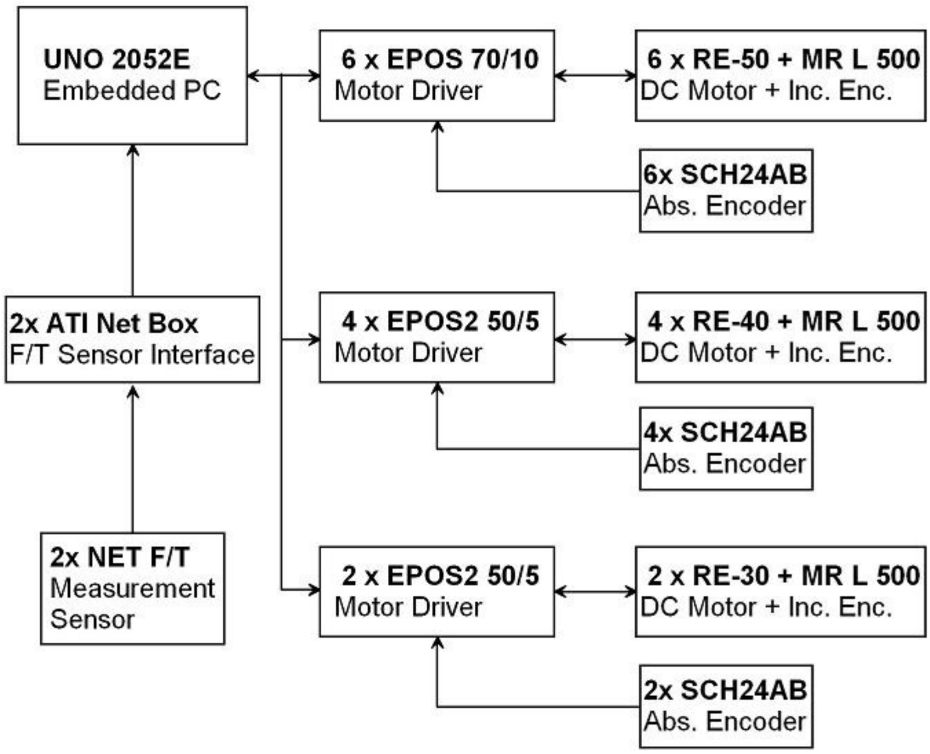


Figure 8: Electrical Design.

B. Actuators

The actuation system of the biped robot consists of 12 Brushed RE Series Maxon DC motors driven by 12 EPOS Motor drivers. The motor drivers will drive the DC motors with respect to trajectory data to be received from the computer via CANbus.

C. Computer System

The embedded computer (UNO 2052E) receives data from absolute encoders and force-torque sensors and transmits data to motor drivers via CANbus. The computer which is installed Embedded Real Time Linux OS will provide the trajectories and control inputs for both static and dynamic walking.

4. Simulation Environment

The Msc.Adams and MATLAB-Simulink software packages are used to simulate the dynamic behaviour of the biped robot with proposed control algorithms. The Msc.Adams is used to solve the dynamics of biped walking and Simulink runs the pattern generation algorithms and control laws. An illustrative diagram for the simulation process is shown on *Fig. 9*, where MPC is the model predictive controller, LIPM the linear inverted pendulum model and ZMP the zero moment point.

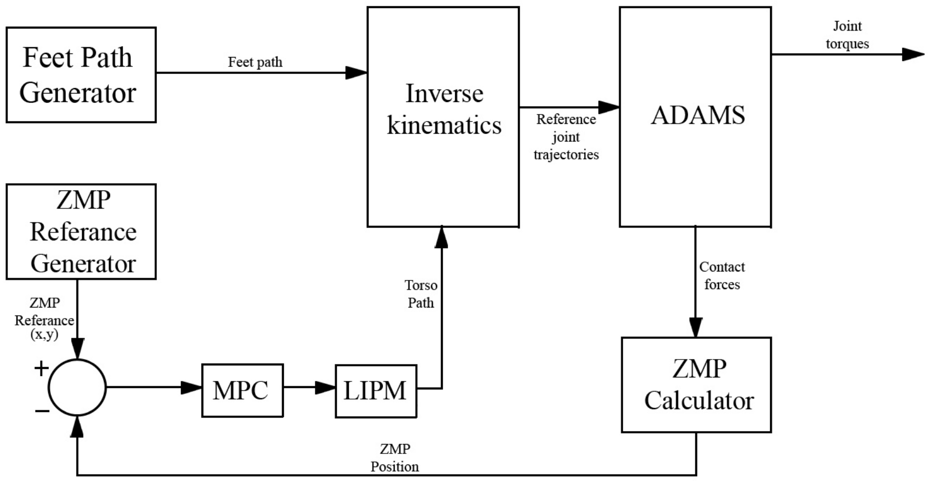


Figure 9: Control System.

A. Simulink Side

Simulink is responsible for all process excepting the computation of forward and inverse dynamics. Desired feet trajectories and ZMP references are generated off-line. To solve the inverse kinematics problem of the biped robot model, feet and torso paths must be known simultaneously. The torso path is generated to maintain the dynamic balance of the robot. In torso path generation, model predictive control with linear inverted pendulum model is used [14].

The ZMP calculator algorithm provides the actual ZMP position during walking by using the contact forces measured in Msc.Adams.

Feet and torso trajectories are used to solve the inverse kinematics problem. The reference joint trajectories given by the inverse kinematics block are transmitted to Msc.Adams.

B. Msc.Adams Side

Msc.Adams is used for the solution of forward and inverse dynamics problems. In this study, the Msc.Adams is provided with the reference joint trajectories as inputs and computes the joint torques and contact forces as outputs. The CAD model of the prototype is imported into Msc.Adams and all material properties and constraints are defined in Msc.Adams environment. Screenshots from a walking simulation is shown on *Fig. 10*.

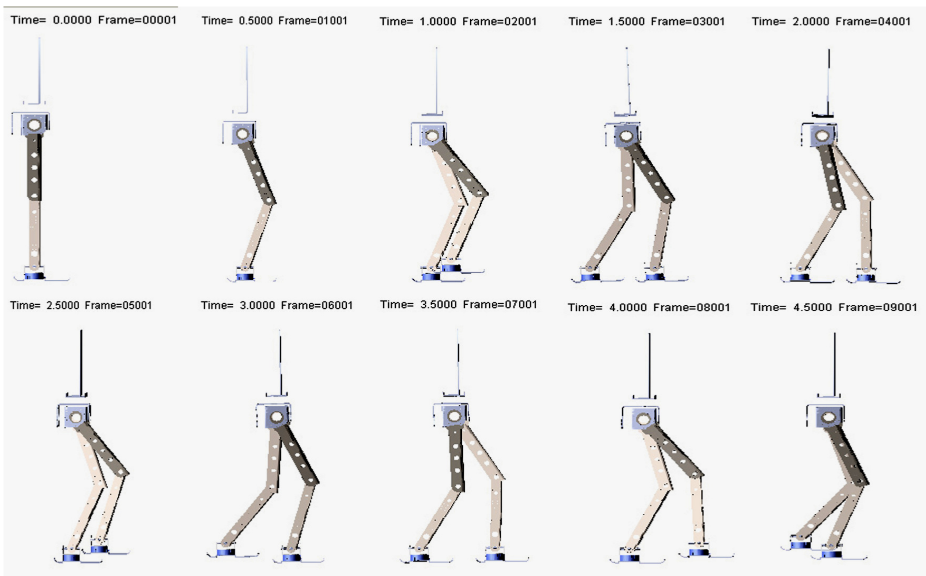


Figure 10: Screenshots from a walking simulation.

C. Simulation Results

Motor and reducer sizing of the robot are based on the results of the dynamic simulations. Required joint torques and velocities are the useful outputs of the simulations for dimensioning of the system. Contact forces are also used for force-torque sensor selection. Obtained joint torques for 1[km/h] walking speed are shown on *Fig. 11*.

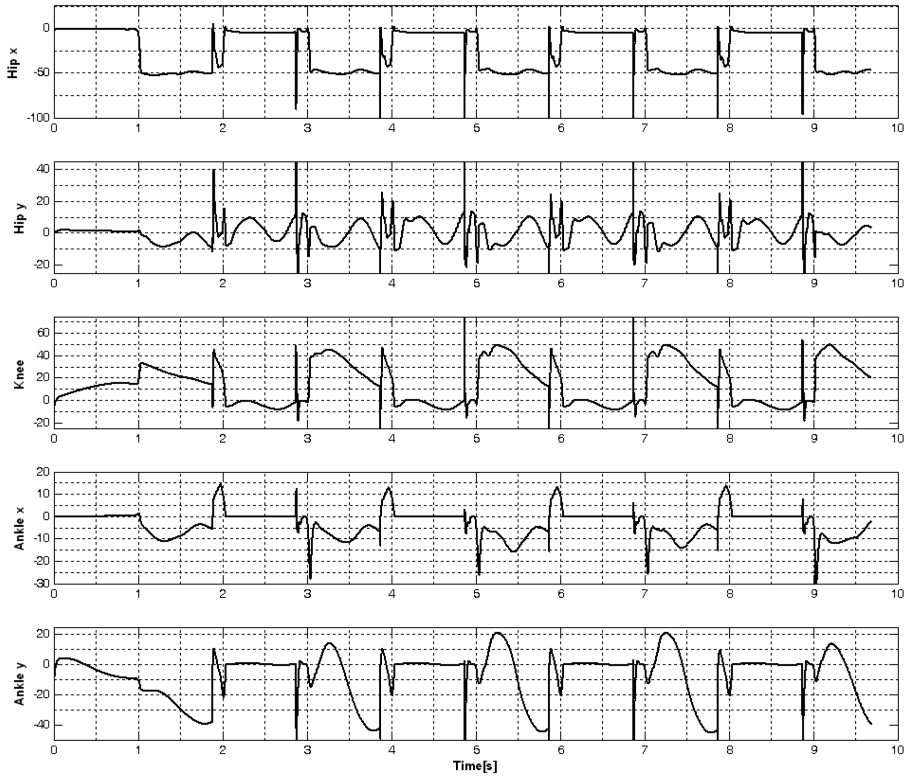


Figure 11: Joint Torques [Nm].

D. Selection of Actuators and Transmissions

In order to obtain the required characteristics for the actuation system including DC motors and transmissions, joint torques and velocities should be converted to the motor torques and velocities. At hip joints, this conversion is between parallel axes and its ratio given by a constant is indeed a parameter of the harmonic drives. However, at knee and ankle joints, the power transmission occurs between perpendicular rotation axes through ball-screw mechanisms. Therefore, nonlinear algebraic equations have to be solved in order to compute the required motor torques and velocities. The reduction process gives the required motor load characteristics. Hip-x motor load characteristic is given as an example on *Fig. 12*.

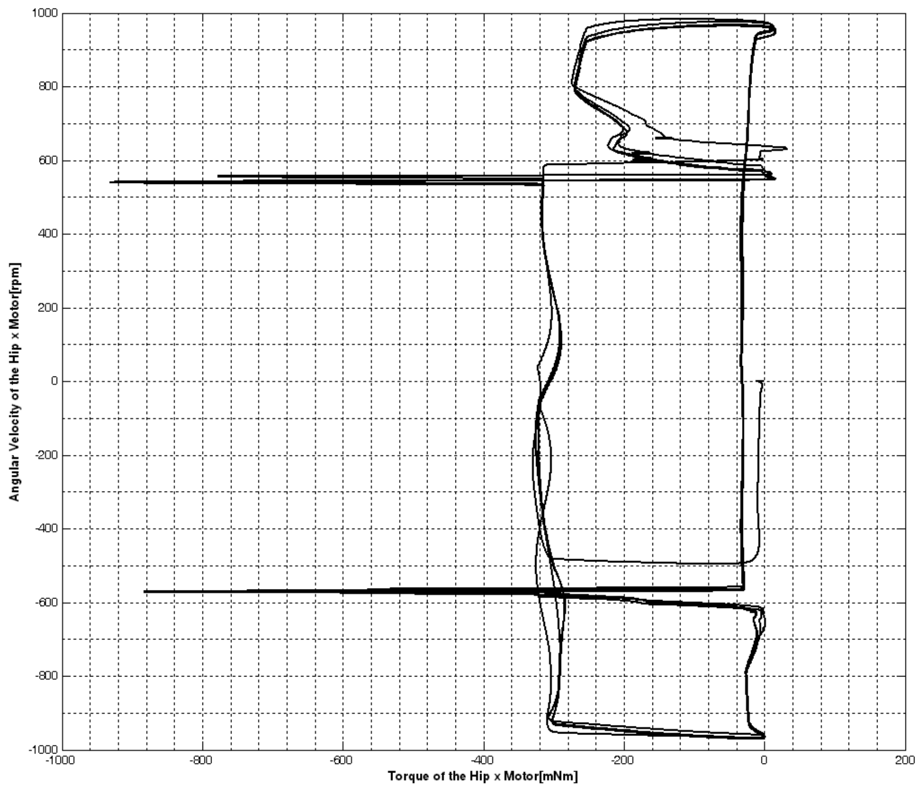


Figure 12: Hip-x motor load characteristic.

5. Future Works

The ongoing work on the subject is the construction of the biped robot design presented in this paper. After construction and assembly, control algorithms will be embedded to the computer and experimental studies will start. Main objectives of the experimental work are statically and dynamically balanced walking.

Acknowledgements

This research is supported by the TUBITAK Research Project No.106M340.

References

[1] Hirai, K., 1997, Current and Future Perspective of Honda Humanoid Robot. In *IEEE IROS'97*, *Intelligent Robots and Systems*.
[2] Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., Fujimura, K., 2002. The intelligent ASIMO: System overwiew and integration. *IEEE/RSJ*, *Intelligent Robots and Systems*
[3] Hirai, K., Hirose, M., Haikawa, Y., Takenaka, T., 1998. The Development of Honda Humanoid Robot, In *IEEE ICRA'98*, *Robotics and Automation*.
[4] Hirukawa, H., Kanehiro, F., Kaneko, K., Kajita, S.,2004. Humanoid Robotics Platforms Developed in HRP, In *Robotics and Autonomous Systems*
[5] Akachi, K., Kaneko, K., Kanehira, N., 2005. Development of Humanoid Robot HRP-3P, *IEEE-RAS, International conference on Humanoid Robots*
[6] Kaneko, K., Kanehiro, F., Kajita, S., Yokoyama, K., Akachi, K., Kawasaki, T., Ota, S., Isozumi, T., 2002. Design of Prototype Humanoid Robotics Platform for HRP. In *IEEE RSJ, Intelligent Robots and Systems*
[7] Inoue, H., Tachi, S., Nakamura, Y., Hiari, K., 2001. Overview of Humanoid Robotics Project of METI. In *ISR, International symposium on Robotics*
[8] Löffler, K., Gienger, M., Pfeiffer, F., 2003. Sensor and Control Design of a Dynamically Stable Biped Robot. In *IEEE ICRA'03, Robotics and Automation*.
[9] Lohmeiner, S., Löffler, K., Gieber, M., Ulbrich, H., 2004. Computer System and Control of Biped "Johnnie". *IEEE, International conference on Robotics and Automation*
[10] Gienger, M., Löffler, K., Pfeiffer, F., 2001. Towards the Design of a Biped Jogging Robot. *IEEE, International conference on Robotics and Automation*
[11] Pfeiffer, F., Löffler, K., Gienger, M., 2002. The Concept of Jogging JOHNNIE. *IEEE, International conference on Robotics and Automation*
[12] Lohmeier, S., Buschmann, T., Ulbrich, H., Pfeiffer, F., 2006. Modular Joint for Performance Enhanced Humanoid Robot Lola, In *IEEE ICRA'06, Robotics and Automation.*
[13] Raibert, M.H.,1986. *Legged Robots that Balance*.MIT press, Cambrige.
[14] Kajita , S., Kanehiro, F., Kaneko, K., Fujiwara, K., Harada, K., Yokoi, K., Hirukawa, H., 2003. Biped Walking Pattern Generation by using Preview Control of Zero-Moment Point. In *IEEE ICRA'03, Robotics and Automation*.

On Some Pecularities of Paloid Bevel Gear Worm-Hobs.

Dénes HOLLANDA[1], Márton MÁTÉ[2]

[1] Department of Mechanical Engineering, Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, e-mail: hollanda@ms.sapientia.ro
[2] Department of Mechanical Engineering, Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, e-mail: mmate@ms.sapientia.ro

Manuscript received April 14, 2010; revised April 25, 2010.

Abstract: The paper discuss the generating surfaces of the paloid bevel worm hob used for paloid gear cutting. A pitch modification is required by this type of tool in order to ensure the optimal contact pattern by gearing. As a consequence of this modification results the flank line of the plain gear tooth as a paloid- a more general shaped curve like the theoretical involute of the basic circle. Equations of the generating surfaces result as a generalized arhimedic bevel helical surfaces, presenting those modifications that arise from the variation of the tooth thickness.

Keywords: paloid bevel worm hob, tooth thickness modification, generating surface.

1. General description

Paloid bevel gears are realized on Klingelnberg type teething machine-tools, using paloid bevel gear worm-hobs. These tools present a straight shaped edge in their axial section. As a conclusion, the origin surface of the paloid worm hob is an Arhimedic bevel worm having the half taper angle of 30° as shown in *Fig. 1*. The chip collecting slots are axially driven. In order to realize the clearance angle on all edges, a helical relieving, perpendicular oriented to the bevel generator is allowed.

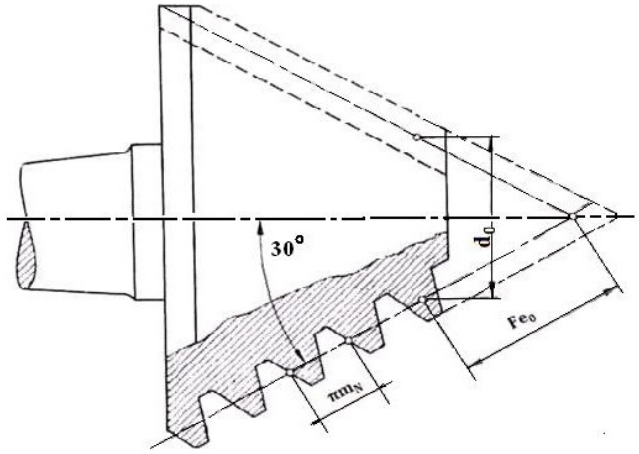


Figure 1: The paloid worm hob characteristic dimensions

In order to ensure the well positioning of the contact patch by paloid gears, the pitch of tool is variable but the tooth thickness (measured on the rolling taper generator) must increase at the extremities of the tool. Using this principle, the tooth thickness on the rolling tape generator will increase in both directions starting from the point N (*Fig. 2*).

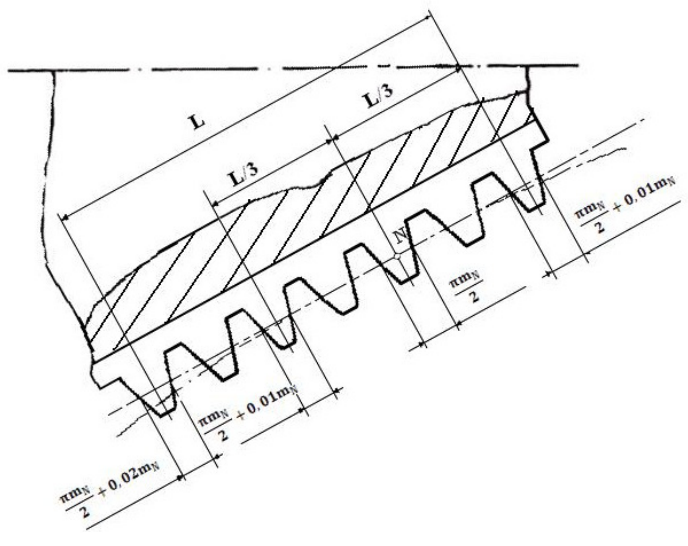


Figure 2: The tooth thickness distribution along the cone generator.

The thickness of teeth using the worm-hob described above results smaller at the tooth extremities and larger in the middle of it. This localizes the contact pattern in the middle of the tooth-flank when gearing. This modification causes the deformation of the tooth line on the plain gear too, transforming the theoretical involute into a more general shaped curve named paloid.

Both the threading and the relieving operation is realized on a relieving machine, where the axis of the worm-hob is declined with a 30° angle reported to the axis of the chuck, in the axis horizontal plain. This way the taper generator becomes parallel with the carriage of the relieving lathe.

The curved generator of the worm-hob is realized through leading the cutter slider by a profiled turning-template, identical by threading and relieving operations.

Bevel worm hobs present one thread, and allow the cutting of any teeth number by a fixed module and pressure angle. The pitch is normalized along the rolling cone generator and can be calculated using the formula: $p_N = \pi \cdot m_N$. The axial pitch (defined on the worm hob's axis) as following, is determined by $p_A = \pi . m_N \cdot cos\, 30^o$.

2. The geometry of the tooth thickness modification

Considering the bevel worm-hob as a helical bevel surface, like any bevel thread, it is to mention that next to the axial pitch, a radial pitch can be defined, depending on the axial pitch and the rolling cone 30^o matching half angle. If the generator of the worm hob would be a straight line, the radial pitch value would be satisfy the expression $h = p_A .tg30^o$. However the generator is curved, radial pitch will be calculated respecting that the endpoint of the characteristic radius must be on that generator. Considering the tool's pressure angle α, the distance between the curved and the straight generators (based on the dimensions of the tool profile indicated in [2]) can be calculated using the formula $\Delta_1 = (p_N + 0{,}01.m_N - p_N)/2.tg\alpha = 0{,}01.m_N / 2.tg\alpha$ applicable for point A (figure 3). The same distance is indicated related to point C (*Fig*. 3). But in point D, the distance discussed above will increase to double as follows from the formula $\Delta_2 = (p_N + 0{,}02.m_N - p_N)/2.tg\alpha = 2\Delta_1$.

The curved generator is usually an arc of ellipse who's major axis (superposed to the Ox) is parallel to the rolling cone generator, and it's minor axis (superposed to Oy) passes through point B (*Fig. 3*), situated at a distance of $F_{eo} + 2p_N$ to the top of the rolling cone.

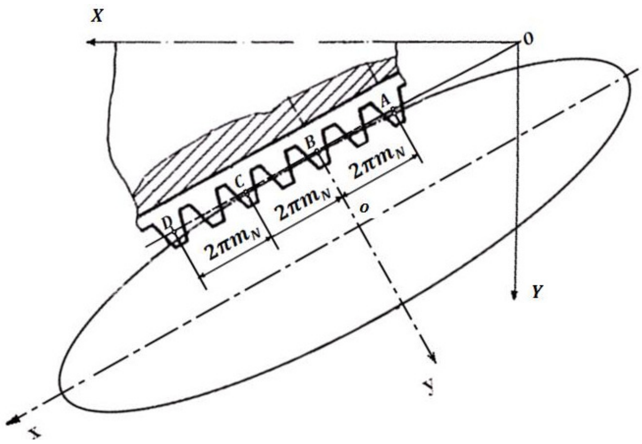


Figure 3: The elliptical curved generator.

It is to remark that the ellipse passes through the following points:

$A(-p_N, a-\Delta_1)$;

$B(0, a)$;

$C(-2.p_N, \Delta_1)$

$D(-4.p_N, a-\Delta_2)$.

The canonic equation of an arbitrary shaped ellipse, reported to its axes, is:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-1=0 \tag{1}$$

If writing the equation of the ellipses passing through points A, B, C and D a system of equation will result, whose solution leads to the reference radius endpoint function e.g. $h = f(p_N / \cos 30^o)$. In this paper the radial pitch is accepted having a general form as explained through the formula before. As a consequence, the relations deduced in the next can be used for other forms of the curved generator too. Using this form, the essential aspects of the dependences will be not influenced by the very sophisticated expressions of the radial pitch.

The matrix equation of the bevel worm hob's edge, reported to the self coordinate system, is:

$$r_t = \left\| \begin{matrix} (F_{eo} - \frac{\pi . m_N}{4})[\cos 30^O + tg30^O \sin(30^O - \alpha)] + \lambda \sin(30^O - \alpha)] \\ \lambda \cos \alpha \\ 0 \\ 1 \end{matrix} \right\| \tag{2}$$

This edge is bounded with the mobile coordinate system $O_M X_M Y_M Z_M$ superposed at the initial moment with the fixed coordinate system $OXYZ$. The parametric form of the edge expressed in the mobile system will be denoted as r_M but it keeps the form of r_t given by (2).and the

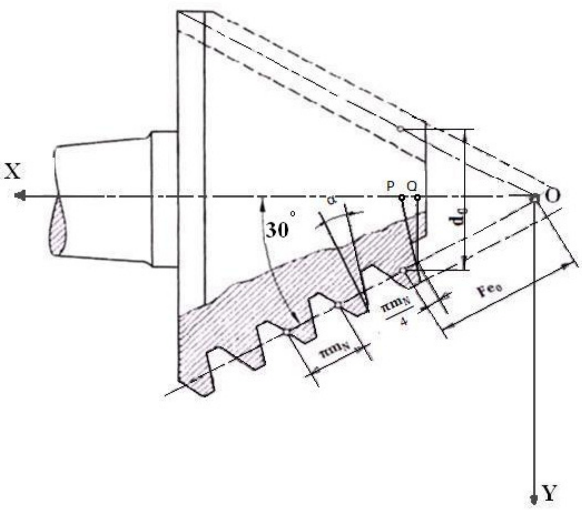


Figure 4: The position of the right edge related to the used reference system.

The surface of the generator flank of the bevel worm hob can be obtained considering the followings: the edge is fixed to the mobile coordinate system (*Fig. 4*) that executes a helical motion reported to the stationary system. The origin moves by the direction of axis with an amount of, simultaneously with a rotation by angle of the worm-hob (representing the work piece in the case above).

First the edge must be reported to an auxiliary coordinate system. Between this and the auxiliary system $O_M X_M Y_M Z_M$ exist only translations by values

$p_A.\varphi/2\pi$ and $h.\varphi/2\pi$ respectively. A transfer matrix describing the translations above is:

$$M_{aM} = \begin{Vmatrix} 1 & 0 & 0 & \frac{p_A}{2\pi}\varphi \\ 0 & 1 & 0 & \frac{h}{2\pi}\varphi \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix} \tag{3}$$

After that the edge's equation will be transferred to the stationary system, mentioning that auxiliary system executes only a rotation around axis, characterized by the matrix

$$M_{Oa} = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi & 0 \\ 0 & \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix} \tag{4}$$

The pointing vector of the bevel helical surface described by the edge, reported to the stationary reference system is to obtain from the following matrix equation

$$r = M_{Oa}.M_{aM}.r_M = M_{OM}.r_M \tag{5}$$

Multiplying $\mathrm{M_{Oa}}$ by $\mathrm{M_{aM}}$ results

$$M_{OM} = \begin{Vmatrix} 1 & 0 & 0 & \frac{p_A}{2\pi}\varphi \\ 0 & \cos\varphi & -\sin\varphi & \frac{h}{2\pi}\varphi.\cos\varphi \\ 0 & \sin\varphi & \cos\varphi & \frac{h}{2\pi}\varphi.\sin\varphi \\ 0 & 0 & 0 & 1 \end{Vmatrix} \tag{6}$$

Finally, the matrix expression of the pointing vector is:

$$r = \begin{Vmatrix} 1 & 0 & 0 & \frac{p_A}{2\pi}\varphi \\ 0 & \cos\varphi & -\sin\varphi & \frac{h}{2\pi}\varphi.\cos\varphi \\ 0 & \sin\varphi & \cos\varphi & \frac{h}{2\pi}\varphi.\sin\varphi \\ 0 & 0 & 0 & 1 \end{Vmatrix}.r_M \tag{7}$$

Expression (7) realizes the matrix form of the right flank of the bevel worm-hob tooth, reported to the stationary system OXYZ, attached to the hob.

Analytical expression will be obtained after multiplying the matrices in the equation before. Similarly results the equations of the opposite flank, if starting the calculus with the equations of the opposite edge.

Acknowledgements

Publishing of this Conference Proceeding is supported by the "Szülőföld Alap" foundation.

References

[1] Krume,W., “Klingelnberg-Spiralkegelräder Dritte neubearbeitete Auflage”, Springer Verlag, Berlin, 1967.

[2] *** „MASINOSTROIENIE”- ENCYCLOPEDY, Vol.VII., Masghiz Moskou.

[3] Máté, M., Hollanda, D., „The Enveloping Surfaces of the Paloid Mill Cutter”, 18th International Conference on Mechanical Engineering, Baia Mare, 23-25 April 2010, ISSN2068-1267, pag.291-294.

The Electro-Hydraulic Analogy

András KAKUCS

Department of Mechanical Engineering, Faculty of Technical and Human Sciences,
Sapientia University, Tg. Mureş, e-mail: kakucs@rdslink.ro

Manuscript received April 14, 2010; revised April 25, 2010.

Abstract: There is an analogy between electric RLC circuits and hydraulic circuits. There also exists an analogy between electric and damped elastic structures, so there must exist an analogy between the hydraulic circuits and mechanical structures. This paper shows this analogy and on its base establishes a new model and method in analyzing hydraulic circuits.

Keywords: Hydraulic circuits, electro-hydraulic and mechanic analogy.

1. Introduction

Any hydraulic system is composed by some discrete elements interconnected by some pipes. It is well known that all these objects manifest a resistance against flow. We can conclude that the pressure drop caused by this resistance is proportional with the square of the volumetric flow rate:

$$\Delta p(t) = R_h \cdot Q^2(t)\,, \tag{1}$$

where the R_h hydraulic resistance depends on the flow regime and on the flow rate itself.

Inside of the hydraulic elements and the pipes a certain quantity of fluid is found. This quantity varies with the pressure and for an infinitesimal process the accumulated volume is proportional with the variation of the pressure:

$\mathrm{d}V = C_h \cdot \mathrm{d}p$, where C_h is the hydraulic capacity. Because $\mathrm{d}V = Q(t) \cdot \mathrm{d}t$, when the flow rate is varying in time, with zero initial conditions we obtain

$$\Delta p(t) = \frac{1}{C_h} \cdot \int_0^t Q(t)\,\mathrm{d}t \,. \tag{2}$$

Due to the mass of the fluid it manifests certain inertia to the change of the flow rate:

$$\Delta p(t) = L_h \cdot \frac{\mathrm{d}Q(t)}{\mathrm{d}t} \,, \tag{3}$$

where L_h is the hydraulic inductivity.

In general all three phenomena take place and the pressure drop is the sum of the three quantities above:

$$\Delta p(t) = L_h \cdot \frac{\mathrm{d}Q(t)}{\mathrm{d}t} + R_h \cdot Q^2(t) + \frac{1}{C_h} \cdot \int_0^t Q(t)\,\mathrm{d}t \,. \tag{4}$$

This last equation is analogous with Ohm's law written for a serial RLC circuit, which is:

$$u(t) = L \cdot \frac{\mathrm{d}i(t)}{\mathrm{d}t} + R \cdot i(t) + \frac{1}{C} \cdot \int_0^t i(t)\,\mathrm{d}t \,, \tag{5}$$

the only difference being the second power of the flow rate. This observation is the base of the electro-hydraulic analogy.

If we take the first derivative of Eq 4 that is (when the coefficients are constants):

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta p(t) = L_h \cdot \frac{\mathrm{d}^2 Q(t)}{\mathrm{d}t^2} + 2 \cdot R_h \cdot Q(t) \cdot \frac{\mathrm{d}Q(t)}{\mathrm{d}t} + \frac{1}{C_h} \cdot Q(t) \,, \tag{6}$$

Another analogy can be established, this time with the single degree of freedom vibrating mechanical system. We can define a new, flow-rate dependent coefficient as $R_h^* = 2 \cdot R_h \cdot Q(t)$ – this equation depicts nonlinear damping of the motion. The equation of the motion of such a system is:

$$F(t) = m \cdot \frac{\mathrm{d}^2 D(t)}{\mathrm{d}t^2} + c \cdot \frac{\mathrm{d}D(t)}{\mathrm{d}t} + k \cdot D(t)\,. \tag{7}$$

Based on this triple analogy the matrix stiffness method can be introduced in analyzing hydraulic circuits.

2. The equivalent electric circuit

The electric circuits usually are built up from some elements with concentrated parameters, but pipes are elements with distributed parameters. Given this, a problem arises: how to make a model with discrete elements with concentrated parameters when we have in fact distributed ones? The answer is given in electric engineering, when transmission lines are studied: these lines are modeled as quadripoles. We can use this discretization of the continuous elements in the case of pipes, too, and we obtain some equivalent circuits such as those in *Fig. 1*. The resistance and the inductivity are divided into two equal parts and the capacity is concentrated in the middle of the line element. In the figure Δp_1 and Δp_2 are the relative pressures in the nodes and Δp is the pressure drop along the pipe. It can be observed that the quantity of the fluid that enters is not equal to the quantity that leaves this pipe because of the presence of the capacitor.

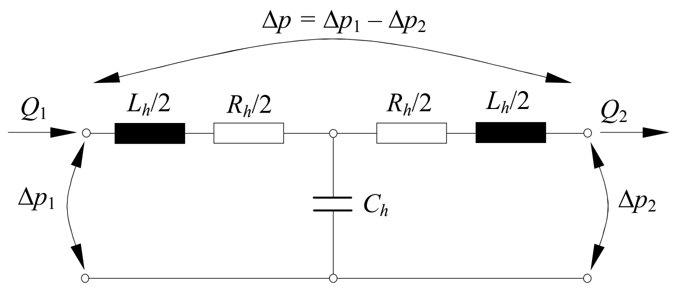


Figure 1: The “T” quadripole as the model of a pipe.

With these quadripoles we can analyze any hydraulic circuit with the methods of analysis of electric RLC circuits. Hydraulic organs (like valves, pressurized tanks) can be modeled by identifying their behavior. To illustrate all these things let us take a simple example.

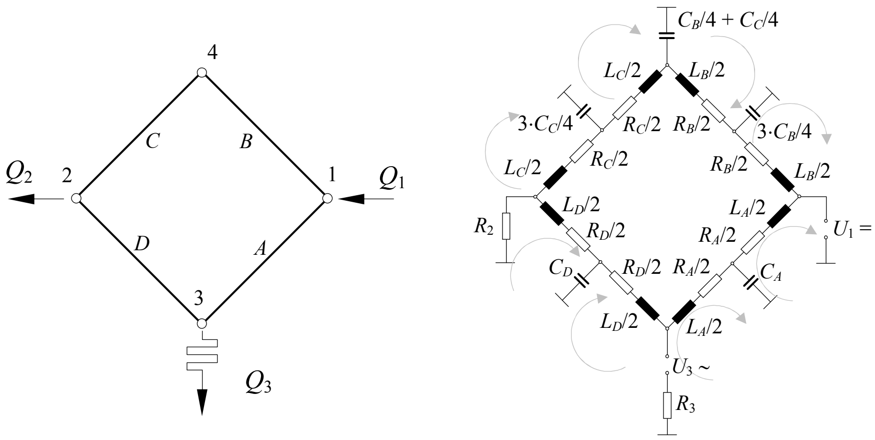


Figure 2: An example.

Fig. 2 represents a simple hydraulic circuit made from four pipes. In two nodes the fluid flows out and is supplied in another node. In one of the consumption points a hydraulic organ generates some vibrations: we want to find the effects of these vibrations on the other nodes. We can observe that almost all pipes form a pair of loops and those are independent loops of the circuit. A minor exception is caused by the fourth node where there is no discrete element linked to the mass (ground). These exceptions become annoying when three or more elements are linked in such points because in such cases the process of establishing the independent loops is not as simple as it was in the other cases. A solution to this problem is the slight modification of the "T" quadripole: we move a part (e.g. one quarter) of the hydraulic capacity of the pipe into that node; in this way this node will be linked to the mass by the capacitor moved there (the capacities are summed up because they are linked in parallel).

3. The solution of the problem

If we will write Kirchhoff's laws for the loops of the equivalent circuit we will get a system of equations of the form

$$[\boldsymbol{L}_h]\cdot\{\ddot{\boldsymbol{Q}}\}+[2\cdot\boldsymbol{R}_h]\cdot\{\dot{\boldsymbol{Q}}\}+[1/\boldsymbol{C}_h]\cdot\{\boldsymbol{Q}\}=\{\Delta\dot{\boldsymbol{p}}\}, \tag{8}$$

which is analogous with the equations of the motion of a vibrating mechanical system:

$$[\boldsymbol{M}]\cdot\{\ddot{\boldsymbol{D}}\}+[\boldsymbol{C}]\cdot\{\dot{\boldsymbol{D}}\}+[\boldsymbol{K}]\cdot\{\boldsymbol{D}\}=\{\boldsymbol{F}\}. \tag{9}$$

With this analogy we can imagine that all numerical procedures and algorithms used in solving the equations of motions of the vibrating systems can be directly adapted to solve RLC circuits.

Acknowledgements

Publishing of this Conference Proceeding is supported by the "Szülőföld Alap" foundation.

References

[1] BATHE, K. J. Finite Element Procedures. Prentice - Hall Inc, USA, 1996.
[2] LÁTRÁNYI J., ZALKA A. Válogatott fejezetek hidraulikus körfolyamatok tervezéséből (Selected Chapters of Design of Hydraulic Circuits). Budapest, Műszaki Egyetem, 1982
[3] SHABANA A. A. Vibration of Discrete and Continuous Systems. Springer-Verlag, New York, USA, 1997.

Basic Trends in the Development of the Holonic Manufacturing Concept

Levente BAKOS[1]

[1] Department of Mechanical Engineering, Faculty of Technical and Human Sciences, Sapientia University, Tg. Mureş, e-mail: bakos@ms.sapientia.ro

Manuscript received May 04, 2010; revised May 10, 2010.

Abstract: The paper presents some of the most relevant trends in the development of HMS. After a few years of "camouflage" it can be seen that in the past 1-2 years new achievements and development appeared in one of the most promising manufacturing concept of the last years of the past century. Being in accordance to the requirements of the modern industrial systems, more and more practical applications arise, and the concept itself evolves. The holonic concept became, once again, "trendy".

Keywords: Holonic Manufacturing Systems, e-manufacturing, distributed control system, ERP.

1. The HMS concept

The Holonic Manufacturing System (HMS) had its first period of popularity in the last years of the past century. The HMS was probably the most promising manufacturing concept developed during the IMS (Intelligent Manufacturing Systems) Programme [9]. The IMS is one of the most relevant industrial research programme launched ever. Holonic manufacturing was trying, and it still does, to overcome the limitations in flexibility of manufacturing systems by autonomous, decentralized and cooperative approach. The model was situated somewhere between the traditional manufacturing approach and one of the most influential model emerging at that time: MAS (Multi Agent Systems). Even till today the comparison of these two models is a subject of scientific work.

The HMS systems transpose the concepts developed by Koestler [6] for living organisms and social organizations to the manufacturing world. Holonic manufacturing is characterized by holarchies of autonomous and cooperative entities, called holons, which represent the entire range of manufacturing entities. The basic idea of holon is a double-face effect, which means that every holon is an autonomous whole and also a part of a larger holon at the same time.

A holon, as devised by Koestler, is an identifiable component of a (manufacturing) system that has a unique identity, yet is made up of subordinate components, and in turn is part of a larger whole [6]. The HMS concept is widespread, particularly due to the success of the reference architecture, PROSA, proposed by a group of researchers form Katholieke Universitat Leuven, led by Van Brussel [2].

Similarly to the MAS, in this concept the products and production resources are assumed to be decisional, non passive entities. They communicate, decide or act during production process. Beside that, the holonic concept has many other similarities with the Multi-Agent System concept. Probably the main difference consists of the notion of holarchy. The holons develop collaborative algorithms which guide the holons to form, more or less, temporary hierarchical structures. The holarchy is the main tool that will move the system coherently toward the ultimate desired goal state. A holon can be distinguished from an agent not only because it is composed of recursive holons, but almost each type of them has both an informational and a physical component. In the recent studies, a holon in the context of manufacturing can correspond to a physical part (for instance a workstation), having three responsibilities: self-assertion, integration and canon maintenance [7] In this context a holon can become a part of the MAS and the agent part of the HMS. See *Fig.1*

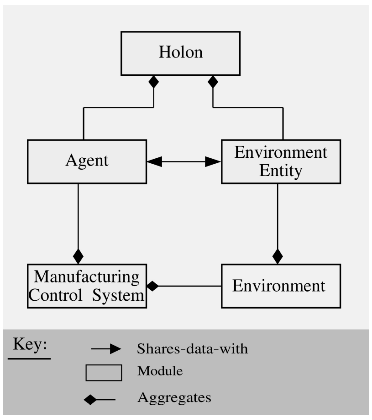


Figure 1: Holon versus agent (source [7]).

3. Holonic Production Control

In the early period the research of the HMS was focused mostly on production control. Inside that category the research field was extremely wide: scheduling, shop floor control, process planning, task management, fault recovery, etc. The general topic of the research activity was the development of a holonic based, non-centralized control system. The holonic production control system is not fully hierarchical, which means that non-hierarchical relationships can and do exist in the system. We must notice here the difference between the shop floor control system (that includes manual and automatic labor, production reporting, online inquiries, provides links to tasks that take place on the production floor) and the (holonic) manufacturing control system (production scheduling, schedule execution, on-line resource allocation decisions, responsiveness towards breakdowns, etc.)

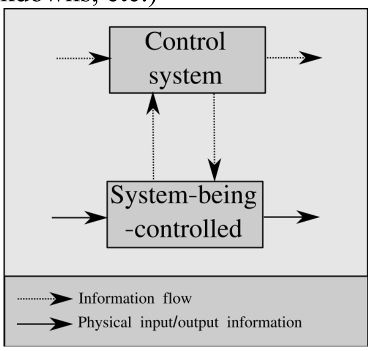


Figure 2: The difference between the control system and the manufacturing system. (Source [7]).

The performance of the control system is largely determined by its capability to influence the manufacturing system. The manufacturing control system, if it is holonic-based, consists of several holons, starting with the scheduling holon till the resource holons. Usually the process of creating these holons consists of the following steps:

- developing an abstract model of the shop floor reality,
- creating functional holons according to a certain application domain,
- identifying the usual relationships among the holons,
- defining the holarchies and the messages, the information flow, in the holarchies,
- creating the frame of the disturbance handling and the control system,

- later, if necessary, the number of holons is set according to the needs.

Today these researches are still representing one of the key research topics, mostly connected to the topics presented in the next paragraphs. This branch of the research seems to achieve a certain maturity, the model proved its functionality, robustness and its applicability in numerous industrial applications.

4. Supply Chain Management and HMS

The objective of the supply chain is to deliver the right product at the right time, to the right customer at minimum cost. A network of facilities is designed to perform tasks such as procuring materials and distributing the finished products to the customers. The supply chain is designed to achieve timely and economical delivery of desired products for accomplishing the cycle of O2D (Order to Delivery). The main activities are:

- demand planning,
- allocation planning,
- capacity modeling,
- allocation management,
- order management,
- available to promise,
- output planning.

In the recent period the holonic research topics exceeded the physical boundaries of the enterprise. Being in the e-manufacturing era, the holonic

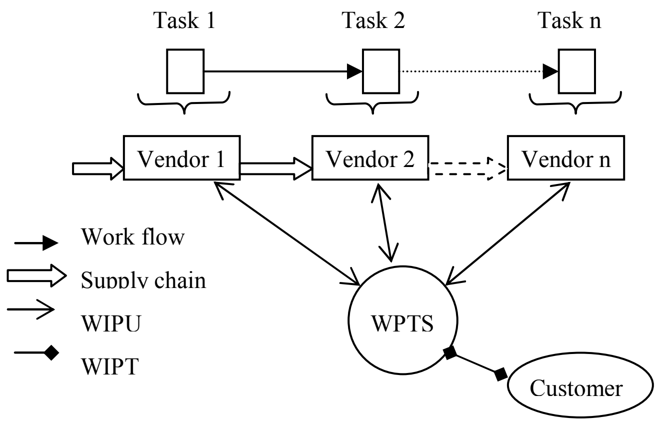


Figure 3: Supply chain with adaptive WIP tracking system.

concept is successfully developed outside of the shop-floor control level. Let us recommend just two papers appeared in the last months [5], [3]. These holonic projects integrate not only the ERP (Enterprise Resource Planning) or CRM (Customer Relationship Management) applications, but they are dealing with the similar applications of their industrial partners and/or customers. The purchase order process, the real time tracking work-in-progress system, the demand forecast process or the demand planning are more and more in the focus of the researches.

In *Fig.3* a short form of a supply chain system is presented, in which holonic solutions can be used. In the figure, being constrained by the space, the following abbreviations have been used: WIP- Work in Progress, WIPU- Work in Progress Update/Notices, WIPTR- Work in Progress Tracking Report. Usually the management of the supply chain requires large scale system modeling, object oriented systems. By introducing the holonic approach in the supply chain management, the optimization effort decreases considerably [4]. The supply chain holonic handling methodology can be extended also in the Engineering Chain System (ECS). The ECS is similar to the supply chain network, and means a network of facilities and distributed services that performs device design, verification of design, manufacturing pilot run, assembly and test operations, yield improvements, final release for mass production. For short product life cycle it is essential that the data exchange among the departments and companies to be with the speed of the light.

5. Stigmergy and HMS

The recent observation is that both MAS and HMS software for real-world manufacturing control is enormously complex, both from conceptual and pragmatic point of view, and needs to be integrated. A second observation is that the environment has a prominent role. While previous research primarily focused on modeling the holons and their interactions, later research discovered that the environment plays an important role in the dynamics of the multi-agent system. In the latest research the notion of stigmergy arises extremely often. [1], [8]. The basic PROSA design has been augmented to support stigmergy. A French entomologist (Grassé) introduced the term “stigmergy” to describe the mechanism by which termites coordinate their mound-building activities. In such activities, many individuals participate in a collective task, and the stimuli provided by the emerging structure are used to coordinate the individual actions. A similar mechanism is used by ants laying down pheromone trails between a food source and their nest. So, stigmergy refers to a mechanism of spontaneous, indirect coordination between entities, in which the marks left in the environment by an action stimulates the completion of other subsequent actions. In HMS it is used an artificial pheromone in order to obtain similar effects to the natural pheromone. In the real world, three basic operations have been associated with the stigmergic process: information fusion, information removal and local information distribution. In the first, deposits from individual entities are aggregated to allow the easy fusion of information. In the second, pheromone evaporation over time is used to remove obsolete or inconsistent information. In the last, information is provided according to pheromone diffusion in the immediate (local) neighborhood. Comparatively to the ants, the holons have to face other challenges such as multiple objectives, individual goals, conflicts with other holons [1].

6. Conclusion

Holonic manufacturing executing systems today still need to make their way into the market. Achieving a large enough number of implementations of the manufacturing execution system is essential to its real-world success. Having a recent conversation with Paul Velckenaers, from K.U. Leuven, one of the key figures in the holonic research in Europe, the author has the conclusion that at global level holonic concept has a dynamic developing period, and it is not the same like few years before. Holonics became "trendy" in many research fields, with all the advantages and disadvantages which may result about this fact. A lot published materials appeared in which the concept is used, but the topic in fact has nothing in common with the main holonic idea. This may lead to the "dilution" of the concept, making it "cheap". Without a well developed mathematical model, communication protocol, disturbance handling procedures, testbed, industrial or practical relevance any attempt to develop the concept will remain just an attempt. Beside "being popular", the Holonic Manufacturing concept little by little reenters in the attention of the researchers and new horizons arise.

References

[4] Bakos L., "Contribuţii la modelarea sistemelor de fabricaţie holonice", *PhD Thesis*, Universitatea Tehnica Cluj Napoca, 2006

[5] Brussel, H., Wyns, J., Valckenaers, P., Bongaerts L., Peeters P., "Reference architecture for holonic manufacturing systems: PROSA.", *Computers in Industry* 37 3, 1998

[6] Cheng, F.T *et al*, „Advanced e-Manufacturing Model. The significance of Large-Scale, Distributed and Object Oriented Systems", *IEEE Robotics and Automation Magazine*, vol.17. No.1, 2010

[7] Dávid L., Bakos L., "Solving MPC Strategies Used In Manufacturing Process Planning Based On Holonic Concept*", MTEM2009,* Cluj Napoca, 2009

[8] Jafari, D., Husseini, S. M. Moattar, Fazel Zarandi, M. H., Farahani, R. Zanjirani, "Coordination of order and production policy in buyer–vendor chain using PROSA Holonic

architecture", *The International Journal of Advanced Manufacturing Technology*, volume 45, nr.9-10, Springer-Verlag London Limited 2009
[9] Koestler, A., "*The Ghost in the Machine",* Hutchinson & Co, London 1967, (Second Edition: Arkana Books, London, 1989
[10] Verstraete,P., „Integrating existing scheduling techniques into the Holonic Manufacturing Execution System", *PhD Thesis*, K.U.Leuven, 2009
[11] Valckenaers, P, Holvoet, T., „The Environment: an Essential Abstraction for Managing Complexity in MAS-based Manufacturing Control", *Environments for Multi-Agent Systems II*, SpinerLink, 2006
[12] www.ims.org

MACRo'2010- International Conference on Recent Achievements in
Mechatronics, Automation, Computer Science and Robotics

Structural Analysis of 6 DOF 3-PRRS Parallel Mechanism

Zoltán FORGÓ

Department of Mechanical Engineering, Faculty of Technical and Human Sciences,
Sapientia University, Tg. Mureş, e-mail: zforgo@ms.sapientia.ro

Manuscript received April 04, 2010; revised April 18, 2010.

Abstract: The number of parallel mechanism applications in the industry is growing and the interest of the academia to find new solutions and applications to implement such mechanisms is present all over the world. The lower degree of freedom mechanisms are suited for some specific tasks are preferred because of the architecture simplicity and therefore the easy mathematical modeling and finally, but not at least for economical reasons. In this paper a symmetrical six degrees of freedom mechanism (3-PRRS) will be modelled, with double actuated limbs. The easy way of mathematical modeling is given by the fact that the mechanism can be considered as an extended well known planar Delta manipulator. The driven P joints from each limb ensures (planar Delta not owned) third translation and two rotations of the moving platform. The aspects of singular configurations are considered also in the paper.

Keywords: parallel mechanism, kinematics, group theory, Lie algebra.

1. Introduction

The number of applications in the industry witch uses parallel mechanisms are growing and the interest of the academia to find new solutions and applications to implement such mechanisms is present all over the world. The lower degree of freedom mechanisms witches are suited for some specific tasks are preferred because of the architecture simplicity and therefore the easy mathematical modeling and finally, but not at least for economical reasons.

The 6 degrees of freedom (dof) parallel mechanism is introduced by Steward and Gough [8] and since then many aspects of the mechanism and his application are revealed. During the last decades more attention has been paid to the study of 6 dof parallel mechanisms, including synthesis and analysis on kinematics, dynamics, singularities, error and workspace. Some milestones in the analysis of those mechanisms are made by Earl and Rooney using a method for synthesis of new kinematic structures [1], Hunt studied the manipulators on the basis of screw theory [5], Tsai is using systematic methodology in [9] and Hervé discussed the structural synthesis of parallel robots using the mathematical group theory [2]. More recently Shen proposed a systematic type synthesis methodology for 6 dof kinematic structures enumerating 29 parallel structures [7]. Hereby Shen defines the hybrid single-open chains (HSOC) which are able to generate three translations and three rotation angles. Using those HSOCs four 6 dof manipulators are presented with symmetrical arrangement of the limbs (see No.3-No.6 architectures, Table 2. from [7]). According to Tsai [10], the symmetry implies the use of the same number of actuators on the same positions in each limb. Moreover he says that a parallel manipulator is symmetrical if satisfies the condition that the number of limbs is equal to the number of degrees of freedom of the moving platform. In the case of double actuated limbs (with two actuated joints) the last presented condition can be omitted. So the HSOCs defined by Shen can be replaced by serial chains witch enables three translations and three rotations also.

This paper will present some kinematic structures according the above mentioned criterions without the aim of full discussion about the all possible structures. The geometrical model of one architectures is presented as well.

2. General motion generators

The enumeration of serial topology limbs which enables the spatial motion (three translation and three rotation) is greatly simplified by using the Lie group of rigid body displacement as introduced by Hervé [3]. If each limb of a parallel manipulator generates a subset of possible displacement, which is a Lie subgroup, the intersection set is also a Lie subgroup of the mobile platform. According this statement if the platform undergoes the spatial, general motion, each limb must ensure the three translations and three rotations. According to Table 1 $\{D\}$ denotes the general rigid body motions for the 6 dof mobile platform and $\{L_i\}$ the displacement Lie subgroup of the i^{th} limb. The relation between them is given by:

$$\{L_1\} \cap \{L_2\} \cap \{L_3\} = \{D\} . \tag{1}$$

It is obvious that the only possibility for a true equation (1) is:

$$\{L_1\} = \{L_2\} = \{L_3\} = \{D\} . \tag{2}$$

To obtain simple mechanical structures, better symmetry and good manufacturing for the three limbs the same architecture is considered. That for, the analysis of the $\{L_i\}$ displacement Lie subgroup is carried out. The notations for displacement Lie subgroups are recalled in Table 1 [4].
According to the group theory can be stated:

$$\begin{aligned}\{L_i\} &= \{D\} = \{T\}\{S(N)\} = \\ &= \{T(\boldsymbol{u})\}\{T(\boldsymbol{v})\}\{T(\boldsymbol{w})\}\{R(N,\boldsymbol{u})\}\{R(N,\boldsymbol{v})\}\{R(N,\boldsymbol{w})\} \quad \forall N.\end{aligned} \tag{3}$$

Table 1. List of displacements Lie subgroups [4].

Lie subgroup	*Description of the subgroup*
$\{E\}$	identity
$\{T(\boldsymbol{u})\}$	translations parallel to the ***u*** vector
$\{R(N,\boldsymbol{u})\}$	rotations around the axis determined by N and ***u***
$\{H(N,\boldsymbol{u},p)\}$	helical motions with axis (N,***u***) and the pitch p
$\{T(Pl)\}$	translations parallel to the *Pl* plane
$\{C(N,\boldsymbol{u})\}$	cylindrical motions along an axis (N,***u***)
$\{T\}$	spatial translations
$\{G(\boldsymbol{u})\}$	planar gliding motions perpendicular to ***u***
$\{S(N)\}$	spherical motions about point *S*
$\{X(\boldsymbol{u})\}$	Schoenflies motions
{*D*}	general rigid body motions or displacements

Based on [4] a planar joint has 5 equivalencies as presented bellow:

$$\{G(\boldsymbol{u})\} = \{R(A,\boldsymbol{u})\}\{R(B,\boldsymbol{u})\}\{R(C,\boldsymbol{u})\}; \tag{4}$$

$$\{G(\boldsymbol{u})\} = \{R(A,\boldsymbol{u})\}\{T(\boldsymbol{v})\}\{R(C,\boldsymbol{u})\} \quad \boldsymbol{v} \perp \boldsymbol{u}; \tag{5}$$

$$\{G(\boldsymbol{u})\} = \{T(\boldsymbol{v})\}\{R(B,\boldsymbol{u})\}\{R(C,\boldsymbol{u})\} \quad \boldsymbol{v} \perp \boldsymbol{u}; \tag{6}$$

$$\{G(\boldsymbol{u})\} = \{T(\boldsymbol{v})\}\{R(B,\boldsymbol{u})\}\{T(\boldsymbol{w})\} \quad \boldsymbol{v},\boldsymbol{w} \perp \boldsymbol{u}; \tag{7}$$

$$\{G(\boldsymbol{u})\} = \{T(\boldsymbol{v})\}\{T(\boldsymbol{w})\}\{R(C,\boldsymbol{u})\} \quad \boldsymbol{v},\boldsymbol{w} \perp \boldsymbol{u}. \tag{8}$$

Considering equations (4) and (8) respectively equality $N \equiv C$ the equation (3) becomes:

$$\{L_i\} = \{T(\boldsymbol{u})\}\{R(A,\boldsymbol{u})\}\{R(C,\boldsymbol{u})\}\{R(B,\boldsymbol{u})\}\{R(B,\boldsymbol{v})\}\{R(B,\boldsymbol{w})\} \ \forall A,B,C; \tag{9}$$

$$\{L_i\} = \{T(\boldsymbol{u})\}\{R(A,\boldsymbol{u})\}\{R(C,\boldsymbol{u})\}\{S(B)\} \quad \forall A,B,C. \tag{10}$$

The above defined $\{L_i\}$ displacement Lie group variants are presented in Figure 1.

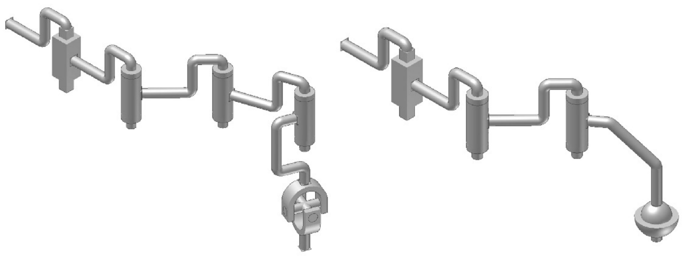


Figure 1 The $\{L_i\}$ displacement Lie group variants incorporating the X-motion generator.

It can be easily observed, from equation (9) and in *Fig. 1*a as well, the X-motion (or Shoenflies motion) generator. Considering primitive Schoenflies-motion generators [6] equivalences can be applied. Extending those generator family members with the universal joint as seen in *Fig. 1*, new generators for {D} displacement Lie group can be introduced. However, this enumeration is out of the topic of this paper. Because of the reduced link number and simplicity, in further investigation, the Figure 1b variant is preferred. Using other geometrical constraints the architecture is presented in [11] also. The schematic design of such a limb for a 6 dof manipulator is presented in Figure 2b. The index *i* is introduced because the same type of limbs are used for moving the manipulator platform.

3. Kinematics of 3-PRRS mechanism

The general setup for the parallel mechanism having three translations and three rotations for the end effector (denoted by P) is presented in *Fig. 2c.* For simplicity the mechanism is presented from top view. The geometrical parameters used in the mathematical modeling are enumerated in sketches b) and c) from *Fig. 2*. Further the real number values x_N, y_N and z_N are introduced as the coordinates of a point N in the Cartesian space $0x_0y_0z_0$.
Using the equivalency between sketches a) and b) from *Fig. 2* can be stated:

$$\boldsymbol{C_i B_i} = \boldsymbol{C_i D_i} + \boldsymbol{D_i B_i} = C_i D_{ix} \cdot \boldsymbol{i} + C_i D_{iy} \cdot \boldsymbol{j} + D_i B_i \cdot \boldsymbol{k}\,, \tag{11}$$

$$\boldsymbol{C_i B_i} = C_i B_{ix} \cdot \boldsymbol{i} + C_i B_{iy} \cdot \boldsymbol{j} + D_i B_i \cdot \boldsymbol{k}\,, \tag{12}$$

where $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ are the unit vectors of the x_0, y_0 and z_0 Cartesian axes. The setup of the mechanism (based on the projection of the manipulator on the $0x_0y_0$ plan – top view from *Fig. 2*) suggests a planar Delta manipulator.

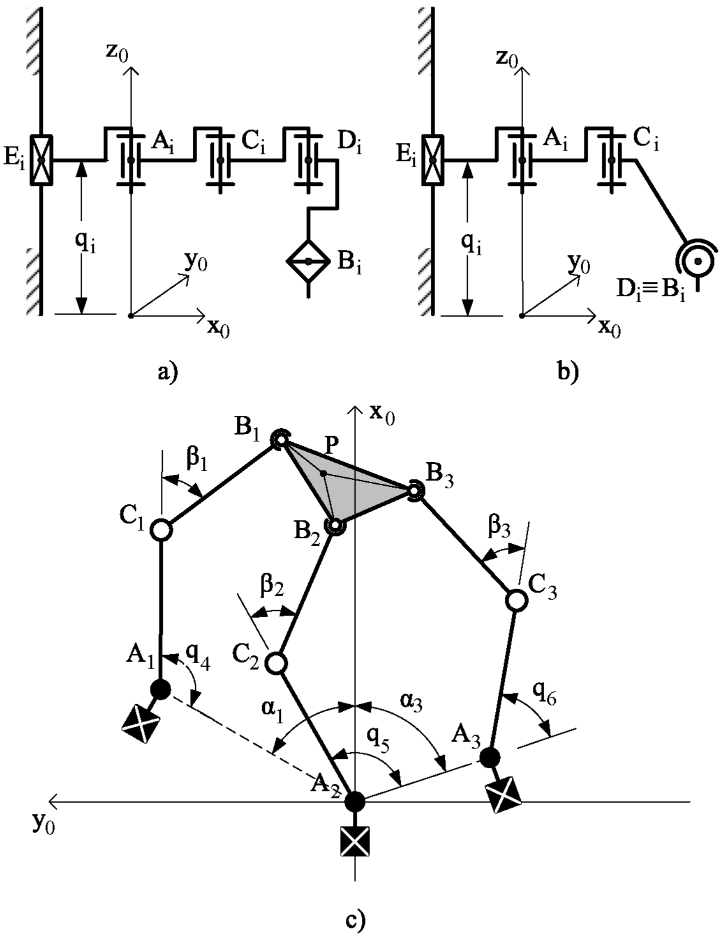


Figure 2 Schematic design of i[th] limb of the 6 dof manipulator (a, b), and the top view of the proposed mechanism (c). The shaded couplings are the active joints (one prismatic and one rotation for each limb), and the white ones are passive bonds.

That for the mathematical modeling of the proposed mechanism is made easy and it is like a well known planar Delta manipulator modeling [10] with some completions. Those must be made due to the fact, that it is possible to rotate the platform around the x_0 and y_0 axes too, and so the projections of B_iB_{i+1} platform length are variable. Through these paragraphs the inverse and direct kinematics of the proposed mechanism is defined, and issues about singular configurations are presented as well.
At the beginning the closure equation is considered for the three limbs:

$$\boldsymbol{OA_i} + \boldsymbol{A_iC_i} + \boldsymbol{C_iB_i} = \boldsymbol{OP} + \boldsymbol{PB_i} \text{ where } i = 1,2,3\,. \tag{13}$$

In the case of inverse kinematic modeling the right side of the (13) equation is given $\boldsymbol{X} = [x_P \; y_P \; z_P \; \alpha \; \beta \; \gamma]^T$ through the characteristic point (denoted by P) coordinates and through the three angles around the axes of $0x_0y_0z_0$ fixed system, and the task is to determine the robot parameters $\boldsymbol{q} = [q_1 \; q_2 \; q_3 \; q_4 \; q_5 \; q_6]^T$ from the left side of the equation. Assuming that vector $\boldsymbol{a}$ have the components $\boldsymbol{a_{xy}}$ parallel to the $0x_0y_0$ plan and $\boldsymbol{a_z}$ parallel to the z_0 axis, the (13) equation will be:

$$\begin{cases} \boldsymbol{OA_{ixy}} + \boldsymbol{A_iC_{ixy}} + \boldsymbol{C_iB_{ixy}} = \boldsymbol{OP_{xy}} + \boldsymbol{PB_{ixy}} \\ \boldsymbol{OA_{iz}} + \boldsymbol{A_iC_{iz}} + \boldsymbol{C_iB_{iz}} = \boldsymbol{OP_z} + \boldsymbol{PB_{iz}} \end{cases} \text{where } i = 1,2,3 . \qquad (14)$$

To determine the q_i translational parameters (i=1,2,3), introduced in Figure 2b, the second equation from (14) is considered:

$$q_i \cdot \boldsymbol{k} + \boldsymbol{C_iB_{iz}} = z_P \cdot \boldsymbol{k} + \boldsymbol{PB_{iz}} \text{, respectively in scalar form} \qquad (15)$$

$$q_i = z_P + PB_{iz} + C_iB_{iz} \qquad (16)$$

Hence C_iB_{iz} is a constant geometrical parameter of the manipulator, the first two terms from equation (16) right side contains the general parameters because $PB_{iz} = PB_{iz}(\alpha,\beta,\gamma)$.

To obtain the q_{i+3} rotation joints parameters (i=1,2,3) the first equation from (14) is recalled and presented in scalar form:

$$\begin{cases} OA_i\cos\alpha_i + A_iC_i\cos(q_{i+3}+\alpha_i-\pi)+C_iB_i\cos(q_{i+3}+\alpha_i-\pi-\beta_i)=x_P+PB_{ix} \\ OA_i\sin\alpha_i + A_iC_i\sin(q_{i+3}+\alpha_i-\pi)+C_iB_i\sin(q_{i+3}+\alpha_i-\pi-\beta_i)=y_P+PB_{iy} \end{cases} ,(17)$$

where $PB_{ix} = PB_{ix}(\alpha,\beta,\gamma)$ and $PB_{iy} = PB_{iy}(\alpha,\beta,\gamma)$ respectively i=1,2,3. To eliminate the β_i parameter belonging to a passive joint, the equations are rearranged, and summing the square of the two equations in (17) yields:

$$e_{1i}\cdot\sin(q_{i+3}+\alpha_i-\pi)+e_{2i}\cdot\cos(q_{i+3}+\alpha_i-\pi)+e_{3i}=0 , \qquad (18)$$

where

$$\begin{cases} e_{1i}=-2\cdot A_iC_i\cdot(y_P+PB_{iy}-OA_i\sin\alpha_i); \\ e_{2i}=-2\cdot A_iC_i\cdot(x_P+PB_{ix}-OA_i\cos\alpha_i); \\ e_{3i}=(x_P+PB_{ix}-OA_i\cos\alpha_i)^2+(y_P+PB_{iy}-OA_i\sin\alpha_i)^2+A_iC_i^2-C_iB_i^2. \end{cases} \qquad (19)$$

Solving equation (18) by using the substitutions:

$$\begin{cases} \sin\left(q_{i+3}+\alpha_i-\pi\right)=\dfrac{2t_i}{1+t_i^2} \\ \cos\left(q_{i+3}+\alpha_i-\pi\right)=\dfrac{1-t_i^2}{1+t_i^2} \end{cases} \text{ where } t_i=\tan\frac{q_{i+3}+\alpha_i-\pi}{2}, \tag{20}$$

the q_{i+3} parameters (i=1,2,3) are given by:

$$q_{i+3}=\pi-\alpha_i+2\tan^{-1}\frac{-e_{1i}\pm\sqrt{e_{1i}^2+e_{2i}^2-e_{3i}^2}}{e_{1i}-e_{2i}} \tag{21}$$

Equations (16) and (21) define the robot parameters in case of inverse kinematics. To obtain the general coordinates of the mechanism it is necessary to calculate the position of the $B_i(x_{Bi},y_{Bi},z_{Bi})$ joints (i=1,2,3) in Cartesian space and knowing the geometrical dimensions of the mobile platform the $\boldsymbol{X}=[x_P\ \ y_P\ \ z_P\ \ \alpha\ \ \beta\ \ \gamma]^{\mathrm{T}}$ vector is obvious. The x_{Bi}, y_{Bi}, and z_{Bi} values are defined through the following nine equations:

$$\begin{cases} \left(x_{Ci}-x_{Bi}\right)^2+\left(y_{Ci}-y_{Bi}\right)^2=C_iB_i^2 \\ \left(x_{Bi}-x_{Bj}\right)^2+\left(y_{Bi}-y_{Bj}\right)^2=d^2-\left(z_{Bi}-z_{Bj}\right)^2 \\ z_{Bi}=z_{Ci}-D_iB_i \end{cases} \text{ for } \begin{array}{l} i=1,2,3 \\ j=\begin{cases} i+1, if\ i=1,2 \\ 1\ , if\ i=3 \end{cases} \end{array} \tag{22}$$

where $x_{Ci}=x_{Ci}(q_{i+3}), y_{Ci}=y_{Ci}(q_{i+3}), z_{Ci}=z_{Ci}(q_i)$ respectively C_iB_i and D_iB_i are constant geometrical dimensions using i=1,2,3. It is important to mention, that in present case the forward kinematics deals with only 8 solutions (as by the planar Delta robot).

To complete the kinematic calculations the relation between the actuated joints and the platform velocities is needed and obtained through:

$$J_x\cdot\dot{\boldsymbol{X}}=J_q\cdot\dot{\boldsymbol{q}}, \tag{23}$$

where the matrices:

$$J_x=\begin{bmatrix} b_{1x} & b_{1y} & b_{1z} & e_{1y}b_{1z}-e_{1z}b_{1y} & e_{1z}b_{1x}-e_{1x}b_{1z} & e_{1x}b_{1y}-e_{1y}b_{1x} \\ b_{2x} & b_{2y} & b_{2z} & e_{2y}b_{2z}-e_{2z}b_{2y} & e_{2z}b_{2x}-e_{2x}b_{2z} & e_{2x}b_{2y}-e_{2y}b_{2x} \\ b_{3x} & b_{3y} & b_{3z} & e_{3y}b_{3z}-e_{3z}b_{3y} & e_{3z}b_{3x}-e_{3x}b_{3z} & e_{3x}b_{3y}-e_{3y}b_{3x} \end{bmatrix}, \tag{24}$$

$$J_q = \begin{bmatrix} b_{1z} & 0 & 0 & a_{1x}b_{1y}-a_{1y}b_{1x} & 0 & 0 \\ 0 & b_{2z} & 0 & 0 & a_{2x}b_{2y}-a_{2y}b_{2x} & 0 \\ 0 & 0 & b_{3z} & 0 & 0 & a_{3x}b_{3y}-a_{3y}b_{3x} \end{bmatrix}, \quad (25)$$

can be written using the notations $\boldsymbol{a}=\boldsymbol{A_i C_i}$, $\boldsymbol{b}=\boldsymbol{C_i B_i}$ and $\boldsymbol{e}=\boldsymbol{PB_i}$. Equation (23) can be considered for calculation of direct and inverse kinematics.

4. Singular configurations

The singularity analysis of this mechanism can be done based on the matrices from (24) and (25). Inverse kinematic singularities occurs in case of $a_{ix}b_{iy}-a_{iy}b_{ix}=0$ (i=1,2,3) which says for workspace boundaries. Other possibility is $b_{iz}=0$ (i=1,2,3) but it can be avoided through geometrical design, because it is a constant value. Direct kinematic singularities occurs when at the same time can be stated that $b_{ix}=0$ or $b_{iy}=0$ (i=1,2,3), what means the C_iB_i links are parallel. The same type of singularities can be found for coexistence of $e_{ix}b_{iy}-e_{iy}b_{ix}=0$ (i=1,2,3) in case of coliniar $\boldsymbol{C_i B_i}$ and $\boldsymbol{PB_i}$ vectors. Both direct kinematic singularity cases can be avoided by carefully geometrical design, which is out of the aim of this paper.

5. Conclusions

This paper deals with a 6 degrees of freedom manipulator architecture using the group theory. The mobile platform is connected to the base through three PRRS limbs, each being double actuated on the first and second joint levels. The inverse geometrical calculations are performed through equations (16) and (21), hence the direct modeling is presented through the equation system (22). The relation between the robot and general velocities is stated by the (23) equality. Some aspects about the singular configurations are introduced in the paper based on the mentioned equation. As can be seen in the figures presented in this paper the architecture is the extension of the well known planar Delta robot to a 6 dof mechanism. The mathematical model of the spatial manipulator reflects that fact very well. The simple setup of the presented mechanism assures a good manufacturability and needs a relative easy control algorithm considering some other 6 dof manipulators.

References

[1] Dasguta, B. and Mruthyunjaya, T.S.: The Stewart platform manipulator: a review. Mechanism and Machine Theory, 35, 15-40 (2000)

[2] Hervé, J.M.: Design of parallel manipulators via the displacement group. In: Proceedings of the 9th World Congress on Theory of Machine and Mechanisms, Miland, pp. 2079-2082 (1985)

[3] Hervé, J.M.: The Lie group of rigid body displacements, a fundamental tool for mechanism design. Mechanism and Machine Theory, 34, 719-730 (1999)

[4] Hervé, J.M.: The planar-spherical kinematic bond: Implementation in parallel mechanisms. Available via http://www.parallemic.org/Reviews/Review013p.html

[5] Hunt, K.H.: Structural kinematics of in-parallel-actuated robot arms. ASME Journal of Mechanical Design, 105, 705-712 (1983)

[6] Lee, C.C. and Hervé, J.M.: Type synthesis of primitive Schoenflies-motion generators. Mechanism and Machine Theory, 44, 1980-1997 (2009)

[7] Shen, H., Yang., T. and Ma, L.: Synthesis and structure analysis of kinematic structures of 6-dof parallel robotic mechanisms. Mechanism and Machine Theory, 40, 1164-1180 (2005)

[8] Stewart, D.A.: Platform with six degrees of freedom. In: Proceedings on Institution of Mechanical Engineering, 180, 371-386 (1965)

[9] Tsai, L.W. and Joshi, S.: Kinematics and optimization of a Spatial 3-UPU Parallel manipulator. ASME Journal of Mechanical Design, 122, 439-446 (2000)

[10] Tsai, L.W.: Robot Analysis – The Mechanics of Serial and Parallel Manipulators. John Wiley & Sons (1999)

[11] Olea G., Plitea N., Takamusa K. Kinematical analysis and simulation of a new parallel mechanism for robotics application. In *ARK*, pages 403-410, Piran, 25-29 June 2000

Industrial Reality session participants

We have started in 1992. Our company was and is an active participant in computer market and a well known solution provider in information and communication technology services. After years of high quality efforts and based on a growing intellectual and infrastructural capacity, we are now well known representatives of our area of specialty.

At this time, we are providers of high-level technology and services. The main areas of interest are:

- Industrial Automation: mainly industrial process control projects and commissioning of these projects
- Programming of SCADA, HMI systems
- Design and manufacturing of low voltage circuits, programmable controllers, power supplies, data acquisition systems and communication devices
- Business and IT services
- Software development, web-design, webhosting & email systems.

NIVELCO
SC NIVELCO Tehnica Măsurării SRL

Nivelco Measurement Techniques Ltd, Romania is a young company. It was founded in 2005 as a unique representative in Romania of Nivelco Process Control, Hungary who is present in almost all countries of the World. The profile of this company is to produce the instrumentation of measurement and industrial control equipments. Our robust, precise and durable products can be classified as limit switches, different type of converters, transducers, sensing elements, controllers, software packages for process monitoring. These products have a very large area of utility starting from industrial applications to systems used to improve every day standard of living.

Contact address: SC NIVELCO TEHNICA MĂSURĂRII SRL
540069 Tg. Mureş, str. Ion Creangă nr.3/1
Telefon/fax: 0265- 30.61.92, E-mail: romania@nivelco.com, Web: www.nivelco.com

Industrial Reality session participants

FUJITSU

Fujitsu Microelectronics Europe (FME) is a major supplier of semiconductor products. The company provides advanced systems solutions to the automotive, digital TV, mobile telephony, networking and industrial markets. Engineers from design centers dedicated to microcontrollers, graphics controllers, mixed-signal, wireless, multimedia ICs, ASIC products and software development, work closely with FME's marketing and sales teams throughout EMEA to help satisfy customers' systems development requirements. This solutions approach is supported by a broad range of advanced semiconductor devices, IP, building blocks and software.

For more information, please see: www.emea.fujitsu.com/microelectronics

Lambda Communications

The power of information

Lambda Communications Ltd. is a Romanian R&D private limited company, committed to supply Romanian industry and education with concepts to find innovative solutions in electronics and related products. Being a "problem-solver" enterprise, Lambda Ltd has technical facilities and knowhow to follow a project from an initial idea to a finished product for clients ranging from individuals to big companies. In the focus of our activities stands the hardware and software development of microcontroller based products and advanced projects for electronics and telecommunications, incorporating 'High Tech' technology. Highly qualified employees guarantee the quality of our activity.

Lambda Communications Ltd.
Tirgu Mureş
Str. Avram Iancu Nr. 37.
Tel. +40-265-215218

Scientia Publishing House
400112 Cluj-Napoca
Matei Corvin street 4.
Tel./fax: +40-264-593694
E-mail: scientia@kpi.sapientia.ro
www.scientiakiado.ro

Proofreading:
László Bakó, András Kelemen, Katalin György

Make-up:
József Domokos

Typography:
The authors

Printed by Digital Color Company S.R.L.
in 100 issues
Director: Andreea Moscviciov