Future Internet Visions and Research Clusters

Gyula Sallai

Future Internet Research Cooperation Centre, Debrecen;
Inter-University Centre for Telecommunications and Informatics;
Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics
Magyar tudósok krt. 2, H-1117 Budapest, Hungary
e-mail: sallai@tmit.bme.hu

Abstract: Telecommunications and the Internet are forming an increasingly integrated and global system for processing, storing, transporting information and managing content. At the same time the identification capacity of the Current Internet is running out, Internet architectures are reconsidered for better managing mobility and quality requirements, data handling and security issues, as well as for exploiting the opportunities derived from technological development. The future of the Internet became an important research and standardization area, focusing on service, resource, content and environmental awareness. The paper summarizes the challenges of the Current Internet, draws up the visions and the recent capabilities of the Future Internet, then identifies and clusters the relevant research topics defining the chapters of Future Internet research activity in a layered model from basic research on Internet Science through the Internet Engineering up to Future Internet applications and experiments.

Keywords: Future Internet; Future Internet vision; Future Internet research; Internet Science; Internet Engineering; Internet technologies; Future Internet applications

1 Introduction

Telecommunications and the Internet are forming more and more an integrated global infrastructure for processing, storing and transporting information and managing content related to locations, people and intelligent devices. This integration process is based on the rapid evolution of the digital technology and the penetration of the Classic Internet concept. The progression towards an integrated telecommunications, information technology and electronic media (TIM) sector has been presented e.g. in [21, 36, 49, 50]:

- The first step was the digitization and integration of network functions within voice communication (telephony), data communication and media communication (broadcasting) sectors, each separated by their content.

- Secondly, a uniform digital communication technology of various contents and an integrated telecommunications (formally referred to as electronic communications) sector was formed.

- The third step presents the convergence of resources and processes for communications, information processing and storing, and content handling, the emergence of the integrated TIM sector and synergic TIM applications based on IPv4/TCP.

- Recently, in the fourth step the content space is expanding by cognitive and sensory contents, the billions of devices (things) are to be interconnected (Figure 1), and an open Digital Ecosystem is being formed [1, 24, 40, 56, 58].
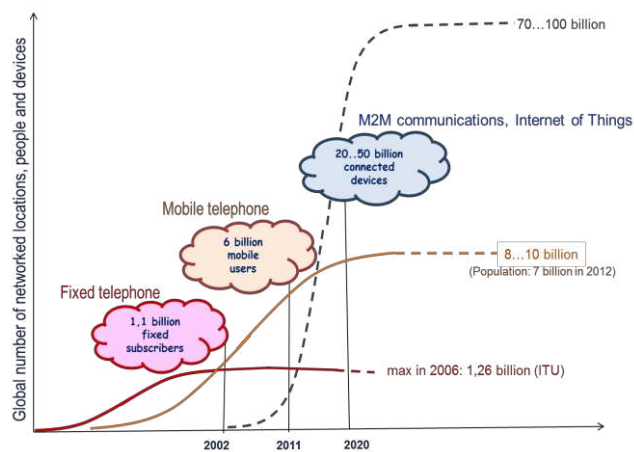
Figure 1

Evolution phases of networking: number of networked locations, people and devices

A horizontally and vertically integrated information (TIM) value chain is shaped (Figure 2), comprising all the layers and services of the intelligent infrastructure of the networked knowledge society, sometimes referred to as *Internet infrastructure*. At the same time the identification capacity of the Current Internet is running out; Internet architectures are reconsidered for better managing mobility functions, quality requirements and security issues, as well as for exploiting the opportunities derived from the technological advancements and the recent data handling and cognitive concepts. This integration process transforms business and bank spheres, administration, production, agriculture, transport, health, education and knowledge systems, etc., our everyday life. The future networked knowledge society is going to be established on Internet base, but the limitations of the Current Internet must be eliminated.

Recognizing the challenges of the Current Internet and the opportunities and requirements for a more advanced Internet, European Commission supported intensively the research activity on the Future Internet (FI) and initiated the

organisation of the Future Internet Assemblies (FIAs). Since 2008 ten FIAs were held and four books were issued on the research results [1, 2, 3, 4]. This paper collects the challenges and opportunities of the Current Internet, the visionary and recent capabilities of the Future Internet, defines the research goals and spheres, and identifies and clusters the relevant research topics into nine research chapters: from basic research on Internet Science through Internet Engineering up to Future Internet applications and experiments.
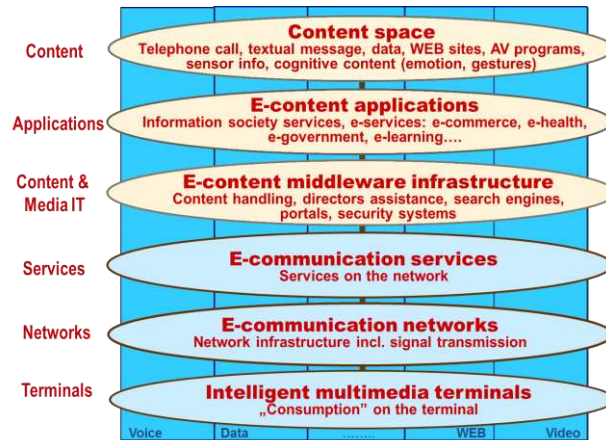

Figure 2

Integrated information value chain:

the layers of intelligent infrastructure of networked knowledge society

2 Challenges of the Current Internet

The Internet concept was born with different conditions and goals by "the fathers of the Internet", Vinton G. Cerf and Robert E. Kahn [12] in the seventies. Since then the Internet has become a large global network and we are currently speaking on the Internet age. The original concept of the Internet (TCP/IPv4) has limitations, and there are new societal requirements (mobility, security), while technological developments provide new opportunities and solutions for the challenges.

Limitations of the Current Internet are [1, 2, 3, 4, 44, 60, 62]:

- The limited identification capacity, the lack of IPv4 address space;

- The essentially private wireline network concept, the lack of an inherent mobile centric network architecture and a scalable efficient network and mobility management;

- The lack of guaranteed and differentiable quality of services and security (the best effort solutions);
- Energy consumption: the emission of CO_2 is increasing due to the increase in network size and usage;
- Application development is slow, inflexible, etc.

New technological opportunities are for managing limitations [1, 2, 3, 4, 38, 42, 44]:

- Advanced wireless/mobile technologies;
- Broadband optical solutions;
- Huge storage capacity, storage efficiency;
- Innovations in material and manufacturing technology, especially in the technology of sensors, CPUs, memories and energy sources;
- Potential opportunities deriving from nanotechnology and biotechnology.

Growth of the Internet's societal role gives rising demands as [2, 3, 4, 23, 33, 58]:

- Anywhere, anytime access (always on);
- Interconnection of devices, objects, sensors (networked 20…100 billion things);
- Expansion of content space with 3D and cognitive contents (gestures, emotion);
- Scalable and customized data and knowledge engineering;
- A lot of human centric, secure, smart applications.

We can conclude that:

- We are witnesses to the radical increase of Internet's size and complexity;
- There is a great technical potential and societal need for significant expansion of applications;
- The penetration of Internet has a fundamental impact on lifestyle and human relations.

3   Visions of the Future Internet

The challenges of the Current Internet, the tangible and potential demands and the technical opportunities determine the critical research issues, research objectives, and needs the reconsideration of the Classic Internet concept and the construction of the vision of the Future Internet.

Figure 3 shows the Future Internet vision based on the scheme of *Japan's National Institute of Information and Communications Technology* (NICT) [40], and the achievements of the FIA 2011 in Budapest and Poznan, FIA 2012 in Aalborg [2, 3, 57]. NICT prepared the first vision for New-Generation Networks (NWGN) in 2008 focusing on the technology requirements for solving social problems (energy shortage, medical care, crime prevention, technology gap, etc.) and creating new values for achieving a future knowledge society, using the NWGN. Five network targets were identified for an NWGN R&D strategy, whose scheme was used for the Future Internet and continuously evolved aiming at an intelligent, sustainable world and an innovative, secure society as a generic goal of the Future Internet [3, 51]. Figure 3 shows two pillars and six targets (objectives), and the desired applications are also indicated. The Classic Internet is aimed at the interconnection of persons and contents. The Future Internet is completed by the interconnection of all sorts of devices and things which involve sensors and actuators to sense or actuate respectively their physical environment. This idea is demonstrated by two pillars referred to as: *Internet of People* (Media Internet) and *Internet of Things* (IoT) [52, 56]. The Future Internet aims at a unified infrastructure of communication, computing and storage systems that involves the Internet of people, content, devices and things. The Future Internet should support:

- The drastic increase of the networked infrastructure characteristics (scale, diameter, bandwidth, service set, etc.);

- The expected wide range of networked services with diverse and evolving service requirements;

- The efficient operation and management of all communication, computing and storage resources;

- The concept of the "*Internet of Everything*", the convergence of the consumer, business and industrial Internets.
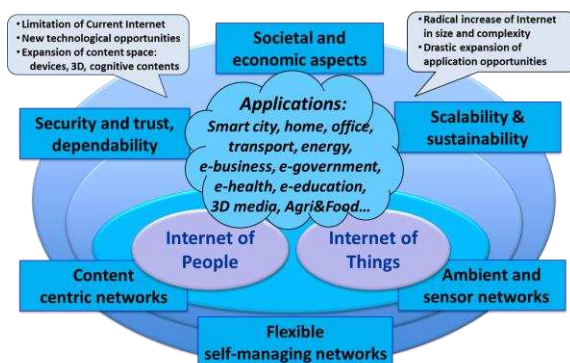

Figure 3
Future Internet vision, based on Japan NICT's NWGN vision, FIA2011 in Budapest and Poznan, and FIA2012 in Aalborg

Therefore Future Internet research activities primarily focus on network architecture issues to solve challenges in scalability, flexibility and manageability, with special attention to ambient (ubiquitous) and sensor networks (IoT), large amounts of various data (Big Data) generated by the IoT, as well as to content centric multimedia networks and networked 3D content. Internet security, dependability (reliability, accountability, verifiability) and trust as well as sustainability referring to energy awareness (green network) and efficient spectrum usage are also priority issues in the deployment of the Future Internet, and are closely related to the Future Internet vision [42, 57]. Drastic expansion of application opportunities (smart city, home, office, intelligent transport, energy, e-business, e-government, e-health, e-education, 3D media, etc.) assisted by the Future Internet Cloud concept [4, 45] and the societal-economic aspects and impacts (e.g. service economy, social interaction, augmented reality, Internet-style innovation) are also significant research targets [57].

A well-established Future Internet vision can be based on the standardization work of the Future Networks (FNs) performed by the *International Telecommunication Union Telecommunication Standardization Sector* (ITU-T). Driven by the research activities in network virtualization, cloud networking and others [3, 22, 28, 59, 60], the ITU-T has started the standardization of FNs as networking systems to be deployed in the 2015-2020 timeframe [4, 37]. FN standardization combined two complementary approaches: a top down method starting from objectives and design goals, and a bottom up method starting from relatively matured candidate technologies. Recommendation ITU-T Y.3001 presents the first standard description of FNs [29, 37]. The Future Network is described, like the Future Internet, as an infrastructure that connects and orchestrates the Internet of people, content and systems (i.e. devices, computers, clouds and things). The Recommendation identifies *four essential objectives* that were not suitably taken into account in the design of current networks. They help to characterize the differences from the current networks. Design goals related to the objectives were also identified as advanced capabilities that are necessary for the realization of FNs. The four objectives described in Y.3001 can be considered the elements of the vision of the Future Internet (Figure 4). They are the followings:
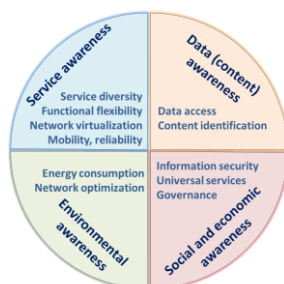

Figure 4

Future Internet vision, based on the standardization work of the Future Networks performed by ITU-T

• *Service awareness:* Future network architectures are expected to support not only current services such as e-mail and web browsing, but also emerging services, including mission critical services, by providing additional functionality without drastic increases in communication, computing and storage resources, deployment and operational costs. Thus, these network resources are to be virtualised to allow for their flexible usage by the services (resources awareness, network virtualization concept) [30]. Network virtualization means the process of partitioning the network resources, abstracting the partitions as virtual network resources, and combining them and network functionalities into logically isolated virtual networks. A unified network resources management is also involved. Such a way, multiple virtual networks can be created in a single physical network. Furthermore the network architecture should support advanced mobility features, enhanced reliability and security requirements.

• *Data (content) awareness:* Communication in current networks is based on the globally unique location identity (ID) and location based routing. If identical contents (data, information) are placed in multiple data locations, they may have the same content ID, and the content can be accessed via a nearest location using content ID based routing. In such a way the handling of a vast amount of data can be optimized [32].

• *Environmental awareness:* The enormous increase in Internet traffic means increase in energy consumption; hence energy awareness is a key objective. To save energy we should optimize the network to reduce the network capacity and traffic loads, as well as improve the energy efficiency using lower power electronic technology and dynamic control techniques [31].

• *Socio-economic awareness:* The Internet becomes an essential infrastructure utility; the right to have access to a global network will be one of the fundamental rights in the future. Hence the Future Internet should consider socio-economic objectives as governance issues, e.g. the barrier to enter the market, the lifecycle cost for sustainability and deployment, service universalization, information security, and personal data protection [29].

Figure 5 suggests *a vision for the Future Internet*, combining the research oriented NWGN vision scheme [40, 51] and the standardization oriented FN vision of the ITU-T [29, 37], taking also into account the achievements of the FIA 2013 in Dublin [4]. The Future Internet vision aims at a unified infrastructure of communication, computing and storage resources by integrating the Internet of people and content, devices and things, computers and clouds. Figure 5 demonstrates three concepts (Internet of People, Internet of Things and their unification and extension) as pillars, indicates the smart, high-value applications provided as services in the Cloud, and shows five strategic objectives. The objectives are based on the ones of the NWGN and FN visions, as follows:

- *Scalable, flexible, service aware network* means a scalable network architecture with functional flexibility, which can accommodate wide range of services with diverse and evolving requirements;

- *Virtual, resource aware network* means the virtualization of combined communication, computing and storage resources and a unified efficient network resources management system;

- *Data and content awareness* embraces goals on content-centric networking, content identification (ID) and efficient handling and usage of the generated Big Data, involving 3D and cognitive contents;

- *Sustainability, environmental awareness* refers to energy awareness, efficient spectrum usage and any other ecological aspects;

- *Intelligent, innovative and secure society,* as a generic target of Future Internet, comprises the societal-economic objectives and aspects.
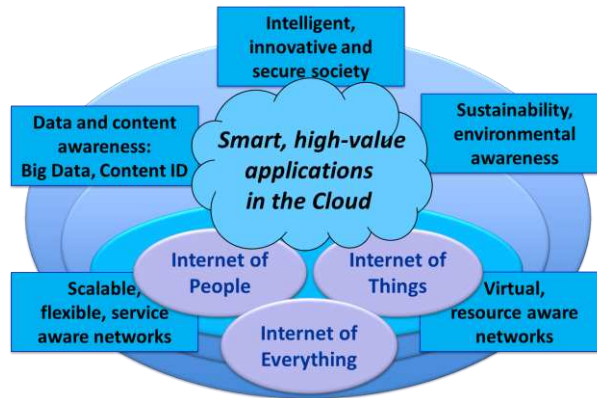

Figure 5
Future Internet vision, combining NWGN and FN visions

4   Recent Future Internet Capabilities

Recently, there has been no accepted definition for Future Internet, rather it is described by some capabilities, which do not exist in, or are not typical of the Current Internet. Some new functions created and certain features became obvious in the last some years that are relevant criteria in separating the Current and Future Internet. In the following, eight functions and four features associated with the Future Internet are listed. The lists as a matter of course are open; recently, one or more functions and some of the features became characteristics of the Future Internet solutions.

A recent list of relevant Future Internet functions is [2, 3, 4, 19, 24, 33, 40, 45, 47, 55, 56]:

1. Identification and interconnection of things, devices, sensors, actuators (Internet of Things);

2. Network architecture intrinsically handling service diversity, mobility, „anywhere, anytime" data collection, tracking and tracing (Service and mobility centric architectures);

3. Networked data bases: real time access and global organization of huge amount of data, information and multimedia contents (Big Data concept);

4. Content-aware technologies: content selection, distribution, outsourcing and management, content centric networks, content mining;

5. Programmability of networks: virtualized, software-defined networks (SDN);

6. Communicating and managing 3D and cognitive contents, virtual and augmented world;

7. Cloud computing and communications: infrastructure, platform, software, network and communication provided as a service (IaaS, PaaS, SaaS; NaaS, CaaS);

8. Remote collaboration, monitoring and control of physical processes (Tactile Internet).

A list of relevant Future Internet features is [2, 3, 4, 37, 44, 57]:

1. Inherent information security, personal data protection;

2. Managed quality, service-aware architectures and application platforms;

3. Managing the energy needed to transmit bits in the device, equipment and system levels;

4. Personalization: customized solutions and presentation profiles.

Recent capabilities are in good coincidence with the visionary ones, which shows the reality of the vision and the awareness of the research governance.

5 Clustering Research Themes

The research goals and spheres can usually be combined into three levels. The Internet is considered a complex networked infrastructure: its common attributes are heavily researched in the frame of Network Science. In general, basic research topics related to the fundamentals of Future Internet, embracing mathematical modelling of large scale networks, cryptography as the theory of security, human and socio-economic characteristics and environmental aspects, legislation and

governance principles, etc. are collected in *Internet Science* [23]. Engineering research issues – such as the creation and elaboration of Future Internet technologies, network architectures and protocols, data and content management methods and design procedures – represent the backbone of the Internet applied research, called *Internet Engineering* [16]. Finally Future Internet experimental research and innovation actions aim at the development of FI-based solutions, smart industrial and community applications, customizable content services, involving their experimentation, demonstration and standardisation issues, referred to as *Internet application development* [2, 3, 4, 24].

Studying the research themes on the Future Internet in the literature, in particularly on FIAs [1, 2, 3, 4, 53], Working Programme of Horizon 2020 [24] as well as the Hungarian Future Internet research activity, including 132 themes [7, 16, 51], the relevant research topics can be identified and arranged into main research areas. We define the main research areas as the chapters of Future Internet research activity in a layered model from basic research on Internet Science through Internet Engineering up to Future Internet applications and experiments. Figure 6 shows the nine layered chapters and their main research goals and spheres. Basic research in Internet Science constitutes Chapter 1, the Internet Engineering applied research is divided into five chapters (Chapter 2….6), the applications and experiments are comprised of three chapters (Chapter 7, 8 and 9). Figure 6 also indicates the possible relations of the chapters to other research goals and spheres. In the following, the research chapters and their relevant topics (topic ranges as subchapters) are presented.
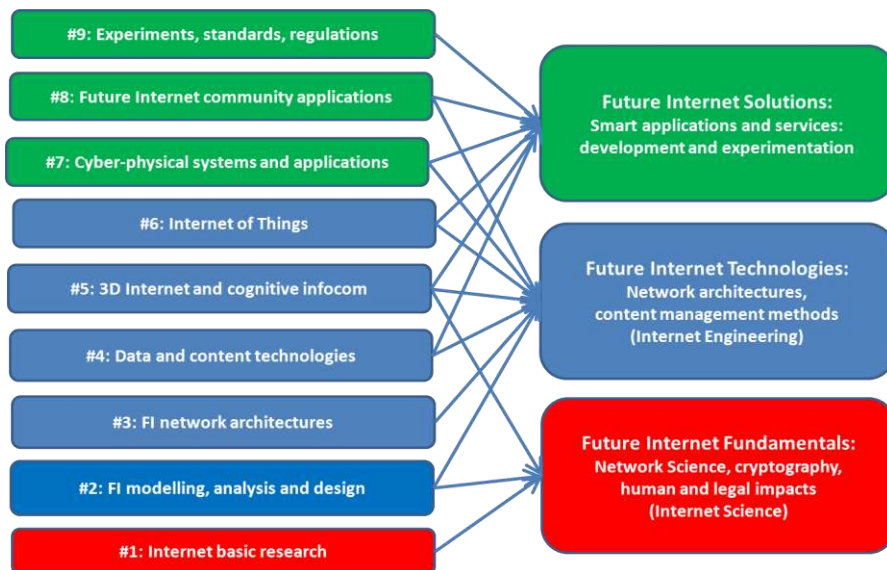

Figure 6

Future Internet research chapters and their relation to research goals

Chapter 1: Internet basic research (Internet Science).

Internet Science aims at an integrated and interdisciplinary scientific understanding of Internet networks and their co-evolution with society, embracing all disciplines that study the Internet from any technological or humanistic perspective. Relevant research topics as subchapters are [6, 8, 18, 23, 25, 26, 35, 36, 58]:

A) Network science: modelling and investigating large scale networks;

B) Computer science, from computational theory to computer architecture;

C) Basic enabling technologies, as quantum and nanotechnologies, etc.;

D) Cryptography, cyber security;

E) Human aspects: social and behavioural attributes, trust and aversion, cognitive processes, cognitive biases, social networks;

F) Network economics, game theory;

G) Legislation and governance (privacy, data protection, net neutrality…).

Chapter 2: Future Internet modelling, analysis and design.

These research topics target the modelling of the Future Internet enabling infocommunication systems, network concepts and technologies, for analysing their performance, scalability, stability, availability, resilience, quality of service, etc. as well as the elaboration of novel modelling and design paradigms. Relevant topics are as [10, 23, 20, 47, 55]:

A) Network modelling and performance analysis;

B) Queuing systems: traffic analysis and design;

C) Communication systems: advanced modulation, coding, access, spectrum usage;

D) Resource allocation and optimization methods;

E) Investigating networked media services: video streaming, VoIP, IPTV;

F) Survivability techniques, monitoring, failure exploration.

Chapter 3: Future Internet network architectures.

Novel architectural solutions are needed to meet the societal challenges, which the Current Internet may not be able to support sufficiently. This central research area focuses on new network architectures and protocols, virtualization of resources, mechanism enhancements, as [1, 2, 3, 4, 22, 24, 28, 37, 38, 42, 44, 45, 54, 59, 60]:

A) Future Internet network requirements (scalability, manageability, flexibility, reliability, resilience, robustness, security, latency, simplicity, etc.) and concepts (network virtualization, etc.)

B) Future Internet routing methods, e.g. transport protocols, path selections;

C) Mobile networking technologies: mobility handling, fixed-mobile convergence (FMC), network concepts, 5G network infrastructures;

D) Future media networks, content-centric networks (CCN), content delivery networks (CDN);

E) Network computing: ubiquitous, grid, cloud computing;

F) Virtual networks, software defined networks (SDN), network function virtualization (NFV);

G) Cloud infocommunications: NaaS, PaaS, SaaS, CaaS, etc.

Chapter 4: Data and content technologies.

The huge volumes of data and multimedia content are essential driving forces for the evolution of the Future Internet. These research topics aim at extracting meaning from data and supporting the collection, searching, managing and publishing of information in any form, as [2, 3, 11, 24, 47, 53]:

A) Data, text and media mining;

B) Big Data challenges (volume, velocity, variety) and solutions;

C) Semantic multimedia search methods, knowledge discovery;

D) Presentation, visualisation;

E) Digital library functions (archives, name spaces, etc.).

Chapter 5: 3D Internet and cognitive infocommunications.

The three dimensional (3D) communications can embrace our cognitive systems, including not only hearing and vision, but touch, gestures, emotion, smell, etc. Along with its enormous potential, 3D Internet opens many research challenges in order to become a reality, as [5, 9, 13, 14, 15, 16, 17, 19, 27, 46]:

A) 3D Internet architecture and content technology (3D media analytics, processing, presentation);

B) Multimodal human-computer interactions;

C) Cognitive infocommunications, cognitive contents and channels;

D) Virtual collaboration, 3D Internet based control and communications;

E) Devices and procedures of 3D and 4D content creation and presentation.

Chapter 6: Internet of Things (IoT).

Internet of Things is seen as a key part of the Future Internet vision, which targets real-time complex interactions and collaborations of billions of heterogeneous devices (objects, sensors, actuators), and requires tackling numerous

technological, connectivity and interoperability issues [1, 2, 3, 4, 33, 40, 62, 52, 56]:

A) IoT enabling technologies (RFID, NFC, etc.), sensors and actuators, energy and spectrum awareness;

B) Identification: naming, addressing, privacy awareness, IPv6-based IoT;

C) Communication architectures for constrained devices, IoT infrastructures, self-aware and organizing networks;

D) Data management, software solutions, security technologies: self-adaptive security mechanism and protocols, self-managed secure IoT;

E) Services support platforms, heterogeneous data collection and processing infrastructures.

Chapter 7: Cyber-physical systems and applications.

Cyber-physical systems refer to next generation embedded ICT systems that are interconnected and collaborate through the Internet of Things, and provide a wide range of innovative applications and services. Relevant research directions are [4, 25, 34, 41, 43, 61, 52, 56]:

A) Embedded and intelligent engineering systems;

B) Intelligent production applications, measuring and controlling physical processes;

C) Intelligent transport, cars and logistics;

D) Smart agriculture-food applications,

E) Smart energy systems, green ICT solutions.

Chapter 8: Future Internet community applications.

Users demand "always on" access to cheap, easy-to-use, secure, mobile, personalized and context-aware applications, which are to be realized over highly interconnected, increasingly complex infrastructures. The Internet of Things is implicated for smart environments and smart spaces. Several cross-disciplinary research challenges should be addressed [1, 2, 3, 4, 39, 52, 56, 61]:

A) Mobile crowd-sensing platform and functions;

B) Smart home and office applications;

C) Smart health and well-being applications;

D) Smart business applications;

E) Smart governance applications;

F) Smart city applications;

G) Other intelligent and cognitive community applications.

Chapter 9: Experimentation, standardization, regulation.

Themes of this practical chapter embrace the requirements and design of comprehensive test facilities, the federation of test-beds, the technical and social experiments, the standardization activity as well as the emerging complex regulatory issues. The subchapters are defined as [2, 3, 4, 29, 30, 31, 32, 36, 37, 48, 56]:

A) Experimental systems, test-beds;

B) Experimental methods, demonstrations, field results;

C) Socio-economic studies, business models;

D) Technical standards, recommendations, standardization issues, e.g.: identification, communications, virtualization, interoperability, security;

E) Technical, economic and content regulatory issues.

Conclusions

This paper collects the challenges of the Current Internet, draws up Future Internet concepts and visions, and identifies the relevant functions and features of the Future Internet as of 2013. The Future Internet visions are based on a concept of unified infrastructure of communication, computing and storage resources by involving the Internet of people and content, devices and things, computers and clouds, and aim at objectives for the service, resource and content awareness, and the societal, economic and environmental awareness. The paper proposes a combined vision scheme with five objectives. The paper also defines the associated research goals and spheres, and clusters the research themes into nine research chapters and 52 subchapters. The identified research chapters are layered from basic research on Internet Science (Chapter 1), through applied research on Future Internet architectures and content engineering (Chapters 2-6) to Future Internet applications and experiments (Chapters 7-9). Within each chapter, the relevant research topics are organized into 5-7 subchapters.

References

[1]   Towards the Future Internet - Emerging Trends from European Research, Future Internet Assembly 2010, Valencia, 15-16 April 2010, Edited by Tselentis, G. et al. ISBN 978-1-60750-538-9/539-6, 2010, IOS Press, Amsterdam

[2]   The Future Internet - Future Internet Assembly 2011: Achievements and Technological Promises, Budapest, 17-19 May 2011, Edited by Dominigue, J. et al. LNCS 6656, ISBN 978-3-642-20898-0, 2011, Springer, Heidelberg

[3]   The Future Internet – Future Internet Assembly 2012: From Promises to Reality, Aalborg, 9-11 May 2012, Edited by Alvarez, F. et al. LNCS 7281, ISBN 978-3-642-30240-4, 2012, Springer, Heidelberg

[4]   The Future Internet - Future Internet Assembly 2013: Validated Results and New Horizons, Dublin, 8-10 May 2013, Edited by Galis, A. and Gavras, A. LNCS 7858, ISBN 978-3-642-38081-5, 2013, Springer, Heidelberg

[5]   Alpcan, T., Bauckhage, C. and Kotsovinos, E.: Towards 3D Internet: Why, What, and How? In International Conference on Cyberworlds, CW'07, pp. 95-99, 2007

[6]   Alpcan, T., Buttyán, L. and J. S. Baras (eds.): Decision and Game Theory for Security, Lecture Notes in Computer Science, No. 6442, Springer 2010

[7]   Bakonyi, P., Sallai, Gy.: Future Internet National Research Program – JINKA2.1, in Hungarian, p. 55, Budapest, February 2014

[8]   Barabási, A-L., Newman, M. and Watts, D. J.: The Structure and Dynamics of Networks, Princeton Studies in Complexity. ISBN 0-691-11357-2, Princeton University Press, 2006

[9]   Baranyi P., Csapó A.: Definition and Synergies of Cognitive Infocommunications. Acta Polytechnica Hungarica, ISSN 1785-8860, Vol. 9, No. 1, 2012, pp. 67-83

[10]   Bíró, J.: Novel Equivalent Capacity Approximation through Asymptotic Loss Analysis. Computer Communications. Special issue for Heterogeneous Networks: Performance Analysis and Traffic Engineering. 33:(1) pp. S152-S156. (2010)

[11]   Boiko, B.: Content Management Bible. Wiley. p. 1176. Nov. 2004. ISBN 0-7645-7371-3

[12]   Cerf, V. G.: The day the Internet age began. Nature 461 (7268): 1202–1203. 2009. doi:10.1038/4611202a. PMID 19865146

[13]   CogInfoCom 2010 (1st International Conf. on Cognitive Infocommunications), 29 Nov. - 1 Dec. 2010, Tokyo, Japan

[14]   CogInfoCom 2011 (2nd International Conf. on Cognitive Infocommunications), 7-9 July 2011, Budapest, Hungary, Print-ISBN: 978-1-4577-1806-9

[15]   CogInfoCom 2012 (3rd IEEEE International Conf. on Cognitive Infocommunications), 2-5 Dec. 2012, Kosice, Slovakia

[16]   CogInfoCom 2013 (4th IEEE International Conference on Cognitive Infocommunications), Dec 2-6, 2013, Budapest, ISBN 978-1-4799-1-1543-9, In CogInfoCom2013: Workshop on Future Internet Science and Engineering

[17]   Csapó A., Baranyi P.: An Interaction-Based Model for Auditory Substitution of Tactile Percepts. In: 14th IEEE Int. Conference on Intelligent Engineering Systems (INES 2010) Paper 5483833. pp. 271-276. Las Palmas, Spain, 5-7- May 2010

[18]   Dányádi, Z., Földesi, P. and Kóczy, L. T.: Fuzzy Search Space for Correction of Cognitive Biases in Constructing Mathematical Models. In: 3rd IEEE International Conference on Cognitive Infocommunications, Kosice, Slovakia, 2012, pp. 585-589

[19]   Daras P. and Alvarez, F.: A Future Perspective on the 3D Media Internet. In: Towards the Future Internet – A European Research Perspective, Edited by Tselentis, G., et al. pp. 303-312, IOS Press, 2009. ISBN 978-1-60750-007-0

[20]   Do, V. T., Chakka, R., Sztrik, J.: Spectral Expansion Solution Methodology for QBD-M Processes and Applications in Future Internet Engineering. In: Nguyen N. T., Do V. T., Hoai A. T.: Advanced Computational Methods for Knowledge Engineering. Genova, Springer-Verlag, 2013, pp. 131-142

[21]   European Commission: Green Paper on the Convergence of the Telecommunications, Media and Information Technology Sectors, and Implications for Regulation. Towards an Information Society Approach. 3 Dec. 1997, COM (1997) 623

[22]   European Commission: Future Media Networks - Research Challenges 2010. Future Media Networks cluster of Networked Media Systems FP7 projects. 2010. ISBN 978-92-79-17393-6 doi:10.2759/37178

[23]   European Commission: International Conference on Internet Science. The FP7 European Network of Excellence in Internet Science (http://internet-science.eu) Brussels, April 9-11, 2013 http://internetscienceconference.files.wordpress.com/2013/04/internet_scie nce_conference_proceedings.pdf

[24]   European Commission: HORIZON 2020 - The Framework Programme for Research and Innovation. Brussels, 2011. Work Programme (2014-2020) 5.i. Leadership in enabling and industrial technologies: Information and Communication Technologies. Annex 6 to Decision. December 2013. p. 107 http://ec.europa.eu/research/horizon2020/pdf/work-programmes/ information_and_communication_technologies_draft_work_ programme.pdf

[25]   Fischer, A., Beck, M. T., and de Meer, H.: An Approach to Energy-efficient Virtual Network Embeddings. In 5th International Workshop on Management of the Future Internet (ManFI 2013) 2013. Google ScholarBibTex Fischer2013benergyefficient.pdf

[26]   Földesi, P. and Botzheim, J.: Computational method for corrective mechanism of cognitive decision-making biases. In: 3rd IEEE International Conference on Cognitive Infocommunications, Kosice, Slovakia, 2012, pp. 211-215

[27]   Galambos, P., Weidig, C., Baranyi, P., Aurich, J. C., Hammann, B., and Kreylos, O.: VirCA NET: A Case Study for Collaboration in Shared Virtual Space, in 3rd IEEE International Conference on Cognitive Infocommunications, Kosice, Slovakia, 2012, No. 42, pp. 273-277

[28]   Galis, A., Denazis, S., at al (eds.): Programmable Networks for IP Service Deployment, p. 450. Artech House Books. June 2004. ISBN 1-58053-745-6

[29]   ITU-T Recommendation Y.3001: Future Network Vision – Objectives and Design Goals, 2011

[30]   ITU-T Recommendation Y.3011: Framework of Network Virtualization for Future Networks, 2012

[31]   ITU-T Recommendation Y.3021: Framework of Energy Saving for Future Networks, 2012

[32]   ITU-T Recommendation Y.3031: Identification Framework in Future Networks, 2012

[33]   Karnouskos, S., Skarmeta, A. F., et al: The Future Internet of Things. Introduction to Chapter on Internet of Things. In: The Future Internet - Future Internet Assembly 2013: Validated Results and New Horizons. pp. xxv-xxvii. Springer, Heidelberg, 2013

[34]   Lee, E. A. and Seshia, S. A.: Introduction to Embedded Systems - A Cyber-Physical Systems Approach, http://LeeSeshia.org, 2011

[35]   Lewis, Ted G.: Network Science: Theory and Applications, Wiley, March 11, 2009. ISBN 0-470-33188-7

[36]   Liu, Yu-li (ed.): Convergence in the Digital Age. Special issue in Telecommunications Policy, pp. 611-685, Editorial pp. 611-614, ISSN 0308-5961, Vol. 37. No. 8, Sept. 2013, Elsevier, Amsterdam

[37]   Matsubara, D., Egawa, T. et al: Open the Way to Future Networks – A Viewpoint Framework from ITU-T. In: The Future Internet - Future Internet Assembly 2013: Validated Results and New Horizons. pp. 27-38, 2013, Springer, Heidelberg

[38]   Meer, H. de, Hummel, K. A. and Basmadjian, R. (eds.): Future Internet services and architectures: trends and visions. Special Issue in Telecommunication Systems, Vol. 51, No. 4, Dec. 2012, pp. 219-303

[39]   Minutolo, A., Esposito, M., De Pietro, G.: Development and Customization of Individualized Mobile Healthcare Applications, in 3rd IEEE Internat. Conf. on Cognitive Infocommunications, Kosice, Slovakia, 2012, No. 49, pp. 321-326

[40]   Nishinaga, N.: NICT New-Generation Network Vision and Five Network Targets. IEICE Trans. on Communications, Vol. E93-B, No. 3, pp. 446-449, March 2010, Online ISSN: 1745-1345, Print ISSN: 0916-8516

[41] Nguyen, K. K., Cheriet, M., Lemay, M., et al: Renewable Energy Provisioning for ICT Services in a Future Internet: In: The Future Internet - Future Internet Assembly 2011: Achievements and Technological Promises, pp. 419-429, Springer, Heidelberg, 2011

[42] Pallot, M., et al: A Tentative Design of a Future Internet Networking Domain Landscape. In: The Future Internet – Future Internet Assembly 2012: From Promises to Reality, pp. 237-249, 2012, Springer, Heidelberg

[43] Papadimitratos, P., La Fortelle, A., Evenssen, K., Brignolo, R. and Cosenza, S.: Vehicular Communication Systems: Enabling Technologies, Applications, and Future Outlook on Intelligent Transportation, IEEE Communications Magazine, Vol. 47, No. 11, pp. 84-95, 2009

[44] Papadimitriou, D., Zahariadis, T., et al: Design Principles for the Future Internet Architecture. In: The Future Internet – Future Internet Assembly 2012: From Promises to Reality, pp. 55-67, 2012, Springer, Heidelberg

[45] Petcu, D., Galis, A. et al: The Future Internet Cloud: Computing, Networking and Mobility. Introduction to Chapter on Computing and Networking Clouds. In: The Future Internet - Future Internet Assembly 2013: Validated Results and New Horizons. pp. xiii-xv. 2013, Springer, Heidelberg

[46] Prekopcsák, Z., Halácsy, P., Gáspár-Papanek, Cs.: Design and Development of an Everyday Hand Gesture Interface. In: 10th Internat. Conf. on Human Computer Interaction with Mobile Devices and Services (Mobile HCI '08). Amsterdam, 2-5. Sept. 2008, ISBN: 978-1-59593-952-4 ACM, pp. 479-480, Paper PS479. New York

[47] Prekopcsák, Z., Makrai, G., Henk, T., Gáspár-Papanek, Cs.: Radoop: Analyzing Big Data with RapidMiner and Hadoop. In: RCOMM 2011: RapidMiner Community Meeting and Conference. Dublin, 8-9 June 2011, pp. 1-12

[48] Sales, B., Darmois, E. et al: A Systematic Approach for Closing the Research to Standardization Gap. In: The Future Internet – Future Internet Assembly 2012: From Promises to Reality, pp. 18-29, Springer, Heidelberg, 2012

[49] Sallai, Gy.: Defining Infocommunications and Related Terms. Acta Polytechnica Hungarica, Vol. 9, No. 6, 2012, pp. 5-15

[50] Sallai, Gy.: From Telecommunications to Cognitive InfoCommunications and Internet of Things - Phases of Digital Convergence. In: 17th IEEE Int. Conference on Intelligent Engineering Systems (INES 2013) ISBN 978-1-4799-0830-1, Paper 2. pp. 13-17, Costa Rica, 19-21 June 2013

[51] Sallai, Gy.: The FIRST Project and the Future Internet National Research Programme. Presentation. Future Internet PPP Workshop on Building an

Eco-System for Delivering Innovative Future Internet Services and Applications. IEEE ICC 2013, Budapest, 13 June 2013
http://www.nih.gov.hu/nemzetkozi-tevekenyseg/jovo-internet/building-an-eco-system

[52]   Smith, Ian G. (ed): The Internet of Things 2012. New Horizons. IERC - Internet of Things European Research Cluster, 3rd edition of the Cluster Book. p. 360, Halifax, UK, 2012, ISBN: 978-0-9553707-9-3
http://www.internet-of-things-research.eu/pdf/IERC_Cluster_Book_2012_WEB.pdf

[53]   Szűcs, G.: Decision Trees and Random Forest for Privacy-Preserving Data Mining. In: Tarnay, K., Imre, S., Lai Xu (ed.): Research and Development in E-Business through Service-oriented Solutions. IGI Global, Hershey, PA, USA, 2013, pp. 71-90, ISBN 978-1-4666-4181-5

[54]   Stratogiannis, D., Tsiropoulos, G. et al (eds.): Mobile Computing and Networking Technologies. Special Issue in Telecommunication Systems, Vol. 52, No. 4, April 2013, pp. 1714-2145

[55]   Tapolcai, J., Gulyás, A., Heszberger, Z., Biró, J., Babarczi, P., Trossen, D.: Stateless Multi-Stage Dissemination of Information: Source Routing Revisited. In: IEEE Globecom. Anaheim, USA, 3-7. Dec. 2012, pp. 2797-2802

[56]   Vermesan, O. and Friees, P. (eds.): Internet of Things – Converging Technologies for Smart Environments and Integrated Ecosystems p. 363, River Publishers, Aalborg, 2013, ISBN: 978-87-92982-96-4 (E-Book)
http://www.internet-of-things-research.eu/pdf/Converging_Technologies_for_Smart_Environments_and_Integrated_Ecosystems_IERC_Book_Open_Access_2013.pdf

[57]   Wainwright, N., Papanikolaou, N.: The FIA Research Roadmap, Priorities for Future Internet Research. Introduction in: The Future Internet – Future Internet Assembly 2012: From Promises to Reality, pp. 1-5, 2012, Springer, Heidelberg

[58]   World Economic Forum: Digital Ecosystem - Convergence between IT, Telecoms, Media and Entertainment: Scenarios to 2015, World Scenario Series, 2007
http://www3.weforum.org/docs/WEF_DigitalEcosystem_Scenario2015_ExecutiveSummary_2010.pdf

[59]   Zahariadis, T. et al: Towards a Content-Centric Internet. In: Towards the Future Internet - Emerging Trends from European Research, Edited by Tselentis, G. et al. pp. 227-236, 2010, IOS Press, Amsterdam

[60]   Zahariadis, T. et al: Towards a Future Internet Architecture. In: The Future Internet - Future Internet Assembly 2011: Achievements and Technological

Promises, Edited by Dominigue, J. et al. pp. 7-18, 2011, Springer, Heidelberg

[61]   Zeller, D., Olsson, M. et al: Sustainable Wireless Broadband Access to the Future Internet – The EARTH Project. In: The Future Internet - Future Internet Assembly 2013: Validated Results and New Horizons. pp. 249-271, 2013, Springer, Heidelberg

[62]   Ziegler, S., Crettaz, C. et al: IoT6 – Moving to IPv6-based Future IoT. In: The Future Internet - Future Internet Assembly 2013: Validated Results and New Horizons. pp. 161-172, 2013, Springer, Heidelberg

Extension of Liskov Substitution Principle and Application to Curriculum Management

Raul Sorin Fântână[1], Nicuşor Minculete[2], Radu-Emil Precup[3]

[1]Department of International Business, "Dimitrie Cantemir" Christian University, Str. Bisericii Romane 107, RO-500068 Brasov, Romania
E-mail: fantana@universitatea-cantemir.ro

[2]Department of Mathematics and Informatics, Transilvania University of Brasov, Str. Iuliu Maniu 50, RO-500091 Brasov, Romania
E-mail: minculete.nicusor@unitbv.ro

[3]Department of Automation and Applied Informatics, Politehnica University of Timisoara, Bd. V. Parvan 2, RO-300223 Timisoara, Romania
E-mail: radu.precup@upt.ro

Abstract: This paper offers an extension of the Liskov substitution principle by means of compatible systems, i.e., right or appropriate or functional systems. The compatible systems are represented by compatible sets. The compatible sets are first defined and the new approach to the mathematical modelling of compatible systems is given. The properties of compatible systems are next presented and applied to the curriculum management in higher education. The theoretical results are exemplified by a case study that discusses the curriculum management of two academic programs of study using the information densities of optional, fundamental and specialized courses. A discussion on the systems compatibility in several fields is included and the importance of our multidisciplinary approach to compatible resources management is highlighted.

Keywords: compatible sets; compatible systems; curriculum management; information density; Liskov substitution principle; mechatronics

1 Introduction

Following the unprecedented development of science and the emergence of interdisciplinary or border branches, specializations in one area can provide explanations and ways to solve the other. Highly contoured borders of some sciences and courses have disappeared. Engineering including mechatronics has multidisciplinary applications in medicine, efficient management forces engineers to have economic knowledge, economists have knowledge of engineering and all students must maximize the outcome of their knowledge.

Researchers and specializations of academic courses must provide each other knowledge and use specific compatible tools in the curriculum management. The multidisciplinary specialized intellectual resources must be used because of the benefits in terms of value of time and need for excellence.

Since the above aspects have not been proved mathematically, this paper provides the definition of compatible sets by the extension of the Liskov substitution principle (LSP) [1]. This is motivated by numerous examples from different fields that relate in similar terms to elements, sets, activities, which not identical, but compatible insofar as each has a density of identical information. With this regard the main purpose of this paper is to offer a modelling approach that characterizes the compatible systems or sets in the same environment and to apply our approach to academic curriculum management.

This paper uses the symbol ω to indicate the compatibility relation. In this context, the notation $x \, \omega \, y$ outlines that elements $x \in X$ and $y \in Y$ are compatible. As shown in [2], if A and B are independent systems, the problems of the system A are solved using the system B given that:

-    both systems contain compatible elements called mirror elements,

-    both systems operate in the same environment.

Under these conditions, two or more systems can be compatible. Therefore, the compatibility is defined in [2] as the capacity of two or more systems or components to carry out all tasks as long as they operate in the same environment.

The new idea of this paper stems from the characterization of systems and of their elements given in [2]: if in the system S_A represented by the set A all elements are considered known and the elements $x_{\omega yi}$ are unknown, and if in the system S_B represented by the set B all elements are considered known including the elements $y_{\omega xi}$, and if the elements $x_{\omega yi}$ are compatible with the elements $y_{\omega xi}$, then, replacing between them the compatible elements or the mirror elements, the system S_A will be considered resolved or functional. This means that the system S_A is compatible with the system S_B, and we use the notation $S_A \, \omega \, S_B$.

This paper extends this result by suggesting operations on compatible systems and their characterization by the information density of a system. Programs of study in higher education are targeted.

This paper is organized as follows: the background is presented in the next section. Section 3 is dedicated to the new approach to modelling of compatible systems. The properties of compatible systems are formulated and proved. Section 4 validates the theoretical approach by a case study focused on the curriculum management of two academic programs of study. The conclusions are pointed out in the final section.

2   Background

Many theories on the connection between the performance and the human resource management have been proposed recently [3-5]. Due to the market changes, the schools and universities have become increasingly managerially in their approach [6].

This paper offers a theory and some technical points of view on how to generally keep the freedom, the autonomy and the identity by offering, at the same time, the possibility to co-operate if the fields of activity are compatible. We focus on the academic field as a representative example due to its importance [7-10].

The concept of compatible is often used in the theory of systems of equations. Thus, a system of equations is called compatible if the system has at least one solution. The probability theory states that two events A and B are compatible if they take place in the same sample, namely, if they have at least one common favourable case.

The LSP is very close to compatibility. This principle states that [1] if A is a subtype of B, then objects of type B may be replaced with objects of type A (i.e., objects of type A may be substituted by objects of type B) without altering any of the desirable properties of the environment where A and B exist. As shown in [1], if these conditions are fulfilled in a computer program, the desirable properties of that program (correctness, tasks performed, etc.) are not altered. In addition, the subtypes should fulfil certain constraints.

This paper proposes the need to treat the compatible systems by an extension of the LSP on the basis of observing the common elements of this principle in various fields. The concept of compatible will be used in the sense of right or appropriate or functional. For example, software compatibility is in question when referring to compatible online games; this concerns the ability of a software program to run in a particular operating system. The new contribution of this paper is important because the literature analysis conducted as follows points out the absence of models to characterize all compatible systems. Moreover, our modelling approach can be uses in human resources management and, more general, in compatible resources management. These aspects are advantageous with respect to the state-of-the-art because we bridge the compatibility gap between systems in different fields by a systematic modelling approach.

The electromagnetic compatibility is discussed in [11, 12] from the general point of view of the ability of an equipment or system to perform without introducing intolerable disturbances to anything in the same environment; the applications can be related to sensitivity or robustness issues. The compatibility in divergent market systems is studied in [13]; it is proved that the compatibility can be slightly modified if the market is served by fully integrated system suppliers. The compatibility and substitutability of roles in multi-agent systems are analysed in

[14]; a formal specification of role-based interactions components along with their composition is suggested. The compatibility in mechanics refers to the conditions under which a displacement field can be guaranteed, and compatibility conditions are considered in [15] as particular cases of integrability conditions in the framework of the linear elasticity theory. The compatibility in medicine is related to medical and/or technical subsystems that should operate together in order to achieve several well defined goals with emphasis on robotics [16-21]. The use of innovation related to the compatibility in higher education is analysed in [22] with focus on team-based learning. Several problems related to the Bologna process are treated in [23, 24]. The need for experiments in control engineering and mechatronics education is pointed out in [8, 25-30].

In the context of the previous section, we consider that an element $x \in X$ is compatible with an element $y \in Y$ if, by replacing x with y, the new system $S_{A'}$ represented by the set $A' = (A \setminus \{x\}) \cup \{y\}$ is functional.

We consider two systems operating in the same environment, namely S_A represented by the set A and S_B represented by the set B, both of them having a finite number of elements that characterize each of the systems:

$$A = I \cup X_{\omega y} \cup X , \tag{1}$$

$$B = I \cup Y_{\omega x} \cup Y , \tag{2}$$

where the subsets in (1) and (2) are expressed as

$$I = \{id_j \mid j = 1...m\}, X_{\omega y} = \{x_{\omega y k} \mid k = 1...p\}, X = \{x_i \mid i = 1...n\},$$
$$Y_{\omega y} = \{y_{\omega x k} \mid k = 1...p\}, Y = \{y_j \mid j = 1...r\}, \tag{3}$$

and the subsets $X_{\omega y}$ and $Y_{\omega y}$ fulfil the following properties [2]:

-   They have the same number p of elements.

-   The elements $x_{\omega y k}$ and $y_{\omega x k}$ are compatible elements or mirror elements, these elements can take the place of, or substitute each other without being identical.

-   The identical element $id_j \in I$ is not compatible, and a compatible element is not necessarily an identical element.

-   The sets A and B are named compatible if the subset $C_{ms} = C_{ms}(A, B)$ consisting of identical elements, together with the compatible elements (mirror elements), delimits a set of minimum or sufficient characteristics.

-   If $x_i \in X \cap Y$, then x_i is called an identical element.

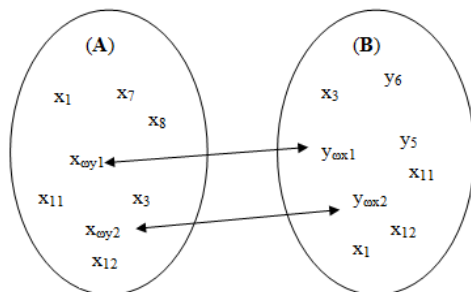An example to illustrate the structure of the two sets A and B is presented in Figure 1.

Figure 1
Structure of sets A and B

3 Mathematical Modelling Approach

The compatible sets A and B are represented by the subset C_{ms} of minimum or sufficient characteristics for which a system S characterised by C_{ms} is functional. The structure of the subset C_{ms} is presented in Figures 2 and 3.
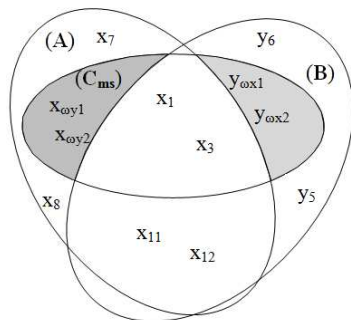

Figure 2
General structure of subset C_{ms}

The identical element $x_i \in X \cap Y$ is not necessary to the two systems S_A and S_B to be functional. This is highlighted in Figure 2 by the two elements x_{11} and x_{12}.
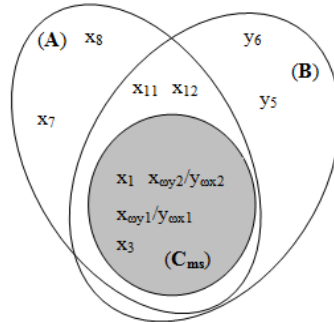
Figure 3

Subset C_{ms} represented as a subset of minimum or sufficient characteristics

The subset $C_{ms} = C_{ms}(A,B)$ is defined as follows [2]: if

$$X_{A/B} \subseteq X_{\omega y} \cup Y_{\omega x}, \ | X_{A/B} | = | X_{\omega y} | = | Y_{\omega x} |, \tag{4}$$

then

$$C_{ms} = C_{ms}(A,B) = I \cup X_{A/B} \cup (X \cap Y), \tag{5}$$

where $| X |$ is the cardinal of the set X.

The additional characteristics C_{ad}:

$$C_{ad} = (A \cup B) \setminus C_{ms} \tag{6}$$

do not define, from the point of view of compatibility, the sets A and B. This is pointed out by the elements x_7, x_8, y_5 and y_6 in Figure 3.

As shown in [2], there is a bijective function f that fulfils

$$f : B \setminus Y \to A \setminus X,$$
$$f(id_j) = id_j \text{ for } id_j \in I, \ f(y_{\omega x}) = x_{\omega y}, \ y_{\omega x} \in Y_{\omega x}. \tag{7}$$

This function and the inclusion of the identical elements lead to a different representation of Figure 2, shown in Figure 4. Figure 4 highlights both the identical elements id_j and the unnecessary elements x_{11} and x_{12}.
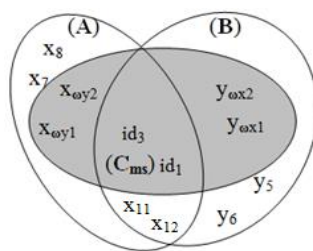
Figure 4
Structure of subset C_{ms} with identical and unnecessary elements included

The properties of the compatibility relation will be presented as follows. These properties are next applied in Section 4.

Proposition 1 (reflexivity). Let the system S_A be represented by the set A. Then $S_A \, \omega \, S_A$.

Proof. Since the system S_A is represented by the set A, there is a minimum set of characteristics $C_{ms}(A, A)$ for which the systems S_A and S_A are compatible, i.e.

$$C_{ms}(A, B) = I \cup X_{A/A} \cup X = A. \tag{8}$$

Proposition 2 (symmetry). If the relation $S_A \, \omega \, S_B$ holds, then $S_B \, \omega \, S_A$.

Proof. Since the system S_A is compatible with the system S_B, there is a minimum set of characteristics $C_{ms}(A, B)$ for which the systems S_A and S_B are compatible, i.e.

$$C_{ms}(A, B) = I \cup X_{A/B} \cup (X \cap Y) = I \cup X_{B/A} \cup (Y \cap X) = C_{ms}(B, A). \tag{9}$$

Equation (9) indicated that the system S_B is compatible with the system S_A.

Proposition 3 (transitivity). If the relations $S_A \, \omega \, S_B$ and $S_B \, \omega \, S_C$ hold, then $S_A \, \omega \, S_C$.

Proof. Since the system S_A is compatible with the system S_B, there is a minimum set of characteristics $C_{ms}(A, B)$ for which the systems S_A and S_B are compatible, and equation (5) is applied. Since the system S_B is compatible with the system S_C, there is a minimum set of characteristics $C_{ms}(B, C)$ for which the systems S_B and S_C are compatible, and the following equation similar to (5) is applied:

$$C_{ms}(B, C) = I \cup X_{B/C} \cup (Y \cap Z), \tag{10}$$

where the set Z is defined similarly to the sets X and Y in (1), (2) and (3), but this set corresponds to S_C.

Using the definitions of the systems S_A, S_B and S_C represented by the sets A, B and C, respectively, it follows that there exist a minimum set of characteristics $C_{ms}(A,C)$

$$C_{ms}(A,C) = I \cup X_{A/C} \cup (X \cap Z), \tag{11}$$

so the systems S_A and S_C are compatible.

These three propositions result in the conclusion that the compatibility relation ω between two systems is an equivalence relation on the set of systems. An equivalence class is given as

$$\hat{S}_A = \{S_B \mid S_A \, \omega \, S_B\}. \tag{12}$$

The antisymmetry property cannot be proved on the compatible sets because these sets also include compatible components that are not identical.

4   Case Study Concerning Curriculum Management

Two programs of study S_{P_1} and S_{P_2}, represented by the sets P_1 and P_2, are considered in order to exemplify relatively easily the two systems compatibility:

$$P_1 = C_o \cup C_f \cup C_s,$$
$$C_o = \{oc_i \mid i = 1...n\}, \; C_f = \{fc_j \mid j = 1...m\}, \; C_s = \{sc_k \mid k = 1...p\}, \tag{12}$$

$$P_2 = C_o^* \cup C_f \cup C_s^*,$$
$$C_o^* = \{oc_i^* \mid i = 1...n^*\}, \; C_f = \{fc_j \mid j = 1...m\}, \; C_s^* = \{sc_k^* \mid k = 1...p\}, \tag{13}$$

where oc_i and oc_i^* are the optional courses, fc_j are the fundamental / imposed courses, sc_k and sc_k^* are the compatible specialized courses, C_o and C_o^* are the sets of optional courses, C_f is the set of fundamental courses, C_s and C_s^* are the sets of compatible specialized courses.

Equations (12) and (13) point out that

$$|C_s| = |C_s^*|, \tag{14}$$

namely the two sets of compatible specialized courses have the same number of compatible elements.

We will evaluate the properties of the sets C_o, C_f, C_s, C_o^* and C_s^* in a case study related to two academic programs of study, an engineering program S_{P_1} and an economic program S_{P_2}.

C_o and C_o^* may also be empty sets, and they are insignificant sets not as a number but in the subject matter. Examples of insignificant sets or courses for these two programs of study are Sports and Languages.

$C_f \neq \varnothing$ for these programs of study. Examples of courses from this set are Statistics, Mathematics and Control Engineering.

Each element $fc_j \in C_f$ is characterized by the information density $\rho(fc_j)$, with

$$\rho(fc_j) > \rho_{acc}(fc), \ j = 1...m, \tag{15}$$

where $\rho_{acc}(fc)$ is the accepted threshold of density of fundamental courses. The parameter $\rho_{acc}(fc)$ is fixed.

$C_s \neq \varnothing$ and $C_s^* \neq \varnothing$ for these programs of study. Examples of courses from these sets are Management and Investments.

Each element $sc_k \in C_s$ is characterized by the information density $\rho(sc_k)$, with

$$\rho(sc_k) > \rho_{acc}(sc), \ k = 1...p, \tag{16}$$

where $\rho_{acc}(sc)$ is the accepted threshold of density of specialized courses. The parameter $\rho_{acc}(sc)$ is also fixed.

The densities can be quantified and measured. They are related as follows to the intersection and union of compatible sets with emphasis on the curriculum management.

If S_{P_1}, S_{P_2}, ..., S_{P_q} are compatible programs of study, the intersection of sets containing the compatible (mirror) elements is defined as the set $P_1 \cap_\omega P_2$

$$P_1 \cap_\omega P_2 = P_1 \cap P_2 \cup X_{P_1/P_2}, \tag{17}$$

where

$$X_{P_1/P_2} = \{sc_k \vee sc_k^* \mid k = 1...p\}. \tag{18}$$

The set X_{P_1/P_2} allows for the definition of a matrix of programs of study.

The academic curriculum results in three types of professions from the point of view of compatible resources management:

- Profession, which is given by the university diploma or licence or certificate.

- Border profession, built by supplementing a core profession with specific sufficient partial knowledge belonging to other professions. This type represents an effect of relevant rules or of special laws, e.g., mediator, assessor of property, industrial property attorney. The border profession is not a quality of the main profession.

- Specialization, which completes the curricular area by increasing the quality and / or the scope of profession. The specialization is not a profession itself. A specialization course without faculty cannot give equal rights with those conferred by the university diploma.

Therefore, we define the notion of admissible information density ρ_{adm} as the value under which complementary or mirror elements are not allowed. The admissible information density ρ_{adm} in higher education can be, for example, the sum of credits of the courses C_f, C_s and C_s^*. In fact, the compatible elements are some sets (i.e., they represent the information within the programs S_{P_1} and S_{P_2} in this case study) of a certain information density.

We define the representativity value associated to the program S_{P_1} as $V(P_1)$

$$V(P_1) = \sum_{j=1}^{m} fc_j + \sum_{k=1}^{p} sc_k , \qquad (19)$$

and the representativity value associated to the program S_{P_2} as $V(P_2)$

$$V(P_2) = \sum_{j=1}^{m} fc_j + \sum_{k=1}^{p} sc_k^* . \qquad (20)$$

Thus, for the programs S_{P_1} and S_{P_2} to be compatible, their representativity values $V(P_1)$ and $V(P_2)$ must fulfil the conditions

$$\begin{aligned} V(P_1) &> \rho_{adm}, \\ V(P_2) &> \rho_{adm}. \end{aligned} \qquad (21)$$

If S_{P_1}, S_{P_2}, ..., S_{P_q} are compatible programs of study, the union of sets containing the compatible (mirror) elements is defined as the set $P_1 \cup_\omega P_2$

$$P_1 \cup_\omega P_2 = C_o \cup C_o^* \cup C_f \cup X_{P_1/P_2}, \qquad (22)$$

The set $P_1 \cup_\omega P_2$ in this case study is the mathematical expression of the curriculum. The set $P_1 \cup_\omega P_2$ also represents all programs generated by P_1 and P_2 to be compatible:

$$S_{P_1} \cup_\omega S_{P_2} = \hat{S}_{P_1} = \hat{S}_{P_2}. \qquad (23)$$

The number of programs of study generated by P_1 and P_2 is $N(P_1, P_2)$

$$N(P_1, P_2) = 2^{n+n^*} \binom{2p}{p}. \qquad (24)$$

As a result, the programs generated by S_{P_1} and S_{P_2} offer an image on the compatible elements through the courses Management and Investments. If these courses respect in both programs an at least above ρ_{adm}, then these will contain complementary or mirror courses or items. Therefore, the exchange of experts or teachers will be not only acceptable but also useful to those national or international educational systems where the optimisation focusing on the cost minimisation is required. An objective function that can be used with this regard is

$$J = |C_o \cup C_f|. \qquad (25)$$

The optimisation problem must be correctly defined and associated with the correct definition of the variables of the objective function J and of the constraints. One such constraint concerns the set C_o that belongs to curricular areas and is given by laws.

The profit optimisation can be targeted as well. But this objective involves a different objective function.

Unfortunately, in education, the profit optimisation does not offer a competitive level of preparation, but just a sufficient one. Therefore national bodies involved in decision making in the educational systems should pay special attention to this subject.

The cost optimisation and the profit optimisation lead to minimum and maximum curricular areas. These areas are used in the assessment of the private higher education. However, the cost optimisation and the profit optimisation can be combined in terms of multi-objective optimisation problems. All these optimisation problems must be solved by appropriate algorithms [31-37].

Fuzzy logic can also be involved in our modelling approach [38-44]. The information density values $\rho(fc_j)$ and $\rho(sc_k)$ can be normalised as follows:

$$\rho_n(fc_j) = \rho(fc_j) / \sum_{\iota=1}^{m} \rho(fc_\iota), \ j = 1...m,$$

$$\rho_n(sc_k) = \rho(sc_k) / \sum_{\tau=1}^{p} \rho(sc_\tau), \ k = 1...p, \tag{26}$$

where the normalised information densities $\rho_n(fc_j)$ and $\rho_n(sc_k)$ fulfil the conditions

$$0 \le \rho_n(fc_j) \le 1, \ j = 1...m,$$

$$0 \le \rho_n(sc_k) \le 1, \ k = 1...p. \tag{27}$$

We define two fuzzy sets. The first fuzzy set is (C_f, μ_1), with the membership function μ_1

$$\mu_1 : C_f \rightarrow [0,1],$$

$$\mu_1(fc_j) = \begin{cases} \rho_n(fc_j), & \text{if } \rho_n(fc_j) > \rho_{acc}(fc), \\ 0, & \text{otherwise,} \end{cases} \ j = 1...m. \tag{28}$$

The second fuzzy set is (C_s, μ_2), where the membership function μ_2 is

$$\mu_2 : C_s \rightarrow [0,1],$$

$$\mu_2(sc_k) = \begin{cases} \rho_n(sc_k), & \text{if } \rho_n(sc_k) > \rho_{acc}(sc), \\ 0, & \text{otherwise,} \end{cases} \ k = 1...p. \tag{29}$$

These two fuzzy sets are an alternative to the modelling of academic curricula. They can be embedded in decision making using the well acknowledged operators specific to fuzzy logic.

Conclusions

This paper has proposed an extension of the LSP by means of compatible systems. A mathematical modelling approach that involves information densities has been proposed and applied to the curriculum management of two academic programs of study.

Our mathematical modelling approach is presented by means of specific properties of compatible systems and elements including the definition of fuzzy sets. This aspect will be treated as a future research direction by offering a clear modelling algorithm organised in terms of clear steps that should highlight the presence of control systems and mechatronics applications [45-56] in the courses of the engineering programs of study. The generalization to an arbitrary number of programs of study will be treated.

The future research will be focused on more convincing educational applications and on the analysis and design of compatible control systems. The further extensions of our modelling approach by fuzzy sets will enable the investigation of tools for the analysis and design fuzzy control systems.

References

[1]   B. H. Liskov, J. M. Wing: A Behavioral Notion of Subtyping, ACM Transactions on Programming Languages and Systems, Vol. 16, No. 6, 1994, pp. 1811-1841

[2]   R. S. Fântână: Human Resources Management. A Mathematical Image of The Profession Compatibility. Proceedings of IXth International Conference on Challenges in Higher Education and Research in 21st Century, Sofia, Bulgaria, 2011, pp. 342-346

[3]   P. Fodor, J. Poór: The Impact of the Economic and Financial Crisis on HRM and Knowledge-Management in Hungary and Slovakia - Empirical Research 2008-2009, Acta Polytechnica Hungarica, Vol. 6, No. 3, 2009, pp. 69-91

[4]   B. Zemková, J. Talašová: Fuzzy Sets in HR Management, Acta Polytechnica Hungarica, Vol. 8, No. 3, 2011, pp. 113-124

[5]   D. E. Guest: Human Resource Management and Performance: Still Searching for Some Answers, Human Resource Management Journal, Vol. 21, No. 1, 2011, pp. 3-13

[6]   M. Waring: All in This Together? HRM and the Individualization of the Academic Worker, Higher Education Policy, Vol. 26, 2013, pp. 397-419

[7]   F. Farkas, Á. Király: What Makes Higher Education Knowledge-Compatible?, Acta Polytechnica Hungarica, Vol. 6, No. 3, 2009, pp. 93-104

[8]   R.-E. Precup, S. Preitl, M.-B. Radac, E. M. Petriu, C.-A. Dragos, J. K. Tar: Experiment-based Teaching in Advanced Control Engineering, IEEE Transactions on Education, Vol. 54, No. 3, 2011, pp. 345-355

[9]   J. Kabók, T. Kis, M. Csüllög, I. Lendák: Data Envelopment Analysis of Higher Education Competitiveness Indices in Europe, Vol. 10, No. 3, 2013, pp. 195-201

[10]   G. Kiss: Teaching Programming in the Higher Education not for Engineering Students, Procedia - Social and Behavioral Sciences, Vol. 103, 2013, pp. 922-927

[11]   R. M. Carlton: An Overview of Standards in Electromagnetic Compatibility for Integrated Circuits, Microelectronics Journal, Vol. 35, No. 6, 2004, pp. 487-495

[12]   S. Midya, R. Thottappillil: An Overview of Electromagnetic Compatibility Challenges in European Rail Traffic Management System, Transportation Research Part C: Emerging Technologies, Vol. 16, No. 5, 2008, pp. 515-534

[13]   J.-H. Hahn, S.-H. Kim: Mix-and-Match Compatibility in Asymmetric System Markets, Journal of Institutional and Theoretical Economics, Vol. 168, No. 2, 2012, pp. 311-338

[14]   N. Hameurlain: Formalizing Compatibility and Substitutability of Rolebased Interactions Components in Multi-agent Systems, Multi-Agent Systems and Applications IV, M. Pěchouček, P. Petta, L. Z. Varga, Eds., Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, Vol. 3690, 2005, pp. 153-162

[15]   C. Amrouche, P. G. Ciarlet, L. Gratie, S. Kesavan: On Saint Venant's Compatibility Conditions and Poincaré's Lemma, Comptes Rendus Mathematique, Vol. 342, No. 11, 2006, pp. 887-891

[16]   D. C. Klonoff: Designing an Artificial Pancreas System to Be Compatible with Other Medical Devices, Journal of Diabetes Science and Technology, Vol. 2, No. 5, 2008, pp. 741-745

[17]   M. J. Hamamura, S. Ha, W. W. Roeck, L. T. Muftuler, D. J. Wagenaar, D. Meier, B. E. Patt, O. Nalcioglu: Development of an MR-Compatible SPECT System (MRSPECT) for Simultaneous Data Acquisition, Physics in Medicine and Biology, Vol. 55, No. 6, 2010, pp. 1563-1575

[18]   T. Haidegger, L. Kovács, R.-E. Precup, B. Benyó, Z. Benyó, S. Preitl: Simulation and Control for Telerobots in Space Medicine, Acta Astronautica, Vol. 181, No. 1, 2012, pp. 390-402

[19]   W. C. Torres, M. Quintana, H. Pinzón: Differential Diagnosis of Hemorrhagic Fevers Using ARTMAP and an Artificial Immune System, International Journal of Artificial Intelligence, Vol. 11, No. A13, 2013, pp. 150-169

[20]   E. G. Christoforou, I. Seimenis, E. Andreou, E. Eracleous, N. V. Tsekos: A Novel, General-Purpose, MR-Compatible, Manually Actuated Robotic

Manipulation System for Minimally Invasive Interventions Under Direct MRI Guidance, The International Journal of Medical Robotics and Computer Assisted Surgery, Vol. 10, No. 1, 2014, pp. 22-34

[21]   H. Nam-Gung, Y.-H. Kim, I.-K. Lim, J.-G. Lee, J.-P. Lee, J.-K. Lee: Assessment of Compatibility between Standard Medical Systems of u-RPMS and HL7, Ubiquitous Information Technologies and Applications, Y.-S. Jeong, Y.-H. Park, C.-H. Hsu, J. H. Park, Eds., Lecture Notes in Electrical Engineering, Springer-Verlag, Berlin, Heidelberg, Vol. 280, 2014, pp. 387-394

[22]   M. Freeman: To Adopt or not to Adopt Innovation: A Case Study of Team-based Learning, The International Journal of Management Education, Vol. 10, No. 3, 2012, pp. 155-168

[23]   G. Alimehmeti, X. Hysa: Why the Europeanization of Bologna Process is a Curved road? Trying a Response through the Viable Systems Approach in Reference to Albania, Procedia - Social and Behavioral Sciences, Vol. 47, 2012, pp. 722-728

[24]   M. Azzi: The New Pedagogical Practices within the LMD System: Perceptions of EFL Faculty Members, Procedia - Social and Behavioral Sciences, Vol. 69, 2012, pp. 1004-1013

[25]   C. Caro, N. Quijano: Low Cost Experiment for Control Systems, Proceedings of 2011 IEEE IX Latin American Robotics Symposium and IEEE Colombian Conference on Automatic Control and Industry Applications (LARC), Bogota, Colombia, 2011, pp. 1-6

[26]   M. Gunasekaran, R. Potluri: Low-Cost Undergraduate Control Systems Experiments Using Microcontroller-based Control of a DC Motor, IEEE Transactions on Education, Vol. 55, No. 4, 2012, pp. 508-516

[27]   M. Roman, D. Popescu, D. Selişteanu: An Interactive Teaching System for Bond Graph Modeling and Simulation in Bioengineering", Educational Technology & Society, Vol. 16, No. 4, 2013, pp. 17-31

[28]   S. Aleksandrov, Z. Jovanović, D. Antić, S. Nikolić, S. Perić, R. Aleksandrov: Analysis of the Efficiency of Applied Virtual Simulation Models and Real Learning Systems in the Process of Education in Mechatronics, Acta Polytechnica Hungarica, Vol. 10, No. 6, 2013, pp. 59-76

[29]   M. Atanasijevic-Kunc, V. Logar, M. Papic, J. Bester, R. Karba: Motivation Experiments for Complex Control Systems Education, Proceedings of 10th IFAC Symposium Advances in Control Education (ACE 2013), Sheffield, UK, 2013, pp. 268-273

[30]   C. Buiu, A. Buga, A. M. Coman: Teaching Robotics and Virtual Reality in a Synergistic Approach, Proceedings of 7th IEEE International Conference

on e-Learning in Industrial Electronics (ICELIE 2013), Vienna, Austria, 2013, pp. 71-75

[31]   S. Preitl, R.-E. Precup, J. Fodor, B. Bede: Iterative Feedback Tuning in Fuzzy Control Systems. Theory and applications, Acta Polytechnica Hungarica, Vol. 3, No. 3, 2006, pp. 81-96

[32]   M.-B. Radac, R.-E. Precup, E. M. Petriu, S. Preitl: Application of IFT and SPSA to Servo System Control, IEEE Transactions on Neural Networks, Vol. 22, No. 12, 2011, pp. 2363-2375

[33]   S. Formentin, P. De Filippi, M. Corno, M. Tanelli, S. M. Savaresi: Data-driven Design of Braking Control Systems, IEEE Transactions on Control Systems Technology, Vol. 21, No. 1, 2013, pp. 186-193

[34]   M. Shen, W.-N. Chen, J. Zhang, H. S.-H. Chung, O. Kaynak: Optimal Selection of Parameters for Nonuniform Embedding of Chaotic Time Series Using Ant Colony Optimization, IEEE Transactions on Cybernetics, Vol. 43, No. 2, 2013, pp. 790-802

[35]   M. Samejima, Y. Saito, M. Akiyoshi, H. Oka: A Predictive Search Method for FAQ-based Question Answering System, International Journal of Artificial Intelligence, Vol. 11, No. A13, 2013, pp. 103-114

[36]   L. G. Guzmán, Alix S. Gómez, C. J. Ardila, D. Jabba: Adaptation of the GRASP Algorithm to Solve a Multiobjective Problem Using the Pareto Concept, International Journal of Artificial Intelligence, Vol. 11, No. A13, 2013, pp. 222-236

[37]   F. Valdez, P. Melin, O. Castillo: A Survey on Nature-inspired Optimization Algorithms with Fuzzy Logic for Dynamic Parameter Adaptation, Expert Systems with Applications, Vol. 41, No. 14, 2014, pp. 6459-6466

[38]   R.-E. Precup, S. Preitl, M. Balas, V. Balas: Fuzzy Controllers for Tire Slip Control in Anti-lock Braking Systems, Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2004), Budapest, Hungary, 2004, Vol. 3, pp. 1317-1322

[39]   R.-E. Precup, S. Preitl: Stability and Sensitivity Analysis of Fuzzy Control Systems. Mechatronics Applications, Acta Polytechnica Hungarica, Vol. 3, No. 1, 2006, pp. 61-76

[40]   R.-E. Precup, S. Preitl: PI-fuzzy Controllers for Integral Plants to Ensure Robust Stability, Information Sciences, Vol. 20, No. 20, 2007, pp. 4410-4429

[41]   R.-E. Precup, M. L. Tomescu, S. Preitl: Fuzzy Logic Control System Stability Analysis Based on Lyapunov's Direct Method, International Journal of Computers, Communication & Control, Vol. 4, No. 4, 2009, pp. 415-426

[42]   Z. C. Johanyák: Fuzzy Modeling of Thermoplastic Composites' Melt Volume Rate, Computing and Informatics, Vol. 32, No 4, 2013, pp. 845-857

[43]   P. Angelov: Autonomous Learning Systems: From Data Streams to Knowledge in Real-time, John Wiley & Sons, Chichester, UK, 2013

[44]   T. Vollmer, M. Manic, O. Linda: Autonomic Intelligent Cyber-Sensor to Support Industrial Control Network Awareness, IEEE Transactions on Industrial Informatics, Vol. 10, No. 2, 2014, pp. 1647-1658

[45]   S. Preitl, R.-E. Precup: On the Algorithmic Design of a Class of Control Systems Based on Providing the Symmetry of Open-loop Bode Plots, Scientific Bulletin of "Politehnica" University of Timisoara, Romania, Transactions on Automatic Control and Computer Science, Vol. 41 (55), No. 1-2, 1996, pp. 47-55

[46]   I. Dumitrache, M. Dragoicea: Some Problems of Advanced Mobile Robot Control, Control Engineering and Applied Informatics, Vol. 7, No. 4, 2005, pp. 11-30

[47]   R.-E. Precup, S. Preitl: PI and PID Controllers Tuning for Integral-type Servo Systems to Ensure Robust Stability and Controller Robustness, Electrical Engineering, Vol. 88, No. 2, 2006, pp. 149-156

[48]   G. Hermann, G. Benkó, A. Nagy: Retrofitting a Length Measuring Machine for Linear Scale Calibration, Proceedings of 13th International Conference on Intelligent Engineering Systems (INES 2009), Barbados, 2009, pp. 79-82

[49]   A. Csapó, P. Baranyi: A Unified Terminology for the Structure and Semantics of CogInfoCom Channels, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012, pp. 85-105

[50]   M. Bošnak, D. Matko, S. Blažič: Quadrocopter Control Using an On-board Video System with Off-board Processing, Robotics and Autonomous Systems, Vol. 60, No. 4, 2012, pp. 657-667

[51]   J. Vaščák, M. Paľa: Adaptation of Fuzzy Cognitive Maps for Navigation Purposes by Migration Algorithms, International Journal of Artificial Intelligence, Vol. 8, No. S12, 2012, pp. 20-37

[52]   F. G. Filip: A Decision-Making Perspective for Designing and Building Information Systems, International Journal of Computers, Communication & Control, Vol. 7, No. 2, 2012, pp. 264-272

[53]   L. Horváth, I. J. Rudas: Active Knowledge for the Situation-driven Control of Product Definition, Acta Polytechnica Hungarica, Vol. 10, No. 2, 2013, pp. 217-234

[54] K. Lamár, J. Neszveda: Average Probability of Failure of Aperiodically Operated Devices, Acta Polytechnica Hungarica, Vol. 10. No. 8, 2013, pp. 153-167

[55] D. Gong. S. Liu: A Holographic-based Model for Logistics Resources Integration, Studies in Informatics and Control, Vol. 22, No. 4, 2013, pp. 367-376

[56] L. Kovács, A. Szeles, D. A. Drexler, J. Sápi, I. Harmati, Z. Sápi: Model-based Angiogenic Inhibition of Tumor Growth using Modern Robust Control Method, Computer Methods and Programs in Biomedicine, Vol. 114, 2014, pp. e98-e110

Novel Incremental Sheet Forming System with Tool-Path Calculation Approach

Imre Paniti[1], János Somló[2]

[1] Institute for Computer Science and Control, Hungarian Academy of Sciences, Kende u. 13-17, H-1111 Budapest, Hungary, imre.paniti@sztaki.mta.hu

[2] Óbuda University, Bécsi út 96/B, H-1034 Budapest, Hungary, somlo@uni-obuda.hu

Abstract: Incremental Sheet Forming (ISF) is a prosperous forming technique since the end of the 20th Century. Research projects are still active in this topic with the goal to understand more deeply this flexible process and to reach the full industrialisation of it. These investigations are mainly focusing on the one-side version of ISF, however only the two side variants of this process could reach the ultimate flexibility. This paper gives an overview of the recent results and inventions in the field of ISF processes. A new solution of the Two-Sided ISF processes realisation is given. A new approach to the tool-path calculation is outlined which is based on a recently patented system of the authors.

Keywords: incremental sheet forming; die-less forming; tool-path calculation; new solution of the Two-Sided ISF

1 Introduction

One of the early sheet metal forming processes used at large scale is spinning. This process is suitable for the manufacturing of rotational parts in low and medium large series. Today, a new technology aroused. This is the Incremental Sheet Forming (ISF) or also called as Asymmetric Incremental Sheet Forming (AISF) process. This is in some respect similar to spinning [1], [2], [3].

ISF can be grouped into two main groups, depending on the number of contact points between sheet, tool and die [4]. The term Single Point Incremental Forming (SPIF) is used when the bottom contour of the part is supported by a backing-/faceplate. In SPIF the forming tool is pressing contours from the outside inwards, moving the inner flat area of the blank gradually downwards.

The term Two Point Incremental Forming (TPIF) is used when a full or partial die is present to support the sheet. In TPIF the blank holder is moving downwards with the forming tool or the die is moving upwards while the forming tool is pressing the sheet on it (see Figure 1).
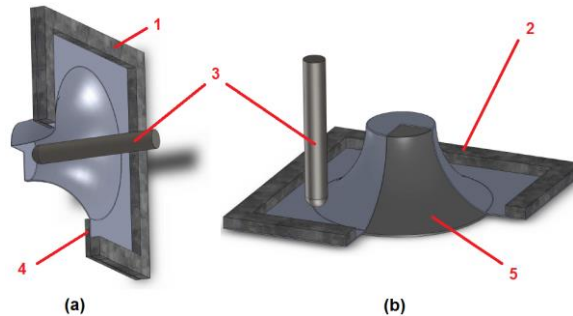
Figure 1

SPIF (a) in a vertical alignment and TPIF (b) in a horizontal alignment with 1: stationary blank holder, 2: moving blank holder, 3: forming tool, 4: backing plate, 5: stationary or moving full or partial die

A special TPIF variant is the two-sided Incremental Sheet Forming or fully kinematic ISF (with supporting tool) [5], which is more flexible but needs synchronised motion of the forming tools. In the following chapters we will discusses the different types of the two-sided ISF (TSISF), the tool-path calculation approaches for the process, a general description of it, and the novel solution of TSISF with a particular example. Furthermore, a tool-path calculation approach of the suggested system is also presented.

2 TSISF Systems

In the previous decades a development of flexible forming technologies (particularly in ISF) could be observed which manifested itself in a number of patents and research projects [6], especially in Europe, in Japan and in the United States.

By analysing the main variants of ISF, it can be seen that SPIF has several limitations compared to two-sided ISF; for example it needs a face-/backing plate, and the forming of convex and concave parts on the same sheet is only feasible if the blank holder is released.

A process with one controller for two synchronised forming tools, with at least three degrees of freedom, corresponding to movements according to axes X, Y and Z, can be found in the patented solution of Rodriguez Gutierrez et al. (TECNALIA) [7].

A prototype of this invention, based on a parallel kinematic machine (PKM) and a coordinate table, can be seen in Figure 2.
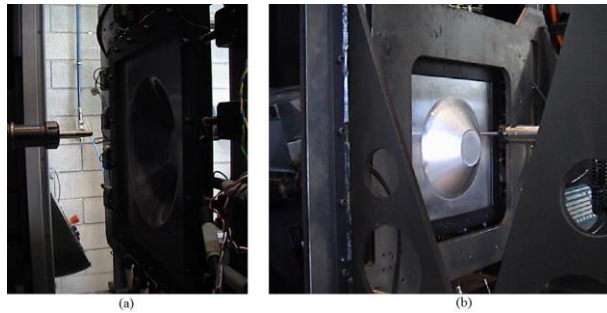
Figure 2

Two-sided Incremental Forming machine of TECNALIA from (a) PKM side and from (b) coordinate table side

This variant is also characterised in that it comprises the continuous detection of the stresses on the sheet due to the movement of the forming tools and the actuation on the means of fastening to alter the gripping force thereof, according to the stresses detected.

In the year of 2012, Carl Frederick Johnson et al. patented a similar method and system in Ford Global Technologies [8]. Figure 3 shows a side view of an exemplary system which can be identified with the set-up of the Ford Freeform Fabrication Technology (F3T).
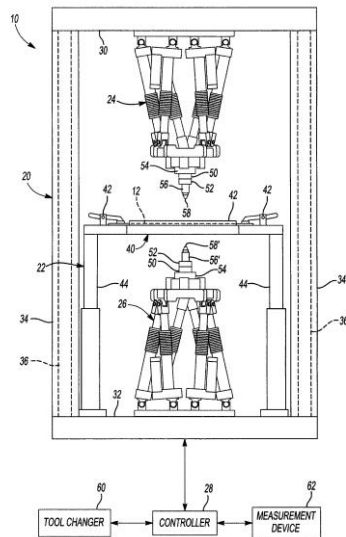
Figure 3 [8]

Side view of Ford's patent variant

This process variant, like in [9] is an obvious extension of TPIF. However, a synchronised control needs additional sensors and is expensive. A set-up, so-called "Dyna(mic)-Die" has been suggested by Franzen et al. [10] but it is restricted to rotational symmetric motions and needs a synchronised control, too.

According to Yongjun Wang et al. [11] Double Sided Incremental Forming (DSIF) requires two forming tools opposing one another and moving simultaneously on both sides of a sheet metal. The device consists of a C frame mounted on the spindle saddle of a CNC machine and the tool heads are mounted on the top and bottom bases of the C frame. The drawback in this case is that a C-frame mounted on the spindle saddle of a CNC machine is limiting the workspace and therefor the freedom of the forming.

The main problem with the previously mentioned devices is that the movement of the lower tool may be influenced by the forces acting on the arm holding the tool and the desired accuracy of the forming may not always be achieved.

The load on the mount frame or other means holding the forming tools may be more, than 1 kN (depending on the desired geometry, the material and the thickness of the sheet), and it is rather difficult to keep the exact position of the tools.

Therefore, the object of the present work is to eliminate the above drawback and to provide a device for two sided incremental sheet forming, which can ensure an accurate forming of sheet materials by stabilised and reliable guiding of the forming tools [12].

2.1 Forming Strategies in TS-ISF

Forming experiments on the prototype of TECNALIA showed that the system allows two main forming strategies (see Figure 4) for the manufacturing of sheet metal parts. Furthermore the moving local support of the second tool gives better results in terms of formability [13].

In the case of "strategy A: peripheral support" the second tool acts as a backing plate, moving synchronised with the first tool, but it does not leave the first contour level of the supporting tool-path. In the case of "strategy B: local support" the second tool moves synchronised with the forming tool, but ensuring continuous local support at each tool-path level.

The thinning of the sheet in ISF is usually estimated (with a raw approximation) by means of the sine-law (1), originated from the metal spinning process [14];

$$t = t_0 \sin(90° - \beta) \tag{1}$$

In (1) the parameter t_0 is the initial thickness of the sheet and β is the wall/draw angle, defining the angle between the un-deformed and the deformed part of the sheet.
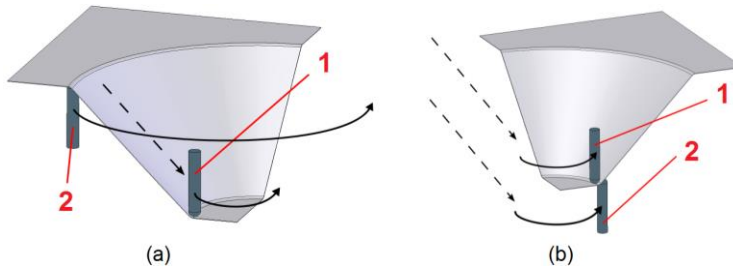
Figure 4

(a) "strategy A: peripheral support", (b) "strategy B: local support" where 1: forming tool, 2: supporting tool

Maidagan et al. [15] defined the tool configuration for the forming strategy with local support (see Figure 5) and created the equation (2).
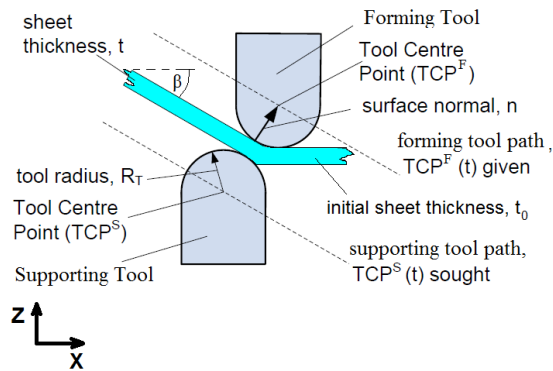

Figure 5

Tool configuration for the forming strategy with local support

$$S = F + \left(2R_T + t\right)\frac{\underline{n}}{|\underline{n}|} \qquad (2)$$

where

$$S = \begin{bmatrix} TCP_X^S \\ TCP_Y^S \\ TCP_Z^S \end{bmatrix}, \qquad F = \begin{bmatrix} TCP_X^F \\ TCP_Y^F \\ TCP_Z^F \end{bmatrix}, \qquad \text{and} \qquad \underline{n} = \begin{bmatrix} n_X \\ n_Y \\ n_Z \end{bmatrix}$$

As it is clear, here (in Figure 5) the y coordinates are not manipulated (n_y=0). The normal vector $\underline{n}$ can be calculated from a generic tool path, which contains the Tool Centre Point and the contact point.

Paniti in [16] adapted the equation (2) to simple rotational symmetric parts and defined the following equation for the forming strategy with peripheral support;

$$
S = \begin{bmatrix} TCP_X^F \dfrac{R}{r} \\[2mm] TCP_Y^F \dfrac{R}{r} \\[2mm] TCP_Z^S \end{bmatrix}
\tag{3}
$$

where TCP_Z^S is calculated only for the first level of the forming. The parameter r is the radius of the last tool-path contour of the forming tool, while R is the radius of the first tool-path contour of the supporting tool. Detailed description with explanatory figures of this adaptation can be found in [16].

The same forming strategies were applied by Meier et al. [17], [18] with two synchronised industrial robots (see Figure 6 and Figure 7), like in a previous European Patent [19] which discloses various configurations, including the use of two robots which needs synchronised control. They also observed better results with the "strategy B: local support"; however their names for the abovementioned strategies were "duplex incremental forming with peripheral support" (DPIF-P) and "duplex incremental forming with local support" (DPIF-L) respectively [18].
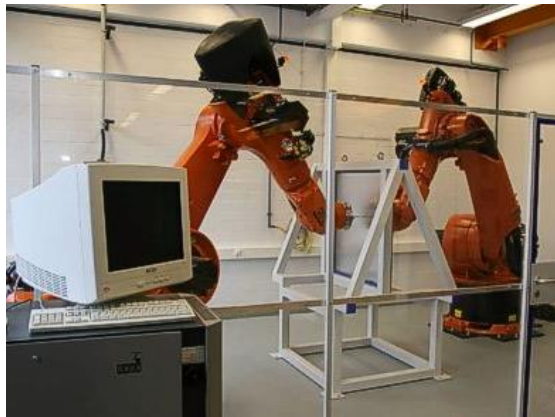

Figure 6 [20]
Two-sided Incremental Forming with synchronised KUKA robots

Figure 7 [21]

Two-sided Incremental Forming with synchronised ABB robots

Meier et al. [18] extended the analytical model of Silva et al. [22] for rotational symmetric SPIF with the so called "superimposed pressure" induced by the supporting tool to analyse the influence of the parameters in forming strategy "B" with local support. They found that superimposed pressure increases the formability in case an optimal supporting contact position (see Figure 8) and optimal force is applied.
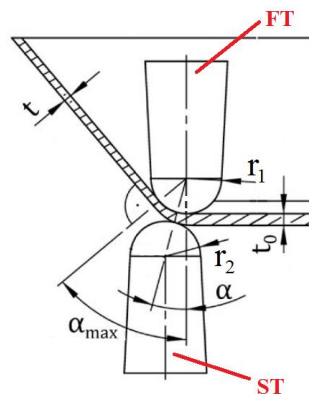

Figure 8 [20]

Contact position of the Supporting Tool (ST) in the localised plastic zone, where FT: Forming Tool, α: shifting angle of the ST, α_{max}: maximum shifting angle of the ST, t: sheet thickness of the deformed sheet, t_0: initial sheet thickness

They introduced a shifting angle α and highlighted that the relative position of the ST to the FT can be varied over the idealised two dimensional forming gap. At an angle α of 0° the two tools are coaxial and facing opposite to each other, like in case of [11]. The maximum shifting angle α is limited by the actual draw angle, so that the force is induced normal to the surface.

Their experiments showed that a higher angle α leads to an unwanted deformation of the sheet and "the applied force always acts in the direction of the connecting line between both tool centre points" [18]. Experiments were carried out on an aluminium alloy sheet (AlMn99.8) with an optimal shifting angel α=30° and optimal contact force equal to 300 N.

Similar results can be reached with the presented solution in this paper, but in contrary to the competitive variants this system does not need to synchronise minimum 3 extra actuators, only one.

2.2   Two-sided (Die-Less) Sheet Forming Strategy in General

In Figure 9 a part of the die-less sheet forming system is shown where the upper tool is the forming tool, the lower one is the supporting tool and between them there is the sheet (fixed by blank holders at the sides) to be formed. Both the forming tool and a supporting tool are applied in a 2.5D to 5D forming machine, which is supposed to be used for manufacturing. It can be a milling machine, a robot or some other device. Except for the "CNC Controlled coupling" solutions below, (when a separate robot or CNC machine is needed to move the supporting tool) both tools are controlled by the same CNC.
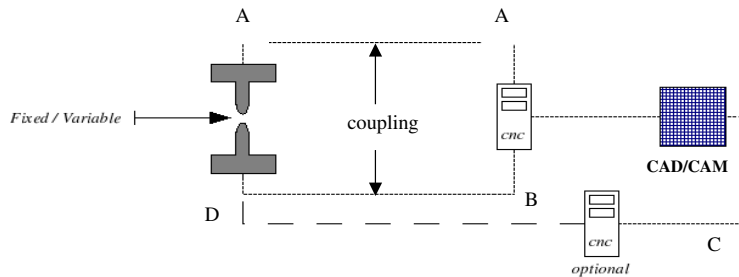

Figure 9

2.5D-5D die-less sheet metal forming

The motion of the forming tool is planned using some proper geometric programming package, CAD/CAM. The package provides CNC program, which is like the one for realising milling process with spherical tool. The forming tool axis should coincide with the normal to the surface in most cases. The axis of the supporting tool can be on the same line as that of the forming tool like in [11].

The connection of the forming and supporting tools may be solved in the following ways:

1. Fixed, mechanical coupling similar to the device presented in [11]:

The two tools are allocated using proper turning, rotating, translating mechanical joints at the beginning of the processing. These joints allow turning in the axis direction. A-A-B-D mechanical connection, the AB CNC is common for both tools.

2. Pressed, mechanical coupling as patented in [12]:

Similar to case 1, just the supporting tool is pressed to the surface by spring mechanism or by pneumatic, hydraulic or other device. A-A-B-D mechanical connection, the AB CNC is common for both tools.

3. Controlled, mechanical coupling as patented in [12]:

Similar to case 2, but the force of the supporting tool is controlled with a proper servo drive implemented in a coupling device. It may be the AB CNC or a device substituting the optional CNC.

4. CNC Controlled coupling/1:

The movement and the force of the supporting tool are controlled with a proper, optional CNC (CD CNC). This CNC gets the info from the AB CNC and controls a second machine (robot or milling machine). The two CNCs may be connected to give a simple parallel movement to the supporting tool.

5. CNC Controlled coupling/2:

The movement and the force of the supporting tool are controlled with a proper, optional CNC (CD CNC). This CNC gets the info from the AB CNC plus from measurements (sheet thickness, forces etc.) and controls a second machine (robot or milling machine). Same as case 4, just the optional CNC gets more input info.

Solutions 1., 2. and 3. suppose a fixed or almost fixed mechanical connection between the forming tool and the supporting tool. The forming tool is controlled by a CNC based on CAD/CAM data. We could define solutions 1a., 2a., 3a. where instead of the mechanical connection the control CNC would control a second machine (robot or milling machine) having the supporting tool. Or even a second (the optional) CNC could be used with the same info as the other CNC and just the sheet material thickness should be taken into account.

Solutions 4. and 5. are more sophisticated and suppose two corresponding, but independent machines with their own controllers for the forming tool and for the supporting tool respectively. If we have the control code for the forming tool, and the CAD-information of the final product we can manage to produce the CNC control code for the supporting tool for every case, when needed.

3 Description of the New TSISF System

The blank holder is mounted on support rods standing on a base plate. The upper forming tool is moved in X, Y and Z directions, according to the programmed tool-path. The lower forming tool is copying the X and Y motion of the upper forming tool. This motion is realised with a mechanical movement copying device (MMCD), but differently from conventional solutions like in [11], the lower

forming tool has a motion in Z direction not together with the copying device but using some linear actuator (mechanic, pneumatic, hydraulic, electrical or the combination of those).

This linear actuator is fixed on a transfer unit, which holds the lower forming tool at standard height with respect to the base plate. The transfer unit in its simplest solution has two degrees of freedom, corresponding to movements according to axes X and Y, on a base plate.
The reaction force in the Z direction acting on the lower forming tool and consequently on the linear actuator from the pressing of the sheet is transferred to the base plate through the transfer unit.

The moving device is preferably passive motion equipment and allows free motion in the X-Y plane, and can stabilise the movement of it. A stabiliser may be constructed according to provide e.g. two rotational motions, one rotation and one translation motion or translation motion in X and Y direction.

The lower forming tool mounting can be stiff, compliant, or regulated. According to a preferred set-up, there is a link mechanism or a lever transmission between the base plate and the transfer unit.

The MMCD may be a mount frame, having a C, O or U shape, or may be a spindle-gear or a belt-pulley slide mechanism.

To better understand the system characteristics, according to preferred practical set-ups thereof, there is a set of illustrative drawings below (see Figure 10 and Figure 11) to the detailed description.
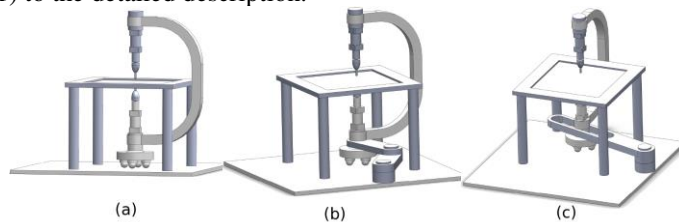

Figure 10

Variants of the TSISF device for small dimensions (a: without stabiliser mechanism, b: with a rotational stabiliser mechanism, c: with a rotational and translational stabiliser mechanism).
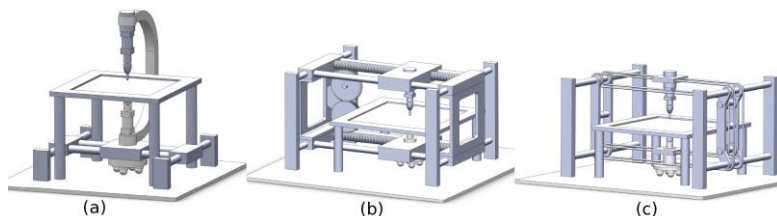

Figure 11

Variants of the TSISF device for large dimensions (a: with linear slides, b: with spindle-gear movement transmission elements, c: with belt-pulley slide mechanism)

This TSISF system enables safe and exact movement of the forming tools without applying sophisticated and expensive constructions. The system elements for moving the lower forming tool on the surface of the base plate, e.g. a transfer unit with rolling balls eliminates the non-desired movement of the forming tools in the direction of the load. The system can easily be mounted and dismounted, and therefore it is especially applicable for producing pieces with complex part geometry even in small series [12]. The proposed TSISF system can be applied as an external axis in some 4D milling machine centres and in robotic cells, too.

4 Realisation of the TSISF System with Linear Slides and Force Control

In order to test a selected variant of the TSISF an experimental system was designed. Figure 12 shows the CAD model of a set-up, similar to the variant presented in Figure 11a.

Figure 12
3D model of the set-up with a Fanuc S430iF type industrial robot

The form of the C-frame is designed to secure the linear actuator against radial forces. FEM simulations have been made in ANSYS with radial forces up to 600 N (see Figure 13) based on previous forming experiments (SPIF of 5 mm thick polyethylene sheets) to validate the C-frame design.

Results of the FEM simulation showed tolerable stresses below 87 MPa in the rib elements. The C-frame is mounted on a FANUC S430iF type industrial robot's 5th axis. The movement of the upper forming tool (which preferred to be mounted eccentric to the axis of the 6th joint of the robot) is realised with a FANUC R-J3 controller. The eccentric mounting allows the realisation of the optimal contact position discussed in chapter 2.1.
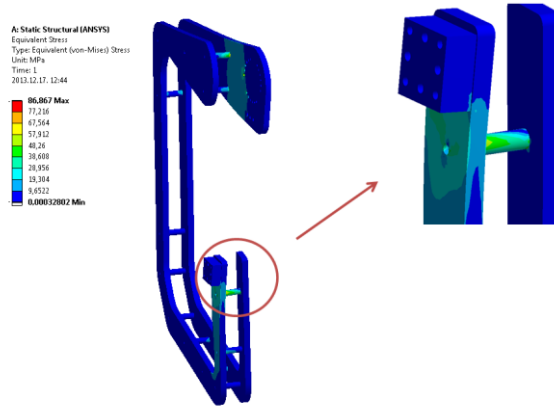
Figure 13
FEM simulation of the C-frame

Both forming tools are mounted on a CLS1000 type measurement cell to give a force feedback to the TSISF control and to ensure an optimal contact force. In this case only the highest component (z) of the reaction force is measured. The linear motion of the lower forming tool is realised with a BSA 20 type ball screw linear actuator with rotational encoder. High level motion commands are sent from the R-J3 controller to the robot and to the linear actuator, too.

4.1　Trajectory Planning

One of the most important problems to be solved at the use of such a new kind of material processing principles and systems is the development of proper process planning methods. Among other problems, trajectory planning (determination of velocities along the paths) should have an effective solution. It seems to us that approaches used at metal processing by machine tools in general can be used as basis for this technology, too (see for example:[23]).

A mathematical model for cutting processes proposed in [23], consist from 3 parts:

　　　　　　　1.)　System of constraints
　　　　　　　2.)　Performance index
　　　　　　　3.)　Tool life equation

In the case of incremental sheet metal forming the tool life issues are less important (but not negligible). So, the determination of the optimal regimes becomes rather simple. It seems that at given depth, maximum feed values should be used. These are determined by the system of constraints. But, the determination of the constraints becomes a new and rather complicated problem. To the first plan the dependence of the surface quality on feeds and also temperatures moves.

Very promising opportunity is the application of high velocity processing. For that an important issue is the realisation of time optimal cruising motion on the given paths. These problems are solved and published in Somlo, Lantos, Cat [24]. The time-optimal cruising trajectory planning supposes that the forming tool passes any length on a given path for minimum time. The knowledge of the limit values opens opportunity for effective utilisation for given equipment (robot).

4.2 Tool-Path Calculation Approach for Concave and Convex Surfaces without Releasing the Blank Holder

In case an eccentric upper tool is applied in TSISF, the following tool-path calculation method is proposed;

When the eccentric tool (mounted on the machine) is acting as forming tool, while the supporting tool on the linear actuator is force controlled:

1. Calculate generic tool-path GTP(X, Y, Z, I, J, K) of the part as in SPIF

     (I, J, K values are coordinates of the contact point).

2. Calculate supporting points with equation (2).

3. Replace Z coordinates in (2) from GTP for eccentric tool-path.

4. Calculate rotational angle γ of the last joint based on the following coordinates: TCP_X^F, TCP_Y^F, TCP_X^S, TCP_Y^S.

5. Solve inverse kinematics problem.

6. Replace the joint position value of the last joint to γ.

When the eccentric tool (mounted on the machine) is acting as supporting tool, the Z-coordinate of it is force controlled and the linear actuator is acting as forming tool:

1. Calculate generic tool-path GTP(X, Y, Z, I, J, K) of the part as in SPIF

     (I, J, K values are coordinates of the contact point).

2. Recalculate Z for linear actuator, based on starting position of the actuator.

3. Calculate supporting points with equation (2).

4. Calculate rotational angle γ of the last joint based on the following coordinates: TCP_X^F, TCP_Y^F, TCP_X^S, TCP_Y^S.

5. Solve inverse kinematics problem.

6. Replace the joint position value of the last joint to γ.

The inverse kinematics problem can be solved with conventional methods (numerically, simulated) or directly by the recording of the joint positions during a

test execution of the GTP without forming the sheet. The eccentricity of the upper forming tool is pre-defined by equation (1). In case the part geometry contains more than one wall angel, the eccentricity have to be adjusted to it.

5   Experimental Results

Experiments have been carried out on a TSISF prototype and on a robotic cell to compare the results with a simplified SPIF. Both set-ups excluded the use of a backing plate. The TSISF prototype formed a truncated cone (see geometry in Figure 14) while the robotic cell created a truncated pyramid (see geometry in Figure 15). Both geometries are commonly used to test the forming limits of Incremental Sheet Forming. Aluminium sheets (Al 1050) with 0.5 mm and 0.6 mm initial thickness were selected with forming speed 50 mm/sec and 300 mm/sec. A proper lubrication was used during the experiments. Forming tools with 10 mm tool diameter were selected in both cases.
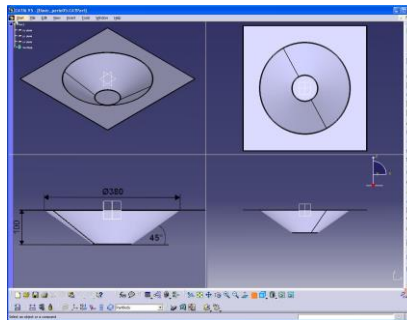

Figure 14
Truncated cone geometry with wall angle β=45° in a CATIA CAD program

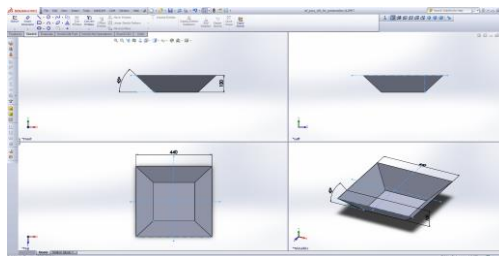

Figure 15
Truncated pyramid geometry with wall angle β=45° in a SolidWorks CAD program

Figure 16 and Figure 17 show the results of the TSISF experiment, while Figure 18 and Figure 19 represent the results of the simplified SPIF process.
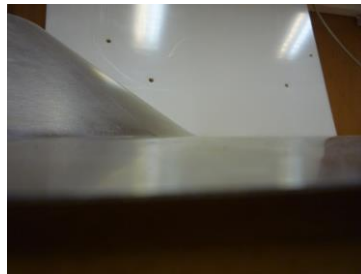

Figure 16
Final part made with TSISF

Figure 17
Peripheral area on the final part made with TSISF

Figure 18
Final part made with SPIF

Figure 19

Peripheral area on the final part made with SPIF with unwanted deflection

No deflection can be seen at the peripheral area on the final part made with TSISF contrary to SPIF where unwanted deflection can be monitored.

Conclusions

A novel Two Sided Incremental Sheet Forming (TSISF) system was presented which can be described as a cost efficient fully kinematic ISF solution. Motion and Finite Element Method simulations proved that the proposed TSISF concept with a C-frame is suitable for creating concave and convex shapes without the release of the sheet. The presented tool-path calculation approach showed that the eccentric mounting of the upper tool allows the realisation of an optimal contact position in TSISF, however the eccentricity have to be adjusted in case of different drawing angles.

Acknowledgement

References

[1]   Edward Leszak (1967): Apparatus and Process for Incremental Dieless forming, US Patent US3342051

[2]   Berghahn Walter G, Murray Jr George F (1967): Method of Dielessly Forming Surfaces of Revolution, US Patent US3316745

[3]   Izeki, Hideo; Kamiide, Takuo; Shimoda, Naoki (1998): Tool Unit for Sequential Stretch Forming and Sequential Stretch Forming Machine, Japanese Patent JP10-180365

[4]   Hirt, G.; Ames, J.; Bambach, M.; Kopp, R. (2004): Modeling and Experimental Evaluation of the Incremental CNC Sheet Metal Forming Process, CIRP Annals - Manufacturing Technology, Vol. 53, No. 1, pp. 203-206

[5]   Tekkaya, A. Erman, ed. Sheet Metal Forming: Processes and Applications, Asm International, 2012

[6]   Emmens, WC; Sebastiani, G; van den Boogaard, AH (2010): The Technology of Incremental Sheet Forming - a Brief Review of the History, J. of Material Processing Technology 210(8):981-997, doi:10.1016/j.jmatprotec.2010.02.014

[7]   Gutierrez Pedro Pablo Rodriguez, Rastrero Maria Asuncion Rivero, Onandia Elixabete Maidagan (2008): Machine for Shaping Sheet Metal and Shaping Method, PCT Patent EP1977842

[8]   Carl Frederick Johnson, Vijitha Senaka Kiridena, Feng Ren, Zhiyong Cedric Xia (2012): System and Method for Incrementally Forming a Workpiece, US Patent US8322176

[9]   Shima, Akio; Yoshikawa, Takenao; Nakamura, Keiichi; Sudo, Yoichi; Suzuki, Shoji (1997): Formation of Successively Expanding Metallic Plate and Apparatus Therefor, Japanese Patent JP09-085355

[10]   Franzen, V., Kwiatkowski, L., Sebastiani, G., Shankar, R., Tekkaya, A. E., Kleiner, M. (2008): Dyna-Die: Towards Full Kinematic Incremental Forming, Proceedings Esaform, Lyon, France, April 23-25 2008, paper #302

[11]   Yongjun Wang, Ying Huang, Jian Cao (2008): Experimental Study on a New Method of Double Side Incremental Forming, in the Proceedings of ASME 2008 Conference, October 7-10, 2008, Evanston, Illinois, USA, pp. 601-607

[12]   Imre Paniti, Janos Somlo (2013): Device for Two-sided Incremental Sheet Forming, EU Patent EP2505279

[13]   Tisza, M.; Paniti, I.; Kovács, P. Z. (2010): Experimental and Numerical Study of a Milling Machine-based Dieless Incremental Sheet Forming, International Journal of Material Forming, Volume 3, Supplement 1, pp. 441-446

[14]   Avitzur, B.; Yang, C. T. (1960): Analysis of Power Spinning of Cones, Journal of Engineering for Industry. Transactions of ASME, Ser. B 82, 1960, pp. 231-245

[15]   Maidagan, E., Zettler, J., Bambach, M., Rodríguez, P. P., & Hirt, G. (2007): A New Incremental Sheet Forming Process Based on a Flexible Supporting Die System, Key Engineering Materials, 344, pp. 607-614

[16]   Paniti, Imre (2010): CAD API-based Tool-Path Control for Novel Incremental Sheet Forming, Pollack Periodica, 5 (2). pp. 81-90

[17]   Meier, Horst, B. Buff, and V. Smukala (2009): Robot-based Incremental Sheet Metal Forming–Increasing the Part Accuracy in an Automated, Industrial Forming Cell, Key Engineering Materials 410 (2009): 159-166

[18]   Meier, H.; Magnus, C.; Smukala, V. (2011): Impact of Superimposed Pressure on Dieless Incremental Sheet Metal Forming with Two Moving Tools, CIRP Annals - Manufacturing Technology, Volume 60, Issue 1, 2011, pp. 327-330

[19]   Timo Tuominen (2005): Method and Apparatus for Forming Three-Dimensional Shapes in a Sheet Metal, PCT Patent EP1560668

[20]   Roboforming Project Documentation, Incremental Sheet Forming with two KUKA robots, http://www.lps.ruhr-uni-bochum.de/ausstattung/labor/roboforming/, available online in Febr. 2014

[21]   Roboforming Project Documentation, ISF with two ABB robots, http://www.dieffenbacher.de/fileadmin/bilder/Sonstiges/Broschueren_PDFs/Forschung_Entwicklung/bericht_roboforming_de.pdf, available online in February 2014

[22]   Silva, M. B.; Skjoedt, M.; Martins, P. A. F.; Bay, N. (2008): Single-Point Incremental Forming and Formability-Failure Diagrams, Journal of Strain Analysis for Engineering Design, 43/1 (2008), pp. 15-36

[23]   Horváth M., Somló J. (1979): Optimization and Adaptive Control of Manufacturing Processes (In Hungarian) Műszaki kiadó, Budapest, p. 260

[24]   J. Somló, B. Lantos, P. T. Cat (1997): Advanced Robot Control, Academy Press, Budapest, p. 425

WaypostEye – a Software Tool for Navigation Improvement Support of Adaptive Websites

Željko Eremić[1], Dragica Radosav[2]

[1] Technical College of Applied Sciences, Djordja Stratimirovića bb, 23000 Zrenjanin, Serbia, zeljko.eremic@vts-zr.edu.rs

[2] University of Novi Sad, Technical faculty "Mihajlo Pupin", Djure Djakovića bb, 23000 Zrenjanin, Serbia, radosav@tfzr.uns.ac.rs

Abstract: One frequently used method for navigation improvement of web sites is to establish a link between resources such as web pages or files, when it is fully justified. Adaptive web sites may change their display and structure based on a previously recorded user's behavior. The software tool that is described in this paper provides the designer with a graphical presentation of established link proposals between resources. In this way, it supports the designer's decisions referring to establishing new links between resources. Therefore, it makes a contribution to the improvement of the navigational structure of such web sites. This software establishes decision support information and Web designers are able to make business decisions in a relatively easy and visual way.

Keywords: adaptive website; waypost; website design

1   Introduction

When designing a website, an important aspect is to check the resource connections (web pages, files) within a site. The navigation structure in the web site is often not in keeping with the wishes and expectations of visitors. It often appears that a user must visit a number of web pages in order to get to the desired content. Well placed links on the pages can make a significant contribution to faster and easier access to the desired content. "Providing a link (i.e., shortcut) between these potential wayposts could assist users by reducing the number of clicks they have to make while browsing, pointing them in the right direction towards a specific target document." [1].

Adaptive web sites can customize their look and structure to meet the aspirations of user for fast and comfortable navigation. Users' behavior (navigation), previously recorded in log files, is used for the detection of the users' objective. By the appropriate processing of these data, it is possible to obtain the estimates of

the resources that need to be connected using the hyperlinks. Each proposed hyperlink is evaluated on the basis of contributions to the navigation improvement. Based on the list of suggested links, the connections can be made between the resources, either automatically or manually added by the designer.

If the resources, which are located in the existing navigation, can be connected by adding links, these links can significantly accelerate the navigation through the website. This paper aims to present a software tool called WaypostEye which can graphically represent the distances between directly related web pages. In this way, the decision support is provided in the re-designing process of adaptive website navigation. Although the software tool is ready for use, its main function is to present the ideas on which it is based.

2   Previous Studies

Adaptive web sites are defined in [12] as "web sites that improve themselves by learning from user access patterns". In the previously mentioned article it is described the improvement in webmaster' understanding the interaction among visitors and the web site. Adaptive web sites often use information about previous user's behavior which is stored in log files. That information is in the standard form which is suitable for computer processing. General details referring to the log files are given in [11].

Shortcutting is a common way to improve the navigation on the web site. The article [1] considers using waypost documents as a relevant factor in process of improving classic shortcutting. Two illustrations of identifying waypost documents are given in [8].

Interaction events could be recorded for various digital objects (e.g. web pages). Those recordings could be used for better understanding of digital objects. Article [9] introduces history-enriched digital objects, describes a few prototypes and discusses privacy issues. The limited set of information stored in log files could be combined with the use of history-enriched digital objects, which has no such limitation. One approach comes from The Footprints project [14], which allows users to create history-rich digital objects in digital environment. Framework and software tools for visualization of navigational structure are developed to support the theory which claims that activities of past users could help new users in navigation through the space of information.

Displaying web site's navigational structure, for its improvement purposes, can be a challenging task. Fisheye technique implemented in hyperbolic browser, presented in [10], can visualize large hierarchies. The method for getting list of proposed shortcuts, based on recordings in log files, is explained in details within [5]. Its output is one of inputs in WaypostEye software tool.

There are many areas where improved web site's navigation can bring benefits, and one of them is education. One approach to reduce omitting relevant documents and avoid user's withdraw from the web site is given in [6]. This approach reduces the number of clicks, avoiding visits to non-interesting documents (web pages) by offering links to potentially interesting documents.

While log files enable a better navigation from past information, there are possibilities to get help in navigation at the moment of browsing. One possibility is given in [7], where one user can ask another for help in a navigation process. That possibility is the combination of adaptive web site and Ajax. History-enriched digital objects in combination with Ajax and database could be used for storing information about the knowledge sharing among users in real time, as described in [4]. Those objects in combination with log files could be great help for designing the efficient navigation of adaptive web site. Article [4] gives a conceptual and physical model, for the database required for recording knowledge in real time.

A parallel could be drawn with an interesting idea of Flexible Graphical User Interface presented in [2]. Proposed intuitive interface significantly reduces the task execution time and the number of interactions in the communication between a man and a machine.

3   WaypostEye

WaypostEye is a software tool that, graphically and appropriately, presents the optimal navigation proposal to the website designer. It uses the data from the input files and the results can be displayed in a number of ways.

The software tool (Figure 1) that is presented in this paper uses two files as the input where each file contains the correct type of information in the prescribed format:

- Structure File - Navigational structure file which contains the information about the existing relationships between resources (web pages or files). The first element of the pair is a resource which has a hyperlink to the second element of the pair.

- Shortcuts File - The file that contains the proposed links between documents based on the previous analyses of user's behavior. The first element of the pair is a resource with a proposed hyperlink to the second element of the pair.
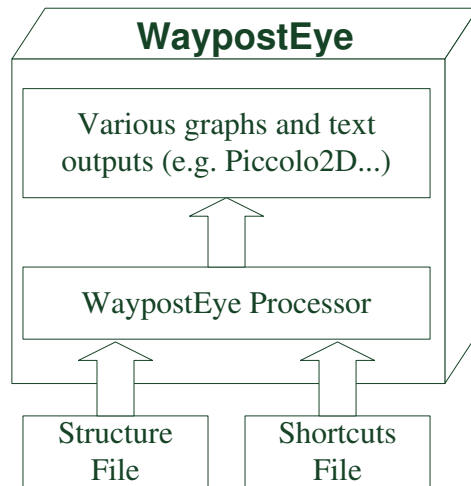
Figure 1
Input files and components of WaypostEye tool

"WaypostEye Processor" is a part of WaypostEye tool which stands between the input data (files) and the particular display of processing results. This part accepts the input data and provides a list of resources possible to perform the analysis of the current ones and recommended connections to other resources. When the user selects the appropriate resource, the graphical illustration of existing connections of this resource to other resources is given, as well as a proposal for the establishment of new hyperlinks.

The results can be graphically and textually presented. In this paper the results are presented using a graph based on a free and open source solution [13], but the list of possible outputs can be easily expanded.

The tool will be presented using the example of a simple web site which contains numerous resources. Each of the resources, as it is shown in the example, is associated with one or more resources, but there is no resource that is connected to all other resources. For the sake of simplicity, the simple rule has been established in this example - if resource A has a link to resource B, then resource B has the link to the resource A. Furthermore, four proposals for establishing connections between the unrelated pairs of resources have been also given. The navigation structure of the observed example and connection proposals are given in Figure 2.
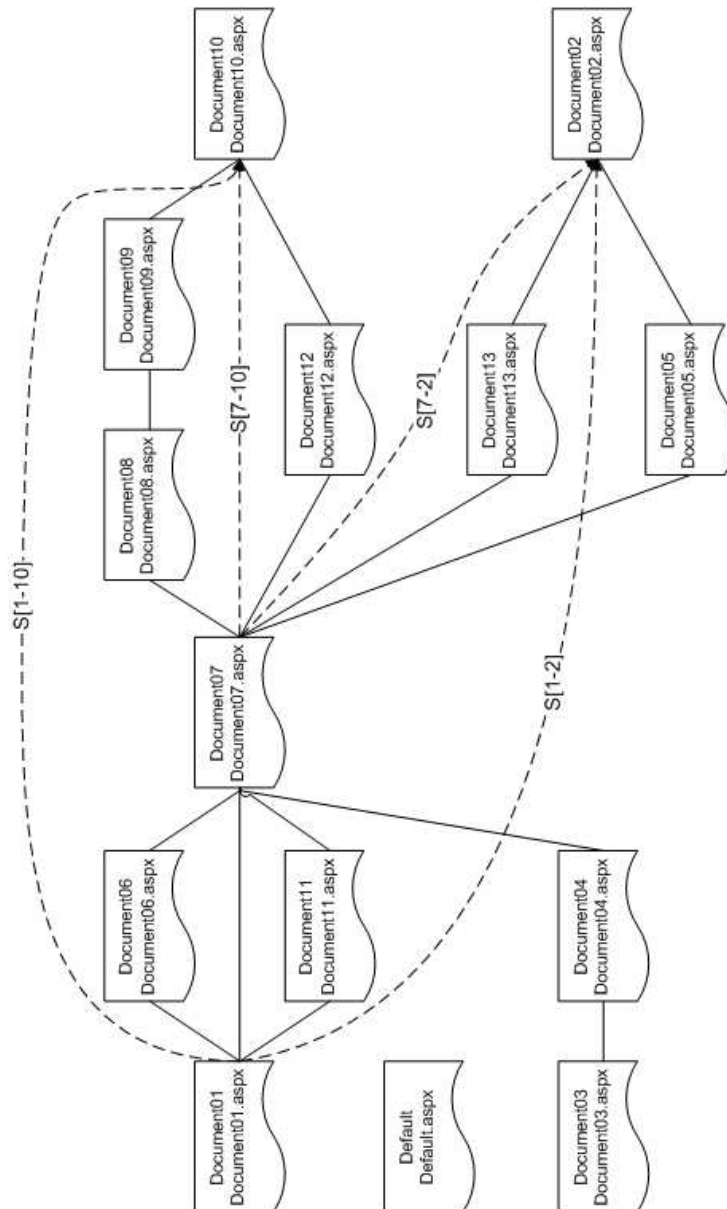
Figure 2

The example of a navigational structure, the connections are represented by solid lines, and suggested links (shortcuts) in dotted lines. Resources are represented by the title and the name of its file.

4   Input Data

The first file (the navigation structure file) contains two kinds of information related to the existing navigation structure:

- Arranged pairs of resources that, are currently directly related.

- Title – Path pairs which connect the resource name with the path to the resource.
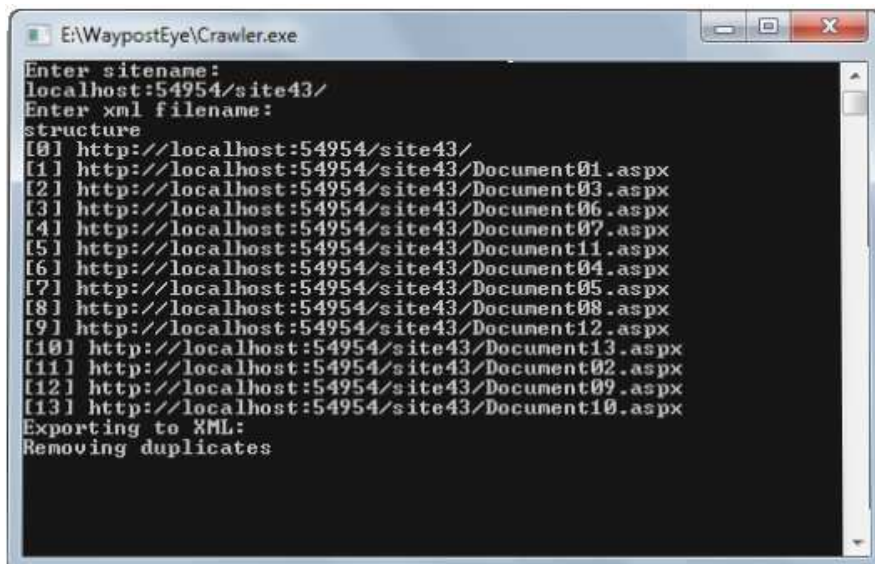
A console application which requests entering a web location has been created for the purpose of this research work (in the example, the experimental web site is located at localhost:54954/site43/). The application also requests entering the name of the file in which the analysis results of the existing navigation structure will be put. Figure 3 shows an example of console applications, and Figure 4 shows the parts of the file "structure.xml" containing the analysis results of console applications. The same result could be obtained by the expression in the command line "Crawler.exe localhost:54954/site43/ structure".

The previously mentioned xml file contains, in its first part, the arranged pairs of resources. By analyzing the HTML code, it was clear that there was a connection between those pairs of resources. The resource with a tag Source contains a link to a resource with a tag Destination. It should be emphasized that, if there is a connection from resource A to resource B, it does not necessarily mean that a link from B to the resource A exists. The connections between the name of the resource and its path are in the second part of the file.

The second file contains output data (in XML format) from the software described in [5]. This file, which example is shown in Figure 5, contains pairs of resources for which establishing connections is recommended.

5   Experiment

The experiment demanded adequate software support. The development environment used was Visual Studio 2008. The test site with the appropriate web pages was made with ASP.NET technology. Web pages were given appropriate headings (Title attribute), and they were associated with other Web pages in accordance with Figure 2. By using the C# programming language, a console application "Crawler" was made. It was used to analyze the structure of the existing website and store the results in an XML file. The software called "WaypostEye" was also made using C#.

Figure 3
The console application that performs analysis of the site (localhost:54954/site43/) and stores the
analysis results in the file (structure.xml)

Running the software takes us to the main window (Figure 6). The previously mentioned two input files need to be loaded. Afterwards, the list of resources of observed site is obtained in the lower part of the window. Appropriate transformations are done in WaypostEye software and finally, the output data are obtained. These data can be presented in different textual and graphical ways. Only graphical output based on the use of Piccolo2D [13], is supported at this stage of software development (version 1.0).

At this point, a resource needs to be selected from the lower section, in order to obtain the display of the existing connections for the resource, as well as, the connections of resources which should be associated with the observed document based on the content of the shortcut file. The result is the graph, in Figure 7. The documents are presented by their names inscribed in the rectangles of different colors. This document is inscribed into a black rectangle. The documents, associated with this selected document, are presented by a light gray rectangle and they are circularly arranged in relation to this document.

The paths towards the documents, which should be connected to the current document, are also presented on the graph. If there are multiple paths between documents, the shortest one is selected using Dijkstra's algorithm, described in [3]. In the given example, the selected "Document-7" should be connected to the "Document-2" (currently connected via the "Document-5"), and "Document-10" (currently connected via the "Document-12").
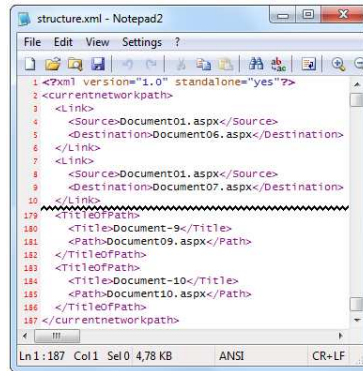
Figure 4

The parts of the file "structure.xml" obtained by an analysis. In the first (upper) part there are links between resources, and in the second (lower) part are connections between the name of the resource and its path

Picolo2D graph [13], allows changing the zoom of the whole graph, moving all the elements of the graph or each of the elements by using a mouse. If you double-click on any graph documents, it becomes a central element of the new graph (Figure 9), and acts as if the document has been selected in the lower section of the form in Figure 6.

6   Directions of Further Development

Previously mentioned software tool represents a convenient way for a website designer to take advantage of the results of the complex analysis of the previous user's behavior, and also to have a possibility to accept or reject a number of the results.
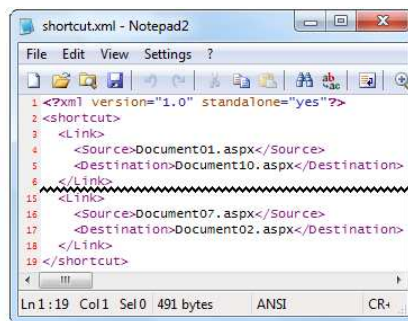

Figure 5

Recommended links between currently unrelated documents
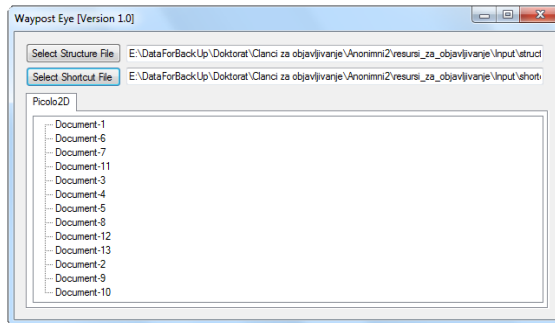
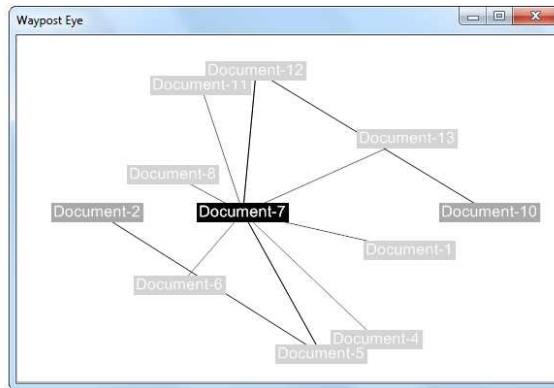Figure 6
The main window of WaypostEye software

Figure 7
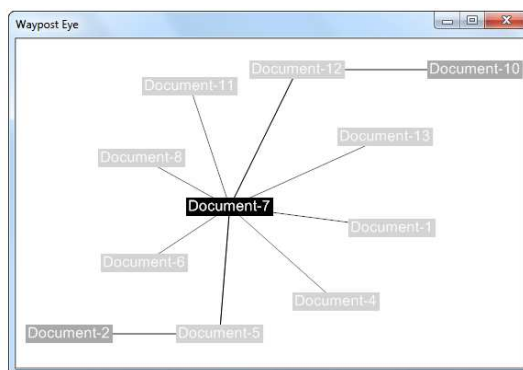Graph which refers to selected document

Figure 8
The graph which is related to the selected document, where by using a mouse, the resources are
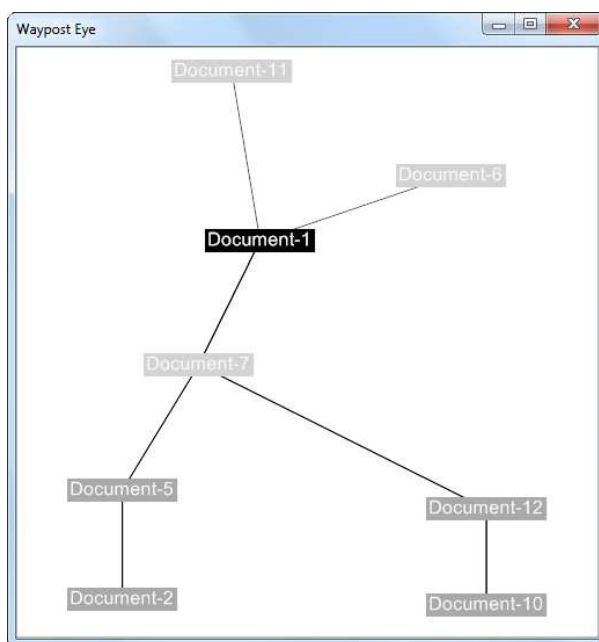arranged in such a way to be more ordered

Figure 9

The graph which is related to the selected document "Document-1", where by using a mouse, the resources are arranged in such a way to be more ordered

In this way, the designer is in a position to make better business decisions in an easier way. There are several directions in which the tool could be developed in the future. The presentation of the results could be extended to various graphic and textual outputs so that the designer could choose the way of presentation that suits him best. If users are given a possibility to leave their comments during the navigation, these comments could appear in the graphs, giving the designer a possibility to read the users' comments and to take some suggestions on board.

Conclusions

One of the tasks in the process of designing a website is to create an efficient and intuitive navigation structure. Web site users often find it difficult to find a resource that they need, or the path to the resource requiring a long series of visits to various other web sites. Adaptive web sites create the possibility of the dynamic addition of links between the resources that are not connected and when there is a justification for these connections. The software, such as the one described in [5], enables us to obtain the connection proposals which should be established between resources based on the user's behavior previously recorded in log files. Such connections are not often obvious, and a serious analysis is required in order to obtain them. Current navigational structure is relatively easily obtained by HTML code analysis, which is performed by custom software.

WaypostEye is a software tool that uses two types of input from the website:

- Arranged pairs of connections for the current navigational structure, with the connection between the resource name and the path is recorded.

- Arranged pairs of connections which should be established between formerly unrelated documents.

This software tool provides, at its output, the resource nodes that are directly linked to the selected document, as well as the documents (including the inter-documents of shortest path by Dijkstra algorithm, described in [3]), to which the selected document should be directly linked. In this version, the tool offers only one possible solution, based on the Piccolo2D [13] solution, but in the future it would be possible to easily add a variety of outputs, either graphical or textual, based on the output data obtained by this tool. The usage of this tool gives adaptive web site designers the possibility to consider the proposals for the new connections which are offered to them. A quick and straightforward business decision can be also made based on this graphical output.

References

[1] Bathumalai G.: Self Adapting Websites: Mining User Access Logs, The Robert Gordon University, Jun, 2008

[2] Dániel B., Korondi P., Sziebig G., Thomessen T.: Evaluation of Flexible Graphical User Interface for Intuitive Human Robot Interactions, Acta Polytechnica Hungarica, Vol. 11, No. 1, 2014, pp. 135-151

[3] Dijkstra E. W.: A Note on Two Problems in Connexion with Graphs, Numerische Mathematik, Vol. 1, No. 1, 1959, pp. 269-271

[4] Eremić Ž.: History-enriched Digital Objects as a Factor of Improvement of Adaptive Educational Web Site Navigation, Journal of Information Technology and Applications, Banja Luka, 2012, pp. 39-43

[5] Eremić Ž.: Navigation Improvement of Adaptive Web Sites by Using Log Files, PhD thesis, Technical faculty "Mihajlo Pupin", University of Novi Sad, Zrenjanin, 2012

[6] Eremić Ž., Radosav D.: Adaptive Web Sites in the Function of Information Access Improvement in Education, Information technology and development of education – ITRO 2011, Zrenjanin, pp. 343-347, 2011

[7] Eremić Ž., Radosav D.: Collaborative User Support as a Contribution to Navigation Improvement of Adaptive Websites for Distance Learning, Information Technology and Development of Education – ITRO 2012, Zrenjanin, 2012, pp. 265-269

[8] Eremić Ž., Radosav D., Markoski B.: Mining User Access Logs to Optimize Navigational Structure of Adaptive Web Sites, Proceedings of the CINTI 2010: 11th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, 2010, pp. 271-275

[9]   Hill W., Hollan, J.: History-enriched Digital Objects, Proceedings of Computers, Freedom: and Privacy (CFP'93), 1993, pp. 139-145

[10]   Lamping J., Rao R., Pirolli P.: A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies, Proceedings of CHI'95 Conference on Human Factors in Computing Systems, ACM Press, 1995, pp. 401-408

[11]   Pamnani R., Chawan P.: Web Usage Mining: A Research Area in Web Mining, Paper presented at the International Conference on Recent Trends in Computer Engineering ISCET 2010, Punjab, India, 2010, pp. 73-77

[12]   Perkowitz M., Etzioni O.: Adaptive Sites: Automatically Learning from User Access Patterns, In Proc. 6th Int. World Wide Web Conf, Santa Clara, California, April 1997

[13]   Piccolo2D, http://www.piccolo2d.org (Accessed on February 18th 2013)

[14]   Wexelblat A., Maes, P.: Footprints: History-Rich Tools for Information Foraging, Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, 1999, pp. 270-277

Universal Fluctuations in Very Short ECG Episodes

Peter Bakucz

Óbuda University, Institute of Mechatronics and Vehicle Engineering
Népszínház u. 8, 1081 Budapest, Hungary
e-mail: bakucz.peter@bgk.uni-obuda.hu

Stephan Willems, Boris A. Hoffmann

University Heart Center, University Hospital Hamburg-Eppendorf, Department of
Cardiology – Electrophysiology
Martinistraße 52, 20246 Hamburg, Germany
e-mails: willems@uke.uni-hamburg.de; b.hoffmann@uke.de

Abstract: We propose a new algorithm for the detection of ventricular fibrillation (VF) in very short surface electrocardiogram (ECG) episodes. Ventricular fibrillation is the most commonly identified arrhythmia in cardiac arrest patients and can lead to syncope, within seconds. The fast detection of ventricular fibrillation is necessary for prompt defibrillation either with an implantable cardioverter/defibrillator or an automated external defibrillator. Ventricular fibrillation generates stochastic waveforms and recently it has been shown that it exhibits characteristics similar to a non-chaotic signal and contains determinism Probability Density Function (PDF), for the different physical fluctuations was described previously. Accordingly, we describe scaling properties of very short shockable, VF and non-shockable ECG episodes and show that a universal PDF exists for the fluctuations of shockable ECG episodes. We compared the proposed algorithm with nine standard VF detection algorithms. The comparison indicated that our algorithm consistently produced more accurate detection results, then with standard algorithm. We conclude that the proposed method, based on fluctuation analysis, provides new information on the dynamics underlying VF, and allows a better detection compared to other algorithms.

Keywords: electrocardiogram; universal fluctuations; ventricular fibrillation; defibrillation; probability density function

1   Introduction

The ventricular myocardium is a dynamic system which can appear synchronized in sinus rhythm or irregular, anomalous in ventricular fibrillation. [1]. The ventricular fibrillation, (VF) is the most common cause of sudden cardiac death [2]. In the non-appearance of defibrillation, VF leads to death within a few minutes, therefore, accurate and timely detection of this arrhythmia is of great clinical importance.

An implantable cardioverter/defibrillator (ICD) or an automated external defibrillator (AED) should reliably decide between *shockable* (e.g. VF)-, and *non-shockable* ECG episodes. Shockable episodes demand the urgent response of electrical defibrillation. The decision is based on different VF detection algorithms, which in the embedded, real-time software environment should reveal sufficient performance. Widely proposed methods for this level are: complexity measure, peak analysis, threshold cutting and narrow band filter algorithms [3].

Historically, it has been accepted that VF generate stochastic ECG waveforms [4]. Using power spectral density, Kaplan et al. [5] have demonstrated, that the VF waveform exhibits characteristic similar to a non-chaotic signal. Most recent studies show that shockable ECG episodes own 80-90% determinism [6].

In this contribution the properties of the very short ECG episodes (4 sec duration, 1000 Hz sampling rate, 12-bit amplitude resolution) are examined to the general attributes of a critical system presented by Bramwell et al. [7]. The authors demonstrated a PDF (Bramwell-Holdsworth-Pinton, BHP PDF) exists for the fluctuations in different physical systems (magnetism, Danube water level [7]). The BHP PDF is always in systems which lack a characteristic time scale. The underlying common thread in all those applications is rooted in the attempt to quantify the fluctuations of different temporal patterns.

2   Method

2.1   Probability Density Function (PDF)

The universal PDF for the very short episodes of ECG recordings *y* can be determined by using the distribution of $h = \dfrac{y - \bar{y}}{\sigma_y}$ in the thermodynamic limit (in the infinite system size), where $\bar{y}$ is the mean and σ_y the standard deviation.

In statistics, the characteristic function $\varphi(t)$ of any variable completely defines its probability distribution. If a variable admits a PDF, then $\varphi(t)$ is the inverse Fourier transform of the PDF [7].

The intensity of the correlations within an ECG signal depends on the degrees of irregularity of the miscellaneous topological rhythm statements, and on the duration of the episodes. In a very short ECG signal the central limit theorem breaks down, consequently the limit function can be non-Gaussian.

Let the sum of the ECG recording $Y = \sum_n y_n$ and each measured ECG value y_n be a gamma variable, with known PDF:

$$PDF(y_n) = \frac{n}{\Gamma(\gamma)} e^{-ny_n} y_n^{\gamma-1} \tag{1}$$

where $\Gamma()$ is the gamma-function, $\gamma = \frac{1}{2}$, and n is the size of the episodes.

The logarithm of the characteristic function $\varphi_n(t)$ of the y_n could be expanded as a series of statistical moments $k_r(y_n)$.

$$\log \varphi_n(t) = \sum_1^\infty \frac{(it)^r}{r!} k_r(y_n) \tag{2}$$

Where the rth statistical moment $k_r(y_n) = \gamma(r-1)! n^{-r}$.

The gamma variables are statistically independent [7], therefore the characteristic function $\Phi(t)$ of Y is the product of the $\varphi(t)$:

$$\Phi(t) = \prod_n \varphi_n(t) \tag{3}$$

Normalizing Y by its standard deviation σ_y, the rth statistical moment

$$k_r\left(\frac{Y}{\sigma_y}\right) = \frac{\frac{1}{2}(r-1)! \sum_n \left(\frac{1}{n}\right)^r}{\left(\frac{1}{2} \sum_n \left(\frac{1}{n}\right)^2\right)^{\frac{r}{2}}} \tag{4}$$

Finally, the PDF of the very short episodes of ECG is the inverse Fourier transformation of the series [4].

3 Results and Discussion

3.1 Probability Density Function

Figure 1 shows the PDF of the fluctuations, $h = \dfrac{y - \overline{y}}{\sigma_y}$ as found from the inverse Fourier transform of (4) and the histograms of the fluctuations applying 2400 short episodes of ECG signals with shockable arrhythmias and 2400 short episodes with non-shockable arrhythmias.
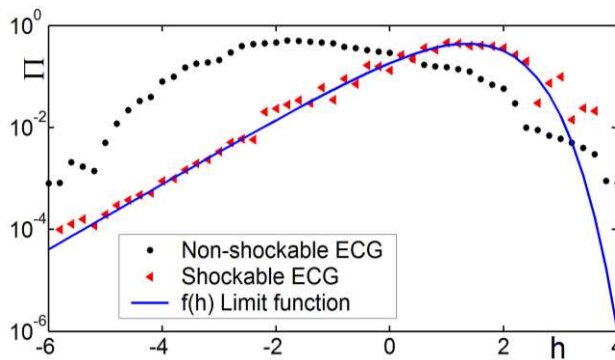

Figure 1

Comparison, without fitting parameters, between the histogram $\sigma P(h)$ of the short episodes of shockable fluctuations (circle symbols), non-shockable fluctuations (triangle symbols) and the universal PDF (line). The data is plotted as $\Pi = \sigma P(h)$ against $h = \dfrac{y - \overline{y}}{\sigma_y}$.

In the case of the shockable episodes, an excellent agreement can be seen with the universal PDF *f(h)*, however Figure 1 contains no fitting parameters. *f(h)* is asymmetric, with the position for fluctuations above the mean exponentially variations, yielding a probability for a positive fluctuation that might be larger than by a normal distribution.

For the non-shockable arrhythmias, the fluctuations curve has a Gaussian characteristic; hence, the normal sinus rhythm system cannot be approximated as a non-linear system.

The specific features of shockable signals are a consequence of strong correlations and self-similarity and do not share the same universal class as the non-shockable ones. These results imply that the degree of organization in short episodes of VF is low, which may explain the lack of convergence of fractal dimension of the VF reported in the literature [4].

To examine the linearity of the episodes, the ECG signals were analyzed with power spectral density (PSD). The PSD is the square of the coefficients in a Fourier series representation and measures the average variation of the arrhythmia at different frequencies. If the adjacent ECG points are uncorrelated, then the power spectrum will be constant, as a function of the normalized frequency. If the adjacent ECG points are correlated relative to points far apart the power spectrum will be large at small frequencies and small at large frequencies.

The PSD response of a non-shockable signal can be seen in Figure 2. The slope of the least squares fit shows clear signs of a weakly correlated origin.

Figure 2

Left Panel: Typical sinus rhythm ECG. Right Panel: Power spectral density with least-squares fit. The fractal dimension of the 2400 signals $\in$ [1.03-1.24].

Figure 3 depicts the typical PSD of the shockable ECG recordings. The relationship between the spectral power of the signal shows fractal power-law scaling similar to a fractal $1/f^m$ framework, which is directly related to non-linearity of the input arrhythmia scale distribution and suggests a lower dynamic complexity for the ECG activity.

The signal shows fractal power - law scaling similar to a fractal structure $1/f^m$.
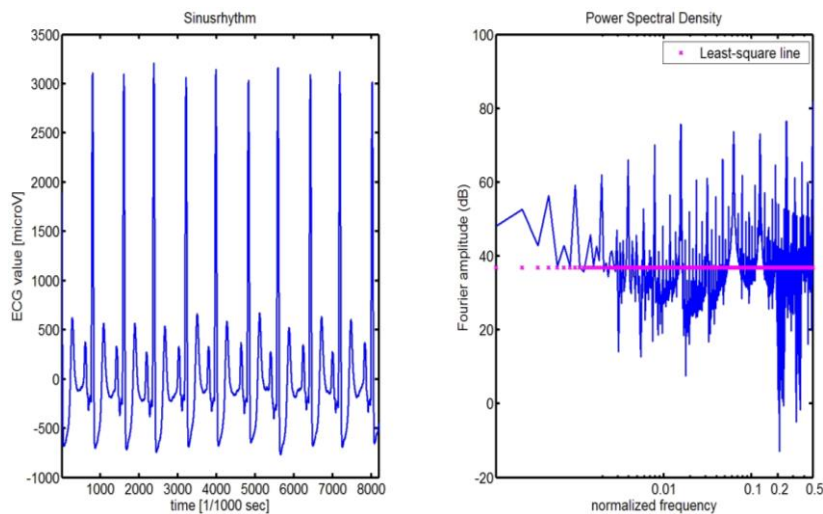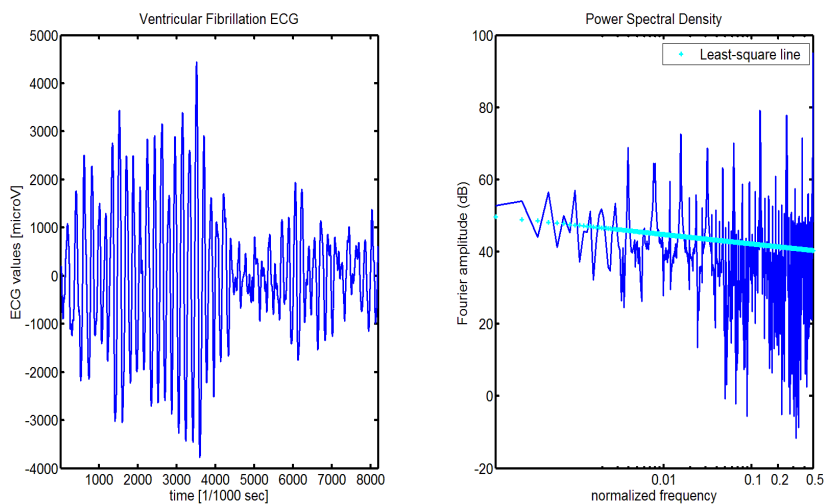
Figure 3

Left Panel: Ventricular fibrillation ECG. Right Panel: Power spectral density with least-squares fit. The fractal dimension of the 2400 signals $\in$ [1.19-1.46].

3.2   Scaling of Higher Moments

In the very short episodes of ECG it is expected that the standard deviation σ_y and the mean value of the order parameter *<m>* scale with system size, so that a rescaled PDF is independent of the episode size.

The order parameter of ECG is a measure of the degree of order in a system; it ranges between zero for shockable arrhythmias and the saturation value for non-shockable ones. This can be expressed as:

$$<m> \propto \sigma_y \propto t^{\alpha} \tag{5}$$

where α is the scaling exponent [13, 14]

The scaling exponents can be used as a function of *n*. The 2400 short episodes of ECG signals with shockable arrhythmias approximately fulfil the Eq.(5) relationship, while the 2400 short episodes with non-shockable arrhythmias show divergences from multiscaling [15].

The multiscaling function is shown in Fig. 4.
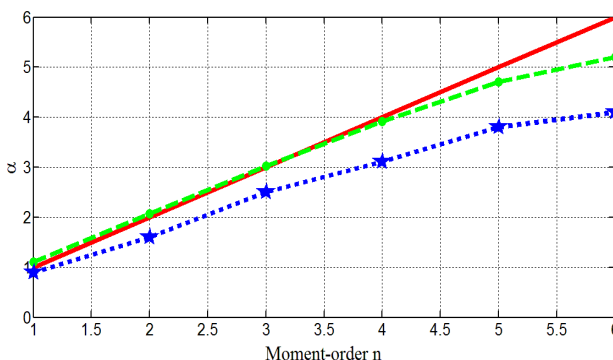
Figure 4

Multiscaling function Eq. 5. Simple scaling – [solid line]; the short episodes of ECG signals with shockable arrhythmias [o] $\alpha(n) = 1.0899 - 0.04955n + 0.00271n^2$ and short episodes with non-shockable arrhythmias [*] $\alpha(n) = 0.91244 - 0.02618n$.

We have investigated the skewness $\dfrac{\langle(m - \langle m \rangle)^3\rangle}{\sigma^3}$ and kurtosis $\dfrac{\langle(m - \langle m \rangle)^4\rangle}{\sigma^4} - 3$ of the ECG episodes.

Theoretically, the skewness and kurtosis might scale with the size of episode [7]. Nevertheless, both the short episodes of ECG signals with shockable and non-shockable arrhythmias, skewness and the kurtosis, as functions of the size, do not adapt to a power-law relationship (Fig. 5) nor is there a multiscale relationship: $kurtosis \propto skewness^{2.08}$ for the shockable and $kurtosis \propto skewness^{1.74}$ for the non-shockable episodes. The ECG episodes change to multiscaling and the skewness and kurtosis do not scale with the size of the epsiode.

Figure 5

Skewness [o] and kurtosis [*] of the short episodes of ECG signals with shockable arrhythmias, skewness [□] and kurtosis [Δ] of the short episodes of ECG signals with non-shockable arrhythmias

3.3 Comparison with Standard Algorithm

To gain insight into the quality of our algorithm it is essential to compare the results with other VF detection algorithms under equal conditions, with very short ECG episodes (4 sec duratio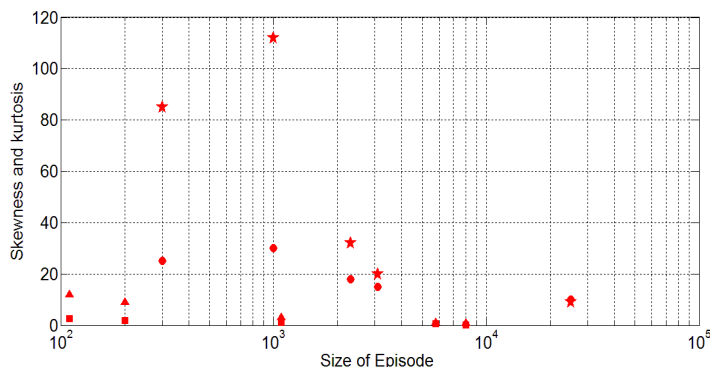n, 1000 Hz sampling rate, 12-bit amplitude resolution). The following methods have been considered, for details refer the noted literature:

- Threshold crossing algorithm (TCIA) [3]

- VF filter algorithm (VFFA) [9]

- Spectral algorithm (SPECA) [10]

- Complexity measure algorithm (CPLXA) [11]

- Standard exponential algorithm (STEA) [3]

- Signal comparison algorithm (SCA) [3]

- Wavelet based algorithm (WBA) [3]

- Li algorithm (LIA) [12]

- Pan-Tompkins QRS based algorithm (QRSA) [15]

Table 1 shows the values for the sensitivity and the specifity of the algorithms using annotated databases: MIT-BIH malignant ventricular arrhythmia database (Massachusetts Institute of Technology (MIT), Beth Israel Hospital (BIH)), Creighton University ventricular tachyarrhytmia database (CU) and American Heart Association database (AHA).

Table 1

Quality of VF detection algorithms. Specificity (SP), Sensitivity (SN) in %

Algorithm	MIT BIH		CU		AHA	
	SP	SE	SP	SE	SP	SE
TCIA	68.4	73.2	49.8	66.1	74.3	70.9
VFFA	86.2	25.9	82.8	19.1	80.6	12.7
SPECA	83.5	30.7	78.3	24.2	76.1	23.9
CPLXA	88.6	17.2	72.7	20.5	82.6	59.4
STEA	77.1	62.8	59.4	54.2	43.1	64.9
SCA	91.3	63.4	88.7	57.1	88.4	71.5
LIA	92.9	16.7	95.7	8.3	87.2	40.4
WBA	94.5	68.2	91.8	54.7	93.3	49.9
QRSA	33.2	50.8	51.5	63.1	48.3	85.2
BHPPDFA	97.3	98.6	96.2	99.4	93.9	97.3

MIT BIH: MIT-BIH malignant ventricular arrhythmia database; CU: Creighton University ventricular tachyarrhythmia database; AHA: American Heart Association database; SE: sensitivity; SP: specifity; VF: ventricular fibrillation; BHPPDFA: Bramwell-Holdsworth-Pinton probability density function.

For the very short episodes the PSD algorithm yields the best value both the sensitivity and specifity, followed by the algorithms SCA and VFFA. All other algorithms present heterogeneous simulations. We can interpret here, that algorithms for QRS detection or those developed in the spectral domain are not adequate for very short episodes.

Conclusions

In this work fluctuation universality of very short ECG episodes was analyzed. The Universal Probability Density (PDF) function is the inverse Fourier transform of the statistical moments of the fluctuations within the thermodynamic limit.

For the non-shockable arrhythmias, e.g. sinus rhythm, the function has Gaussian characteristics, while in the case of shockable, the VF shows an excellent agreement with the BHP PDF, without fitting parameters. The BHP PDF is very well suited to classify the fluctuations of very short ECG episodes.

Our results show, that fluctuation analysis provides new information for the dynamics underlying VF, and allows for a better detection as compared to other algorithms.

References

[1]   Josephson, M. E. Clinical Cardiac Electrophysiology: Techniques and Interpretation. Philadelphia: Lea & Febiger, 1993

[2]   Wang, P. J. Ventricular Arrhythmias and Sudden Cardiac Death. Malden, Mass.; Oxford: Blackwell Futura, 2008

[3]   Amann, A., K. Rheinberger, and U. Achleitner. Algorithms to Analyze Ventricular Fibrillation Signals. Curr Opin Crit Care 7(3), 152-156, 2001

[4]   Garfinkel, A., M. L. Spano, W. L. Ditto, and J. N. Weiss. Controlling cardiac chaos. Science 257(5074), 1230-1235, 1992

[5]   Kaplan, D. T., and R. J. Cohen. Is Fibrillation Chaos? Circ Res 67(4), 886-892, 1990

[6]   Yu, D., M. Small, R. G. Harrison, C. Robertson, G. Clegg, M. Holzer, and F. Sterz. Measuring Temporal Complexity of Ventricular Fibrillation. Physics Letters A 265(1-2), 68-75, 2000

[7]   Bramwell, S. T., K. Christensen, J. Fortin, P. C. Holdsworth, H. J. Jensen, S. Lise, J. M. Lopez, M. Nicodemi, J. Pinton, and M. Sellitto. Universal Fluctuations in Correlated Systems. Phys Rev Lett 84(17), 3744-3747, 2000

[8]   Bohdalová, M., and M. Gregus. Fractal Analysis of Forward Exchange Rates. Acta Polytechnica Hungarica Vol. 7, No. 4. 2010

[9]   Thakor, N. V., Y. S. Zhu, and K. Y. Pan. Ventricular Tachycardia and Fibrillation Detection by a Sequential Hypothesis Testing Algorithm. IEEE Trans Biomed Eng 37(9), 837-843, 1990

[10]   Kuo, S., and R. Dillman. Computer Detection of Ventricular Fibrillation. In: Computers in cardiology Long Beach, California: IEEE Computer Society, 1978, pp. 347-349

[11]   Barro, S., R. Ruiz, D. Cabello, and J. Mira. Algorithmic Sequential Decision-Making in the Frequency Domain for Life Threatening Ventricular Arrhythmias and Imitative Artefacts: a Diagnostic System. J Biomed Eng 11(4), 320-328, 1989

[12]   Zhang, X. S., Y. S. Zhu, N. V. Thakor, and Z. Z. Wang. Detecting Ventricular Tachycardia and Fibrillation by Complexity Measure. IEEE Trans Biomed Eng 46(5), 548-555, 1999

[13]   Li, C., C. Zheng, and C. Tai. Detection of ECG Characteristic Points Using Wavelet Transforms. IEEE Transactions on Biomedical Engineering 42(1), 21-28, 1995

[14]   J. Cardy. Current Physics Sources and Comments, Vol. 2. Finite-Size Scaling. Nort-Holland, Amsterdam 1990

[15]   K. Dahlstedt and H. J. Jensen. Universal Fluctuations and Extreme Value Statistics. Journal of Physics. A. Math. Gen. 34:11193-11200, 2001

[16]   H. J. Jensen. Self Organized Criticality. Cambridge University Press. New York 1998

Contingency-constrained Congestion Management and Transmission Cost Allocation

Oana Pop, Constantin Barbulescu, Stefan Kilyeni, Antheia Deacu

Politehnica University Timisoara, 2 Bd. V. Parvan, Timisoara, Romania,
oana.pop@upt.ro, constantin.barbulescu@upt.ro, stefan.kilyeni@upt.ro

Abstract: Open access to the transmission network represents one of the important tasks for transmission and system operators. Moreover, increases of transactions in markets may lead to transmission network congestion. The approaches proposed for transmission congestion management are: generated power re-dispatching, generators outside the congested area dispatching, or consumed power mitigation for specific buses where it is possible. Transmission costs are able to be computed once the congestion has been solved and it is allocated to generators and consumers, using various methods. This paper analyzes congestion occurrence in case of N-1 contingencies and the allocation of transmission costs, using pro-rata and Bialek methods. The case study is performed for an existing large-scale system – the Southern-Western side of the Romanian Power System. The results are relevant to for the Romanian TSO (Transmission System Operator) – Romanian Power Grid Company Transelectrica.

Keywords: congestion management; transmission system; N-1 criteria; cost allocation

1 Introduction

Currently, electric power systems around the world are subject to radical changes, moving from a monopolistic and regulated industry to a new model characterized by competition and open access to the transmission network. These facts are responsible for an increased number of transactions between market participants, under the optimal use of transmission facilities [1].

An important number of power transactions have been anticipated in the context of increased competition [2], [3] and this may lead to an overloading of transmission lines. These overloads will cause the system to exceed allowable thermal limits, stability and voltage limits, which can affect power transmission equipment.

These limits are power flow constraints and they depend on the power system operating condition at any given moment [4], [5]. Contingency limits ensure that no other system element is overloaded, once another one is disconnected. Thus, *N-1* and *N-2* criteria must be satisfied. These constraints have to be fulfilled by the transmission network operator to ensure safety and power balance in the system.

The problem of congestion also has an economic dimension [6], [7]. If there are no overloads within the system, then marginal costs corresponding to buses have very close values. The differences occur due to transmission losses. In this case, the generated power will be distributed depending on the offered price (classic power flow optimization). The congestion occurrence will lead to an important growth of the marginal costs and the generated power distribution will depend not only on the offered price, but also on the congestion "cost". The system operator will act to eliminate the congestion. Congestion management methods are generally divided into two main groups [28], [29]: preventive and corrective methods. The latter are used to remove the occurred congestion by generated power re-dispatching or consumed power mitigation.

A sensitive issue related to power transmission open access is represented by the transmission cost allocation [8], [9]. The use of transmission network by market participants is associated with transmission pricing, including: operating and capital costs, congestion costs, ancillary service costs, loss compensation costs, balancing system costs and stranded costs. Currently, there are numerous studies that have synthesized transmission-pricing issues with many internationally accepted options and practices. Postage stamp pricing [10] has the simplest design, and hence is the most common in immature power markets. In postage stamp pricing, all points are equivalent in terms of connection and use of the system network. The postage stamp area is generally a country or a controlled area.

Reference [11] and [12] present the power flow based on the MW-mile method. It was firstly proposed by Shirmohammadi. The MW-mile method is an embedded cost method. It computes transaction charges based on the transmission capacity use as a function of transacted power magnitude, the path followed and the distance travelled by transacted power. The MVA-mile method includes charging for reactive power, in addition to the charging for real power. The pro-rata method, presented in [13], allocates costs to generators and consumers according to the sum of real produced power and/or consumed by each generator and/or consumer.

The proportional sharing principle-based methods use Kirchhoff's Laws. These are known as tracing methods, the main versions of which are the Bialek and Kirschen methods. The Bialek tracing method computes the real and reactive generated power percentage supplied to a particular consumer. It includes two algorithms: upstream-looking algorithm (generation-load) or downstream-looking algorithm (load-generation) [14]-[17]. In the case of the first algorithm, the costs for the transmission network usage are allocated to individual generators and real power losses are allocated to consumers. The downstream-looking algorithm allocates the transmission network using costs to individual consumers, and real power losses are allocated to generators. The Kirschen method [18], [19] organizes the network buses and branches in homogeneous groups according to the following concepts: domain of generator, commons and links.

The distribution factor methods are power flow based. Distribution factors are used to determine the impact of generation and load on transmission power flow.

Generally, generation distribution factors have been used in security and contingency analyses. The traditional version of this method was proposed by Ng in [20]. It was extended to AC power flow, being able to evaluate the real and reactive power flow [21]. Allocation method based on the equivalent bilateral exchanges (EBE) [22] does not depend on the slack bus selection. Also, it offers the counter-flow acceptance or exclusion option and network elements dependence. In recent years, methods based on system matrices $\left|\underline{Y}_n\right|$ or $\left|\underline{Z}_n\right|$ have received great attention, since these methods are able to integrate the network characteristics and circuit theories into real and reactive power and transmission losses allocation [24]-[26].

This paper proposes the congestion analysis in case of *N-1* contingencies. The case study is performed on a large-scale power system, including the Western, South-Western and North-Western parts of the Romanian Power System. The necessary measures to eliminate the congestions are indicated. The authors use the Bialek method and pro-rata method to determine the usage costs of each branch allocated to generators and consumers.

The paper is organized as follows. The second section outlines the implemented transmission cost allocation methods. Section three is dedicated to the case study. The occurrence of congestion is presented along with suggested measures to eliminate them. Two operating conditions are considered: the first containing the congestion, and the last with the congestion solved. Using the methods presented within the previous sections, the network usage results are discussed for both operating conditions. The usage cost for generators and consumers are computed using Romanian OTS transmission tariffs. Conclusions are synthesized in Section four.

2   Transmission Cost Allocation Method

Over recent years different proposals have appeared pertaining to allocation and transmission usage costs. This section presents two transmission cost allocation approaches: the Bialek and the pro-rata methods. In both cases it is necessary to decide how much of the cost should be assigned to generators and how much to consumers. For example, in Romania, the share of the transmission cost for system usage is as follows: 20.69% to generators and 79.31% to consumers.

2.1   The Bialek Method

Consider the *i-j* branch connecting the sending bus, *i,* with the receiving bus, *j*. Both buses are connected to the rest of the system. P_{ij} represents the real *i-j* power flow. The *i-j* branch loss is $\Delta P_{ij} = |P_{ji}| - |P_{ij}|$. The gross power is defined as the sum between the consumed power and the part allocated from total transmission losses. The gross real power flow through bus *i*, P_i^b is expressed as [23]:

$$P_i^b = \sum_{j \in N_i} \left| P_{ij}^b \right| + P_{gi} \; ; \quad i \in N \tag{1}$$

where N_i is the subset of buses supplying directly bus i, P_{ij}^b is the gross real branch flow through the network elements i-j and P_{gi} is the real generated power in bus i.

The term $\left| P_{ij}^b \right|$ can be replaced by $(P_{ij}^b / P_j^b) P_j^b$. Considering the real power losses, relation (1) becomes:

$$P_i^b - \sum_{i \in N_i} \left(\frac{\left| P_{ji} \right| - \Delta P_{ij}}{P_j} \cdot P_j^b \right) = P_{gi} \tag{2}$$

where P_{ji} is the real power flow through the i-j network elements and P_j is the real power injected in bus j.

The matrix form of relation (2) is:

$$P_g = A^{\text{g-c}} \cdot P^b \tag{3}$$

where P^b – is the gross bus flow matrix, P_g – is the bus generation matrix, and A^{g-c} is the upstream distribution matrix, having its elements defined by relation:

$$a_{ij}^{\text{g-c}} = \begin{cases} 1 & \text{if} \quad i = j \\ -\dfrac{\left| P_{ji} \right| - \Delta P_{ij}}{P_j} & \text{if} \quad i > j \, , \; j \in N_i \\ 0 & \text{if} \quad i > j \, , \; j \notin N_i \; \text{or} \; i < j \end{cases} \tag{4}$$

Considering relation (3) P_i^b yields for each bus:

$$P_i^b = \sum_{k \in N} \left(ai_{ik}^{\text{g-c}} \cdot P_{gk} \right) ; \quad i \in N \tag{5}$$

The usage of any branch l allocated to the generator at bus k can be expressed as:

$$U_l^{g-c} = \frac{1}{5} \frac{\left| P_{ij} \right|}{P_j} \cdot ai_{ik}^{\text{g-c}} \cdot P_{gk} \tag{6}$$

The net power is defined as the difference between the generated power and the part allocated from total transmission losses. The net real power flow through bus i, P_i^n, can be expressed as:

$$P_i^n = \sum_{j \in N_i} \left| P_{ij}^n \right| + P_{ci} \; ; \quad i \in N \tag{7}$$

where N_i is the subset of buses supplying bus i directly, P_{ij}^n is the net real branch flow through the network elements i-j; P_{ci} is real consumed power at bus i.

As an analogy with the previous algorithm, the relation (7) can be written as:

$$P_i^n - \sum_{i \in N_i} \left(\frac{|P_{ji}| - \Delta P_{ij}}{P_j} \cdot P_j^b \right) = P_{ci} \tag{8}$$

In matrix form this means:

$$P_c = A^{c-g} \cdot P^n \tag{9}$$

where P^n is the net bus flow array, P_c is the consumed power array and $A^{c\text{-}g}$ is the downstream distribution matrix, its elements are defined as follows:

$$a_{ij}^{c\text{-}g} = \begin{cases} 1 & \text{if } i = j \\ -\dfrac{|P_{ji}| - \Delta P_{ij}}{P_j} & \text{if } i < j \, , \, j \in N_i \\ 0 & \text{if } i < j \, , \, j \notin N_i \text{ or } i > j \end{cases} \tag{10}$$

From relation (9), P_i^n results for each bus:

$$P_i^n = \sum_{k \in N} \left(a i_{ik}^{c\text{-}g} \cdot P_{ck} \right); \quad i \in N \tag{11}$$

The usage of any branch l allocated to the generator at bus k can be expressed as:

$$U_l^{c-g} = \frac{4}{5} \frac{|P_{ij}|}{P_j} \cdot a i_{ik}^{c\text{-}g} \cdot P_{ck} \tag{12}$$

The cost of branch k allocated to generators and consumers at bus i and j is:

$$C_k^G = \sum_{i \in G} c_{gi} \cdot U_l^{g-c}, \quad C_k^D = \sum_{j \in D} c_{cj} \cdot U_l^{c-g} \tag{13}$$

where c_{gi} – transmission costs for injected power in bus i [\$/MWh]; c_{cj} – transmission cost for extracted power in bus j [\$/MWh]

2.2   Pro-Rata Method

Interconnected network users adopt the pro-rata method [15]. Transmission costs are allocated proportionally to the power injected by each generator or by each consumer. The branch k network usage allocated to generator i or consumer j is determined by the following relations:

$$UG_{ik} = \frac{1}{5} \frac{P_{gi}}{\sum\limits_{i \in G} P_{gi}} UG_k \,, \quad UD_{jk} = \frac{4}{5} \frac{P_{cj}}{\sum\limits_{j \in D} P_{cj}} UD_k \tag{14}$$

where P_{gi} is the generated power in bus i, P_{cj} is the consumed power in bus j, G is the subset of PV buses and C is subset of PQ buses.

The cost of branch k allocation to generators and consumers at bus i and j is:

$$C_k^G = \sum_{i \in G} c_{gi} \cdot UG_{ik}, \quad C_k^D = \sum_{j \in D} c_{cj} \cdot UD_{jk} \tag{15}$$

3   Case Study

The case study is performed for the Western and South-Western side of the Romanian Power System [28], [29], which has 88 buses and 107 branches. The 35 PV buses are divided into 17 real generating units and 18 equivalent PV buses, obtained by extracting the analyzed part from the Romanian Power System. The system has 42 PQ buses. The buses at medium voltage (real generating groups), 220 kV, 400 kV are represented. At 110 kV level, only the generated and consumed powers are considered.

The operating condition presented in Figure 1 corresponds to the disconnection of the 400/220 kV Rosiori autotransformer [28]. On the 220 kV 28087-28093 overhead line (OHL) there is an overload of 126%. An inadmissible voltage level is recorded for 28094, 28095 and 28093 buses area (182.6 kV, 184.6 kV and 185.4 kV). The hourly cost of the power system is 121478 $/h and the penalty cost is 47363 $/h. As was expected, the maximum value of nodal marginal princes is recorded for 28095 PQ bus (1760 $/MWh) and the lowest value is registered for bus 29102 (31 $/MWh). Other areas with nodal marginal price high values are mentioned in the cases of buses 28093 (1601 $/MWh, 1643 $/MWh, 1663 $/MWh), 28094 (1630 $/MWh) and 28095 (1692 $/MWh).

The power consumption in 28093 and 28095 buses area has been reduced by the TSO: 6 MW in 28484, 28485, 28093 and 28491 buses in order to eliminate congestion. The hourly cost value is 116299.61 $/h. All local marginal prices are around 30 $/MWh.

Furthermore, the generated power at 29169, 29260, 29262, 28036, 29159, 29160, 29119, 29121 and 29238 bus groups has been reduced. In the case of the 29189-29193, 29250, 29232, 29233 and 29162 PV buses, the power has been increased. Another category of groups has the same value (28795, 28709, 28719, 28756, 28562, 29232 and 29233 buses). Figure 2 presents the generated power re-dispatching mechanism.

Mathematica® environment is used to compute the transmission costs allocated to generators and to consumers for both operating conditions. The software tool *Tracing for Real and Reactive Power* (*TAPQ)* has been developed by the authors [29] and is linked with PowerWorld software. The database containing the power system
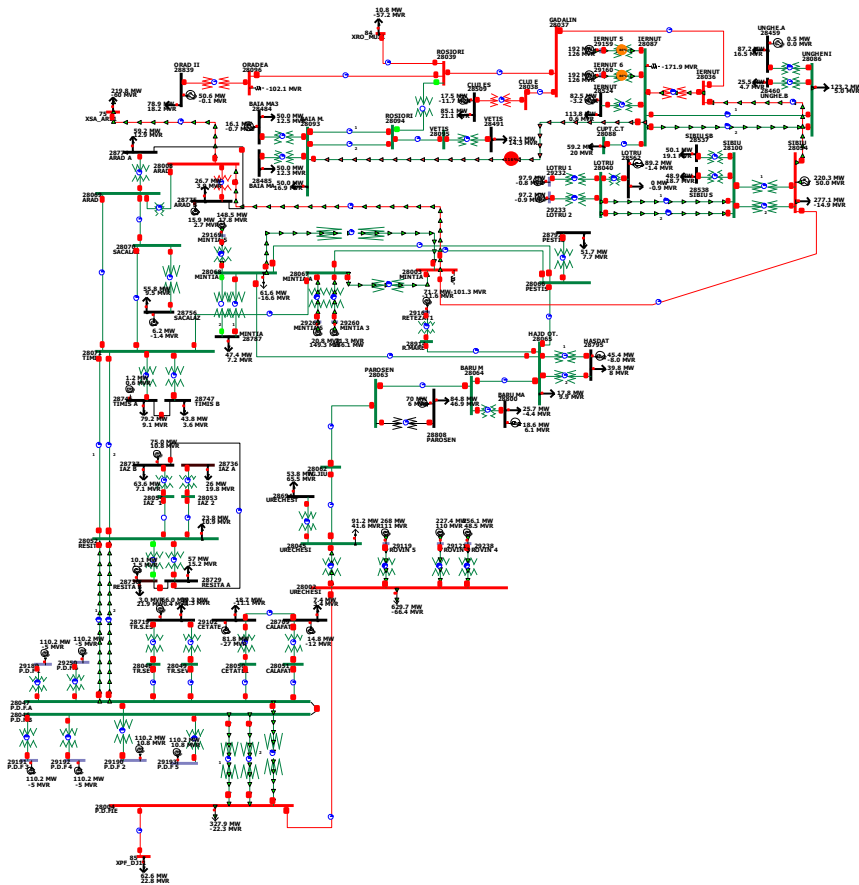
Figure 1
Operating condition with congestion on 28087-28093 OHL

topology and parameters is extracted from Powerworld software and transmission cost allocation methods are launched. In instances of pro-rata method, network usage allocated to generators and consumers and transmission cost allocation are going to be computed. The Bialek method-computing algorithm contains the following steps:

1. Extract the P_{ji}, real power flow and P_j, real injected power;
2. Upstream distribution matrix elements and matrix form determination;
3. Gross bus flows array obtaining;
4. Network usage computing for all branches using relation (6), in case of generators;
5. Downstream distribution matrix elements and matrix form determination;
6. Net bus flow array obtaining;
7. Network usage computing for all branches using relation (12), in case of consumers;
8. Transmission cost allocated to generators and consumers.

The upstream and downstream matrices dimension is 88 lines and columns, in case of the analyzed operating conditions.
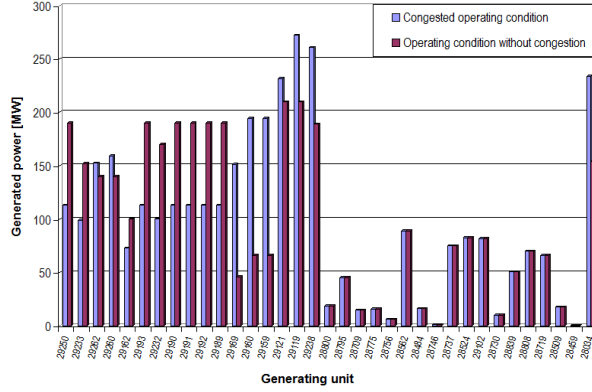
Figure 2
Generated power re-dispatching

Only a computing synthesis is going to be presented in the following. A case study power system detail is shown in Figure 3, which is used to clarify the pro-rata and Bialek methods. This figure provides additional generators, loads and power flow data.
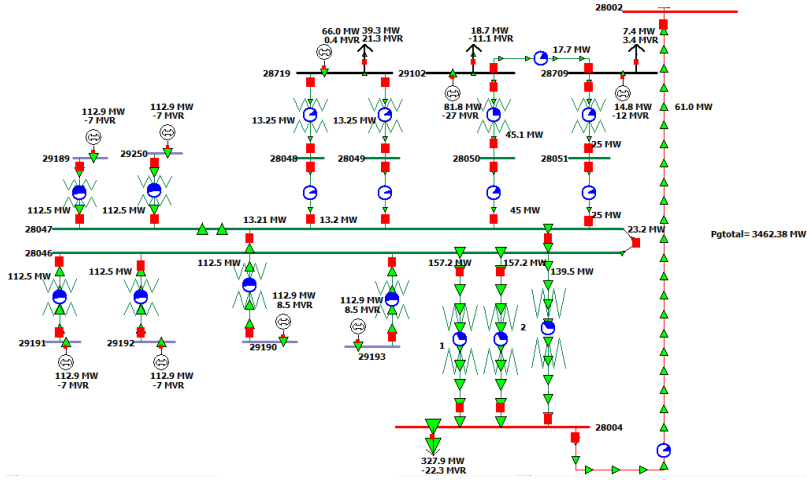
Figure 3
Power system detail. Congestion on 28087-28093 OHL

With the pro-rata method, OHL network usage 28002-28004, allocated to generator 29189 and load bus 28002 is determined as follows:

$$UG_{29189(branch\,28002\text{-}28004)} = \frac{1}{5} \cdot \frac{P_{g\,29189}}{\sum\limits_{i \in G} P_{gi}} UG_{28002\text{-}28004} = \frac{1}{5} \cdot \frac{112.926}{3462.37} \cdot 61 = 0.3979 \; MW \; ,$$

$$UD_{28002(branch\,28002\text{-}28004)} = \frac{4}{5} \cdot \frac{P_{c28002}}{\sum\limits_{j \in D} P_{cj}} UD_{28002\text{-}28004} = \frac{4}{5} \cdot \frac{629.7}{3398.1} \cdot 61 = 0.0431 \; MW$$

The Bialek method upstream distribution matrix elements are determined as (relation 4):

$$a^{\text{g-c}}_{28046\text{-}29191} = -\frac{112.5}{112.9} = -0.9965$$

$$a^{\text{g-c}}_{29191\text{-}28046} = -\frac{112.5}{112.5+112.5} = -0.5000$$

$$a^{\text{g-c}}_{28004\text{-}28047} = -\frac{139.5}{23.2+25+45+13.2+13.2+112.5+112.5+112.5} = -0.2461$$

$$a^{\text{g-c}}_{28004\text{-}28046} = -\frac{157.2+157.2}{112.5+112.5+112.5} = -0.9316$$

$$a^{\text{g-c}}_{29191\text{-}28047} = 0$$

$$a^{\text{g-c}}_{29191\text{-}29191} = 1$$

$$a^{\text{g-c}}_{28002\text{-}28004} = -\frac{61}{157.2+157.2+139.5} = -0.1344$$

Any branch usage allocated to generator and consumer can be determined in two ways: by applying relations (6) and (12) or by directly applying a proportional sharing principle. In the following, the case of proportional sharing principle is presented. As shown in Figure 4, bus 1 is connected to upstream buses 2 and 3 and downstream buses 4 and 5, by four branches. Bus 1 real power flow is denoted by $P_{2\text{-}1}$ and $P_{3\text{-}1}$, respectively, while real power flowing out of bus 1 is denoted by $P_{1\text{-}4}$ and $P_{1\text{-}5}$.
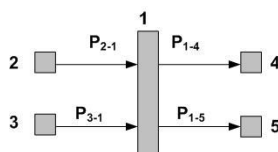

Figure 4
Proportional sharing principle

Buses 2, 3, 4 and 5 can be either PV or PQ buses, supplying or being supplied from bus *1*. According to the proportional sharing principle, inflows are shared proportionally between the outflows in following manner:

- $P_{1\text{-}4}$ is determined by two components:

$$\frac{P_{2-1}}{P_{2-1}+P_{3-1}} \cdot P_{1-4} \text{ coming from } P_{2\text{-}1} \text{ ; } \frac{P_{3-1}}{P_{2-1}+P_{3-1}} \cdot P_{1-4} \text{ coming from } P_{3\text{-}1};$$

- $P_{1\text{-}5}$ is determined by two components:

$$\frac{P_{2-1}}{P_{2-1}+P_{3-1}} \cdot P_{1-5} \text{ coming from } P_{2\text{-}1} \text{ ; } \frac{P_{3-1}}{P_{2-1}+P_{3-1}} \cdot P_{1-5} \text{ coming from } P_{3\text{-}1}.$$

Following power flow direction, the generator and consumer contributions computing process starts from bus 29102. Generator contribution from bus 28719 on branches

28719-28048 ($P_{\text{branch}(28719\text{-}28048)}^{G28719}$), 29102-28049 ($P_{\text{branch}(29102\text{-}28049)}^{G28719}$), 28048-28047 ($P_{\text{branch}(28048\text{-}28047)}^{G28719}$) and 28049-28047 ($P_{\text{branch}(28049\text{-}28047)}^{G28719}$) respectively is the same: 13.25 MW. As shown in Figure 3, generator 29102 uses branches 29102-28050, 28050-28047 and 29102-28709. Its contributions are 45 MW, 45 MW and 17.7 MW.

In the following, the contributions of generators 29102 and 28709 on branch 28709-28051, and bus local consumer 28709 are shown.
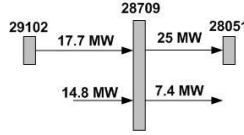

Figure 5
Bus 28709

The contribution of generator 29102 ($P_{\text{branch}(28709\text{-}28051)}^{G29102}$) and consumer 28709 ($P_{\text{branch}(28709\text{-}28051)}^{C28709}$) on branch 28709-28051 is determined as:

$$P_{\text{branch}(28709\text{-}28051)}^{G29102} = \frac{P_{\text{branch}(29102\text{-}28709)}}{P_{\text{branch}(29102\text{-}28709)} + P_{\text{G}_{29102}}} \cdot P_{\text{branch}(28709\text{-}28051)} =$$

$$= \frac{17.7}{14.8 + 17.7} \cdot 25 = 13.6154\,MW$$

$$P_{\text{branch}(28709\text{-}28051)}^{C28709} = \frac{P_{\text{G}_{29108}}}{P_{\text{branch}(29102\text{-}28709)} + P_{\text{G}_{29108}}} \cdot P_{\text{branch}(28709\text{-}28051)} =$$

$$= \frac{14.8}{14.8 + 17.7} \cdot 7.4 = 4.0302\,MW$$

The consumer from bus 28709 is supplied from generator 29109, through branch 29102-28709 (4.0302 MW). The remaining power, 3.3698 MW, belongs to the local generator from bus 29109.
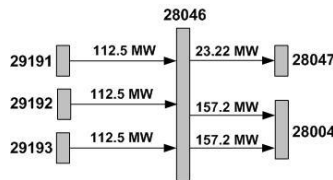

Figure 6
Bus 28046

According to Figure 6, bus 28046 is connected to three upstream generator buses (29191, 29192 and 29193) and another two downstream buses (28047, 28004).

The contribution of generators from buses 29191, 29192, 29193 on branch 28046-28047 has the same value (7.7333 MW):

$$P_{\text{branch}(28046\text{-}28047)}^{G_{29191}} = \frac{P_{G_{29191}}}{P_{G_{29191}} + P_{G_{29192}} + P_{G_{29193}}} \cdot P_{\text{branch}(28046\text{-}28047)} =$$
$$= 7.7333\,MW$$

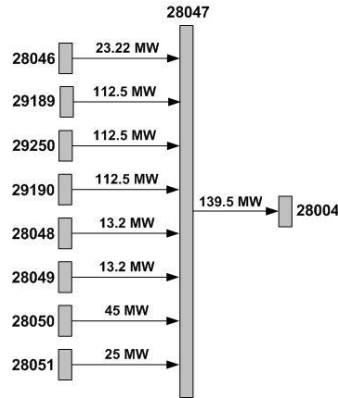104.8 MW is obtained in case of branch 28046-28004.

Figure 7
Bus 28047

In the case of branch 28047-28004, the contribution of generators 29189, 28250, 28190 and 28719 is performed in the same manner (Figure 7). The contribution of generator 29102 on branch 28047-28004, represents the contribution of the sum on branches 28050-28047 and 28051-28047.

$$P_{\text{branch}(28047\text{-}28004)}^{G_{29102}} = 13.6381 + 4.1634 = 17.8015\,MW\,.$$

The 28051-28047 branch remaining value belongs to generator 28709 – 3.481 MW. Generator contribution determination requires an additional computation. Firstly, the contribution of branch 28046-28047 on branch 28047-28004 $P_{\text{branch}(28047\text{-}28004)}^{branch(28046\text{-}28047)}$ has been computed.

$$P_{\text{branch}(28047\text{-}28004)}^{branch(28046\text{-}28047)} = 7.0943\,MW$$

Then, the contributions of the three generators on branch 28047-28004 are obtained.

$$P_{\text{branch}(28047\text{-}28004)}^{G_{29191}} = \frac{P_{\text{branch}(28047\text{-}28004)}^{branch(28046\text{-}28047)}}{P_{\text{branch}(28046\text{-}28047)}} \cdot P_{\text{branch}(28046\text{-}28047)}^{G_{29191}} =$$
$$= \frac{7.0943}{23.22} \cdot 7.7333 = 2.3648\,MW$$

The sum of the generators' contributions on branch 28047-28004 is:

$$P_{\text{branch}(28047\text{-}28004\,)} = P^{G29189}_{\text{branch}(28047\text{-}28004\,)} + P^{G29250}_{\text{branch}(28047\text{-}28004\,)} + P^{G29190}_{\text{branch}(28047\text{-}28004\,)} +$$

$$+ P^{G28051}_{\text{branch}(28047\text{-}28004\,)} + P^{G29102}_{\text{branch}(28047\text{-}28004\,)} + P^{G28709}_{\text{branch}(28047\text{-}28004\,)}$$

$$+ P^{G29191}_{\text{branch}(28047\text{-}28004\,)} + P^{G29192}_{\text{branch}(28047\text{-}28004\,)} + P^{G29193}_{\text{branch}(28047\text{-}28004\,)} =$$

$$=34.4010 + 34.4010 + 34.4010 + 3.4813 + 17.8015 +$$

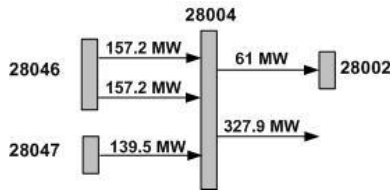$$+8.0728 + 2.3648 + 2.3648 + 2.3648 = 139.65\,\text{MW}$$

Figure 8
Bus 28004

Contributions on branch 28004-28002 and local consumer 28004 can be obtained using branches 28046-28004 and 28047-28004, as is shown in Figure 8.

$$P_{\text{branch}(28004\text{-}28002\,)} = P^{G29189}_{\text{branch}(28004\text{-}28002\,)} + P^{G29250}_{\text{branch}(28004\text{-}28002\,)} + P^{G29190}_{\text{branch}(28004\text{-}28002\,)} +$$

$$+ P^{G28051}_{\text{branch}(28004\text{-}28002\,)} + P^{G29102}_{\text{branch}(28004\text{-}28002\,)} + P^{G28709}_{\text{branch}(28004\text{-}28002\,)}$$

$$+ P^{G29191}_{\text{branch}(28004\text{-}28002\,)} + P^{G29192}_{\text{branch}(28004\text{-}28002\,)} + P^{G29193}_{\text{branch}(28004\text{-}28002\,)} =$$

$$=4.6232 + 4.6232 + 4.6232 + 0.4679 + 2.3924 +$$

$$+1.0849 + 14.402 + 14.402 + 14.402 = 61\,\text{MW}$$

$$P_{C28004} = P^{G29189}_{C28004} + P^{G29250}_{C28004} + P^{G29190}_{C28004} + P^{G28051}_{C28004} + P^{G29102}_{C28004} + P^{G28709}_{C28004} + P^{G29191}_{C28004} +$$

$$+ P^{G29192}_{C28004} + P^{G29193}_{C28004} = 24.8515 + 24.8515 + 24.8515 + 2.5149 + 12.8599 +$$

$$+5.8318 + 77.4164 + 77.4164 + 77.4164 = 328\,\text{MW}$$

Values corresponding to generator contribution on consumers will not be used for transmission cost allocation. For example, local generators only supply consumer 28719. Thus, transmission cost allocation is 0 [$/h]. Only generator 29109 contribution to consumer 28709, through branch 29102-28709, will be used in transmission cost allocation.

In the following, Figures 9-16 present a selection of values corresponding to branch usage allocated to each generator and each consumer. The analysis is performed for OHLs 28003-28034 and 28002-28004, before and after congestion is solved.

It can be observed that in both cases PV buses 29169, 29260 and 29262 have a significant contribution on OHL 28003-28034, on the Bialek method. The contribution

of these groups is increased from 20.34 MW to 40.57 MW. Using the pro-rata method, for the same OHL, 29119, 29121 and PV buses 29238 have the greatest influence on the transmission network usage. Once the congestion is solved the results are: 3.47 MW, 3.47 MW and 3.12 MW.

Figures 13-16 presents the OHL usage values allocated to consumers. The same OHLs (as in case of the generating units) have been considered within the analyses. The highest usage value is recorded for the PQ bus 28034 (Bialek method): 88.27 MW (operating condition with congestion) and 49.59 MW (operating condition with congestion solved). It is explained by the fact that PQ bus 28034 is connected to OHL 28003-28034. With the pro-rata method, the significant contribution has been highlighted for the PQ bus 28002 (629.7 MW).

Figure 9
28003-28034 OHL usage allocated to generating groups using the pro-rata method

Figure 10
28003-28034 OHL usage allocated to generating groups using the Bialek method

Figure 11
28002-28004 OHL usage allocated to generating groups using the pro-rata method

Figure 12
28002-28004 OHL usage allocated to generating groups using the Bialek method

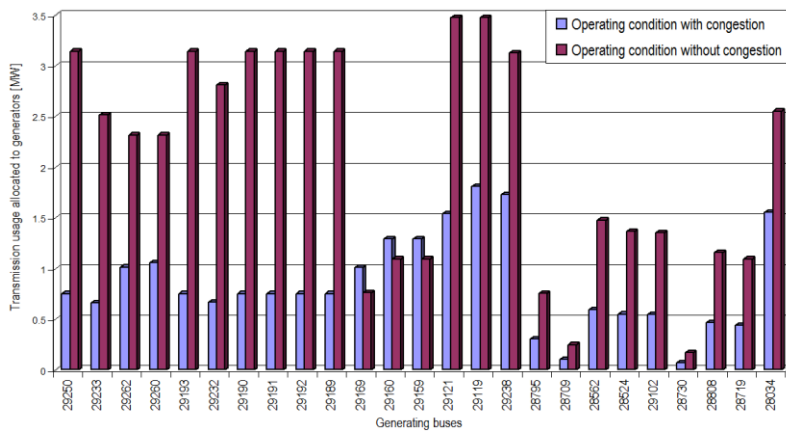
Figure 13
28003-28034 OHL usage allocated consumer buses using the pro-rata method
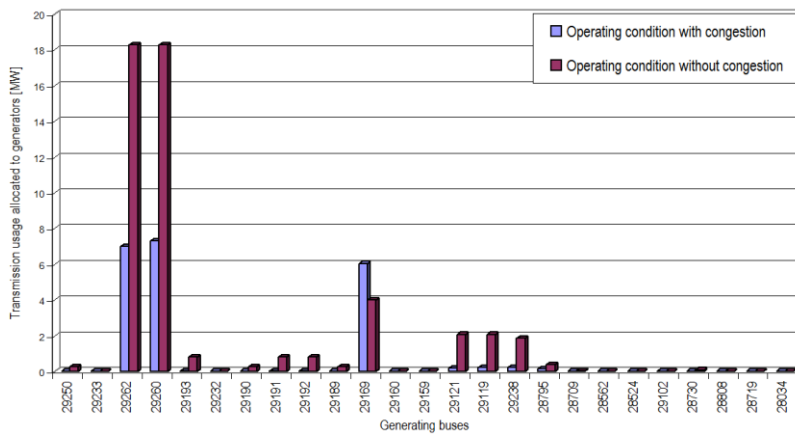
Figure 14
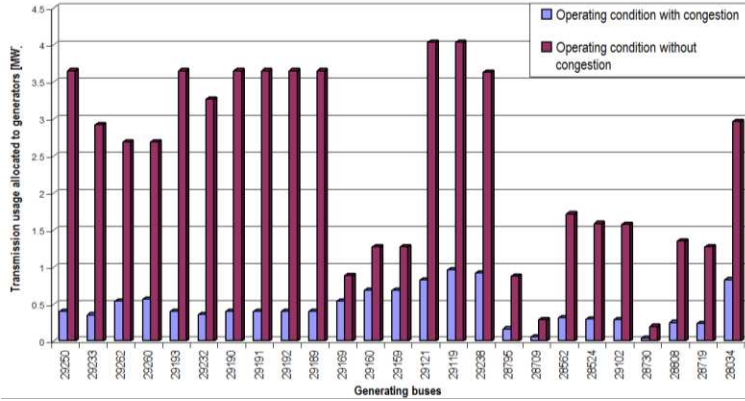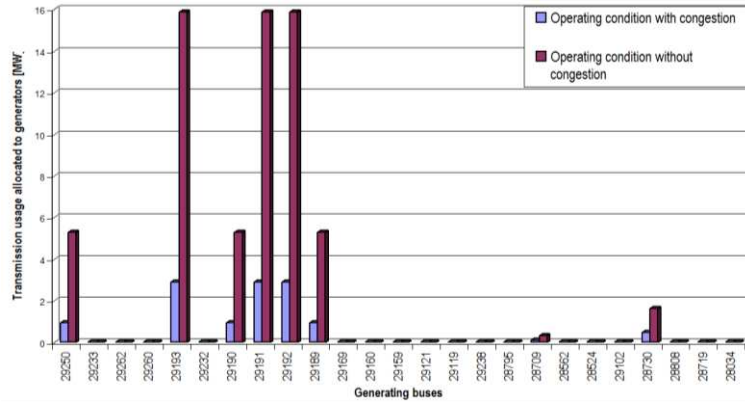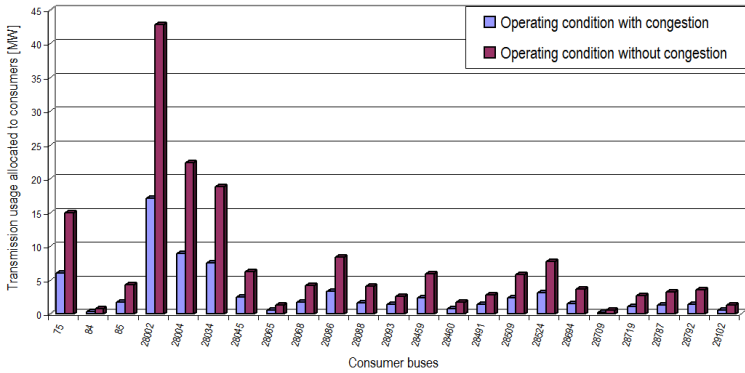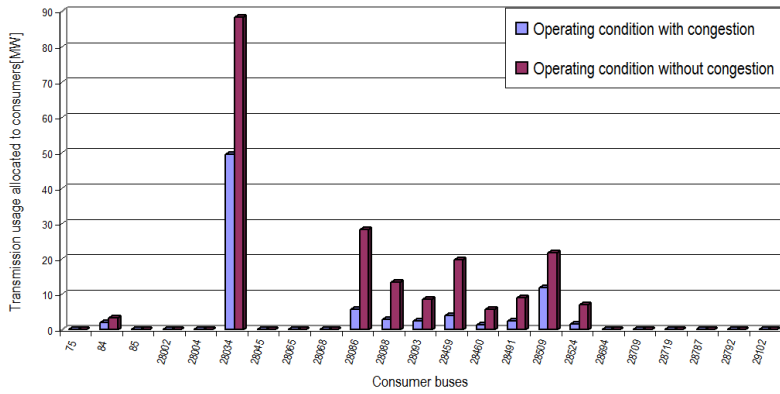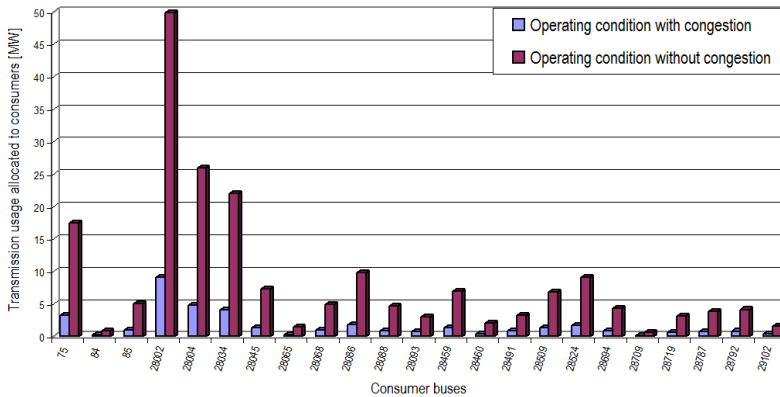28003-28034 OHL usage allocated consumer buses using the Bialek method

Figure 15
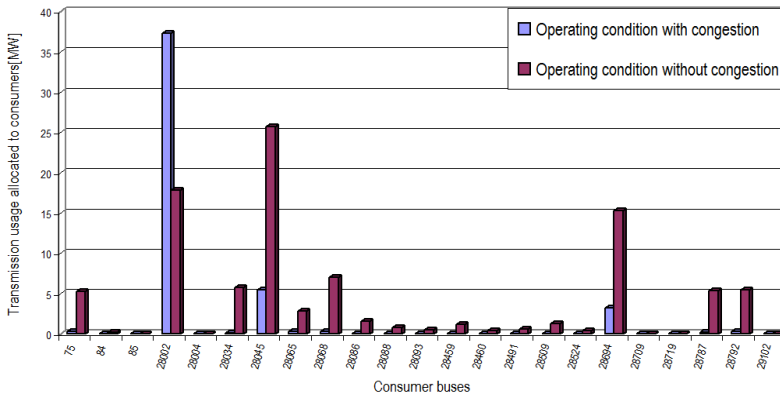28002-28004 OHL usage allocated to consumer buses using the pro-rata method

Figure 16
28002-28004 OHL usage allocated to consumer buses using Bialek method

The transmission costs are computed for generators and consumers. The transmission tariffs used by the Romanian OTS have been considered. These tariffs cover operation, maintenance and development costs. Currently, the transmission tariff system is divided into six areas of generation and eight load areas. The transmission tariff values [27] corresponding to both areas are presented in Tables 1 and 2 respectively, and the results are presented in Tables 3 and 4.

Table 1
Generation area transmission tariff values

Generation area	1G	2G	3G	4G	5G	6G
Transmission tariff value ($/MWh)	21.154	14.998	22.150	30.841	19.193	25.107

Table 2
Load area transmission tariff values

Load area	1L	2L	3L	4L	5L	6L	7L	8L
Transmission tariff value ($/MWh)	35.820	33.527	24.564	31.837	33.436	33.587	35.005	26.345

Using the Bialek method, generators 28756, 28484, 28746, 28730, 28839, 28808, 28509 and 28459 only supply the local consumers. For these generators the transmission cost is 0 $/h. According to the pro-rata method, the transmission cost allocation does not depend on network topology. For example, cost allocation for generator 28524 is 159.7 $/h. Real generated power on buses 29119 and 29121 is 210 MW. Cost allocation to both generators is 835.94 $/h (pro-rata method) and 665.28 $/h (Bialek method). We should expect for both methods (pro-rata and Bialek), that the 29232 generator allocated cost (486.02 $/h and 701.56 $/h) would be much higher than the one allocated to generator 28034 (441.04 $/h and 316.46 $/h), because 29232 uses the transmission network in a higher degree. Real generated power in case of 29232 is 170 MW, and in the case of 28034 it is 154.3 MW.

Table 3
Transmission cost allocated to generators

Generating units	29250	29233	29262	29260	29162	29193	29232	29190	29191
Pro-rata	756.34	434.84	400.26	400.26	285.9	756.34	486.02	543.2	756.34
Bialek	986.5	627.84	560.28	560.28	493.42	813.54	701.56	986.5	813.48

Generating units	29192	29189	29169	29160	29159	29121	29119	29238	28800
Pro-rata	756.34	756.34	131.52	127.76	127.76	835.94	835.94	752.36	53.18
Bialek	813.54	986.5	160.02	103.2	103.2	665.28	665.28	598.86	0.72

Generating units	28795	28709	28775	28756	28562	28484	28746	28737	28524
Pro-rata	129.8	58.92	45.46	17.72	255.02	31.16	3.44	214.42	159.7
Bialek	21.94	73.38	0.62	0	368.58	0	0	18.66	0

Generating units	29102	28730	28839	28808	28719	28509	28459	28034	
Pro-rata	325.62	28.88	144.66	200.12	262.72	33.88	0.96	441.04	0
Bialek	397.74	0	0	0	172.08	0	0	316.46	0
Total [$/h]	Pro rata		11550.12		Bialek			12009.46	

The same analysis is performed with consumers located at buses 28491 (52.1 MW) and 28002 (629.7 MW), the second being characterized by an accentuated transmission network usage. Although PQ bus 28002 has a higher consumed power value, by using the Bialek method, the transmission cost allocated value (17196.48 $/h) is smaller than bus 28034 (17927.2 $/h). A cost reduction is due to its location within the power system. PQ bus 28002 is located near PV buses 29119, 29121, 29238, 29189, 29190, 29191, 29192, 29193 and 29250, with all generators having a significant contribution in supplying the area. According to the results presented in Table 4, transmission cost allocated to PQ buses 28719, 28795 and 29102 is 0 $/h, because consumers are only supplied by local generators. The smallest transmission cost value is allocated to PQ 84 (1.76 $/h, 3.92 $/h), due to the smallest transmission network usage.

Table 4
Transmission cost allocated to consumers

Consumer bus	75	84	85	28002	28004	28034	28045	28052	28065
Pro-rata	564	1.76	172.88	32846.48	17104	19674.64	4757.2	1609.04	1203.36
Bialek	1072.72	3.92	162.72	17196.48	10769.68	17927.2	3916.8	1168	1707.76

Consumer bus	28068	28086	28088	28093	28459	28460	28484	28485	28491
Pro-rata	4164.48	8747.44	4222.32	2710.24	6191.36	1810.56	2710.24	2710.24	2860
Bialek	5959.12	14265.92	6497.12	4321.68	11042.08	3179.2	3590.56	3978.08	6866.56

Consumer bus	28509	28524	28537	28538	28562	28694	28709	28719	28729
Pro-rata	6069.52	8080	3557.2	3472	0	2806.32	386	2050	3853.52
Bialek	10160	3215.04	3265.52	3188.48	0	3166.56	63.36	0	3281.92

Consumer bus	28730	28736	28737	28746	28747	28756	28774	28775	28787
Pro-rata	1480.56	1757.76	4299.76	5354.4	2961.12	3772.4	4002.24	1805.04	3204.48
Bialek	894.56	2264.88	606.56	7225.84	3921.92	5432.96	6853.76	1800.4	5569.52

Consumer bus	28792	28795	28800	28808	28839	29102			
Pro-rata	3495.2	2690.72	1737.44	5732.96	5627.36	975.44			
Bialek	5694.16	0	990.56	1736.8	5047.52	0			
Total [$/h]	Pro rata	193231.9			Bialek	188005.7			

Conclusions

We suggest a simultaneous *N-1* contingencies congestion management approach and transmission cost allocation among market participants. Two congestion solving corrective methods have been used in this case study: generated power re-dispatching and consumed power mitigation. It finds that the power generation reserves' availability and power system structure play a positive role in solving congestion.

Transmission costs allocated to generators and consumers are computed using a developed software tool based on the Bialek and pro-rata methods. Power losses have been considered within this approach. Both methods are highly dependent on total system bus power injection. The values corresponding to the analyzed methods are different, due to different premises. Using the pro-rata method, transmission usage is distributed across all generating groups and PQ buses. With regard to the Bialek method, transmission usage is distributed to each generator and each consumer, through the power flow tracing mechanism.

The main contributions of this paper are as follows: base operating conditions computed by the authors, correspond with the ones managed by the Romanian Power Dispatcher; transmission tariff values are the ones used by the Romanian TSO, who is the main beneficiary of this study; methods have been implemented within a software tool developed in Mathematica environment; relations used for the pro-rata method have been adapted for the needs of Transelectrica (taking into consideration the ratio between the generating units / consumers contribution).

Acknowledgements

References

[1]   Azadeh A., Saberi M., Nadimi V., Iman M., Behrooznia A.: An Integrated Intelligent Neuro-Fuzzy Algorithm for Long-Term Electricity Consumption: Cases of Selected EU Countries, Acta Polytechnica Hungarica, Vol. 7, No. 4, 2010, pp. 71-90

[2]   Tao S., Gross G.: A Congestion-Management Allocation Mechanism for Multiple Transaction Networks, IEEE Transactions on Power Systems, Vol. 17, No. 3, August 2002, pp. 826-833

[3]   Fu J., Lamont J. W.: A Combined Framework for Service Identification and Congestion Management, IEEE Transactions on Power Systems, Vol. 16, No. 1, February 2001, pp. 56-61

[4]   Layden D., Jeyasurya B.: Integrating Security Constraints in Optimal Power Flow Studies, IEEE Transactions on Power Systems, Vol. 20, No. 2, May 2005, pp. 675-683

[5]   Belmadani A., Benasla L., Rahli M.: The Dynamic Economic Dispatch including Wind Power Injection in the Western Algerian Electrical Power System, Acta Polytechnica Hungarica, Vol. 8, No. 5, 2011, pp. 191-204

[6]   Kumar A., Srivastava S. C., Singh S. N.: A Zonal Congestion Management Approach Using Real and Reactive Power Rescheduling, IEEE Transactions on Power Systems, Vol. 19, No. 1, February 2004, pp. 554-562

[7]   Fang R., David A.: Transmission Congestion Management, IEEE Transactions on Power System, Vol. 14, No. 3, August 1999, pp. 877-833

[8]   Abhyankar A. R., Soman S. A., Khaparde S. A.: Optimization Approach to Real Power Tracing: An Application to Transmission Fixed Cost Allocation, IEEE Transactions on Power System, Vol. 21, No. 3, August 2006, pp. 1350-1361

[9]   Pan J., Teklu Y., Rahman S., Jun K.: Review of Usage-based Transmission Cost Allocation Methods under Open Access, IEEE Transactions on Power System, Vol. 15, No. 4, November 2000, pp. 1218-1224

[10]   Ilic M. D., Yoon Y. T., Zobian A., Paravalos M. E.: Toward Regional Transmission Provision and its Pricing in New England, Utility Policy, Vol. 6, No. 3, 1997, pp. 245-256

[11]   Shirmohammadi D., Gribik P. R., Law E. T. K., Malinowski J. H., O'Donnel R.: Evaluation of Transmission Network Capacity Use for Wheeling Transactions, IEEE Transactions on Power Systems, Vol. 4, No. 4, October 1989, pp. 1405-1413

[12]   Shirmohammadi D., Rajogopalan C., Alward E. L., Thomas C. L.: Cost of Transmission Transactions: An Introduction, IEEE Transactions on Power Systems, Vol. 6, No. 4, November 1991, pp. 1546-1560

[13]   Conejo A. J., Arroyo J. M., Alguacil N., Guijarro A. L.: Transmission Loss Allocation: A Comparison of Different Practical Algorithms, IEEE Transactions on Power Systems, Vol. 17, No. 3, August 2002, pp. 571-576

[14]   Bialek J.: Topological Generation and Load Distribution Factors for Supplement Charge Allocation in Transmission Open Access, IEEE Transactions on Power Systems, Vol. 12, No. 3, November 1997, pp. 1185-1193

[15]   Bialek J.: Allocation of Transmission Supplementary Charge to Real and Reactive Loads, IEEE Transactions on Power Systems, Vol. 13, No. 3, August 1998, pp. 749-754

[16]   Bialek J.: Tracing the Flow of Electricity, IEE Proceedings C, Generation, Transmission and Distribution, Vol. 143, No. 4, July 1996, pp. 313-320

[17]   Bialek J., Kattuman P.A.: Proportional Sharing Assumption in Tracing Methodology, IEE Proceedings C, Generation, Transmission and Distribution, Vol. 151, No. 4, July 2004, pp. 526-532

[18] Strbac G., Kirschen D., Ahmed S.: Allocating Transmission System Usage on the Basis of Traceable Contributions of Generators and Load Flows, IEEE Transactions on Power Systems, Vol. 13, No. 2, May 1998, pp. 527-534

[19] Kirschen D., Allan R., Strbac G.: Contributions of Individual Generators to Loads and Flow, IEEE Transactions on Power System, Vol. 12, No. 1, February 1997, pp. 52-66

[20] Ng W. Y.: Generalized Generation Distribution Factors for Power System Security Evaluations, IEEE Transactions on Power Apparatus and Systems, Vol. PAS-100, No. 3, March 1981, pp. 1001-1005

[21] Kilyeni S., Pop O., Slavici T., Craciun C., Andea P., Mnerie D.: Transmission Cost Allocation Using the Distribution Factors Method, Proceedings of the 15th IEEE Mediterranian Electromechanical Conference MELECON, April 25-28, 2010, Valletta, Malta, pp. 1093-1098

[22] Galiana F. D., Conejo A. J., Gil H. A.: Transmission Network Cost Allocation Based on Equivalent Bilateral Exchanges, IEEE Transactions on Power Systems, Vol. 18, No. 4, November 2003, pp. 1425-1431

[23] Pop O., Barbulescu C., Andea P., Jigoria-Oprea D., Coroiu F., Tirian O.: Comparison of Power System Tracing Cost Allocation Methods, Proceedings of the IEEE International Conference EUROCON, April 27-29, 2011, Lisbon, Portugal, pp. 1-4

[24] Conejo A. G., Galiana F. D., Kockar I.: Z-bus Loss Allocation, IEEE Transactions on Power System, Vol. 16, No. 1, February 2001, pp. 105-110

[25] Conejo A. J., Contreras J., Lima D. A., Padilha-Feltrin A.: Z-bus Transmission Network Cost Allocation, IEEE Transactions on Power System, Vol. 22, No. 1, February 2007, pp. 342-349

[26] Kilyeni S., Pop O., Prostean G., Craciun C.: Transmission Cost Allocation Based on Power Flow Tracing Using Z Bus Matrix, Proceedings of the 14th International Conference on Harmonics and Quality of Power ICHQP 2010, Bergamo, Italy, September 26-29, 2010, pp. 1-6

[27] http://www.transelectrica.ro/5Piete/administrare.php#

[28] Barbulescu C., Congestion Management in Free Energy Market Conditions, PhD Thesis, Politehnica University Timisoara, 2009

[29] Pop O., Contributions on the Tariff Access Assessment to the Transport System, PhD Thesis, Politehnica University Timisoara, 2009

Mechanical Behavior of Multiple-forged Al 7075 Aluminum Alloy

Tareg S. Ben Naser, György Krallics

Dept. of Materials Science and Engineering, Budapest University of Technology and Economics, Bertalan Lajos út 7, H-1111 Budapest, Hungary
taregnaser@ eik.bme.hu, krallics@eik.bme.hu

Abstract: The mechanical behavior of any material can be described using several tests, such as, compression tensile, hardness test, etc. In this work, cold and hot compression tests and hardness measurements were utilized. The subject material of this study was the Al 7075 alloy in the initial state (IS) and the multiple forged (MF) state. The cold compression test at room temperature was used to measure the deformation anisotropy on the MF specimen, while the hot compression test results were used as a source data in order to establish the constitutive equation in the wide ranges of working temperature (from 250 to 450°C) and strain rate (from 0.002 to 2 s-1). The homogeneity and structure of the material were evaluated using the Vickers hardness measurements and optical microscopy images.

Keywords: 7075 aluminium alloy; multiple forging; constitutive equations; hot compression test, cold compression test

1 Introduction

The Al 7075 alloy is a very interesting material because of its mechanical properties, namely, low density, high strength, moderate ductility and toughness. Due to these properties, the alloy is mainly used for highly stressed structural parts. This material has a wide range of applications such as aircraft fittings, gears and shafts, fuse parts, meter shafts and gears, missile parts, regulating valve parts, worm gears, keys and various other parts of commercial aircrafts and aerospace vehicles [1].

Multiple forging (MF) is one of the severe plastic deformation (SPD) techniques [2-4], which is used to refine the grain size down to nanostructure range. The principle of this technique is to perform multiple repeats of open-die forging operations, while changing the axis of the load by 90° at each pass. The strain magnitude in one pass of multiple forging is around 0.6; three passes of MF can push the accumulated strain to a high level. For instance, the strain magnitude

during the first pass of equal channel angular press (ECAP) is around 1, and this is where the grain refinement occurs. The MF technique provides less homogeneity compared to other SPD techniques, however, it produces larger specimen dimension. The grain refinement has high effect on several material properties such as strength, fatigue and superplasticity [5].

In order to measure the mechanical behavior of the MF Al7075, some basic tests were performed, including the cold compression test, the hot compression test and the hardness test.

The hot compression test is usually performed to produce a data that can be used as a source to determine the constitutive equation constants [5-9]. The constitutive model for hot deformation proposed by Sellars and McTegart [10] is one of the most widely used for summarizing and extrapolating experimental data:

$$\dot{\varepsilon} = A \, sinh(\alpha\sigma)^n exp\left(\frac{Q}{RT}\right) \tag{1}$$

$$\dot{\varepsilon} = A \, exp(\beta\sigma) \, exp\left(-\frac{Q}{RT}\right) \tag{2}$$

From the equations, $\dot{\varepsilon}$ is the strain rate s^{-1}, σ is the true stress (MPa), Q is the activation energy of deformation, R is the gas constant and T is the absolute temperature of deformation (K). Q, A, β, n, and α are material parameters.

Nowadays, finite element methods (FEM) are most widely used in engineering applications such as the simulation of metal forming in order to define the optimum values of the deformation parameters. The constitutive equation, which represents the flow behavior of the materials, is used as an input variable to the FEM application to simulate the material's response under certain loading conditions [5]. The accuracy of FEM simulation has relatively high dependency on the accuracy of prediction of the constitutive equation. In these equations, Eq. 1 and Eq. 2, the flow stress depends on the working temperature and the strain rate, and also on the material which constants considered as a function of equivalent strain [5, 9, 11-13]. In this study, the two constitutive models (Eq. 1 and Eq. 2) were compared in terms of the prediction of the hot flow behavior of the considered alloy. The present work results were compared with the results of our previous work [14], where we investigated only IS material using Eq. 2.

Materials usually gain anisotropy behavior after severe plastic deformation, which might differ from one another according to the axis of deformation [15]. In our case the material was deformed in three axes, therefore it was interesting to see the effect of the MF on the deformation anisotropy on all the axes of deformation.

2 Materials and Experiments

2.1 Test Material

The material was produced at SWA (Southwest Aluminum (group)) Co. Ltd. The chemical composition is listed in Table 1.

Table 1

Material	Si	Fe	Cu	Mn	Mg	CR	Zn	Ti	Al
WT%	0.12	0.37	1.76	0.06	2.66	0.23	5.66	0.02	balance

The IS material was received as a rod, 40 mm in diameter and 3m in length. By IS we mean the material in the initial state of our work, which was received as a cylindrical rod; this shape was produced by extrusion at SWA Co. Ltd. The received rod was annealed at 450° C for 30 minutes, and then it was cooled to room temperature in a furnace in order to allow the precipitation process to complete. This results in a low cooling rate, which ensures a stable base material for further tests.
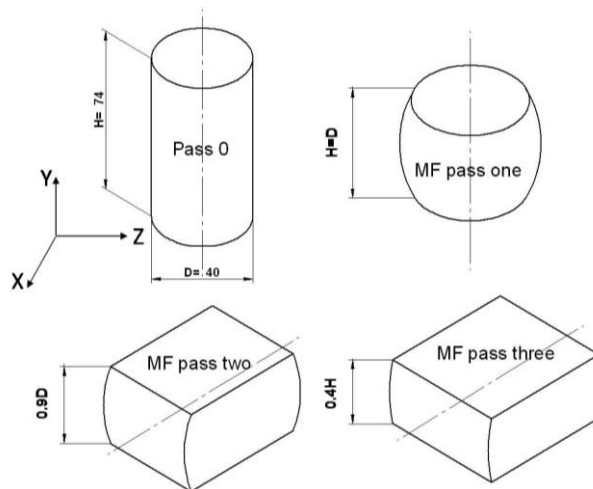
Figure 1

MF operation steps (specimen dimensions are D= 40 mm, H= 74 mm)

2.2 Multiple Forging

Fig. 1 shows the MF operation steps, which consist of performing a multi-repeat of open die forging operations, whereby the axis load is changed by 90° in each pass. Some parameters have to be determined before performing the MF process. One of the most important parameters is the forming temperature, since it has significant effect on forming ultrafine grain structure. The ultrafine grain structure

development requires lower working temperatures than those commonly used in conventional manufacturing of semi-finished products [9]. In order to determine the forming temperature, several trial and error tests were conducted; the results have shown that the lowest temperature which can be applied for all process steps without any undesirable behavior was 250 °C. The diameter-to-height ratio is another parameter that has to be taken into consideration. According to the experimental result it was found that 0.54 (D/H) is a suitable ratio. Thus, the specimens for the MF experiments were cut to a diameter of 40mm and a height of 74 mm, as can be seen in Fig. 1. The process was performed using a hydraulic press machine with a maximum force of 200 tons, at constant cross head speed (5 mm/min).

2.3   Cold Compression Test

Some specimens were subjected to the cold compression test at room temperature and constant cross head speed; the compression tests were performed only for MF materials. The specimens were subjected to this test to investigate the anisotropy behavior by measuring the flow stress curve at room temperature, and by measuring two perpendicular diameters on the specimen's cross section. The IS material wasn't interesting for this test because it didn't show any notable deformation anisotropy.

2.4   Hot Compression Test

The compression test specimens were machined to have cylindrical shape with 7 mm diameter and 8.4 mm height. Isothermal compression tests at constant strain rates and constant temperatures were conducted. The maximum compression equivalent strain of 0.55 was achieved in the test. Before beginning the test, the specimens were held in the die for some minutes to allow the material to reach steady state. The temperature was measured using thermocouples. Carbon powder was used as a lubricant to decrease the effect of friction and barreling. The deformation temperature and strain rate were automatically controlled by the machine control unit. The experimental temperatures were between 250 and 450°C with an interval of 50°C. At each deformation temperature four constant strain rates were applied (0.002, 0.02, 0.2 and 2 s^{-1}). All compression tests were carried out using an Instron 5982 electromechanical universal material testing equipment with a maximum load of 100 kN with high a temperature furnace.

2.5   Hardness Measurements

Vickers hardness measurements (HV5) were used to evaluate the structure homogeneity and the influence of the MF on the surface hardness of the material. The main load during the test was 5 kg.

2.6   Metallographic Work

Some samples were prepared to monitor the structure, starting with IS material and MF steps. The samples were cut from the original specimen by handsaw, were subsequently grinded in several steps, and the mechanical preparation was finished by polishing with aluminum powder. Immediately before the start of the optical microscope investigation, electro-polishing was done for all samples with different polishing parameters; the aching liquid was 700 ml ethyl alcohol, 120 ml distilled water, 78 ml perchloric acid 70% and 100 ml butylglycol; the used voltage was 30 V for 10 sec.

3   Results and Discussion

The discussion of the results will cover four points as following:

1.   Compression test

2.   Constitutive equation

3.   Hardness measurements

4.   Optical microscopic investigation

3.1   Compression Test

The force-height data was used to define the true stress-true strain curves using some basic equations as given by Eq. 3, Eq. 4, and Eq. 5:

$$\varepsilon = ln\left(\frac{h_o}{h}\right) \tag{3}$$

$$\sigma = \frac{F}{A} \tag{4}$$

$$A = \frac{A_o h_o}{h} \qquad \text{(Volume constancy)} \tag{5}$$

where ε is the equivalent strain, h_o is the initial height (mm), h is the current height (mm), σ is the flow stress (MPa), F is the force (N), A is the current area cross section (mm^2) and A_o is the initial area of the cross section (mm^2).

3.1.1   Cold Compression Test

The step-by-step compression tests were carried out in the three axes, namely, X, Y and Z of MF material, as illustrated in Fig. 2. In all axes, the anisotropy phenomenon appears where the cross section is changed from cylindrical to elliptical shape. Using the following equations: Eq. 6, Eq. 7, the compressive and tensile strains were calculated:

$$\varepsilon_1 = ln\left(\frac{h_i}{h_o}\right) \tag{6}$$

$$\varepsilon_a = ln\left(\frac{a}{D_o}\right) \tag{7}$$

$$\varepsilon_b = ln\left(\frac{b}{D_o}\right) \tag{8}$$

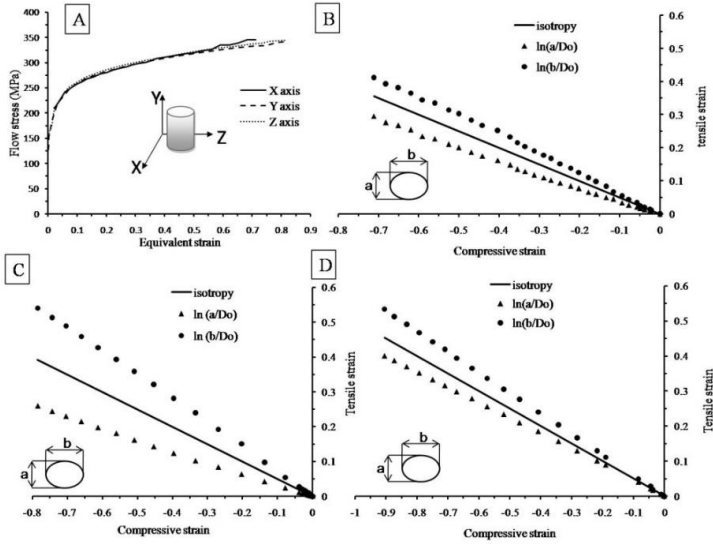$$\varepsilon_1 + \varepsilon_a + \varepsilon_b = 0 \tag{9}$$

Figure 2

A: the flow curves in the three axes. B, C and D are the logarithmic ratio between initial diameter and ellipse axes on X, Y and Z axes respectively as scatter points and the line for isotropy material

At the free surfaces of the compressed cylindrical test specimens, the strains consist of two types of strain: tension strain at the circumference and axial compression strain. Because the final cross-sectional area of the deformed specimens is elliptical, there are two values for tensile strain in the direction of the diameter: a and b. That means ε_1 is a compression strain and ε_a ε_b is a tensile strain. The strain components were calculated using Eq. 7 and Eq. 8, and the results also satisfied the equation 9.

According to the experimental results, the Y axis has the highest deformation anisotropy compared to the other axes because it is the direction of forward extrusion for the IS material rod. Fig. 2 shows the deformation anisotropy on all axes of deformation. All flow curves along the various axes are close to each other.

3.1.2   Hot Compression Test

In Fig. 4, the flow curves of the IS material are drawn as scatter points, which were obtained from the hot compression tests at various working temperatures (250, 300, 350, 400 and 450 °C) and different strain rates (0.002, 0.02, 0.2 and 2 s^{-1}). At strain rates of 0.002-0.02 s^{-1} and working temperatures of 350, 400 and 450 °C, the flow curve has almost no influence from equivalent strain, especially when the strain rate was 0.02 s^{-1}. Strain softening behavior was observed in the IS material at working temperatures of 300, 350, 400 and 450 °C and strain rates of 0.2 and 2 s^{-1}. In contrast, the IS material at working temperature of 250 °C and strain rates of 0.002, 0.02 and 0.2 s^{-1} has shown a clear strain hardening behavior.

In Fig. 6, the flow curve of the MF material is shown as scatter points, which were obtained from a hot compression test under the following working temperatures: 250, 300, 350, 400 and 450°C and the following strain rates: 0.002, 0.02, 0.2 and 2 s^{-1}. The flow curves of the MF behave in the same manner as the flow curves of the IS material under the same experimental conditions; however, there were slight differences in the behavior where the strain softening is more visible in the MF specimens at the highest strain rate.

The effects of the anisotropic deformation in hot compression test were negligible compared to the cold forming result.

The results obtained from the hot compression tests under different strain rates and working temperatures were used to determine the material parameters of the constitutive equations.

3.2   Constitutive Equation

The stress-strain data obtained from the hot compression tests under different strain rates and working temperatures were used to determine the material parameters of the constitutive equations for the IS and MF material. In the present work, two results are presented:

1. Constitutive equation using sin hyperbolic function for IS material

2. Constitutive equation using exponential function for MF material

The results for the IS material using the exponential equation can be found in [14]. The sin hyperbolic function was used for the MF material; however, the result of approximation wasn't good. Although the verification of the parameters was good, the flow curve using this constitutive equation has a zigzag shape.

The sin hyperbolic constitutive equation that was derived by Sellars and McTegart [10] was used to describe the plastic deformation behavior of the IS material; see Eq. 1. Thus, the flow stress can be expressed as follows:

$$\sigma = \frac{arcsinh\left(exp\left(\frac{(RT(ln(\dot{\varepsilon})-ln(A))+Q)}{RTn}\right)\right)}{\alpha} \tag{10}$$

The exponential constitutive equation that was derived by Sellars and McTegart, Eq.2 [10], was used for the MF material; the flow stress was expressed as shown in Eq.11:

$$\sigma = \frac{ln\dot{\varepsilon}}{\beta} + \frac{Q}{\beta RT} - \frac{lnA}{\beta} \tag{11}$$

where $\dot{\varepsilon}$ is the strain rate (s^{-1}), σ is the true stress (MPa), Q is the activation energy of deformation, R is the gas constant and T is the absolute deformation temperature (K). Q, A, β, n, and α are the material parameters. All material parameters have been obtained by solving the equations as an over-determined problem, where we have a higher number of equations than the number of the unknown parameters. The obtained parameters were plotted with equivalent strain between 0.05 and 0.55 with an interval 0.05. A fourth-order polynomial function was applied to describe the relationship between the parameters and equivalent strain, to fit to the series of obtained data.

3.2.1 Initial State Material

$$ln(A) = 12.28 + 18.15\varepsilon + 96.77\varepsilon^2 - 211.4\varepsilon^3 + 100\varepsilon^4 \tag{12}$$

$$Q = 12486 - 16484\varepsilon + 10^6\varepsilon^2 - 3 \times 10^6\varepsilon^3 + 2 \times 10^6\varepsilon^4 \tag{13}$$

$$n = 2.232 + 19.28\varepsilon - 48.03\varepsilon^2 + 71.86\varepsilon^3 - 4549\varepsilon^4 \tag{14}$$

$$\alpha = 0.055 - 0.416\varepsilon + 1.84\varepsilon^2 - 3.536\varepsilon^3 + 2.475\varepsilon^4 \tag{15}$$

The R square value of the fitting of polynomial equations for lnA, Q, n, and α were 0.999, 0.997, 0.999 and 0.989 respectively. These values indicate good fitting. The observation from Fig. 3 is that lnA, Q and n increase with the equivalent strain; also, they are independent of changes in the working temperature and the strain rate at fixed strain. In contrast, α values decrease with the equivalent strain. The parameter lnA was between 13.46 and 25.49, Q was between 120554 and 170336 J/mol, n was varying from 3.067 to 6.096 and α was between 0.039 and 0.01994. Equations 10, 12, 13, 14 and 15 are the constitutive equations for the IS material.
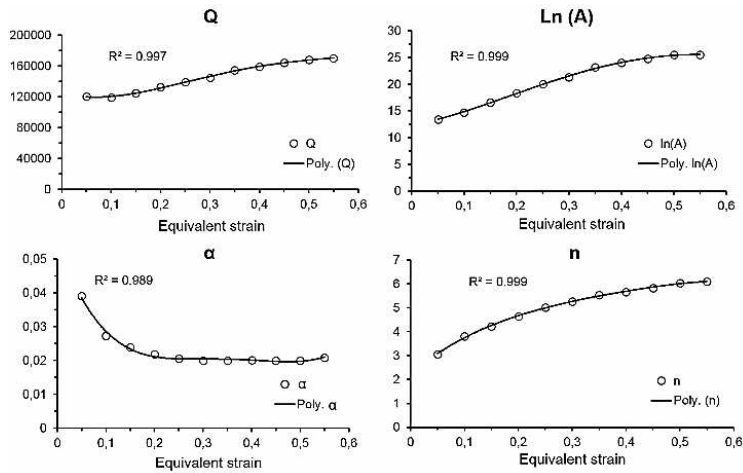
Figure 3

The values of IS material parameters lnA, Q, n, and α obtained for different strain levels and the fitted curves using fourth-order polynomial
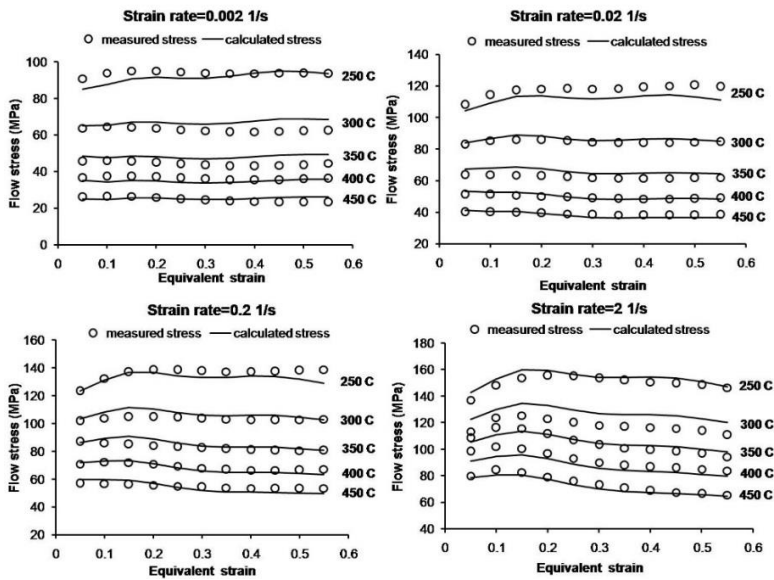

Figure 4

Measured IS material flow curves (scatter) and calculated flow curves (solid lines)

3.2.2   MF Material

$$ln(A) = 13.397 - 29.758\varepsilon + 275.959\varepsilon^2 - 545.449\varepsilon^3 + 391.007\varepsilon^4 \qquad (16)$$

$$\beta = 0.144 - 0.579\varepsilon + 3.181\varepsilon^2 - 6.333\varepsilon^3 + 4.405\varepsilon^4 \qquad (17)$$

$$Q = 131.31 - 244.40\varepsilon + 1761.4\varepsilon^2 - 3448.7\varepsilon^3 + 2314\varepsilon^4 \qquad (18)$$

The R square values of the fitting of the polynomial equations for lnA, β and Q were 0.999, 0.983 and 0.997 respectively. The A, β and Q parameters are varying with equivalent strain and are independent of the working temperature and the strain rate at a fixed strain. The obtained parameter values are: lnA from 12.4 to 20.61, β from 0.1099 to 0.1437, and Q between 119.9 and 173.1 kJ/mol. All of these parameters tend to increase with equivalent strain (Fig. 5). Eq. 11, 16, 17 and 18 are the constitutive equations for the MF material.
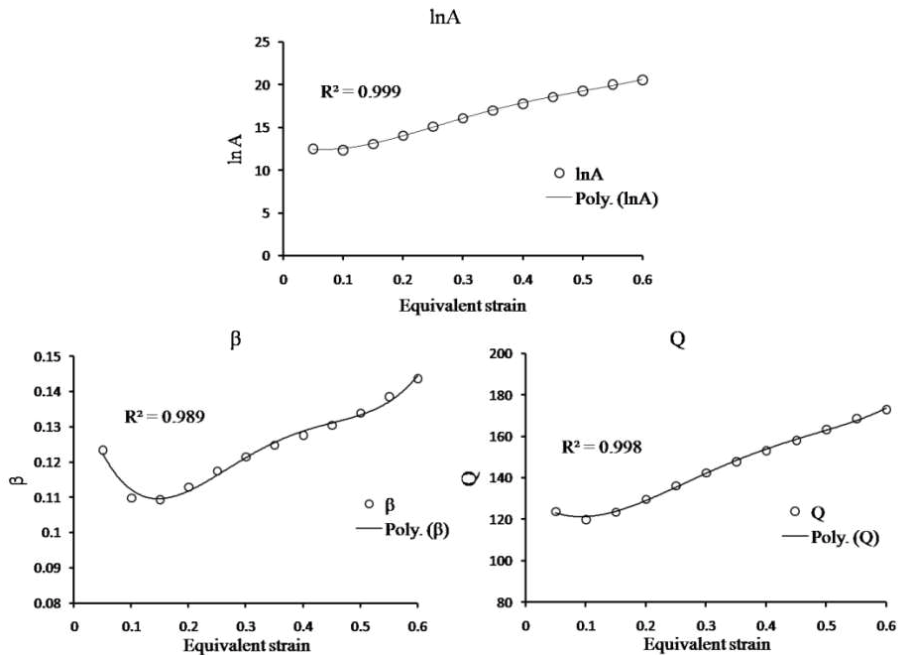

Figure 5

Values of the MF material parameters, lnA, Q, and β, which were obtained for different strain levels and the fitted curves

In order to verify the constitutive equations for the IS and MF materials, stress-strain curves were calculated using the constitutive equations at the same working temperatures and strain rates that were used in the experiments. The measured true stress as scatter points and the calculated true stress as solid lines are plotted in Fig. 4 and 6, respectively.

The average absolute relative error (AARE) and correlation coefficient (R^2) were calculated for all the data using Eq. 19 and Eq. 20:

$$AARE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\sigma_m - \sigma_c}{\sigma_m}\right| 100 \qquad (19)$$

$$R = \frac{\sum_{i=1}^{n}(\sigma_m - \bar{\sigma}_m)(\sigma_c - \bar{\sigma}_c)}{\sqrt{\sum_{i=1}^{n}(\sigma_m - \bar{\sigma}_m)^2(\sigma_c - \bar{\sigma}_c)^2}} \tag{20}$$

where n the number of data, σ_m measured stress and $\bar{\sigma}_m$ mean value of σ_m, σ_c is a calculated stress and $\bar{\sigma}_c$ is the mean value of σ_c. The verification parameter values of the present work are listed in Table 2.
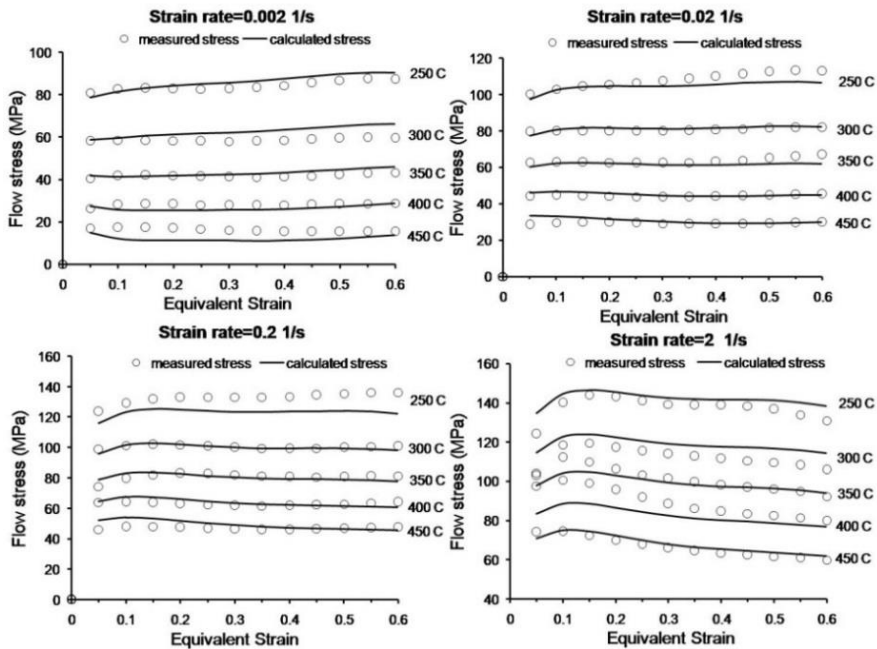
Figure 6
The measured and calculated flow curves of the MF material

Table 2
Verification parameter values

Material type	Used function	AARE	R^2
IS material	sin hyperbolic function	4.55907	0.99854
MF material	exponential function	5.123815	0.998685

The results in Table 2 show the reliability of the equations in predicting the behavior of the experimental material.

Table 3 shows the comparison between the verification parameters for the two applied constitutive functions. The value of AARE and R^2 parameters shows the better prediction capabilities of the sin hyperbolic function compared to the exponential function.

Table 3

Comparison of the verification parameters for the two used functions

Material type	Used function	AARE	R^2
IS material	sin hyperbolic function	4.55907	0.99854
IS material	exponential function	6.194949	0.998621
MF material	exponential function	5.123815	0.998685

The flow curves at hot compression turn out very similar to those of the IS and MF samples. In the attempt to understand the process and find an explanation for such interesting results, we shared our hot compression measurements with the dynamic recrystallization study group, and they stated the following:

The reason for this result was attributed to the stored energy, which is higher in the MF material compared to the IS, therefore the dynamic recrystallization occurs at a lower stress level in the case of MF [16].

3.3   Hardness Measurements

Fig. 7 plots the histogram of the HV5 hardness measured on the IS and MF material specimens. The number of measurements on each specimen was 30, distributed randomly on the surface of specimens. The distribution of the hardness for both the IS and MF materials conforms to a normal distribution with different mean and standard deviations. The mean hardness value of the IS material was HV 118.25, whereas that of the MF material was HV 81.6; we attribute these unexpected hardness results for the MF material to the effects of warm deformation during the MF process. The standard deviation of the hardness measurements for the MF material was less than that for the IS material, which can be clearly seen from the width of the normal distribution of the histogram. These results show an improvement in the structural homogeneity.
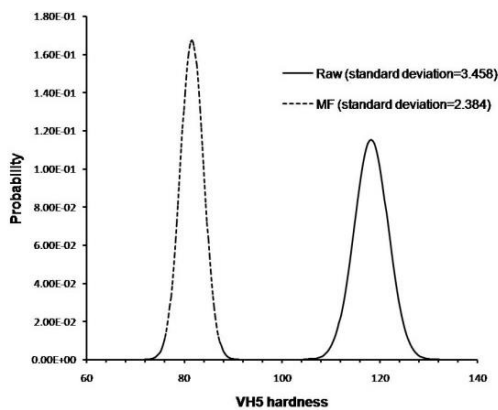

Figure 7

The normal distribution of hardness measurements

3.4 Optical Microscopic Investigation

Some specimens were prepared for micrographic investigations. The preparation involves mechanical grinding and polishing followed by electro polishing. The optical microscope photos show a microstructure change during the MF processes, where the second-phase particle size at the third pass of MF was decreased, compared to the IS material, as is illustrated in Fig. 8.
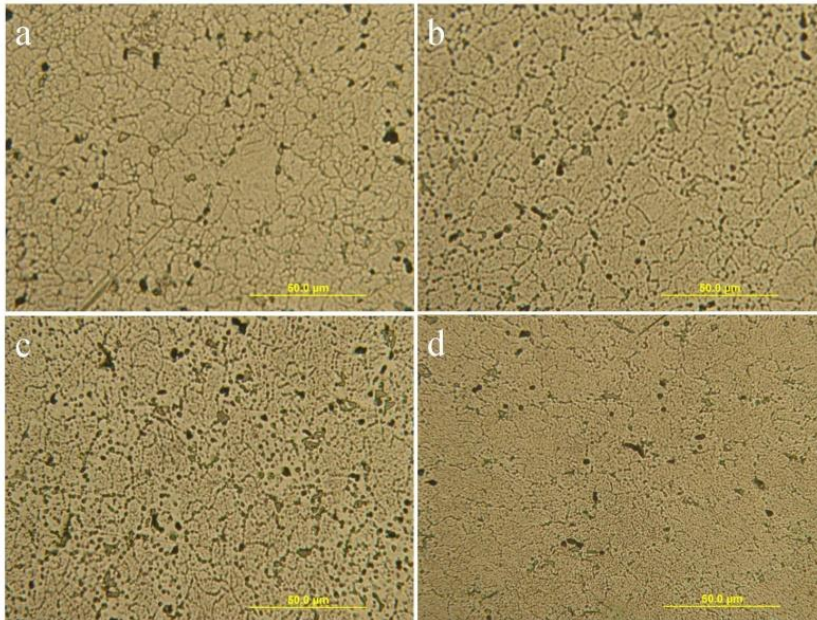
Figure 8
(a) IS material; (b) MF first pass; (c) MF second pass; (d) MF third pass

Summary

- A set of constitutive equations was obtained for the IS material, where the flow stress depends on temperature and strain rate according to a sin hyperbolic function. The material parameters, A, Q, n and α are considered as functions of equivalent strain.

- A set of constitutive equations was obtained for the MF material, where the flow stress depends on the temperature and strain rate according to an exponential function. The material parameters, lnA, β and Q are considered functions of equivalent strain.

- The constitutive equations were found to predict flow stress precisely over a wide range of working temperatures and strain rates, where the overall AARE and the correlation coefficient R^2, were good for both the IS and MF materials.

- Comparing the present IS material result of *sin* hyperbolic constitutive equation with our previous results [14], it can be stated that the *sin* hyperbolic function is more reliable than the exponential function to predict the experimental behavior of the material.

- The deformation anisotropy has significant effect during cold deformation, while it has negligible effect during hot deformation.

Acknowledgement

References

[1]   George E. Totten, D. ScottMacKenzie. Handbook of aluminum Vol. 1 Physical Metallurgy and Processes (2003) 185

[2]   Oleg Sitdikov, Taku Sakai, Alexandre Goloborodko, Hiromi Miura and Rustam Kaibyshev, Effect of Pass Strain on Grain Refinement in 7475 Al Alloy during Hot Multidirectional Forging. Materials Transactions, 2004, Vol. 45 (2232-2238)

[3]   Zhang Hui, Lin Gao-yong, Peng Da-shu, Yang Li-bin, Lin Q-quan. Manufacturing of Aluminum Alloy Ultra-Thick Plates by Multidirectional Forging and Subsequent Rolling. Trans. Nonferrous Met. Soc. China. 2002, Vol. 12 (218-221)

[4]   T. Sakai, H. Miura, X. Yang. Ultrafine Grain Formation in Face-centered Cubic Metals during Severe Plastic Deformation. Materials Science and Engineering A, 499 (2009) 2-6

[5]   D. Samantaray, S. Mandal, A. K. Bhaduri. Constitutive Analysis to Predict High-Temperature Flow Stress in Modified 9Cr–1Mo (P91) Steel, Materials and Design, 2010, Vols. 31 (981-984)

[6]   Juan Liu, Zhenshan Cui, Congxing Li, Modelling of Flow Stress Characterizing Dynamic Recrystallization for Magnesium Alloy AZ31B, Computational Materials Science, 2008, Vols. 41 375-382

[7]   P. Ponge, M. Bredehöft, G. Gottstein, Dynamic Recrystallization in High Purity Aluminium, ScriptaMaterialia. Vol. 37 (1997) 1769-1775

[8]   Quan Guo-zheng, Liu Ke-wei, Zhou Jie, Chen Bin, Dynamic Softening Behaviors of 7075 Aluminum Alloy, Trans. Nonferrous Met. Soc. China, 2009, Vols. 19 (s537-s541)

[9]   Z. Zeng, S. Jonsson, Y. Zhang, Constitutive Equations for Pure Titanium at Elevated Temperatures, Materials Science and Engineering A. 505 (2009) 116-119

[10] C. M. Sellars, W. J. McTegart. On the Mechanism of Hot Deformation. Acta Metallurgica, Volume 14, Issue 9, September 1966, pp. 1136-1138

[11] Y. C. Lin, Ming-Song Chen, JueZhong. Effect of Temperature and Strain Rate on the Compressive Deformation Behavior of 42CrMo Steel. Journal of Materials Processing Technology Vol. 205 (2008) 308-315

[12] Woei-Shyan Lee, Ming-Tong Lin. The Effects of Strain Rate and Temperature on the Compressive Deformation Behavior of Ti-6A1-4V Alloy. Journal of Materials Processing Technology Vol. 71 (1997) 235-246

[13] Jun Cai, Fuguo Li, Taiying Liu, Bo Chen, Min He. Constitutive Equations for Elevated Temperature Flow Stress of Ti–6Al–4V Alloy considering the Effect of Strain. Materials and Design Vol. 32 (2011) 1144-1151

[14] Tareg S. Ben Naser, György Krállics, The Constitutive Behavior of Al 7075 Aluminium Alloy under Hot Compression Test, Materials Science Forum, Vol. 752 (2013) 69-74

[15] Ahmed S. M. Agena, A Study of Flow Characteristics of Nanostructured Al-6082 Alloy Produced by ECAP under Upsetting Test, Journal of Materials Processing Technology, 209 (2009) 856-863

[16] Támas Mikó, Peter Barkóczy, Determination of the Onset of the Dynamic Recrystallization of a 7075 Al Alloy, Material Science Fourm, 752 (2013) 105-114

Economic Aspects of Animal Welfare

Szilvia Vetter*, László Vasa**, László Ózsvári*

*Szent István University, Faculty of Veterinary Science, Department of State Veterinary Medicine and Agricultural Economics, István u. 2, H-1078 Budapest, Hungary, e-mail: szilvia.vetter@maat.hu, ozsvari.laszlo@aotk.szie.hu

**Szent István University, Faculty of Economics and Social Sciences, Institute of Regional Economics and Social Sciences, Páter Károly u. 1, H-2100 Gödöllő, Hungary, e-mail: vasa.laszlo@gtk.szie.hu

Abstract: Economic activities are not conducted in a vacuum; external factors may influence production efficiency, and the activity itself may also result in positive or negative impacts. Although disregarding animal welfare aspects appears to hurt only animals that are harmed as a consequence, banning animal torture is as significant a social interest as combating environmental pollution. Therefore, countries and relevant organisations of certain countries (such as the European Union) regulate via provisions the enforcement of animal welfare aspects with regard to economic activities. Below, economic activities covered by animal welfare regulations and the impacts of animal welfare requirements on economic efficiency will be analysed. In most cases, fulfilling such requirements imposes higher costs for the company at issue and, thus, has an inflationary effect on prices similar to levying a tax. As a consequence, animal welfare regulations generally jolt enterprises from the usual minimum cost-maximum return intersection, so animal protection may appear costly at first. However, in the long term in most cases, they do not bring lower revenues because applying the new – often more expensive – method or technology boosts productivity and because the loss of competitors due to compliance failure may increase the market share for complying companies. The consumers' behaviour is a paradox: on one hand, they are becoming more and more aware of the environmental impacts of their daily lives, whereas on the other hand, concern for the ethical treatment of animals does not always mean changes in purchasing habits. If we look at the production side, animal healthcare statistics prove that the large majority of losses in livestock breeding (mortality, compulsory slaughtering, diseases, poor reproduction and body mass index (BMI) results, medical expenses, etc.) are not caused by obligate pathogens. Most losses are the direct result of diseases due to unfavourable conditions related to animal breeding, feeding and raising or other external factors (power failure, damages from hail, etc.). Through the appropriate keeping and adequate care of animals, sensible animal welfare attitudes and practices may prevent material losses.

Keywords: animal welfare; economics; regulations; negative externality

1 Effects of Animal Welfare: Theoretical Basics and Advantages for Society

Since WWII, animal husbandry has undergone profound change, as demand for food from animals has increased sharply. Traditional extensive farming and breeding has been replaced by intensive, profit-oriented systems. Over-crowdedness, a stressful environment and extreme separation cause behavioural and psychological stress symptoms, which eventually lead to problems that are also economically calculable, such as the lack of controllability and predictability [44]. What was previously part of average farm life is now aggregated into commercial enterprises, which have very little concern for individual animals. Society uses animals in many ways to support our own interests and well-being. The consumption of meat is rising sharply worldwide. Particularly fast growth can be observed in the demand for poultry meat, which has consistently increased at approximately three times the rate of population growth over each of the past five decades. Growth of world egg and milk production is less drastic, but the trend is evident here as well (Figure 1, [20]). The food industry plays a significant role in the countries' national economies. All in all, concern for animals is evident now throughout many societies.
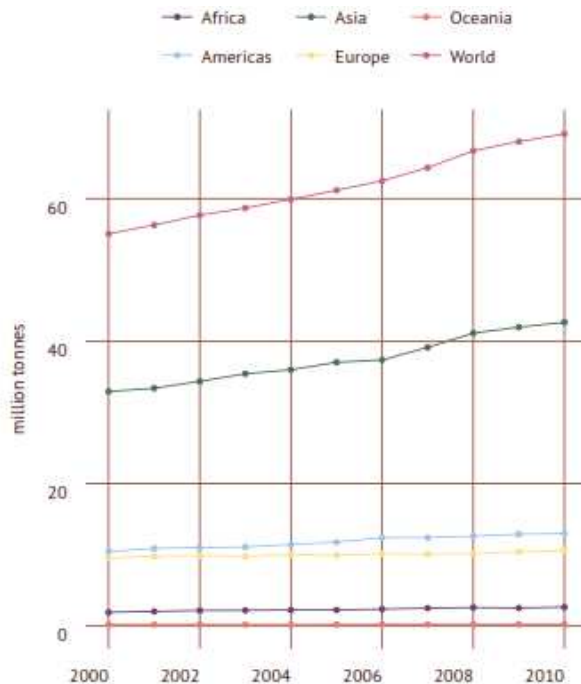

Figure 1
World egg production 2000-2010 [20]

There is a conceptual scale on which the level of animal welfare can be measured. Below a certain point, the welfare standard would be regarded as totally inconsistent with the ethical values of society, and all reasonable people would feel bad if farm animals were treated in such fashion. Animal welfare up to this threshold is therefore a 'public good' — a benefit that government has a responsibility to ensure. At the upper end of the scale are levels of animal welfare that only a minority of people would considers important; the economic value attached to it can be treated as a private good that government has no responsibility to provide [31].

Violation of animal welfare norms, disregarding animal torture and animal welfare – just like any other social and criminal deviations – are not isolated phenomena, and they do not take place in a vacuum. The way we treat animals has a complex relationship with morality, religion, crime, societal development and, last but not least, the economy. Economic principles can and should have an important role when new, market-driven and other approaches are developed to improve farm animal welfare, as there is a limit to the improvements capable of being secured by tightening legal improvements [11].

1.1   Legal Background

The utilitarian view indicates that livestock have primarily a 'use value'. However, societies accept ethical presumptions that confer other values on many features of the biological world. Thus, farm animals can have an additional 'non-use value'.

The main methods for imposing minimum necessary standards are by legislation, regulation and enforcement. These are the only means of ensuring all livestock keepers provide the public good values [31], which is the reason most countries have introduced animal welfare regulations. These regulations differ significantly regarding their contents, forms and execution mechanisms.

In certain countries (e.g., Germany), animal welfare has been made a constitutional issue. Primary legislation belongs to the sphere of competence of the Parliament or the Congress, whereas in the course of secondary legislation, the relevant government bodies frame detailed regulations [6].

Legislation depends also on the traditions and history of a certain country. Austria, the Netherlands, Sweden and Switzerland are among the countries with the most highly developed animal protection systems.

The issue of animal welfare in Hungary has achieved the regulatory status it deserves after the Animal Protection Act was adopted in 1998, and animal torture was defined as a criminal offence in 2004 [1-2]. Hungary's accession to the European Union necessitated the introduction and application of the EU regulatory system, which prompted substantial changes in domestic breeding technologies.

Animal welfare has great significance in the European Union, partly due to the love of animals and the undisputable economic weight of the issue. Since the 1970s, several treaties have been adopted in Europe aimed at safeguarding the living conditions of animals on the grounds of the Five Freedoms of Animal Rights principle [43]. Although the five freedoms were originally developed from a UK Government report on livestock husbandry in 1965, the international animal welfare legislation is based primarily on the "Five Animal Freedoms" set out by Britain's Farm Animal Welfare Council in 1979 [19]. On its grounds, animals must be protected from:

- hunger and thirst;
- pain, injuries and diseases;
- fear and negative stress;
- conditions limiting natural behaviour; and
- discomfort due to insufficient space, improper facilities or overheating.

In 1978, the European Council concluded a treaty on the protection of farm animals. Animals shall be treated in accordance with their natural needs and causing any unnecessary pain shall be avoided [12]. In the EU, time and again, new regulations are framed to be adopted and implemented. These aim to broaden and more precisely define animal welfare regulations, on one hand, and contain concrete requirements for law enforcers, stockbreeders and enterprises. Special animal welfare regulations have been framed, among others, for the caging of egg-laying hens [13] and the keeping of calves [15] and pigs [14]. If the provisions – as key determining factors of supply and demand – are modified and alter production figures, they will also result in a supply-demand shift; market prices and production quantity will change, which will subsequently impact society.

Concerning the issue of the necessity of animal welfare regulation, the key concern is not first and foremost intentionally inflicting pain, as such acts are atypical, but activities motivated by false interpretations of economic or cost efficiency. With regard to stockbreeding, the key problems may be the following:

a) Confinement and an environment lacking stimulation, with subsequent effects, such as deformed feet or burns due to ammonia.

b) Over-crowded environment, with all its adverse effects. The British ethologist John Calhoun examined the effects of an over-crowded environment in experiments with white rats, which showed dramatic alterations as a consequence: aggression, sadism, unrestrained sexuality and a high mortality rate [37].

c) Physical abnormalities resulting from forced maturing (i.e., heart and lung deficiencies).

d) Constant restlessness (i.e., animals cannot relax).

e) Marked separation versus crowdedness. Separation deprives animals of advantages resulting from social bonding, which can lead, for example, to aberrational behaviour and diseases associated with breeding [21].

The most uncomplicated definition of welfare focuses on an animal's bodily functions and its reactions; accordingly, the welfare of animals is optimal when they do not display signs of stress [9]. The stress of farm animals has grave consequences also on the entire stock, as reproduction rates decline, production indicators deteriorate and production costs rise. Therefore, the timely recognition of stress and the introduction of countermeasures are essential.

1.2 Violation of Animal Welfare Regulations as a Special Negative Externality

The expression 'externality' (external effect) means that an activity by an economic stakeholder unintentionally and without legal consequences influences the position of another economic stakeholder. This also implies that externalities impact not only the costs or profits of the stakeholder at issue but also everybody else around him. Externalities may be categorised on the basis of their effect; they can be positive or negative, depending on their impact [10].

Environmental pollution is considered a negative externality, being an unwanted by-product of profit-oriented processes that has grave social consequences. By definition, environmental pollution is a development resulting from an activity or process that alters the composition or mechanism of environmental components (water, air, etc.). For example, if the quantity of some foreign substance in the air is already harmful for living beings and/or the share of original components (i.e., oxygen, nitrogen) changes significantly, then it is called air pollution. Environmental pollution may be physical (e.g., noise pollution), chemical (e.g., soil pollution) or biological (e.g., GMO) [25].

Although the failure to take animal welfare into consideration appears only to harm animals that fall victim as a consequence, banning animal torture is actually in the interest of the entire society, just like combating environmental pollution. In addition to moral reasons, this proposition also has a legal foundation, as the ultimate object that animal welfare regulation aims to safeguard is not animals but human beings and society. Hungary's Criminal Code, for example, currently regulates the crime of animal torture in Chapter XVI, among crimes against the public order and within that, among crimes against public safety [1]. Consequently, as animals are not legal entities, instead of the victimised animals, the regulations qualify the norms of social coexistence as the victim. Therefore, breaching animal welfare regulations and environmental pollution have something in common: both of them are negative externalities.

The violation of animal welfare regulations must, as a matter of fact, be listed among special negative externalities; as opposed to several cases of environmental pollution, its social impact – which is mainly a moral one – is hard to put down in numbers, and its economic damages are indirect. (Such a factor is, for example, the impact of reducing animals' suffering.) Furthermore, the reverse is true:

animal welfare is evidently a 'public good' externality, and there is an obvious role for government policy in establishing and enforcing standards. Farm animal welfare provides an economic value that is not adequately handled through the normal market processes surrounding livestock farming [24, 31].

Another issue worth mentioning is that levying per-unit taxes on environmental pollution has proven to be ineffectual in practice, as evaluating the damage caused faces serious obstacles. Due to the high number of variables and their complicated interactions in nature, a precise scientific evaluation of ecological alterations and their subsequent damages in financial terms is often impossible. Thus, the notion is becoming widespread that instead of a complicated damage assessment doomed to fail anyhow, there should be a socio-political consensus on financing the improvement of environmental indicators [30]. Thus, one can once again draw a parallel between environmental pollution and the violation of animal welfare regulations.

2   Effects of Animal Welfare: Economic Advantages and Disadvantages for Enterprises

"People, planet, profit", also known as the triple bottom line, are the key factors that should be practiced in every move a company makes. 'People' refers to fair business practices toward the community. 'Planet' refers to sustainable environmental practices and environment-friendly solutions. 'Profit' is the economic value created by the organisation after subtracting the cost of all inputs [42]. From the point of view of economic analysis, farm animals are simply one of the resources of livestock production, subject to the same considerations as all other resources.

Before Hungary joined the European Union, it was a widely held view in the country that complying with animal welfare regulations and implementing related investment projects do not result in extra profits for farmers, as these have no direct economic advantages. However, even before EU accession took place, certain foreign markets only allowed the sale of goods produced in accordance with these regulations at prices that included domestic costs. As a result, compliance with animal welfare regulations became a precondition of entering a market. In addition, several large food chains (e.g., McDonald's) pay special attention to the welfare of animals from which their products are made, and thus, they oblige their suppliers to observe a number of extra quality or technological requirements. Producers often endeavour to comply with the subscriptions of an animal welfare quality standard system because such systems increase the value of marketable products. Complying with certain standards is often rewarded with trademarks by food safety organisations (e.g., Global Animal Partnership, Tierschutz Geprüft). In Hungary, unfortunately, no such system has been introduced [39].

2.1   Paradoxes and Difficulties

2.1.1   Welfare-Productivity (McInerney) Model

The welfare-productivity model shows that beyond a point, higher welfare standards involve some sacrifice in livestock productivity cost. The shape of this curve shows that basic welfare improvements can be gained at little cost but moves towards 'high' welfare standards become increasingly expensive.

Figure 2 shows the relationship between livestock productivity (and human benefit indirectly) on the horizontal axis and the level of animal welfare on the vertical axis. The point labelled 'A' represents an initial point. As inputs are given to raise the standard of welfare, economic productivity increases as well. This is only true up to point 'B'. From that point, increasing intensity of production is associated with a decreasing standard of animals' perceived welfare. Point 'C' refers to the level of animal welfare that equals the initial standard but is associated with a much higher level of productivity. Point 'D' is the point where the treatment of animals becomes unacceptable by society. If this process is pursued far enough, it is likely that point 'E' will be reached, where the animals are driven to their limits and the system collapses [31]. It would be very favourable if every farmer could calculate where the figure's turning points are located in the case of his own livestock. Unfortunately, this is a conceptual model that originated from general principles that are almost impossible to quantify.
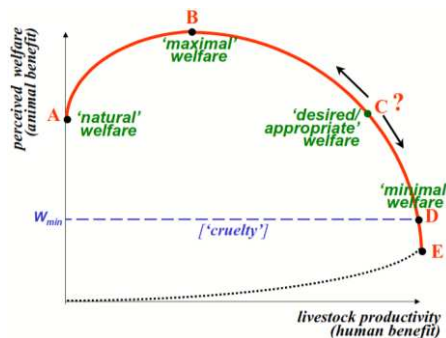
Figure 2
Welfare-productivity (McInerney) model

2.1.2   Aspect of Consumers: Sales Side

Contrary to popular belief, in the developed countries, price is not the primary determinant behind food purchases any more than it is with cars, clothes or any other goods. People seek not the cheapest food items but those with characteristics they want, so they are searching for maximum consumer benefit. There is an exception: when choosing among identical items (eggs, long-life milk, etc.),

consumers will rationally select those with the lowest prices. The quality characteristics associated with food can be real or imagined as well. One of the quality characteristics becoming more and more important to the consumer is the welfare of livestock [24, 31].

Examining the related surveys and statistics about consumers' behaviour, we find a paradoxical situation. On one hand, there is a clear tendency: consumers do care about farm animal welfare in connection with brand reputation. The rise of ethical consumerism over the last two decades has been observed. As global population increases, the limited natural resources are more and more important and valuable [23]. As consumers become more aware of the environmental impacts of their daily lives, hopefully this tendency is the beginning of an ethical consumerism-based society. According to some recent research [40], consumers are concerned with animal welfare, even if they do not necessarily change their eating habits as a result. Seventy-four percent of European citizens think that buying animal welfare-friendly products could have a positive impact on the protection of farm animals. A majority of European Union citizens (55%) state that animal welfare/protection does not receive enough importance in the agricultural policy of their individual countries [18].

Unfortunately, on the other hand, a slight majority of European Union citizens does not take animal welfare into consideration when buying food. (These rates are much worse in the Visegrad 4 (V4) countries; for example, 51% of consumers in Poland never think about the welfare of the animals when purchasing meat. In contrast, nearly two-thirds of Swedes (67%) and Luxembourgers (64%) seem concerned by the conditions under which these animals are reared and consider this when purchasing meat.)

It is also important to add that animal welfare improvements do not increase sales in all cases, but they reduce the chance of a sales loss [26].

Nonetheless, as a result of increased consumer interest, a number of large companies are including animal welfare in their Corporate Social Responsibility (CSR) commitments. Although apparently CSR distracts from the economic role of businesses, it usually helps firms making more long-term profits. Businesses may not be looking at short-run financial returns when developing their CSR strategy [4]. According to a Europe-wide survey [18], the majority of citizens in the V4 countries cannot identify from the label whether the product is sourced from an animal welfare-friendly production system. These identifications are much easier on the other side of Europe, especially in the Germanic and Scandinavian countries. Marketing efforts and CSR are helping attach the label of animal friendliness to certain products or firms, so consumers can distinguish between different products more easily.

It is quite easy to recognise the negative effects of animal welfare regulations on profits because the extra costs speak for themselves. In the long term, it is harder to determine the measurable, quantifiable positive effects of the improved

infrastructure because it is not easy to decide which factor (i.e. more effective medicines, better quality feed, more effective equipment or improved animal welfare) causes the enhanced results. In any case, it can be stated that in the European Union, most of the efficiencies (mortality rate and live births of the livestock, milk production per cow, etc.) have shown progress in recent decades paralleling tighter regulation of animal welfare both at the Community and Member State levels. For example, in Hungary between 2000 and 2010, milk production per cow rose from 5335 litres to 6696 litres, the rate of live-born animals rose from 78% to 84%, and the mortality rate decreased from 5.2% to 4.6% [28]. These changes are not specific to Hungary, as the average cow's milk yield increased by 20% within ten years across the EU-27. This yield was approximately 6 692 kg per dairy cow in 2011 (5 585 kg in 2001); the range extended from more than 8 000 kg per cow in Denmark, Spain, Finland and Sweden to less than 4 000 kg per cow in Romania and Bulgaria [32].

Moreover, experiments focusing on different elements of the connection between animal welfare and productivity show clear correlation related to the investigated phenomenon.

2.2 The Negative Effects of Animal Welfare Measures from the Aspect of Profits

Changes in farm animal welfare standards impact production costs first. Animal welfare regulations mostly require alterations that carry substantial extra costs. Therefore, statutory deadlines for adaptation are often quite long. Even under such conditions, enterprises are usually against the changes, as they believe that the extra production costs incurred due to animal-friendly technologies make employees feel unwanted and inflate the selling price of a product. Farmers would rather focus on immediate productivity instead of non-market externalities such as animal welfare.

The facilities that have failed to comply with the requirements or did not even want to implement them must eventually be closed down. In addition, extra costs may easily cause some enterprises to go bankrupt. It has been an employment policy concern that several enterprises are under-capitalised and will probably go bankrupt, as they cannot provide even the basic funds required for the improvement. This, on the other hand, means a larger market share for competitors that comply with the requirements.

The EU decree of 1999 on egg-laying hens [13] and the Hungarian regulations [34-35] on its implementation are an ample example of this. In the 1990s, the egg-laying hen breeding technologies of that period were often criticised from the aspect of animal welfare. Several proposals and studies had been made on the protection of egg-laying hens, and the European Commission concluded that animal welfare conditions were inadequate with regard to breeding hens in battery

cages and other rearing systems. The regulation stated that its provisions should be applied by facilities that keep more than 350 egg-laying hens and have hens producing eggs for commercial sale. The regulation [13] set out provisions that are more restrictive than the former one regarding the minimum space per animal, the dimension of perches and other technological factors essential for satisfying the physiological-ethological needs of animals. The regulations stipulated that unaltered cages be banned as of 1 January 2012 in the EU, meaning that farmers had 12 years to adjust their systems. Even so, similar to several other EU member countries, in Hungary, not everyone could replace the cages, prompting the EU to adopt an action plan allowing a seven-month transitory period for final compliance. In accordance with this extension, Hungary endeavoured that as of 1 January 2012 through 31 July 2012, traditional cages would be permitted only in facilities that have commenced the transformation procedure but in such cages, a minimum space of 750 square centimetres per hen shall be provided.

The costs of change due to animal welfare regulation may be calculated either in absolute figures or percentages. In the aforementioned example, at a facility formerly with a stock of 10,000 hens, only up to 6,200 animals are permitted to be bred according to the new regulation; this means that either costs per hen rise significantly or space must be increased by 37 percent [33]. These changes must eventually appear in the price of eggs; according to the 2005 Statistical Yearbook of Hungary [28], a minimum price increase of 7.4 percent was observed, and prices were nearing that level at the time when the regulation entered into effect. However, based on the Hungarian Central Statistical Office's flash report, in January 2012 [27], egg prices were up 22.5 percent in comparison to the previous year (it must be noted that prices are also influenced by the overall economic situation).

A study in March 2012 by the European Social and Economic Committee outlined conclusions in light of the assessment of relevant EU policy, revealing that welfare regulations result in higher costs for stockbreeding and animal experiments [33]. The study also concludes that despite all efforts so far, consumer decisions are basically determined by prices, and animal welfare often plays no part or is only one factor among many influencing the product choice. EU animal welfare regulations are difficult to implement, as according to the Commission, they do not enhance producer competitiveness. Stockbreeders are having a tough time anyway, and the extra costs of complying with the regulations exacerbate their woes. Higher costs and the lack of implementation subsidies are aspects of the current policy that undoubtedly deserve to be improved. The document makes it clear that EU animal welfare policy must be more market-oriented. It is crucial that producers recover their extra costs and that consumers be aware of their responsibility and are willing to pay for food produced in line with the European model. In the coming period, the financing of EU animal welfare policy must be boosted by an extent that keeps up with the increasing demands of this policy and meets the claims outlined in the study [16, 33].

2.3 Positive Effect of Animal Welfare Regulations on Individuals and Stocks: Increased Productivity

The key objective of every economic unit, including breeding facilities, is to increase profits. Stockbreeding enterprises are rational economic stakeholders with logical goals. For them, of the numerous options, the most profitable one is the most desirable. Production statistics and costs of a breeding facility are significantly influenced by the animal health indicators of their stocks [8].

Animal healthcare statistics also prove that the large majority of losses of livestock breeding (mortality, compulsory slaughtering, diseases, poor reproduction and body mass index (BMI) results, medical expenses, etc.) are not caused by obligate pathogens. Most losses are the direct result of diseases of animals due to unfavourable conditions related to animal breeding, feeding and raising or other external factors (power failure, damages from hail, etc.). Through the appropriate keeping and adequate care of animals, a sensible animal welfare attitude and practices may prevent material losses from being far larger than expenditures.

Adjustment to the environment requires adaptation energy from animals, which may reduce their performance. The life processes of farm animals are also genetically pre-coded, and technologies should in every case serve the needs of animals by adapting to them. Most breeding technologies, however, cannot ensure conditions that benefit the genetically determined life processes of animals because these must take several other aspects into consideration. Farm animals are capable of tolerating modified conditions up to a certain point, but technological development may reach a level with which animals are unable to cope [7]. The animal's general wellbeing would be affected, and it may become anxious and even feel pain.

Generally speaking, performance is an indicator of contentment. Stress adversely affects the processing of fodder, production of milk and eggs, weight gain and reproductive indicators as well; consequently, its impact on farm animals is negative. Only animals that are content are capable of reproduction and delivering excellent production figures. (It may happen, as a rare exception, that top performing animals are deprived of basic living conditions, as in the case of battery cage systems for egg-laying hens. Although the conditions for an animal to produce 300+ eggs per year are in place, its wellbeing is quite questionable [7]).

Diseases may hamper the production of an animal stock by draining production resources and/or limiting production output (direct effects). The term 'disease' refers to aggregate harmful processes resulting from the disruption of homeostasis, which may either have overt (i.e., death) or covert effects (e.g., reduced milk production). Harmful processes may cause the following calculable losses from production:

- Weaker reproduction figures are one of the most significant loss factors;

- Death;

- Deterioration of quantity or quality of animal produce; and

- Losses due to lower capacity utilisation [8].

The animals' diseases may also affect other elements of the economic system (indirect effects). Examples of this phenomenon are slower growth of the agricultural sector and export restrictions [8].

Accordingly, the following stock indicators are applicable for assessing the performance of a stockbreeding facility:

- Death rate at the facility;

- Frequency of certain diseases or the subsequent death rate;

- Maturation rate;

- Reproduction indicators; and

- Life expectancy [39].

Consequently, regarding the planning of a breeding technology, aspects that may sometimes appear irrelevant or costly also must be taken into account because they eventually contribute to efficient production. Although there is no European Council directive about dairy cows especially, there are protocols relevant in the non-regulated questions. The Welfare Quality Assessment Protocol for Cattle [41] shows standardised ways to develop animal welfare in the case of dairy cattle. Ease of movement is one of the three welfare criteria of the good housing of the animals; they should have enough space to be able to move around freely. In the "Expression of other behaviours" part of the same Protocol, the hours spent at pasture are taken into consideration; the more hours the animal spends at pasture, the higher animal welfare score the farm reaches. In accordance with these criteria, there are several experiments that underline the advantages of open yard housing for cattle: better reproduction index [5], significantly less udder infection and less mastitis disease, and fewer animals falling out of production [22]. Open yard housing not only contributes significantly to the welfare of cattle but also provides conditions favourable to extending the useful lifetime of animals. Furthermore, we are back to the profit-oriented point of view as well. The long useful lifetime of animals is one of the key components of economical production [36].

The successful welfare-based livestock farm should meet four criteria of economic success: it should maintain or improve levels of health, improve the economics of the production system, be practical to employ, be sustainable, but not least, it should increase species-specific behaviour [38].

The measures and changes taken to fulfil the animal welfare requirements are often accompanied by technological improvements and modernisation, which can optionally keep a husbandry company competitive. In Hungary, the production of turkey meat has begun to decrease drastically since 2010. While purchase prices have risen slightly, they have not been able to offset the growth of energy and feed prices. The majority of the market participants have therefore suffered a loss in 2012, except those who had already modernised their ventilation, feeding and water supply systems [3].

These indicators shed light on certain aspects of wellbeing; therefore, multiple assessments and their evaluation can help determine the wellbeing of an individual animal. However, associations identified at the animal level and in the experimental setting might not appear at the farm level and in common practice; therefore, De Vries et al. [17] investigated the associations between variables of routine dairy herd data and the welfare indicators used in the Welfare Quality Assessment Protocol for Cattle to estimate the levels of animal welfare in dairy farms. Their conclusion was that cross-sectional studies using integrated welfare scores at the farm level are needed to more accurately determine the potential for variables of routine herd data to estimate animal welfare on dairy farms.

Conclusions

The economic effects of animal welfare regulations on a company could be examined basically from two points of view: from the aspect of productivity changes or from the aspect of the consumers — the sales' side.

In most cases, the rates of productivity are improving, especially in the long run. Species that are allowed to live according to their nature are healthier, live longer and produce more — a win-win situation for the owner, the animal and society.

From the sales' side, there are difficulties that must be overcome. Public interest towards animal welfare is rising but – because of financial reasons or the lack of information – in most cases, consumers do not buy more animal-friendly products, especially because they are more expensive.

Animal welfare regulations are slowly but surely re-organising livestock production and therefore, the food market itself. In the long term, only some of the market participants are able to remain competitive, whereas the others fall out of the market. However, hopefully the satisfaction of consumers and the remaining companies will rise along with the animals' better living conditions.

Although natural ecosystems and modern-day, large-scale market systems are likely not able to coexist in the long term, partial improvements are possible by developing market methods. Experts of environmental economics are of the opinion that progress can be made by raising public awareness of the problems Nature is facing [29]. Paying attention to animal welfare, including creating more restrictive animal welfare regulations, belongs to this stream of thought.

References

[1]   Act IV. of 1978. on the Penal Code

[2]   Act XXVIII of 1998 on the Protection and Humane Treatment of Animals

[3]   Agrárgazdaságtani Kutatóintézet: A hazai pulykavertikum egy évtizedes fejlődési pályájának értékelése [Evaluation of the development of the Hungarian turkey sector in the last decade] Agrárgazdasági Figyelő, Vol. 5, 2013, pp. 2-20

[4]   Baker, M.: Corporate social responsibility - What does it mean? 2004, http://www.mallenbaker.net/csr/definition.php

[5]   Bakken, G., Ron, I. & Osteras, O.: Clinical disease in dairy cows in relation to housing systems. Environment and animal health. Proceedings of the 6th International Congress on Animal Hygiene, 14-17 June 1988, Skara, Sweden. Vol. 1, pp. 18-22

[6]   Basic Law for the Federal Republic of Germany

[7]   Béri, B.: Tartástechnológia. [Housing systems] Debreceni Egyetem, Nyugat-Magyarországi Egyetem és Pannon Egyetem, Debrecen, Hungary, 2011

[8]   Bíró, O. & Ózsvári, L.: Állat-egészségügyi gazdaságtan. [Animal Health Economics] Szent István Egyetem, Budapest, 2006

[9]   Broom, D. M. & Johnson, K. G.: Stress and animal welfare. Chapman and Hall, London, 1993

[10]   Caplan, B.: Externalities. 2003, http://www.econlib.org/library/Enc/Externalities.html

[11]   Christensen, T., Lawrence, A., Lund, M., Stott, A., Sandoe, P.: How can economists help to improve animal welfare? Animal Welfare, Vol. 21, 2012, pp. 1-10

[12]   Council Decision 78/923/EEC of 19 June 1978 concerning the conclusion of the European Convention for the protection of animals kept for farming purposes

[13]   Council Directive 1999/74/EC of 19 July 1999 laying down minimum standards for the protection of laying hens

[14]   Council Directive 2001/88/EC of 23 October 2001 amending Directive 91/630/EEC laying down minimum standards for the protection of pigs

[15]   Council Directive 2008/119/EC of 18 December 2008 laying down minimum standards for the protection of calves

[16]   Council Regulation (EC) No 1290/2005 of 21 June 2005 on the financing of the common agricultural policy

[17]   De Vries, M., Bookers, E. A. M., Dijkstra, T., Van Shaik, G., De Boer, I. J. M.: Invited review: Associations between variables of routine herd data and dairy cattle welfare indicators. J. Dairy Sci. Vol. 94, 2011, pp. 3213-3228

[18]   European Commission: Attitudes of consumers towards the welfare of farmed animals. Special Eurobarometer, Brussels, 2005

[19]   Farm Animal Welfare Council: Five freedoms, 2013, http://www.fawc.org.uk/freedoms.htm

[20]   Food and Agriculture Organisation of the United Nations: Statistical Yearbook, Rome, 2013

[21]   Gere, G. & Csányi, V.: Gazdasági állatok viselkedése. Általános etológia. [The behaviour of farm animals, General aspects of ethology] Mezőgazdasági Szaktudás Kiadó, Budapest, 2001

[22]   Grabowski, R., Empel, W. & Zdziarski K.: The influence of housing and feeding systems on health, longevity and life-time productivity of dairy cows. Book of abstracts of the 48th Annual meeting of the European Association for Animal Production, Vienna, Austria, 25-28 August 1997

[23]   Grace, D. & Cohen, S.: Business Ethics: Australian Problems and Cases. Oxford University Press, Oxford, 2005

[24]   Hanzséros, F.: Az állatjóllét és az állatvédelem hatása az állattenyésztésben. [The impact of animal welfare and wellbeing on animal-breeding] Animal welfare, ethology and housing systems, Vol. 1, 2005, pp. 14-19

[25]   Horváth, I.: A környezetvédelem fogalma, környezetvédelmi ismeretek jelentősége a gyakorlatban. [The definition and the practical importance of knowledge of environmental protection] Nemzeti Fejlesztési és Szakképzési Intézet, Budapest, 2008

[26]   Hughes, D.: Animal Welfare: The Consumer and the Food Industry. British Food Journal, Vol. 97 (10) 1997, pp. 3-7

[27]   Hungarian Central Statistical Office: Mezőgazdasági termelői árak, 2012. január. [Agricultural producer prices, January 2012] 2013, http://www.ksh.hu/docs/hun/xftp/gyor/mar/mar21201.pdf

[28]   Hungarian Central Statistical Office: Statistical Yearbook of Hungary, 2011. Hungarian Central Statistical Office, Budapest, 2012

[29]   Kerekes, S. & Szlávik, J.: A környezeti menedzsment közgazdasági eszközei. [The economic tools of environmental management] Közgazdasági és Jogi Könyvkiadó, Budapest, 1996

[30]   Kerekes, S.: A környezetgazdaságtan alapjai. [The basics of environmental economics] 1998, http://mek.oszk.hu/01400/01452/html/

[31]   McInerney, J.: Report on a study undertaken for the Farm & Animal Health Economics Division of Defra, 2004,

http://archive.defra.gov.uk/evidence/economics/foodfarm/
reports/documents/animalwelfare.pdf

[32]   Office of the European Union: Agriculture, Fishery and Forestry Statistics. Main Results 2010-2011. Eurostat Pocketbooks. Luxembourg, 2012

[33]   Opinion of the European Economic and Social Committee on the Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the European Union Strategy for the Protection and Welfare of Animals 2012-2015 COM(2012) 6 final

[34]   Regulation of Ministry of Agriculture and Rural Development (MARD) 139/2004 (IX.24.)

[35]   Regulation of Ministry of Agriculture and Rural Development (MARD) 86/2005 (IX.27.)

[36]   Strandberg, E.: Breeding for longevity in dairy cows. In: Progress in Dairy Science, USIP Press Books, USA, 1996

[37]   Szinák, J. & Veress, I.: Ismeretlen ismerősök. [Unknown acquaintances] Kozmosz könyvek, Budapest, 1980

[38]   Van De Weerd, H. A. & Day, J. E. L.: A review of environmental enrichment for pigs housed in intensive housing systems. Applied Animal Behaviour Science, Vol. 116, 2009, pp. 1-20

[39]   Végh, Á.: ÁROP-2216 - Jogalkalmazás javítása a mezőgazdasági szakigazgatásban 2. 14. számú tananyag. [How to improve the implementation of legal provisions in the state agricultural administration. Curriculum No. 14] Nemzeti Élelmiszerlánc-biztonsági Hivatal, Budapest, 2012

[40]   Velde, H. T., Aarts N. & Woerkum C. V.: Dealing with Ambivalence: Farmers' and Consumers' Perceptions of Animal Welfare in Livestock Breeding. Journal of Agricultural and Environmental Ethics, Vol. 15. 2002, pp. 203-219

[41]   Welfare Quality® Consortium: Welfare Quality® assessment protocol for cattle, Lelystad, Netherlands, 2009

[42]   Willard, B.: The Sustainability Advantage: Seven Business Case Benefits of a Triple Bottom Line. New Society Publishers, Canada, 2002

[43]   World Society for the Protection of Animals. 2013, http://www.wspa-international.org/Images/Module_2_Welfare_assessment_and_the_Five_Freedoms_FINAL_tcm25-17558.pdf

[44]   Zayan, R.: Social space for domestic animals. Martinus Nijhoff Publishers, Dordrecht, 1985, pp. 3-11

Two-Area Power System Stability Improvement using a Robust Controller-based CSC-STATCOM

Sandeep Gupta, Ramesh Kumar Tripathi

Department of Electrical Engineering
Motilal Nehru National Institute of Technology Allahabad - 211004, India
E-mail: ree0951@mnnit.ac.in, rktripathi@mnnit.ac.in

Abstract: A current source converter (CSC) based static synchronous compensator (STATCOM) is a shunt flexible AC transmission system (FACTS) device, which has a vital role in stability support for transient instability and damping support for undesirable inter-area oscillations in an interconnected power network. A robust pole-shifting based controller for CSC-STATCOM with damping stabilizer is proposed. In this paper, pole-shifting controller based CSC-STATCOM is designed for enhancing the transient stability of two-area power system and PSS based damping stabilizer is designed to improve the oscillation damping ability. First of all, modeling and pole-shifting based controller design, with damping stabilizer for CSC-STATCOM, are described. Then, the impact of the proposed scheme in a test system with different disturbances is demonstrated. The feasibility of the proposed scheme is demonstrated through simulation in MATLAB and the simulation results show an improvement in power system stability in terms of transient stability and oscillation damping ability with damping stabilizer based CSC-STATCOM. So good coordination between damping stabilizer and pole-shifting controller based CSC-STATCOM is shown in this paper for enhancing the power system stability. Moreover, the robustness and effectiveness of the proposed control scheme are better than without damping stabilizer in CSC-STATCOM.

Keywords: CSC; PSS; STATCOM; transient stability; oscillation damping

1   Introduction

The continuous enhancement of electrical loads due to the growing industrialization and modernization of human activity results in transmission structures being operated near their stability restrictions. Therefore, the renovation of urban and rural power network becomes necessary. Due to governmental, financial and green climate reasons, it is not always possible to construct new transmission lines to relieve the power system stability problem for existing overloaded transmission lines. As a result, the utility industry is facing the

challenge of efficient utilization of the existing AC transmission lines in power system networks. So transient stability, voltage regulation, damping oscillations etc. are the most important operating issues that electrical engineers face during power-transfer at high levels.

In the above mentioned power quality problems, transient stability and oscillation damping are the two most important factors during power-transfer at high levels. According to the literature, transient stability of a power system is its ability to maintain synchronous operation of machines, when subjected to a large disturbance [1]. And oscillation damping is the positive damping of electromechanical modes or oscillations among the interconnected synchronous generators in a power system [2]. While the generator excitation system with PSS (power system stabilizer) can maintain excitation control and stability, but it is not adequate to sustain the stability of power system due to faults or overloading near to the generator terminals [1]. Therefore, researchers have been working on this problem for a long time trying to discover a solution. One of the powerful methods for enhancing the transient stability is to use flexible AC transmission system (FACTS) devices [3-5]. But oscillation damping is improved by the use of damping stabilizer [6]. So power system stability is increased with the help of damping stabilizer (PSS) based FACTS devices [7-9]. Even though the prime objective of shunt FACTS devices (SVC, STATCOM) is to maintain bus voltage by absorbing (or injecting) reactive power, they are also competent of improving the system stability by diminishing (or enhancing) the capability of power transfer when the machine angle decreases (increases), which is accomplished by operating the shunt FACTS devices in inductive (capacitive) mode.

In the cited research papers [4, 7, 10-13], different types of these devices and/or damping stabilizer with different control techniques are used for improving transient stability & enhancing oscillation damping. In these, researchers have investigated the co-ordination of PSS and FACTS based controller [7, 11, 13]. So the PSS based FACTS devices are playing an important role for improving the oscillation damping with transient stability. Among these FACTS devices, the STATCOM is valuable for enhancement of power system dynamic stability and frequency stabilization due to the rapid output response, lower harmonics, superior control stability and small size etc. [14]. By their inverter configuration, basic type of STATCOM topology can be realized by either a current-source converter (CSC) or a voltage-source converter (VSC) [14, 15]. But recent research confirms several advantages of CSC based STATCOM over VSC based STATCOM [16, 17]. These advantages are high converter reliability, quick starting, inherent short-circuit protection, the output current of the converter is directly controlled, in low switching frequency this reduces the filtering requirements compared with the case of a VSC. Therefore CSC based STATCOM is very useful in power systems rather than VSC based STATCOM in many cases. Hence, coordination of PSS and CSC-STATCOM can be used for enhancement of power system dynamic stability.

Presently, the most used techniques for controller design of FACTS devices are Proportional Integration (PI) [12], PID controller, pole placement and linear quadratic regulator (LQR) [18]. But, LQR and pole placement algorithms give quicker response in comparison to PI & PID algorithm. LQR controller Gain (K) can be calculated by solving the Riccati equation and K is also dependent on the two cost function (Q, R). So Riccati equation solvers have some limitations, which relate to the input arguments. But pole shifting method does not face this type of problem. So pole shifting method gives a better and robust performance in comparison to other methods.

The main contribution of this paper is the application of proposed pole-shifting controller based CSC-STATCOM with damping stabilizer for improvement of power system dynamic stability (in terms of transient stability and oscillation damping) by injecting (or absorbing) reactive power. In this paper, the proposed scheme is used in two-area power system with dynamic loads under a severe disturbance (three phase fault or heavy loading) to enhance the power system stability and observe the impact of the CSC-based STATCOM on electromechanical oscillations and transmission capacity. Further, the results obtained from the proposed algorithm-based CSC-STATCOM are compared to that obtained from the conventional methods (without CSC-STATCOM device and without damping stabilizer in CSC-STATCOM).

The rest of the paper is organized as follows. Section 2 discusses the circuit modeling & pole-shifting controller design for CSC based STATCOM. A two-area tow-machine power system is described with a CSC-STATCOM device in Section 3. Coordinated design of pole-shifting based CSC-STATCOM with Damping Stabilizer is proposed in Section 4. Simulation results, to improve power system dynamic stability of the test system with & without CSC based STATCOM (and/or damping stabilizer) for severe contingency are shown in Section 5. Finally, Section 6 concludes this paper.

2   Mathematical Modeling of Pole-shifting Controller-based CSC-STATCOM

2.1   CSC-based STATCOM Model

To verify the response of the CSC-based STATCOM on dynamic performance, the mathematical modeling and control strategy of a CSC-based STATCOM are presented. The design of controller for CSC based STATCOM, the state space equations from the CSC-STATCOM circuit are introduced. To minimize the complexity of mathematical calculations, the theory of dq transformation of

currents has been applied in this circuit, which makes the d and q components as independent parameters. Figure 1 shows the circuit diagram of a typical CSC-based STATCOM.
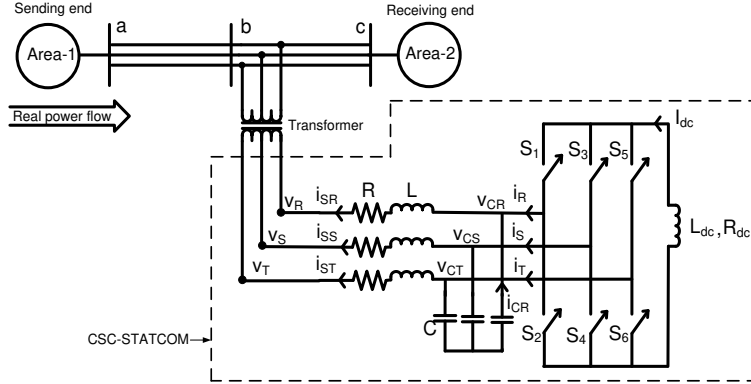

Figure 1

The representation of CSC based STATCOM

Where

i_{SR}, i_{SS}, i_{ST}	line current
v_{CR}, v_{CS}, v_{CT}	voltages across the filter capacitors
v_R, v_S, v_T	line voltages
I_{dc}	dc-side current
R_{dC}	converter switching and conduction losses
L_{dc}	smoothing inductor
L	inductance of the line reactor
R	resistance of the line reactor
C	filter capacitance

The basic mathematical equations of the CSC-STATCOM have been derived in the literature [17]. Therefore, only a brief detail of the test-system is given here for the readers' convenience. Based on the equivalent circuit of CSC-STATCOM shown in Figure 1, the differential equations for the system can be achieved, which are derived in the abc frame and then transformed into the synchronous dq frame using dq transformation method [19].

$$\frac{d}{dt}I_{dc} = -\frac{R_{dc}}{L_{dc}}I_{dc} - \frac{3}{2L_{dc}}M_d V_d - \frac{3}{2L_{dc}}M_q V_q \tag{1}$$

$$\frac{d}{dt}I_d = -\frac{R}{L}I_d + \omega I_q - \frac{1}{L}\frac{E_d}{n} + \frac{1}{L}V_d \tag{2}$$

$$\frac{d}{dt}I_q = -\omega I_d - \frac{R}{L}I_q + \frac{1}{L}V_q \tag{3}$$

$$\frac{d}{dt}V_d = -\frac{1}{C}I_d + \omega V_q + \frac{1}{C}M_d I_{dc} \tag{4}$$

$$\frac{d}{dt}V_q = -\frac{1}{C}I_q - \omega V_d + \frac{1}{C}M_q I_{dc} \tag{5}$$

In above differential equations M_d and M_q are the two input variables. Two output variables are I_{dc} and I_q. Here, ω is the rotation frequency of the system and this is equal to the nominal frequency of the system voltage. Equations (1 to 5) show that controller for CSC based STATCOM has a nonlinear characteristic. So this nonlinear property can be removed by accurately modeling of CSC based STATCOM. From equations (1 to 5), we can see that nonlinear property in the CSC-STATCOM model is due to the part of I_{dc}. This nonlinear property is removed with the help of active power balance equation. Here, we have assumed that the power loss in the switches and resistance R_{dc} is ignored in this system and the turns ratio of the shunt transformer is n:1. After using power balance equation and mathematical calculation, nonlinear characteristic is removed from equation (1). Finally we obtain the equation as below:

$$\frac{d}{dt}\left(I_{dc}^2\right) = -\frac{2R_{dc}}{L_{dc}}\left(I_{dc}^2\right) - \frac{3E_d}{L_{dc}n}I_d \tag{6}$$

In the equation (6) state variable (I_{dc}) is replaced by the state variable (I^2_{dc}), to make the dynamic equation linear. Finally, the better dynamic and robust model of the SATACOM in matrix form can be derived as:

$$\frac{d}{dt}\begin{bmatrix}\left(i_{dc}\right)^2 \\ i_d \\ i_q \\ v_{cd} \\ v_{cq}\end{bmatrix} = \begin{bmatrix} -\frac{2R_{dc}}{L_{dc}} & \frac{3E_d}{L_{dc}n} & 0 & 0 & 0 \\ 0 & -\frac{R}{L} & \omega_o & \frac{1}{L} & 0 \\ 0 & -\omega_o & -\frac{R}{L} & 0 & \frac{1}{L} \\ 0 & -\frac{1}{C} & 0 & 0 & \omega_o \\ 0 & 0 & -\frac{1}{C} & -\omega_o & 0 \end{bmatrix} * \begin{bmatrix}\left(i_{dc}\right)^2 \\ i_d \\ i_q \\ v_{cd} \\ v_{cq}\end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \frac{1}{C} & 0 \\ 0 & \frac{1}{C} \end{bmatrix} * \begin{bmatrix} I_{id} \\ I_{iq} \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{L} \\ 0 \\ 0 \\ 0 \end{bmatrix} * E_d \tag{7}$$

Above modeling of CSC based STATCOM is written in the form of modern control methods i.e. State-space representation. For state-space modeling of the system, section 2.2 is considered.

2.2   Pole-Shifting Controller Design

The pole-shift technique is one of the basic control methods employed in feedback control system theory. Theoretically, Pole shift technique is to set the preferred pole position and to move the pole position of the system to that preferred pole position, to get the desired system outcomes [20]. Here poles of system are shifted because the position of the poles related directly to the eigenvalues of the system, which control the dynamic characteristics of the system outcomes. But for this method, the system must be controllable. In the dynamic modeling of systems, State-space equations involve three types of variables: state variables (x), input(u) and output (y) variables with disturbance (e). So comparing (7) with the standard state-space representation i.e.

$$\dot{x} = Ax + Bu + Fe \tag{8}$$

$$y = Cx \tag{9}$$

We get the system matrices as:

$$x = \begin{bmatrix} I_{dc}^2 & I_d & I_q & V_d & V_q \end{bmatrix}^T ; \; u = \begin{bmatrix} I_{id} & I_{iq} \end{bmatrix}^T ; \; e = E_d ; \; y = \begin{bmatrix} I_{dc}^2 & I_q \end{bmatrix}^T$$

$$A = \begin{bmatrix} -\dfrac{2R_{dc}}{L_{dc}} & -\dfrac{3E_d}{L_{dc}} & 0 & 0 & 0 \\[2mm] 0 & -\dfrac{R}{L} & \omega & \dfrac{1}{L} & 0 \\[2mm] 0 & -\omega & -\dfrac{R}{L} & 0 & \dfrac{1}{L} \\[2mm] 0 & -\dfrac{1}{c} & 0 & 0 & \omega \\[2mm] 0 & 0 & -\dfrac{1}{c} & -\omega & 0 \end{bmatrix} ; \; B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \dfrac{1}{c} & 0 \\ 0 & \dfrac{1}{c} \end{bmatrix} ; \; C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^T ; \; F = \begin{bmatrix} 0 \\ -\dfrac{1}{L} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In above equations (8, 9) five system states, two control inputs and two control outputs are presented. Where x is the state vector, u is the input vector, A is the basis matrix, B is the input matrix, e is disturbance input.

If the controller is set as:

$$u = -Kx + Ty_{ref} + Me \tag{10}$$

Then the state equation of closed loop can be written as

$$\dot{x} = (A - BK)x + Ty_{ref} + BMe + Fe \tag{11}$$

Where $T=(C*(-(A-B*K)-1)*B)-1$ and $M=((C*(-A+B*K)-1*B))-1*(C*(-A+B*K)-1*F)$, these values are find out from mathematical calculation. Here K is the state-feedback gain matrix. The gain matrix K is designed in such a way that equation (12) is satisfied with the desired poles.

$$\left| sI - (A - BK) \right| = (s - P_1)(s - P_2)........(s - P_n) \tag{12}$$

Where P_1, P_2, …..P_n are the desired pole locations. Equation (12) is the desired characteristic polynomial equation. The values of P_1, P_2, …..P_n are selected such as the system becomes stable and all closed-loop eigenvalues are located in the left half of the complex-plane. The final configuration of the proposed pole-shifting controller based CSC-STATCOM is shown in Figure 2.
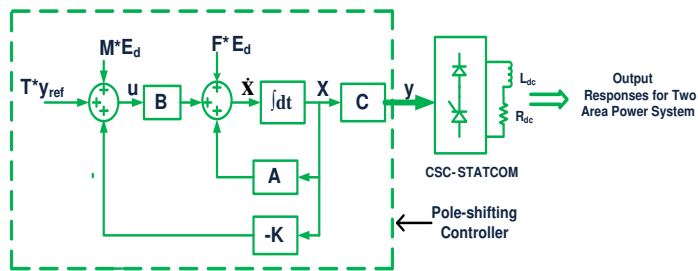
Figure 2

Control Structure of pole-shifting controller based CSC-STATCOM

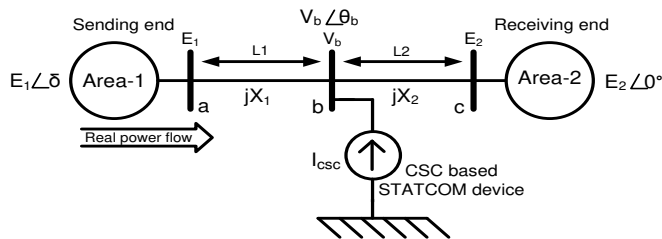# 3   Two-Area Power System with CSC-STATCOM FACTS Device

Figure 3

A single line diagram of two-area two-machine power system with CSC-STATCOM

Firstly consider a two-area two-machine power system with a CSC-STATCOM at bus b is connected through a long transmission system, where CSC-STATCOM is used as a shunt current source device. Figure 3 shows this representation. The

dynamic model of the machine, with a CSC-STATCOM, can be written in the differential algebraic equation form as follows:

$$\dot{\delta} = \omega \tag{13}$$

$$\dot{\omega} = \frac{1}{M}\left[P_m - P_{eo} - P_e^{csc} \right] \tag{14}$$

Here ω is the rotor speed, δ is the rotor angle, P_m is the mechanical input power of generator, the output electrical power without CSC-STATCOM is represents by P_{eo} and M is the moment of inertia of the rotor. Equation (14) is also called the "swing equation". The additional factor of the output electrical power of generator from a CSC-STATCOM is P_e^{csc} in the swing equation. Here for calculation of P_e^{csc}, to assume the CSC-STATCOM works in capacitive mode. Then the injected current from CSC-STATCOM to test system can be written as:

$$I_{csc} = I_{csc} \angle(\theta_{bo} - 90^o) \tag{15}$$

Where, θ_{bo} is the voltage angle at bus b in absentia of CSC-STATCOM. In Figure 3, the magnitude (V_b) and angle (θ_b) of voltage at bus b can be computed as:

$$\theta_b = \tan^{-1}\left[\frac{X_2 E_1 \sin\delta}{X_2 E_1 \cos\delta + X_1 E_2} \right] \tag{16}$$

$$V_b = \left(\frac{X_2 E_1 \cos(\delta - \theta_{bo}) + X_1 E_2 \cos\theta_{bo}}{X_1 + X_2} \right) + \left(\frac{X_1 X_2}{X_1 + X_2} I_{csc} \right) \tag{17}$$

From equation (17), it can be said that the voltage magnitude of bus b (V_b) depends on the STATCOM current I_{csc}. In equation (14), the electrical output power P_e^{csc} of machine due to a CSC-STATCOM, can be expressed as

$$P_e^{csc} = \frac{E_1 V_b}{X_1} \sin(\delta - \theta_b) \tag{18}$$

Finally, using equations (17) & (18) the total electrical output (P_e) of machine with CSC-STATCOM can be written as

$$P_e = P_{eo} + P_e^{csc} \Longrightarrow P_e = P_{eo} + \frac{X_1 X_2 E_1}{(X_1 + X_2) X_1} I_{csc} \sin(\delta - \theta_b) \tag{19}$$

All above equations are represented for the capacitive mode of CSC-STATCOM. For the inductive mode of operation negative value of I_{csc} can be substituted in equations (15), (17) & (19) in place of positive I_{csc}. With the help of equation (14), the power-angle curve of the test system can be drawn for stability analysis as shown in Figure 4.
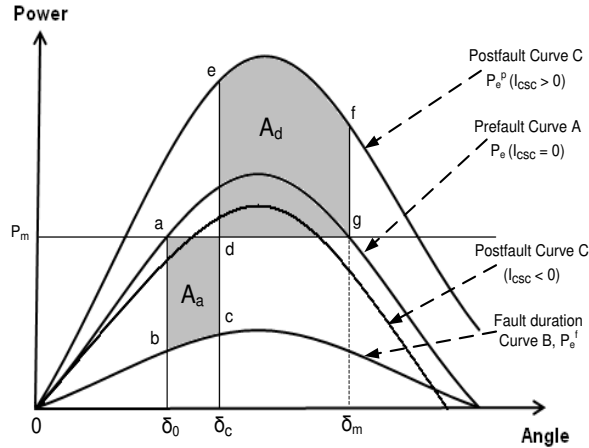
Figure 4
Power-angle characteristic of the test system with a CSC-STATCOM

The power-angle (P-δ) curve of the test system without a CSC-STATCOM is represented by curve A (also called Prefault condition) in Figure 4. Here the mechanical input is P_m, electrical output is P_e and initial angle is δ_0. When a fault occurs, P_e suddenly decreases and the operation shifts from point a to point b on curve B, and thus, the machine accelerates from point b to point c, where accelerating power P_a [= (P_m-P_e)] >0. At fault clearing, P_e suddenly increases and the area a-b-c-d-a represents the accelerating area A_a as defined in equation (20). If the CSC-STATCOM operates in a capacitive mode (at fault clearing), P_e increases to point e at curve C (also called postfault condition). At this time P_a is negative. Thus the machine starts decelerating but its angle continues to increase from point e to the point f until reaches a maximum allowable value δ_m at point f, for system stability. The area e-f-g-d-e represents the decelerating area A_d as defined in equation (20). From previous literature [1], equal area criterion for stability of the system can be written as:

$$\int_{\delta_0}^{\delta_c} \left(P_m - P_e^f \right) d\delta = \int_{\delta_c}^{\delta_m} \left(P_e^p - P_m \right) d\delta \Longrightarrow A_a = A_d \tag{20}$$

This equation is generated from Figure 4, where δ_c is critical clearing angle. P_e^p is an electrical output for post-fault condition. P_e^f is an electrical output during fault condition. From Figure 4, it is seen that for capacitive mode of operation (I_{csc}>0), the P-δ curve is not only uplifted but also displaced toward right and that endues more decelerating area and hence higher transient stability limit. But pole-shifting controller based CSC-STATCOM is not given to sufficient oscillation damping stability. So additional controller with pole-shifting controller based CSC-STATCOM is essential for oscillation damping in the power system. The additional controller is detailed in the next section.

4   Coordinated Design of Pole-Shifting Controller-based CSC-STATCOM with Damping Stabilizer

In this section, a damping stabilizer with pole-shifting controller is proposed for CSC-STATCOM to improve the oscillation damping and transient stability of the system. Modeling of a pole-shifting controller based CSC-STATCOM is explained earlier in section 2.2. So design of a damping stabilizer for pole-shifting controller based CSC-STATCOM is explained in this section. Here PSS-based damping controller is used for damping stabilizer designing. The basic function of power system stabilizer (PSS) is to add damping to the generator rotor oscillations by controlling its excitation using auxiliary stabilizing signals [6]. These auxiliary signals are such as shaft speed, terminal frequency and power to change and adding these output signals of damping stabilizer with a reference signal of pole-shifting controller based CSC-STATCOM. Here coordination between PSS-based damping stabilizer and pole-shifting controller based CSC-STATCOM is very necessary and important.
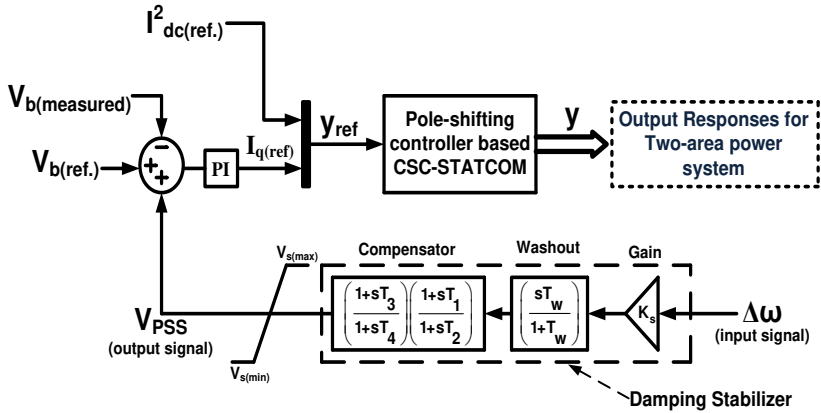

Figure 5

Configuration of damping stabilizer for pole-shifting controller-based CSC-STATCOM

So the damping stabilizer is designed carefully with respect to pole-shifting controller based CSC-STATCOM. A typical structure of damping stabilizer is taken as shown in Figure 5. In this paper, IEEE ST1-Type excitation based PSS is used [1]. The damping stabilizer structure contains one washout block, one gain block and lead-lag compensation block. The number of lead-lag blocks required depends on the power system configuration and PSS tuning. Here the washout block works like as a high pass filter which removes low frequencies from the input signal of the damping stabilizer. The ability of phase lead-lag compensation block is to give the required phase-lead characteristics to compensate for any phase lag between the input and the output signals of damping stabilizer. Hence, transfer function of the damping stabilizer is obtained as follows:

$$S_{out} = K_s \left(\frac{sT_w}{1+T_w} \right) \left(\frac{1+sT_3}{1+sT_4} \right) \left(\frac{1+sT_1}{1+sT_2} \right) S_{in} \qquad (21)$$

Where S_{in} is the damping stabilizer input signal. S_{out} is the damping stabilizer output signal. K_s is the PSS gain. T_w is the washout time constant. T_1, T_2, T_3, T_4 are the compensator time constants. In this arrangement T_w, T_2 and T_4 are generally predefined values. The value of washout time constant (T_w) is not critical issue and may be in the range 1–20 s [1]. The PSS gain K_s and values of T_1 & T_2 are to be found from simulation results and some previous Artificial intelligence techniques based papers [21, 22]. In Figure 5, $V_{s(max)}$ & $V_{s(min)}$ are the maximum & minimum values of damping stabilizer respectively which are predefined values for the test-system. Hence, all the data required for designing of damping stabilizer based controller are given in Appendix 1. In this paper, the input signal of the proposed PSS based damping stabilizer is the rotor speed deviation of two machines (M1 & M2), $\Delta\omega = \omega1 - \omega2$, which is mentioned in equations (13) & (14). Now in the following section the test-system stability in terms of transient stability and oscillations damping ability is analyzed and enhanced using the proposed damping stabilizer based pole-shifting controller with CSC-STATCOM.

5   Simulation Results

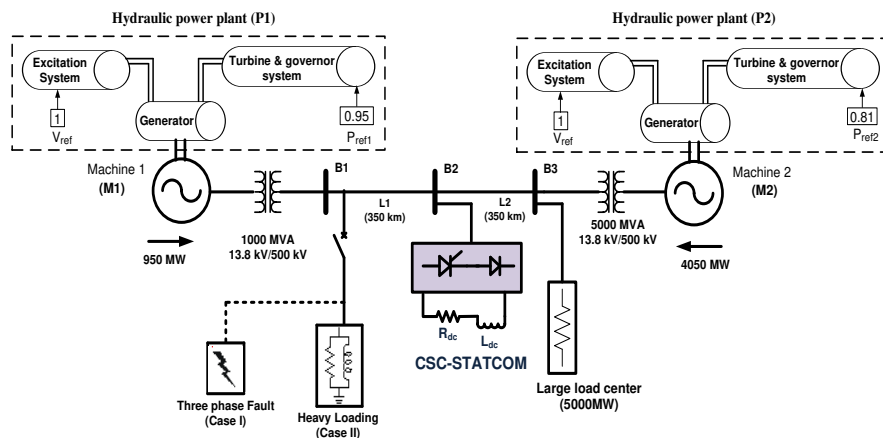## 5.1   Power System under Study

Figure 6

The single line diagram of the test-system model for power system stability study of two power plants

(P1 & P2)

In this section, two-area power system is considered as a test system for study. For this type of test system, a 500kV transmission system with two hydraulic power plants P_1 (machine-1) & P_2 (machine-2) connected through a 700 km long transmission line is used, as shown in Figure 6. Rating of first power generation plant (P1) is 13.8 kv/1000 MVA, which is used as PV generator bus type. The electrical output of the second power plant (P2) is 5000 MVA, which is used as a swing bus for balancing the power. One 5000 MW large resistive load is connected near the plant P2 as shown in Figure 6. To improve the transient stability and increase the oscillation damping ability of the test-system after disturbances (faults or heavy loading), a pole-shifting controller based CSC-STATCOM with damping stabilizer is connected at the mid-point of transmission line. To achieve maximum efficiency; CSC-STATCOM is connected at the mid-point of transmission line, as per [23]. The two hydraulic generating units are assembled with a turbine-governor set and excitation system, as explained in [1]. All the data required for this test system model are given in Appendix 1.

The impact of the damping stabilizer based CSC-STATCOM has been observed for maintaining the system stability through MATLAB/SIMULINK. Severe contingencies, such as short-circuit fault and instant loading, are considered.

5.2   Case I—Short-Circuit Fault

A three-phase fault is created near bus B1 at t=0.1 s and is cleared at 0.23 s. The impact of system with & without CSC based STATCOM (and/or damping stabilizer) to this disturbance is shown in Figures 7 to 14. Here simulations are carried out for 9 s to observe the nature of transients. From Figures 7 to 10, it is observed that the system without CSC-STATCOM is unstable even after the clearance of the fault. But this system with pole-shifting controller based CSC-STATCOM (and/or damping stabilizer) is restored and stable after the clearance of the fault from Figures 9 to 12.
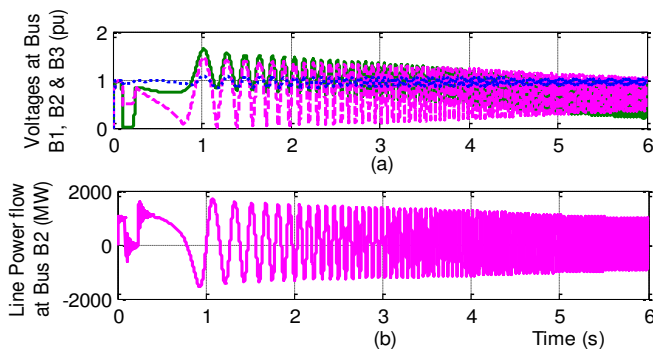

Figure 7

System response without CSC-STATCOM for a three phase fault (Case-I). (a) Positive sequence voltages at different buses B1, B2 & B3 (b) Power flow at bus B2
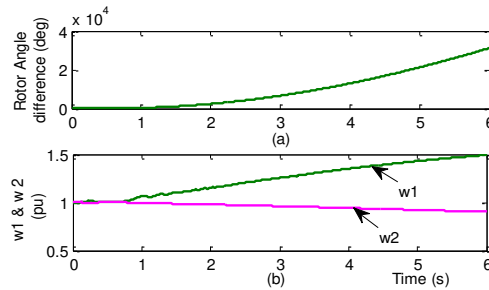
Figure 8

System response without CSC-STATCOM for case-I (a) Difference between Rotor angles of machines M1 & M2 (b) w1 & w2 speeds of machine M1 & M2 respectively

From the responses in Figures 9 and 10, it can be seen that, without damping stabilizer based CSC-STATCOM, the system oscillation is poorly damped and takes a considerable time to reach a stable condition. And with the damping stabilizer based CSC-STATCOM, the oscillation is damped more quickly and stabilized after about 3-4s as shown in Figures 9 to 12. Synchronism between two machines M1 & M2 is also maintained in these figures. The output of the damping stabilizer is shown in Figure 11, which is not rising above their respective limits.
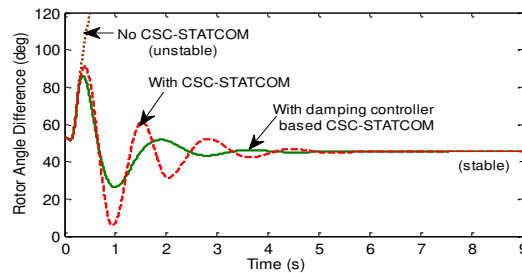

Figure 9

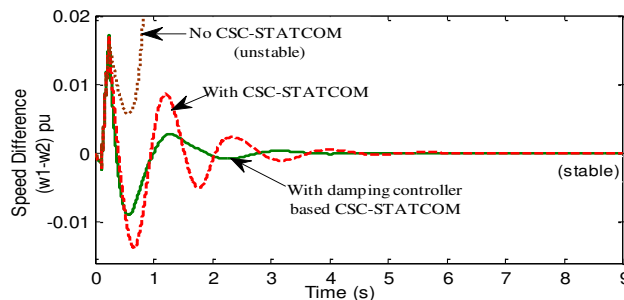Variation of rotor angle difference of machines M1 & M2 for case-I

Figure 10

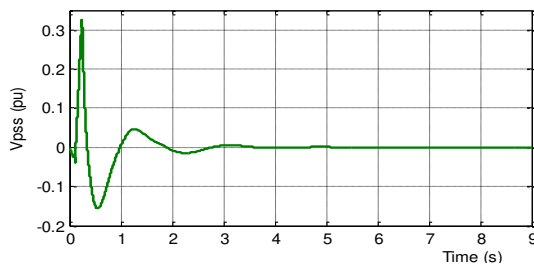Speed difference variation of machines M1 & M2 for case-I

Figure 11

Variation of output signal (V_{PSS}) of damping stabilizer (case-I)
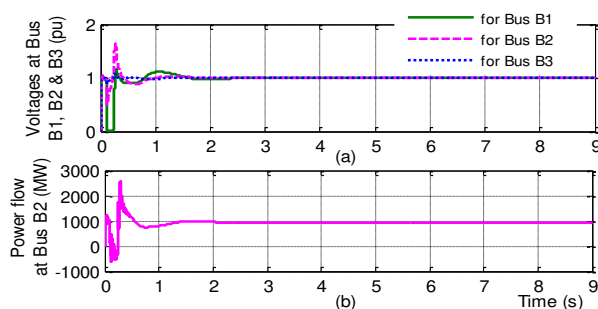

Figure 12

Test system response with damping stabilizer based CSC-STATCOM for a three phase fault (Case-I).
(a) Positive sequence voltages at different buses B1, B2 & B3 (b) Power flow at bus B2

If the fault is applied at t=0.1 and cleared at 0.29 s. Then Figure 13 shows the variation of the rotor angle difference of the two machines for controller without the damping stabilizer and the controller with the damping stabilizer. It is clear that the system without damping stabilizer in CSC-STATCOM is unstable upon the clearance of the fault from Figure 13 & 14. But damping stabilizer based CSC-STATCOM is maintaining the transient stability and oscillation damping ability of the system at this crucial time. CCT is defined as the maximal fault duration for which the system remains transiently stable [1]. The critical clearing time (CCT) of fault is also found out for the test system stability by simulation. CCT of the fault for system with & without CSC-STATCOM (and/or damping stabilizer) are shown in Table I. It is observed that CCT of fault is also increased due to the impact of damping stabilizer based CSC-STATCOM. Clearly, Waveforms show that damping stabilizer based CSC-STATCOM is more effective and robust than that of the system without damping stabilizer based CSC-STATCOM, in terms of oscillation damping, settling time, CCT and transient stability of the test-system.
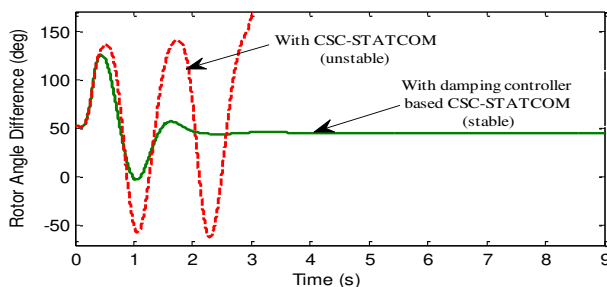
Figure 13

Variation of rotor angle difference of machines M1 & M2 for Case-I (3-phase fault for 0.1s to 0.29s)
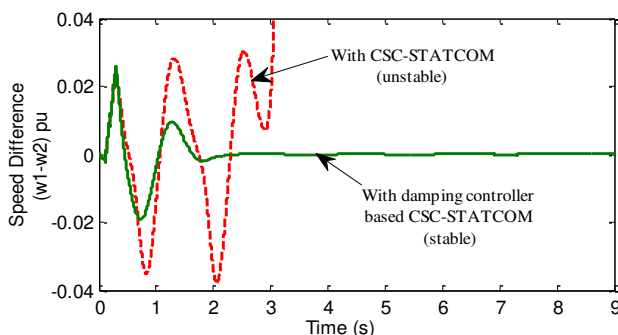

Figure14

Speed difference variation of machines M1 & M2 for Case-I (3-phase fault for 0.1s to 0.29s)

Table I

CCT of disturbances for the system stability with different topologies (Case-I)

S. No.	System with different topologies	Critical Clearing Time (CCT)
1	Without CSC-STATCOM	100 ms – 224 ms
2	With CSC-STATCOM	100 ms – 285 ms
3	With damping stabilizer-based CSC-STATCOM	100 ms – 303 ms

5.3   Case II—Large Loading

For heavy loading case, a large load centre (10000 MW/5000 Mvar) is connected at near bus B1 (i.e. at near plant P1) in Figure (6). This loading occurs during time period 0.1 s to 0.5 s. Due to this disturbance, the simulation results of test system with & without CSC-STATCOM (and/or damping stabilizer) are shown in Figures 15 to 20. Clearly, the system becomes unstable in the absence of the pole-shifting controller based CSC-STATCOM device due to this disturbance as in Figures 15 to 17. But system with pole-shifting controller based CSC-STATCOM (and/or

damping stabilizer) continue to operate under stable condition as observed in
Figures 16 & 17. These figures also show that damping stabilizer based CSC-
STATCOM gives better oscillation damping ability in comparison to without
damping stabilizer in CSC-STATCOM. So that damping stabilizer based CSC-
STATCOM device is preferred. System voltages at different bus B1, B2 & B3
with proposed scheme are shown in Figure 18. Figure 19 represents the output of
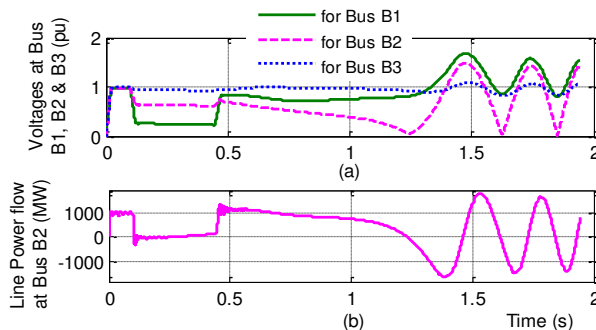the damping stabilizer.

Figure 15

Test-system response without CSC-STATCOM with a heavy loading (Case-II). (a) Positive sequence
voltages at different buses B1, B2 & B3 (b) Power flow at bus B2

Figure 16

Variation of rotor angle difference of machines M1 & M2 in case-II

Figure 17

Speed difference variation of machines M1 & M2 in case-II

Figure 18

Test-system response with damping stabilizer based CSC-STATCOM for a heavy loading (case-II). (a)
Positive sequence voltages at different buses B1, B2 & B3 (b) Power flow at bus B2

Figure 19

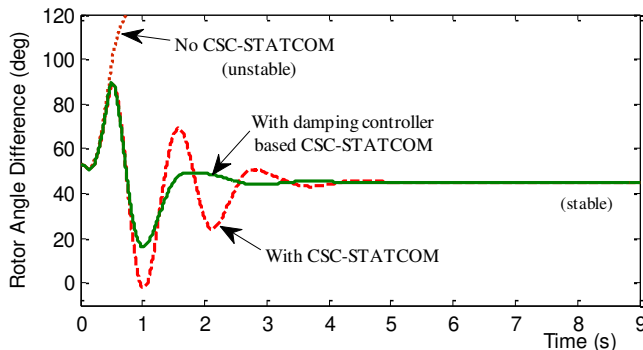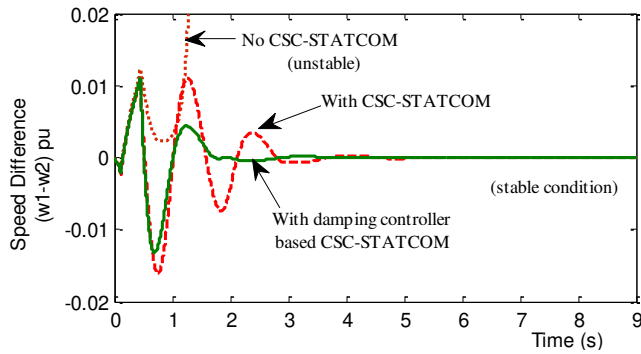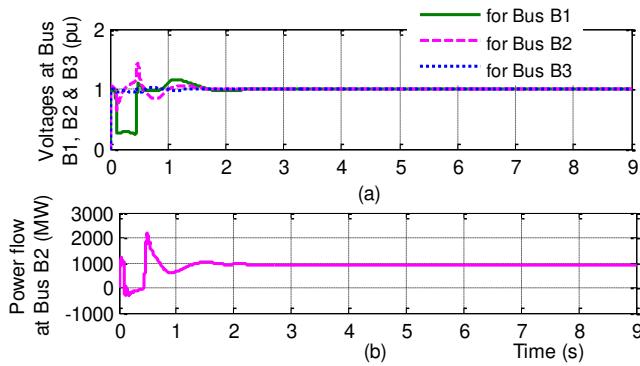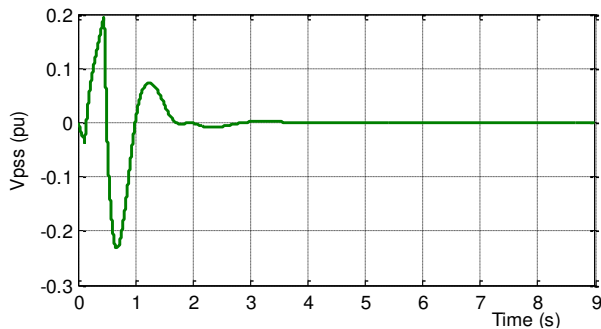Variation of output signal (V_{PSS}) of damping stabilizer in case-II

If the large loading duration is increased from 0.1 s to 0.59 s then, the system without damping stabilizer becomes unstable as shown in Figure 20, but the damping stabilizer based CSC-STATCOM still maintains the power system stability. The CCT for the system with & without CSC-STATCOM (and/or damping stabilizer) are shown in Table II. It clearly shows that CCT for the test system is better due to the impact of pole-shifting controller based CSC-STATCOM with damping stabilizer. Hence, the performance of the proposed scheme is satisfactory in this case also.

Table II
CCT of disturbances for the system stability with different topologies (Case-II)

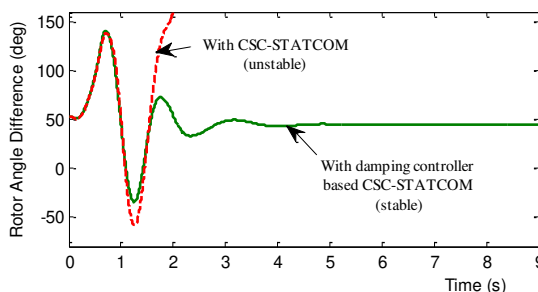| S. No.	System with different topologies	Critical Clearing Time (CCT)
1	Without CSC-STATCOM	100 ms – 440 ms
2	With CSC-STATCOM	100 ms – 583 ms
3	With damping stabilizer based CSC-STATCOM	100 ms – 594 ms

Figure 20
Variation of rotor angle difference of machines M1 & M2 in Case-II (large loading for 0.1 s to 0.59 s)

Conclusions

In this paper, the dynamic modeling of a CSC based STATCOM is studied and pole-shifting controller with damping stabilizer for the best input-output response of CSC-STATCOM is presented in order to enhance the system stability of the power system with the different disturbances. The novelty in proposed approach lies in the fact that, transient stability and oscillation damping ability of a two-area two-machine power system are improved and the critical clearing time of the disturbance is also increased. The coordination between damping stabilizer and pole-shifting controller-based CSC-STATCOM is also shown in the proposed topology. The proposed scheme is simulated and verified with MATLAB software. This paper also shows that a damping stabilizer based CSC-STATCOM is more reliable and effective than a system without damping stabilizer-based CSC-STATCOM, in terms of oscillation damping, critical fault clearing time and transient stability of a two-area power system. Hence, CSC based STATCOM can be regarded as an alternative FACTS device to that of other shunt FACTS devices.

Appendix 1

Parameters for various components used in the test system configuration of Figure 6. (All parameters are in pu unless specified otherwise):

For generator of plant (P1 & P2):

V_G=13.8 kV; R_s= 0.003; f=50 Hz; X_d= 1.305; $X_d^{'}$= 0.296; $X_d^{''}$= 0.252; $X_q^{'}$= 0.50; $X_q^{''}$= 0.243; $T_d^{'}$=1.01 s; $T_d^{''}$=0.053 s; H=3.7 s

(Where R_s is stator winding resistance of generators; V_G is generator voltage (L-L), f is frequency; X_d is synchronous reactance of generators; $X_d^{'}$ & $X_d^{''}$ are the transient and sub-transient reactance of generators in the direct-axis; $X_q^{'}$ & $X_q^{''}$ are the transient and sub-transient reactance of generators in the quadrature-axis; $T_d^{'}$ & $T_d^{''}$ are the transient and sub-transient open-circuit time constant; H the inertia constant of machine.)

For excitation system of machines (M1 & M2):

Regulator gain and time constant (Ka & Ta): 200, 0.001 s; Gain and time constant of exciter (Ke & Te): 1, 0 s; Damping filter gain and time constant (Kf & Tf): 0.001, 0.1 s; Upper and lower limit of the regulator output: 0, 7.

For pole-shifting controller based CSC-STATCOM:

System nominal voltage (L-L): 500 kV; R_{dc}= 0.01Ω; L_{dc}=40 mH; C = 400 µF; R=0.3 Ω; L = 2 mH; ω=314; $V_{b(ref)}$= 1.

For damping stabilizer:

K_s = 25; T_w = 10; T_1 = 0.050; T_2 = 0.020; T_3 = 3; T_4 = 5.4; $V_{s\,(max)}$ = 0.35; $V_{s\,(min)}$ = -0.35

References

[1]   Prabha Kundur, "*Power System Stability and Control*", McGraw-Hill, 1994

[2]   G. Rogers, *Power System Oscillations*, 344 p, Kluwer Power Electronics & Power System Series, Springer 2000

[3]   A. Tahri, H. M. Boulouiha, A. Allali, T. Fatima, "A Multi-Variable LQG Controller-based Robust Control Strategy Applied to an Advanced Static VAR Compensator", *Acta Polytechnica Hungarica*, Vol. 10, No. 4, 2013, pp. 229-247

[4]   H. Tsai, C. Chu and S. Lee, "Passivity-based Nonlinear STATCOM Controller Design for Improving Transient Stability of Power Systems", *Proc. of IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific Dalian*, China, 2005

[5]   M. H. Haque "Improvement of First Swing Stability Limit by Utilizing Full Benefit of Shunt Facts Devices", *IEEE Trans. Power Syst.*, Vol. 19, No. 4, pp. 1894-1902, 2004

[6]   P. Kundur, M. Klein, G. J. Rogers, M. S. Zywno, "Application of Power System Stabilizers for Enhancement of Overall System Stability," *IEEE Transactions on Power Apparatus and Systems*, Vol. 4, No. 2, pp. 614-626, May 1989

[7]   M. A. Abido, "Analysis and Assessment of STATCOM-based Damping
        Stabilizers for Power System Stability Enhancement", *Electric Power
        Systems Research*, Volume 73, Issue 2, pp. 177-185, February 2005

[8]   M. Aliakbar Golkar and M. Zarringhalami, "Coordinated Design of PSS
        and STATCOM Parameters for Power System Stability Improvement
        Using Genetic Algorithm", *Iranian Journal of Electrical and Computer
        Engineering*, Vol. 8, No. 2, Summer-Fall 2009

[9]   M. Mahdavian, G. Shahgholian, "State Space Analysis of Power System
        Stability Enhancement with Used the STATCOM", *IEEE/ECTI-CON*, pp.
        1201-1205, Chiang Mai, Thailand, May 2010

[10]   N. Mithulananthan, C. A Canizares, J. Reeve and G. J. Rogers,
        "Comparison of PSS, SVC and STATCOM Controllers for Damping Power
        System Oscillations", *IEEE Transactions on Power Systems*, Vol. 18, No.
        2, pp. 786-792, May 2003

[11]   Y. L Abdel-Magid & M. A Abido, "Coordinated Design of a PSS & a
        SVC-based Controller to Enhance Power System Stability", *Electrical
        Power & Energy System*, Vol. 25, pp. 695-704, 2003

[12]   N. C. Sahoo, B. K. Panigrahi, P. K. Dash and G. Panda, "Multivariable
        Nonlinear Control of STATCOM for Synchronous Generator
        Stabilization", *International Journal of Electrical Power & Energy
        Systems*, Volume 26, Issue 1, pp. 37-48, January 2004

[13]   L. J. Cai and I. Erlich "Simultaneous Coordinated Tuning of PSS and
        FACTS Damping Controllers in Large Power Systems", *IEEE Trans.
        Power Syst.*, Vol. 20, No. 1, pp. 294-300, 2005

[14]   N. G. Hingorani and L. Gyugyi, *Understanding FACTS: Concepts and
        Technology of Flexible ac Transmission Systems*, IEEE Press, New York,
        1999

[15]   Bilgin, H. F., Ermis M., Kose, et al, "Reactive-Power Compensation of
        Coal Mining Excavators by Using a New-Generation STATCOM ", *IEEE
        Transactions on Industry Applications,* Vol. 43, No. 1, pp. 97-110, Jan.-feb.
        2007

[16]   M. Kazearni and Y. Ye, "Comparative Evaluation of Three-Phase PWM
        Voltage- and Current-Source Converter Topologies in FACTS
        Applications", *Proc. IEEE Power Eng. Soc. Summer Meeting*, Vol. 1, p.
        473, 2002

[17]   D. Shen and P. W. Lehn, "Modeling, Analysis, and Control of a Current
        Source Inverter-based STATCOM", *IEEE Transactions Power Delivery,*
        Vol. 17, p. 248, 2002

[18] Sandeep Gupta and R. K. Tripathi, "An LQR and Pole Placement Controller for CSC-based STATCOM," *in Proc. Inter. Conf. IEEE Power, Energy and Control (ICPEC)*, pp. 115-119, 2013

[19] C. Schauder and H. Mehta "Vector Analysis and Control of Advanced Static VAR Compensators", *IEE Proceedings of Generation, Transmission and Distribution*, Vol. 140, pp. 299, 1993

[20] Katsuhiko Ogata, "*Modern Control Engineering*", 5th Edition, Prentice Hall, 2010

[21] K. R. Padiyar, V. Swayam Prakash, "Tuning and Performance Evaluation of Damping Controller for a STATCOM", *International Journal of Electrical Power & Energy Systems*, Vol. 25, Issue 2, pp. 155-166, February 2003

[22] Ahmad Rohani, M. Reza Safari Tirtashi, and Reza Noroozian, "Combined Design of PSS and STATCOM Controllers for Power System Stability Enhancement", *Journal of Power Electronics*, Vol. 11, No. 5, September 2011

[23] B. T. Ooi, M. Kazerani, R. Marceau, Z. Wolanski, F. D. Galiana, D. McGillis and G. Joos "Mid-Point Sitting of Facts Devices in Transmission Lines", *IEEE Trans. Power Del.,* Vol. 12, No. 4, pp. 1717-1722, 1997

Future Internet-based Collaboration
in Factory Planning

**Christian Weidig[1], Péter Galambos[2], Ádám Csapó[2],
Péter Zentay[3], Péter Baranyi[2], Jan C. Aurich[1],
Bernd Hamann[4], Oliver Kreylos[4]**

[1] Institute for Manufacturing Technology and Production Systems,
    University of Kaiserslautern,
    Gottlieb-Daimler-Str. D-67663 Kaiserslautern, Germany
    {weidig,aurich}@cpk.uni-kl.de

[2] Institute for Computer Science and Control, Hungarian Academy of Sciences
    Kende u. 13-17. H-1111 Budapest, Hungary
    {galambos,csapo,baranyi}@sztaki.mta.hu

[3] Antal Bejczy Center for Intelligent Robotics, Óbuda University,
    Bécsi út 96/B. H-1034 Budapest, Hungary
    zentay.peter@bgk.uni-obuda.hu

[4] Institute for Data Analysis and Visualization (IDAV), Dept. of Computer Science, University of California, One Shields Ave, CA 95616 Davis, USA
    {hamann,kreylos}@cs.ucdavis.edu

Abstract: As the design, development and execution of manufacturing processes continue to spread out across the world, globally distributed enterprises demand new paradigms. Distance collaboration tools are becoming increasingly important in order to maintain synergies between spatially distributed entities enabling effective cooperation over large distances. Virtual Reality (VR) technologies offer unique possibilities for the exchange of planning stages as well as for the identification and collective resolution of problems. This paper discusses the requirements of manufacturing engineers for distance collaboration tools and the challenges associated with the creation of such systems through a working prototype implemented in the VirCA NET framework. This pilot solution is a novel application in the field of VR-enhanced spatially distributed collaboration via shared virtual spaces using immersive visualization. Typical scenarios are provided to highlight the capabilities offered by VirCA NET. The paper identifies and classifies the challenges of distance collaboration from a cognitive infocommunications (CogInfoCom) perspective. The challenges are presented with respect to the theoretical background of CogInfoCom engines and channels emphasising the link between virtual collaboration and the aim of CogInfoCom.

Keywords: Future Internet; 3D Internet; Virtual reality; Remote collaboration; Factory design; Digital factory

1   Introduction

Globalized companies have to adapt continuously to variations in product life cycles, emerging product complexities, agile worldwide markets and supply networks [1]. The inclusion of multiple stakeholders to adaptation processes is crucial in maintaining sustainable factory life-cycles. Several divisions and employees must be involved, so that the effects of non-optimal planning can be alleviated and special planning expertise and professional knowledge can be exploited from many sources at the same time [2]. As spatially distributed entities become involved in this process, the task of achieving such goals is rendered increasingly complex [3, 4]. Especially in the case of multinational corporations, it is necessary to ensure the operation of multi-site manufacturing systems, and the use of appropriate collaborative VR supporting tools should ideally be considered [5]. Information technology within mechanical engineering – mainly expressed as part of the Digital Manufacturing approach in the past decade – has led to the creation of a wider set of software tools and technologies. The main objective of digital manufacturing is to master crucial challenges such as shorten design cycles, increasing product variants and complexity, fast changing market needs, etc. (for details see, e.g. [6]). The spatial distribution of planning participants, which is evolving today, even requires additional support for long-distance collaboration [7].

Although this field has been investigated from different points of view in the past decade, a comprehensive approach to tackling spatially distributed factory planning by means of immersive, collaborative Virtual Reality (VR) is still not available [3].

Cognitive infocommunications (CogInfoCom) is an emerging research paradigm which focuses on the merging process occurring between natural and artificial cognitive systems in modern ICT theories and applications alike [8, 9]. Inspired by CogInfoCom, results of the cognitive sciences – which investigate biological cognitive capabilities – are expected to converge towards ICT through modern applications [10].

In this paper, – extending and summarizing the previous works of the authors [11, 12] – the relevance of CogInfoCom for augmented/virtual collaboration is considered in terms of emerging demands coming from worldwide distributed manufacturing companies. The discussions are based on the Virtual Collaboration Arena (VirCA) framework and a recent extension to the system. The extension (referred to as VirCA NET from here on) allows users to view and collaborate through a shared virtual reality [13, 14]. Usually, CogInfoCom systems such as VirCA are based on technologies and components that are individually well-known to the wider community of engineers. However, the purpose of this paper is to show that the unique combination of these components gives rise to something that is more powerful than the sum of its parts: a system that strongly supports Future Internet research and encapsulates the philosophy behind CogInfoCom. At the same time, new perspectives are opened up and new design challenges are introduced, which will inevitably change the way we relate to our interactions with infocommunications devices.

The paper is organized as follows. Section 2 briefly introduces the VirCA NET platform. Section 3 gives a comprehensive overview of the collaborative processes in the field of Manufacturing Engineering, while focusing especially on factory planning and the supporting digital tools. Section 4 discusses the distance collaboration from the point of view of the end-user requirements, and provides example scenarios covering the main issues of the topic and identifying the key challenges. Following the general discussion, the pilot implementations of two scenarios are introduced and evaluated in terms of the previously identified requirements. A thorough discussion of the challenges from the CogInfoCom aspect is provided in Section 6. The final section concludes the paper.

2   Introduction to VirCA NET

VirCA (Virtual Collaboration Arena) [1] is a software platform — developed and maintained by MTA SZTAKI [2] — that supports various types of collaboration involving shared 3D virtual environments. VirCA consists of a 3D Virtual Reality core engine responsible to maintain and display the virtual environment where the collaboration takes place, and a web-based editor where the content and the actual operation of the virtual environment can be constructed from building blocks representing 3D objects and/or functionalities. This modular behaviour relies on the RT-Middleware (RTM) standard [15] and its open source implementation OpenRTM-aist [16, 17]. RTM is originally developed for networked robotic systems, however it serves well for a much wider field of applications [18].

As a framework, VirCA enables the developers to build customized and flexibly reconfigurable 3D environments and implements the paradigm of augmented virtuality by the synchronization of real world objects and processes with the virtual reality. In this way a virtual augmentation of real environments can be created. VirCA facilitates the so-called "Knowledge plug and play" since different alreadz existing hardware components (e.g. input devices, sensors, etc.) and computational technologies (e.g. speech synthesis and recognition, machine vision, semantic reasoning, etc.) can be integrated into the VirCA-based applications thanks to the RTM-based component interoperation. Exploiting the virtual sensing capabilities of VirCA, it is possible to "virtualize" the not yet existing or unreasonably expensive technologies and investigate whether its addition to a physical system would bring the anticipated advantages.

These capabilities make VirCA appropriate to be a pilot framework for CogInfoCom applications and/or a candidate collaboration tool of the Future Internet.

The recently introduced VirCA NET extension [3] provides further enhancement en-

---

[1]   http://www.virca.hu
[2]   Institute for Computer Science and Control, Hungarian Academy of Sciences (http://www.sztaki.mta.hu)
[3]   Integrated into VirCA version 0.2 and higher

abling the connection between multiple VirCA instances (possibly located far from each other) to share and manipulate 3D virtual environments collaboratively. The collaborative session is managed by a VirCA Master instance that serves as a hub in the synchronization of the distributed virtual world.

Let us define some VirCA related concepts which are often used within the remaining part of the paper:

Collaborator

In the VirCA NET context, a Collaborator is understood as a person or a group of people who are interacting with the shared virtual reality at the same location. For example a team of 4 persons working in the CAVE system at the University of Kaiserslautern are considered as a single Collaborator in the VirCA NET session.

Master VirCA

The VirCA instance to which all other VirCA instances are connected. Master VirCA acts as a hub for the interconnected VirCAs.

Slave VirCA

A VirCA instance that is connected to the Master VirCA. Multiple slaves can be connected to a single master. Slave VirCA instances provide the cloned versions of the virtual space maintained by the master VirCA.

VirCA NET connection

The communication channel between the master VirCA and a slave VirCA. This channel is responsible for the synchronization of the virtual space and for the transmission of user actions in order to ensure that all collaborators receive the same information from the shared virtual reality.

RT-Component or RTC for short

A reusable software component that complies with the Robot Technology Component (RTC) Specification [15]. Different types of capabilities (e.g. machine vision, speech technologies, etc.) and hardware interfaces (e.g. sensors, actuators, complex devices) can be implemented in the form of RTCs. In the VirCA context, RTC is a reusable building block that does not directly appear in the virtual space but operates as background knowledge or as an interface between virtual and real-world objects.

Cyber Device or CD for short

Cyber Devices are special types of RTCs which appear as 3D objects in the shared virtual space. For example, machine tools, robots and other visually instantiated parts of a manufacturing system are implemented as Cyber Devices.

Input Device

Input devices are software components interfacing different UI hardware to VirCA (e.g. pointing devices, MS Kinect, etc.).

System Editor

> The VirCA System Editor is a web application for the management of collaboration sessions. Collaborators can build systems from Input Devices, RTCs and CDs as building blocks.

3   Manufacturing Engineering Review

3.1   Manufacturing planning as collaborative process

The overall system complexity faced by manufacturing companies is continuously rising. The ability to react to changes in globalized, uncertain markets within short periods of time requires sophisticated design, manufacturing and business processes. To handle this complexity with success, manufacturing companies not only have to design flexible technological solutions and products, but also have to focus increasingly on developing and managing complex socio-technical systems [19]. This means that modern products and manufacturing systems involve not only complex mechanical and electrical components, but also include software systems, control modules and advanced user-machine interfaces. To achieve real-time feedback, such systems are and will be connected to the World Wide Web and the Internet of Things, which further increases complexity [20].

Complex systems, such as a manufacturing system, are characterized by multi-dimensional interrelations between a large number of affected participants, elements and agents, which interact with each other as well as their environment [21]. When such systems are re-configured and evolved to new processes, it is necessary to consider the entire life-cycle of the system as a whole. Hence, a manufacturing system must be understood as a comprehensive product that needs to be designed, manufactured and assembled with a holistic view in mind [22]. Typical for such wide and complex problem definitions is the requested interaction between various divisions such as those dealing with engineering, operations management, logistics and IT [23]. The inclusion of a number of different planning fields and specialists is one of the most crucial points of current factory planning projects. Interdependencies must be respected; therefore the participation of employees from several divisions is recommended [2, 24].

To detail out the collaborative aspect of such team-oriented problem solving environments, [25] introduced a formal definition of collaborative engineering as a "socio-technical engineering discipline, which facilitates the communal establishment of technical agreements among a team of interdisciplinary stakeholders, who work jointly towards a common goal with limited resources or conflicting interests". Manufacturing planning in this sense is a task that requires a high level of collaboration. Only comprehensive and efficient communication can result in short planning cycles, improved planning quality and reduced effort during factory planning projects [26]. Sharing experience and knowledge between several co-designers is therefore a key

requirement for a successful factory planning process [27].

3.2   Digital tools supporting collaborative manufacturing planning

In the range of manufacturing planning, industrial companies are using a wide set of software tools to achieve improvements of planning quality and reduction of planning time [28]. Nevertheless, the exchange of planning states and results in industrial projects is often realized with only little support of digital collaborative tools. Even if digital data is provided for each separate planning aspect, approval of comprehensive solutions and decision-making is still paper-based. This is grounded in the fact that visual analytic tools are typically insufficient for collaborative work. They are designed for single user operation on standard desktop systems [29]. Support of whole planning teams, involving several experts with differing technical skills out of different domains is not guaranteed [30]. Even the so called "collaborative solutions" which are available at the market do not fully cover the challenges of collaborative engineering. One open point that is crucial, but not yet covered is that stakeholders perceive problems differently, due to their expertise and personal objectives [25].

To overcome such problems VR is proposed as a beneficial tool to consider different viewpoints and personal objectives of several involved planning specialists and their relationships [31]. The fact that every planning participant has a subjective perception of the common model, which is shaped by his or her own experience and professional know-how adds additional value to VR supported manufacturing planning. This makes VR an ideal platform to foster the cooperation between different technical specialists, reducing interface problems and increasing planning quality [32].

Besides these considerations, the upcoming worldwide spatially distribution of planning participants requires also for collaborative support over long-distance [7]. Tailoring of factories and organizational structures to fulfil the needs of worldwide markets requires the consideration of specific local constraints, cultural diversities and existing factory conditions. To merge distributed expertise for optimised planning results, the support of factory planning by distance collaboration tools is proposed [3, 4].

In general three main challenges could be identified, which must be addressed by distance collaboration tools [25].

- The problems to be solved must be defined in a clear way. The problem definition must be fully shared between all planning participants [25].

- All planning participants need access to the whole set of information behind the problem. There must be a shared understanding of the problem, highlighting all aspects with an equal priority [33].

- The rules for decision-making must be clear for all the stakeholders [34].

VirCA NET will therefore be a contribution and help bringing planning participants together to one shared virtual environment and one shared problem understanding.

4   Distance Collaboration supporting Manufacturing Engineering

In this section, we outline the requirements against the effective collaboration from the viewpoint of the manufacturing engineering and sketch two related collaborative scenarios of different complexity. The first scenario is currently implemented using VirCA NET, while the second scenario is more futuristic and has yet to be implemented.

4.1   Requirements from Manufacturing Engineering perspective

The human factory plays a key role for reconfiguring and optimizing manufacturing systems [35]. To support multi-national corporations, operating multi-site manufacturing systems distance collaboration VR systems should be taken into account for configuring and reconfiguring manufacturing systems. These processes require features for supporting multi stakeholder decision-making and joint discussion on planning results [5, 36]. The extension of communication and cooperation beyond organizational and divisional boundaries will speed up planning processes and reduce the complexity during work. This can be realized by interconnected but spatially distributed VR systems [7]. Because VR systems are identified as suitable tools allowing an intutitive and fast identification with the planned factory, even for participants which are normally uncommon with specialised digital planning tools [37].

The idea of collaborative engineering is to support engineering tasks which are intrinsically human-centers by socio-technical means, to pursue their initial character [25]. One core objective is to support knowledge transfer between remote partners in a social oriented way. In detail, knowledge related to personal skills, that cannot be formulized easily because of its tacit nature must be considred. This knowledge can hardly transfered without direct personal interaction [38]. Even if some projects have been initialized, a comprehensive approach to support spatially distributed manufacturing planning by means of immersive, collaborative VR is still not available [3]. To develop a distance collaboration system in a target driven way, requirements coming from manufacturing planning must be identified. Following SMPAROUNIS' approach, the key features are identified as [39]:

- quick and easy data storage and sharing

- synchronous and asynchronous communication

- cooperation in designing and manipulating of geometrical models

- multi-user visualization and interaction

- decision support

Not to cope with all aspects of manufacturing planning at once, the scope of this article will be focused on factory planning in detail. Factory planning as one key activity during manufacturing planning is identified as one potential domain to be supported by collaborative VR tools in a beneficial way [40]. The key tasks during factory planning, which need to be enhanced by distance collaboration tools are summarized as follows: designing of the factory, validation of planning stages and overview of optimization in the planned and/or existing manufacturing system [40].

4.2   Typical collaborative way of working

Out of the preliminary considerations three main requirements have been derived, that need to be integrated into a VR distance collaboration tool supporting factory planning as:

- shared model visualization

- model interaction

- user interaction, knowledge exchange

Based on a digital factory model the future shop floor gets designed, analysed and later on discussed within a collaborative session. The 3D model which is used for this purpose usually consists of the factory layout (geometric descriptions of the shop floor), its direct environment and additional manufacturing related objects. Information regarding the operation of the manufacturing system, such as process descriptions for example, enrich the factory model [41]. Sharing the digital factory model among all involved planning participants and allowing a comprehensive view for all users is a key requirement of a distance collaboration tool [3]. By implementing a joint virtual environment, VirCA NET provides such an common workspcae which is identical at all remote sites.

As well as the shared visualization, interactive means have to be implemented into VirCA NET. Comprehensing of manufacturing processes and analyzis of simulation results is facilitated by bidirectional interaction between the virtual model and the user. Collaborative features for interaction between users and virtual agents have to be developed as well as interconnection means between several virtual agents.

Since the major target of distance collaborative sessions in factory planning is not to design a new layout or new product, the analyzis of proposed layouts and problem solving afterwards in a cooperatively way is central. Hence, features supporting the

cognitive tasks of engineers and other stakeholders must be provided by distance collaboration tool; social interaction beyond the data visualization must be enabled [42, 43]. In detail tacit knowledge which is hardly transferable without direct personal interaction, must be considered [38]. Therefore the capability to achieve and support user-user interaction must be integrated into the VirCA NET.

The core idea while supporting factory planning is to adopt an established (everyday) working situation, in detail in-place, domain-overlapping factory planning into a distributed, virtual environment [44]. Worldwide distributed employees should work together as if they were in-place and use their well-known problem solving strategies. The shared digital model should allow a simultaneous analyzis, while users are free in navigation and can investigate the model on their own purpose. Due to the interconnected VR systems users will have the appearance as if they were co-located, because the VR systems will react as one distributed unit. By the extension of the VR setup with social user-user interaction means the perception of direct and "real" collaboration will be fully covered.

Because known working situations can be correlated in a direct way to this new way of cooperation, the results which are based on the discussions among the digital model, should be close to the optimizations developed in co-located sessions. The approach is beneficially integrated in existing VR-enhanced work-flows of factory planners. Users are enabled to use established processes and known problem solving strategies, even extended by the distance collaborative aspect.
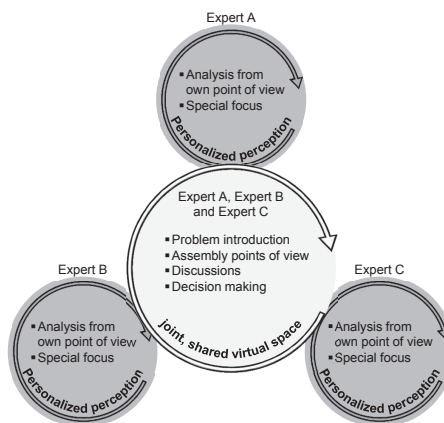

Figure 1
Workflow during collaborative VR sessions

Figure 1 is illustrating the general concept of the proposed collaborative VR sessions operated by VirCA NET. A joint, shared virtual environment is provided to which all planning participants access in the same way. There, several experts can introduce problems out of their own profession to the other involved participants. The discussion within the planning team is the main task in this shared space. In

the collaborative discussion solutions will be developed and decisions, based on the entire expertise of all stakeholders will be made. In addition to this shared space, the personal analysis for each participants is provided by VirCA NET as well. Due to the differing professional expertise and experience of the stakeholders their perception of the factory layout problem will differ. Due to the free navigation and interaction capabilities of VirCA NET they are allowed to investigate the factory regarding their professional interest and figure out exactly the details they need for their personal analysis. Due to this dualism of personal analysis and joint problem solving a comprehensive solution for the factory layout problem should be achieved in a beneficial way, even if planning participants are spread all over the world. To prove this concept idea two scenarios dealing with factory planning have been developed and executed. The following sections will introduce the performed test in detail.

4.3   Collaborative design of automatic workpiece feeding

The objective of this basic scenario is to design an automatic workpiece feeding process for a lathe in cooperation with an industrial robot, which are both not yet physically available in the shop floor. Therefore, the position of the lathe and the robot need to be defined in an existing shop floor. Three remote connected sites are involved representing each a different planning participant. Kaiserslautern (Germany) is representing the factory where the shop floor to be adapted is located. Budapest (Hungary) is representing an division specialised in robot control and Kosice (Slovakia) represents the head office of the fictive company where the central industrial engineering department providing knowledge among the operation and implementation of the lathe is located. Each of the locations VR-systems are connected to VirCA NET, and thus are able to interact and collaborate using the hardware and software components running at different locations (figure 2).

In the collaborative scenario the industrial robot (physically located in Budapest) is connected to VirCA NET. The virtual representations of the robot can be controlled from all planning participants using VirCA NET (the indirect control of the physical robot is also provided through the manipluation of the digital model). The objective of the collaborative session is to optimize the position of the lathe and the robot in a way that the workpiece feeding process can be guaranteed. They have to respect constraints occurring out of the already existing shop floor layout and special basic conditions regarding the knowledge and know-how of the three departments concurrently. The collaborative VR system must therefore provide the capability to involve intrinsic know-how and personal skills of the planning participants to reach the complex objectives of the factory planning project. Every Participants are allowed to change the layout of the virtual shop floor (e.g. position of the robot, position of the lathe) and give commands to the robot (e.g. take workpiece, move workpiece to the machine tool). By doing this, one participant can propose a layout for the lathe and the robot that fits best to his personal targets (e.g., insert the machine tool and the robot without major variance to the existing layout). Afterwards each participant can
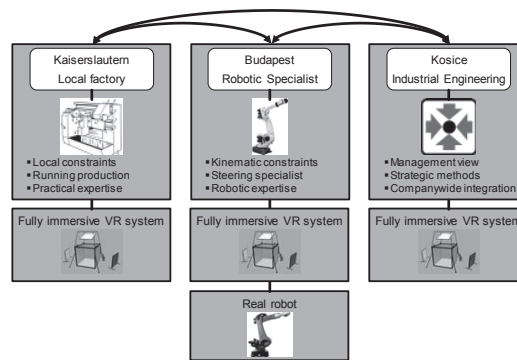
Figure 2
Schematic setup of the basic scenario

investigate the proposed solution according his own point of view and recommend
changes (figure 3). This process ensures the achievement of the best solution for
each point of view without neglecting requirements from other departments.
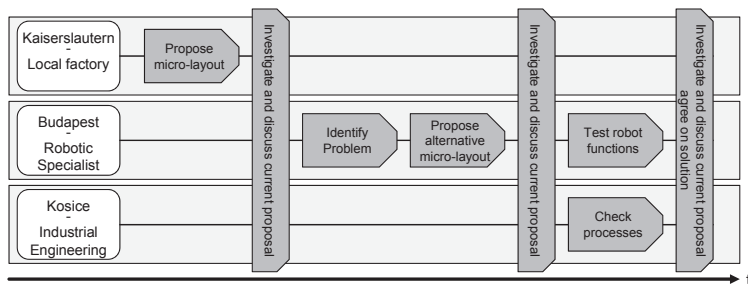

Figure 3
Example workflow for the collaborative design of automatic workpiece feeding scenario

4.4   Collaborative shop floor layout design

The target of this more complex scenario is to allow participants to design the
complete layout of an industrial shop floor collaboratively. The scenario of the
previous subsection is a step in this direction, but there are additional challenges
when extending the example to a fully detailed industrial environment. Geometric
information and ad-hoc impressions of the participants are often not enough for un-
derstanding complex manufacturing processes. Participants need to access additional
information sources as for example simulation results, process plans and other meta-
data during a collaborative session to completely cover the impacts of changes to

the shop floor layout. Also technical advancements are necessary in order to enable truly flexible collaboration. For designing a complete shop floor layout participants are allowed to change the position of all potentially movable machines and other infrastructural elements (e.g., shelves for storing tools and workpieces, pathways for mobile robots and workers, etc.). Further simulations of manufacturing processes and material flows should be accessible during the collaborative session, to estimate the impact of changes and investigate weak points. So the system must take into consideration a variety of constraints while the shop floor is being edited:

- the availability of supply sources for the industrial machines to be able to operate (e.g. power, water etc.)

- dependencies among machines and other infrastructure, dictating that they be placed in close proximity or far away from one another

- the workspace of various robots in order to ensure a safe working environment

- the manufacturing process itself and the material flow through the shop floor

- connections to neighboring workstations and subsequent working areas

- connections to the logistic system in general

Additionally to these hard facts which are extending the sheer geometric representation of the shop floor layout by overlaying information, also the support of soft skills and working methods need to be considered by the complex scenario. Therefore, features must be implemented and tested among their impact factor. To communicate not only the intermediate results to planning participants, the adaptation process itself need to be visualized and transferred to the remote partners. Voice communication and intelligent control protocols to avoid babylonian speech chaos need to be introduced. Highlighting objects and transmission of virtual pointers are only two potential feature which can underline the focus on the object or point of interest.

4.5   Challenges associated with the scenarios

The scenarios serve as typical examples for factory planning problems, they cover the key points and illustrate typical tasks which are executed during factory planning. These crucial challenges mus be considered while designing VirCA NET for factory planning purposes and executing the scenarios. Thereby the detailed characteristics of the key challenges are slightly differing due to distinct factory planning projects, but their main characteristics can be identified in advance. *Shared model visualization*, *model interaction* and *means for user interaction* are identified as three main challenges.

When using full immersive VR systems, the shared model visualization is a key requirement for distance collaboration tools. Thus VirCa NET is allowing a consistent visualization of a joint digital model at all connected sites, independent which

users is in possession of the model parts. In the range of factory planning this gets important to allow all users the simultaneous access to the current planning results in a equal way.

The user-machine interaction is a second major challenge. Since the reorganisation of factory elements is a basic operation when optimizing the shop floor layout, the interaction with the factory model must be enabled for each user. To express ideas and emphasize layout proposals, adaptation of the layout must be allowed for all users. The interaction with intelligent agents such as virtual robots is another example for user-machine interaction. Predefined model behaviour, like material flow simulation is facilitating the understanding of the manufacturing processes.

As third key-challenge the user-user interaction is identified. A natural interaction on a social level must be provided.Even if ideas can be transferred via the reorganization of the layout related models, the verbal and non-verbal expression of ideas is a key element for introducing and solving problems.

5   Implementation details

This section gives a technical insight into the VirCA-based implementation of the previously discussed automatic workpiece feeding design scenario. Three VirCA instances are involved in the collaborative session according to the three locations (Figure 4). The procurer company (LOCATION 1) hosts the master VirCA and the cell controller Cyber Device, while the robot and machine tool responsible persons (LOCATION 2 and LOCATION 3) are running the slave VirCA instances and the components related to the devices to be tested.

Each collaborators have a special expertise and role in the design and testing procedure and accordingly, the different software components are managed by the stakeholder who is responsible for the given physical device. In the concrete example, the robot integrator operates a Robot Cyber Device and a Robot RTC while the machine tool specialist provides the machine tool Cyber Device and the corresponding RTCs. The role of the Cyber Devices is the visual representation of the real or pure virtual entities, while the RTCs establish interfaces between the virtual and real devices allowing access to the device's functions e.g. moving commands of the robot arm and the so-called M-code functions of the machine tools.

Figure 5 illustrates the shared scene from the viewpoint of a collaborator. This screen-shot displays the virtual shop floor where the robot application should be placed in accordance with the decision of the collaborators.

To support the effective interaction between the stakeholders, VirCA NET provides various visualization features for example, within the virtual scene, collaborators can see each other as a symbolic human head representing the gaze direction and can see the 3D pointer of each other. Voice and text based communication is also possible using appropriate software tools (e.g. Skype) parallel to the VirCA session.
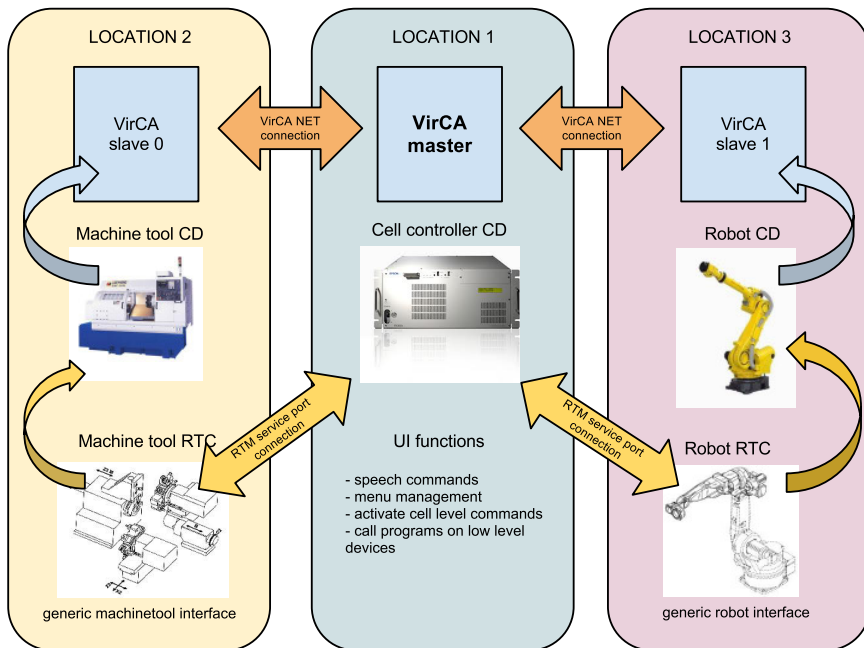
Figure 4
The VirCA NET structure of the proposed collaborative planning session

The integration of such features in VirCA will be the target of further development.

An other crucial point is the way of manipulation of virtual objects. Since users should not be burdened with having to wear complicated haptic devices, CogInfo-Com methods takes place that could provide unencumbered yet situation aware user experience [45].

Because of its extensible modular structure, the proposed scenario serves as good basis for the experimental investigation of different interaction methods. The working prototype of this scenario has been successfully demonstrated in pubic events and conferences [46].
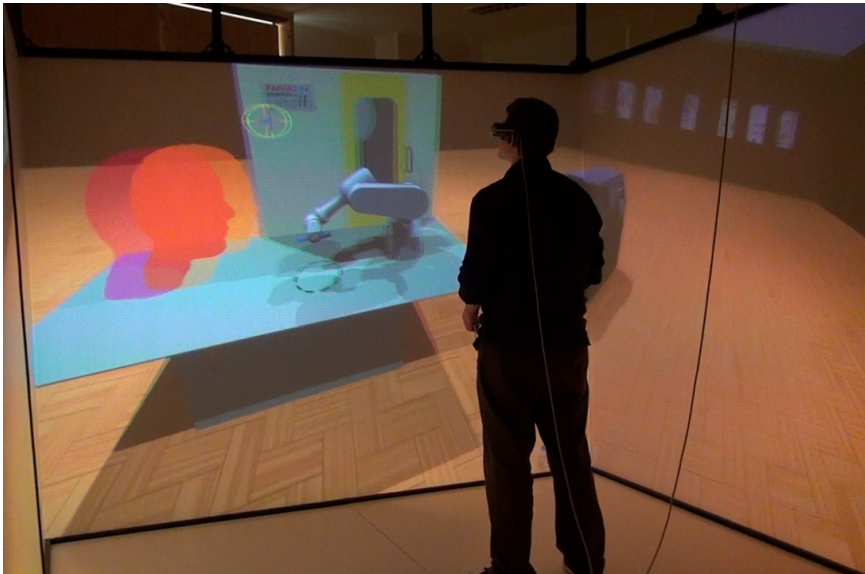
Figure 5
The design session from the viewpoint of a collaborator using immersive projection system

6 Challenges During Collaboration from a CogInfoCom Perspective

The challenges discussed earlier bring to light various issues from two points of view:

- design questions related to supporting collaboration, e.g. in terms of triggering signals that encourage communication and collaboration in the first place

- design question related to supporting feedback through multi-sensory channels

The difficulties associated with these points of view are considered in the remainder of this section.

6.1 Issues related to supporting collaboration

Setting guidelines for the dynamic characteristics of the communication process between systems and users is a key challenge [47]. VirCA is naturally affected by such considerations, given the abundance of interaction modes which can be used by users to communicate with various objects and computational processes.

An important aspect of the challenges which pertain to supporting collaboration is the question of what information VirCA should communicate and when. Considering an industrial environment, for example, the question of when and how a robot communicates its internal state to users is a crucial one. A key concept related to this issue is the concept of CogInfoCom triggers, which can be designed with a number of different characteristics depending on the level of volition and directness they represent in communication [48].

Related to the above, the following questions may arise with respect to the collaborative planning scenario:

- **Visualization aspects**: should each user be notified of possible discrepancies within the visualization of the shared model? For example, if a given region on the shop floor is being manipulated by a user while other users are working in different areas, the set of dependencies which are influenced by those actions may be important to the entire community; thus, it can be argued that the user should be notified. On the other hand, if only a single user is working at a given time, notifications might be spared and/or postponed to a later time, as the user's actions will not directly contradict other users' actions.

- **Feedback aspects**: during interactions between users and the system, how should volition and directness of communication be treated? How can indirect communication be complemented so that users can obtain enough information reflecting the context of interaction?

- **User-user interaction**: how can both contextual communication (e.g., on users' relative locations) and flexible user-to-user communication be supported? For example, how can a user obtain information on how busy other users are at any given time, so as to be able to refrain from disturbing others?

6.2   Issues related to CogInfoCom channels

CogInfoCom channels are multi-sensory signals that reflect information on semantic concepts [47]. A key issue related to CogInfoCom channels is that the cognitive capabilities available to users can be utilized to different extents depending on how the channels are designed. Hence, the following questions may arise:

- **Visualization aspects**: how should users be notified of possible discrepancies within the shared model? If a given area is under active manipulation, what kinds of visual cues should be used? What cues can be used to signal the level of difficulty being tackled by other users?

- **Feedback aspects**: how should the physical parameters of objects be represented? How should concurrency among user actions be handled through feedback?

- **User-user interaction**: how should users be allowed to influence what part of their actions should be communicated to others using the system?

Conclusion

In this paper, the Virtual Reality based multi-user remote collaboration was studied through realistic use-case scenarios in the field of manufacturing system design. The requirements against the collaboration system were identified from the viewpoint of manufacturing engineers as end-users. Furthermore, some of the challenges which arise during augmented/virtual collaboration has been identified and classified from the perspectives of CogInfoCom. The basis of our discussions was taken from the VirCA NET extension built on the Virtual Collaboration Arena (VirCA) developed at MTA SZTAKI. We demonstrated how VirCA and VirCA NET belong to the CogInfoCom paradigm and how they are related to the Future Internet by identifying key features of infocommunication in collaborative scenarios related to production system design. Via the presented use-case study it is shown that the VirCA NET approach can be beneficially utilized in the investigated scenarios wherein the component selection, the shop floor layout design and process simulation/evaluation are integrated in one versatile collaboration system. Based on the features of VirCA NET and the specifics of the scenarios, we drew a list of challenges — some of which are already solved by the system to some extent and some of which remain to be solved in future work. The challenges were presented in relation to the theoretical aspects of CogInfoCom engines and CogInfoCom channels in order to further strengthen the link between augmented/virtual collaboration and the goals of CogInfoCom.

References

[1] E. Westkämper, C. Constantinescu, and V. Hummel, "New paradigm in manufacturing engineering: Factory life cylce," *Production Engineering Research and Development*, vol. 13, no. 1, pp. 143–146, 2006.

[2] B. Denkena, P. Woelk, and A. Brandes, "Flexible process chains by template based configuration," *Annals of the German academics society for production engineering*, vol. 12, no. 2, pp. 81–84, 2005.

[3] N. Menck, X. Yang, C. Weidig, P. A. Winkes, C. Lauer, H. Hagen, B. Hamann, and J. C. Aurich, "Collaborative factory planning in virtual reality," in *The 45th CIRP Conference on Manufacturing Systems*, Athens, Greece, May 2012.

[4] T. Liebeck, T. Meyer, and E. Abele, "Production technology: Adapting to maximize local advantage," *Global Production: a handbook for strategy and implementation, Springer, Berlin*, pp. 192–235, 2008.

[5] P. Pedrazzoli, D. Rovere, C. Constantinescu, J. Bathelt, M. Pappas, P. Depince, G. Chryssolouris, C. R. Boer, and E. Westkämper, "High value adding vr tools for networked customer-driven factory," in *Proceedings of the 4th International Conference on Digital Enterprise Technology*, Bath, United Kingdom, 2007.

[6] G. Chryssolouris, D. Mavrikios, N. Papakostas, D. Mourtzis, G. Michalos, and K. Georgoulias, "Digital manufacturing: History, perspectives, and outlook," *Journal of Engineering Manufacture*, vol. 222, pp. 451–462, 2008.

[7] C. Redaelli, G. Lawson, M. Sacco, and M. D'Cruz, "Difac: Digital factory for human oriented production system," *InTech*, pp. 339–354, 2009.

[8] P. Baranyi and A. Csapo, "Cognitive Infocommunications: CogInfoCom," in 11th *International Symposium on Computational Intelligence and Informatics (CINTI)*, 2010, pp. 141–146, Budapest, Hungary.

[9] ——, "Definition and Synergies of Cognitive Infocommunications," *Acta Polytechnica Hungarica*, vol. 9, pp. 67–83, 2012.

[10] G. Sallai, "The Cradle of Cognitive Infocommunications," *Acta Polytechnica Hungarica*, vol. 9, pp. 171–181, 2012.

[11] C. Weidig, A. Csapo, J. C. Aurich, B. Hamann, and O. Kreylos, "Virca net and coginfocom: Novel challenges in future internet based augmented/virtual collaboration," in *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, 2012, pp. 267–272.

[12] P. Galambos, C. Weidig, P. Baranyi, J. C. Aurich, B. Hamann, and O. Kreylos, "Virca net: A case study for collaboration in shared virtual space," in *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, 2012, pp. 273–277.

[13] P. Galambos and P. Baranyi, "VirCA as Virtual Intelligent Space for RT Middleware," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2011, pp. 140–145.

[14] MTA SZTAKI, VirCA (Virtual Collaboration Arena). [Online]. Available: http://www.virca.hu

[15] Robot technology component specification. [Online]. Available: http://www.omg.org/spec/RTC/

[16] AIST, OpenRTM-aist. [Online]. Available: http://www.openrtm.org

[17] N. Ando, T. Suehiro, K. Kitagaki, T. Kotoku, and W. Yoon, "RT-middleware: distributed component middleware for RT (robot technology)," in *2005*

IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alta., Canada, 2005, pp. 3933–3938.

[18] P. Galambos and P. Baranyi, "VirCA as Virtual Intelligent Space for RT-Middleware," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2011, pp. 140–145.

[19] W. ElMaraghy, H. ElMaraghy, T. Tomiyama, and L. Monostori, "Complexity in engineering design and manufacturing," *CIRP Annals - Manufacturing Technology*, vol. 61, no. 2, pp. 793–814, 2012.

[20] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[21] B. Colwell, "Complexity in design," *IEEE Computer*, vol. 38, no. 10, pp. 10–12, 2005.

[22] H. ElMaraghy, "A complexity code for manufacturig systems," in *ASME 2006 International Manufacturing Science and Engineering Conference*, Ypsilanti, MI, United States, oct 2006, pp. 625–634.

[23] R. P. Smith and J. A. Heim, "Virtual facility layout design: the value of an interactive three-dimensional representation," *International Journal of Production Research*, vol. 37, no. 17, pp. 3941–3957, 1999.

[24] T. Gase and U. G. und A. Krauss, "Integrierte struktur- und layoutplanung," *werkstattstechnik online*, vol. 5, no. 96, pp. 314–320, 2006.

[25] S.-Y. Lu, W. Elmaraghy, G. Schuh, and R. Wilhelm, "A scientific foundation of collaborative engineering," *CIRP Annals - Manufacturing Technology*, vol. 56, no. 2, pp. 3941–3957, 2007.

[26] U. Wagner, D. Oehme, M. Clauß, R. Riedel, and E. Müller, "Cooperative design, manufacturing and assembly of complex products," *wt Werkstattstechnik online*, vol. 102, no. 4, pp. 193–199, 2012.

[27] G. Seliger, H. Karl, and H. Weber, "Cooperative design, manufacturing and assembly of complex products," *CIRP Annals - Manufacturing Technology*, vol. 46, no. 1, pp. 67–70, 1997.

[28] P. Ebbesmeyer, J. Gausemeier, M. Grafe, and H. Krumm, "Designing flexible production systems with virtual reality," in *Proceedings of the 2001 ASME Design Engineering Technical Conference and Computers and Informations*, Pittsburgh, USA, sept 2001.

[29] R. M. Baecker, *Readings in Groupware and Computer-Supported Cooperative Work*, 1st ed. Morgan Kaufmann, dec 1993.

[30] J. Gausemeier, J. Fründ, and C. Matyszok, "Ar-planning tool âĂŞ designing flexible manufacturing systems with augmented reality," in *Proceedings of*

the 8th Eurographics Workshop on Virtual Environments, Barcelona, Spain, sept 2002, pp. 19–25.

[31] N. Menck, C. Weidig, and J. C. Aurich, "Virtual reality as a collaboration tool for factory planning based on scenario technique," in *Proceedings of the Forty Sixth CIRP Conference on Manufacturing Systems 2013*, Setubal, Portugal, May 2013.

[32] H. Wiendahl, T. Harms, and C. Fiebig, "Virtual factory design - a new tool for a co-operative planning approach," *Internatioanl Journal of Computer Interagted Manufacturing*, vol. 16, no. 7-8, pp. 535–540, 2003.

[33] G. Stasser and W. Titus, "Pooling of unshared information in group decision making: Biased information sampling during discussion," *Journal of Personality and Social Psychology*, vol. 48, no. 6, pp. 1467–1478, 1985.

[34] N. P. Suh, *Axiomatic Design: Advances and Applications*, 1st ed. Oxford Univ Pr, may 2001.

[35] X. Yang, E. Deines, C. Lauer, and J. C. Aurich, "Virtual reality enhanced human factor - an investigation in virtual factory," in *Proceedings of the Joint Virtual Reality Conference of EGVE - ICAT - EuroVR 2010*, Paris, France, May 2010.

[36] G. Vigano, L. Greci, and M. Sacco, "Giove virtual factory: the digital factory for human oriented production system," in *Proceedings of the 3h CIRP International Conference on Changeable, Agile, Reconfigurable and Virtual Production*, Munich, Germany, oct 2009, pp. 748–757.

[37] J. C. Aurich, D. Ostermayer, and C. Wagenknecht, "Improvement of manufacturing processes with virtual reality-based cip workshops," *International Journal of Production Research*, vol. 47, no. 19, pp. 5297–5309, 2009.

[38] Y. Cheng, E. S. Madsen, and J. Liangsiri, "Transferring knowledge in the relocation of manufacturing units," *Strategic Outsourcing: An International Journal*, vol. 3, no. 1, pp. 5–19, 2010.

[39] K. Smparounis, K. Alexopoulos, V. Xanthakis, M. Pappas, D. Mavrikios, and G. Chryssolouris, "A web-based platform for collaborative product design and evaluation," in *15th International Conference on Concurrent Enterprising (ICE)*, Leiden, Netherlands, jun 2009.

[40] A. Nee, S. Ong, G. Chryssolouris, and D. Mourtzis, "Augmented reality applications in design and manufacturing," *CIRP Annals - Manufacturing Technology*, no. 61, pp. 657–679, 2012.

[41] X. Yang, R. C. Malak, C. Lauer, C. Weidig, H. Hagen, B. Hamann, and J. C. Aurich, "Virtual reality enhanced manufacturing system design," in *Proceedings of the 7th CIRP International Conference on Digital Enterprise Technology*, Athens, Greece, sept 2011, pp. 125–133.

[42] D. Russell, M. Stefik, P. Pirolli, and S. Card, "The cost structure of sensemaking," in *Proceedings of the Conference on human Factors in Computing Systems (CHI)*, Amsterdam, The Netherlands, april 1993, pp. 269–276.

[43] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen, "Collaborative visualization: Definition, challenges, and research agenda," *Information Visualization*, vol. 10, no. 4, pp. 310–326, 2011.

[44] O. Kreylos, G. Bawden, T. Bernardin, M. I. Billen, eric S. Cowgill, R. D. Gold, B. Hamann, M. Jadamec, L. H. Kellogg, O. G. Staadt, and D. Y. Sumner, "Enabling scientific workflows in virtual reality," in *Proceedings of the ACM International Conference on Virtual Reality Continuum and Its Applications*, Hong Kong, 2006, pp. 155–162.

[45] P. Galambos, "Vibrotactile feedback for haptics and telemanipulation: Survey, concept and experiment," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 41–65, 2012.

[46] P. Galambos, C. Weidig, P. Zentay, A. Csapa, P. Baranyi, J. C. Aurich, B. Hammann, and O. Kreylos, "VirCA NET: a collaborative use case scenario on factory layout planning." Kosice, Slovakia: IEEE, Dec. 2012, pp. 467–468.

[47] A. Csapo and P. Baranyi, "CogInfoCom channels and related definitions revisited," in *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, Sept 2012, pp. 73–78.

[48] ——, "A Taxonomy of CogInfoCom Trigger Types in Practical Use Cases," in *3rd IEEE International Conference on Cognitive Infocommunications (in press)*, 2012.

An Intermediate Level Obfuscation Method

Dmitriy Dunaev, László Lengyel

Budapest University of Technology and Economics
Department of Automation and Applied Informatics
Magyar tudósok krt. 2, H-1117 Budapest, Hungary
dunaev@aut.bme.hu, lengyel@aut.bme.hu

Abstract: The essence of obfuscation is to entangle the code and eliminate the majority of logical links in it; that is, to transform the code so that it becomes complex enough for analysis and unauthorized modification. The developed theoretical apparatus allows us to describe an entangled program using concatenation of operational logics of the routines. Consequently, this approach considers not only the instructions or routines themselves, but the actions, or results, they produce. This allows us to consider obfuscation as the process of adding excessive functionality to the program. This paper is unique in presenting an obfuscation method at intermediate code level that is based on the theory of optimizing transformations. The focus is set on generation of fake intermediate level code, suppression of constants, and meshing of control flow transitions.

Keywords: obfuscation; software protection; entangling transformations; fake context; intermediate code

1 Introduction

The general purpose of obfuscating techniques is to prevent, or at least hamper, interpretation, decoding, analysis, or reverse engineering of software. We may further state that more particularly, although not exclusively, the obfuscating techniques relate to methods and apparatus for increasing the structural and logical complexity of software. All that is done by inserting, removing, or rearranging identifiable structures of information from the software in such a way as to exacerbate the difficulty of the process of decompilation or reverse engineering [1].

The introduction of a non-black-box simulation technique by Boaz Barak [2] has been a major landmark in obfuscation. It has been proven that universal obfuscator does not exist [2, 8], since there exists a class of programs for which the virtual black box property is not feasible. According to [2], program obfuscation is an efficient transformation O of a program P into an equivalent program P' such that

P' is far less understandable than *P* (i.e. *P'* protects any secrets that may be built into and used by *P*). A virtual black box property states that any information that can be extracted from the text of *P'* can be also extracted from the input-output behavior of *P'*.

In the last years, Barak's techniques were subsequently extended, e.g. by solutions based on semi-honest oblivious transfer that do not rely on collision-resistant hashing [3], or by new applications of obfuscation for network coding techniques, such as fountain code that is a rateless erasure code [4].

There are many different practical approaches to obfuscation, which are described and summarized in [5]. Most of them are based on compiler technologies, and some methods require the presence of a source code of the obfuscated program [6]. Others operate at intermediate level or at machine code on the target platform [7]. Usually, one of three directions is followed: source code obfuscation, which is achieved through source code transformations; intermediate level obfuscation through transformations on some precompiled code; or machine level obfuscation through binary rewriting.

Intermediate level obfuscation deals with a target-platform independent intermediate code. Such code is usually a description of the high-level statements with some simpler instructions that accurately represent the operations of the source code statements. It is important that this code will not be executed in a real processor, it is only an internal representation of a program. Since intermediate code uses simpler constructs than the high-level language, it is much easier to determine the data and control flow. In addition, this is very important for obfuscation algorithms.

An advantage of intermediate level obfuscation is that we can create a target-independent infrastructure. It means that for each platform that needs to be supported we only have to write the "machine code to intermediate code" and "intermediate code to machine code" translators, and the intermediate level obfuscator does not change. If we need to port our obfuscator to another platform, we only need to write another translator for a new processor.

The rest of the paper is organized as follows. In Section 2, we discuss the related work by pointing out negative and positive results in the state-of-the-art, and justifying the concepts of our research. In Section 3, we present the intermediate level obfuscation method. The focus is set on dynamic calculation of constants, generation of fake instructions, meshing of control flow transitions, and basic blocks partitioning. Finally, we draw conclusions and outline further work.

2   Background

Negative results

The essence of obfuscation is to entangle the code and eliminate the majority of logical links in it; that is, to transform the code so that it becomes complex enough for analysis and unauthorized modification. A general method for obfuscating programs would solve many open problems in cryptography. However, Boaz Barak has presented families of functions that cannot be obfuscated, since there exists a predicate that cannot be computed from black-box access to a random function in the family, but can be computed from a non black-box access to a circuit implementing any function in the family [2,8]. A later paper of Goldwasser and Kalai [9] shows the impossibility and improbability of obfuscating more natural functionalities.

Positive results

The classes of functions for which obfuscation was ruled out in [2] and [9] are somewhat complex. Quite another issue is the fact that obfuscation can be performed for simpler circuits [10]. We see that in spite of negative results for general-purpose obfuscation, there are positive results for simple functionalities, such as point functions. Canetti [11] shows that under a very strong Diffie-Hellman assumption, point functions can be obfuscated. Further works of Wee [12], Dodis and Smith [13] relax the assumptions required for obfuscation and consider other related functionalities.

Our work

In our approach, we do not restrict ourselves to point functions and do not assume simpler circuits. Obfuscation is understood as a program transformation technique, which attempts to convolute the low-level semantics of routines and aims to counteract the reverse engineering. We have shown in [14] that by restricting ourselves to automatic generation of additional fake operations, we cannot guarantee the absence of effectively optimized algorithm, which could restore the original sequence and deobfuscate the routine. However, the problem can be solved if we neglect Barak's functionality principle; that is, let the functionality of obfuscated routine $O(M)$ be different from the functionality of the original routine M. The solution lies in introduction of a global fake context.

With respect to a routine, we define two contexts: local and global. Local context is private to a particular routine and expires (disappears) when the routine execution is finished. An example of such context is local variables stored on the local stack. Global context may be shared across routines and does not expire immediately after a routine execution. It can be composed from different global parameters, such as pointers to memory buffers, control flow graph parameters, and initializing values, provided as input to a routine. The problem of mixing contexts has been discussed in [17].

We have proven that the problem of determining the significance of operational logic in such obfuscated routine is NP-complete [14]. We have also worked out a general approach to intermediate level obfuscation method and presented a bird-eye view of an obfuscation algorithm, pointing out general problems and proposing solutions [15].

3   Contribution

For intermediate representation, we use a three-address code (TAC), since TAC is not specific to a language being implemented (unlike P-code for Pascal and Bytecode for Java). In addition, the TAC instruction set is sufficient in translation of assembly code [16]. However, there remain a number of problems, especially with input data analysis.

The main problem is a proper selection (and isolation) of different kinds of data sets. It is obvious that e.g. constant values, abstract memory regions, and dead variables must be detected and separately processed by the obfuscating algorithm.

For successful input data analysis at intermediate level, we need the following information about the routine to be obfuscated:

1) three-address code representation of the routine;

2) information about abstract memory cells accessed by instructions of three-address instruction code;

3) information about arguments passed to the routine (estimated values, number of parameters, etc.)
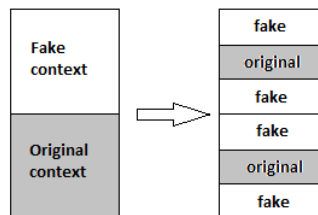

Figure 1
Proper mixing of contexts

We believe that the restriction of input data to just two types – (1) that can be moved to a fake context, and (2) that cannot be moved there – is very limited. Our task is to ensure the non-optimizability of the obfuscated routine, and for that a more thorough input data analysis is required. Otherwise, we will not be able to ensure the proper context mixing (Fig. 1) and optimization resistance. Below we introduce the input data analysis process that includes the following steps:

1) Separation of a set of "dead" local variables (set *D*). That means isolation of such variables that after a certain point are no longer used in the routine. For each point of the routine, we can associate a set *D* that reasonably can change its composition from point to point.

2) Separation of a set of simple type variables (set *C*). By the simple types, we define the following (C programming language) base data types: *char, short, long, longlong*. Some of the variables from the set *C* will be moved to additional fake context for better mixing original and fake contexts (Fig. 1).

3) Separation of a set of complex type variables (*CO*). Some of these variables will also be moved to additional fake context.

4) Separation of abstract memory regions, which constitute indivisible memory elements that have already been used in the subroutine as a unit and will not be used as a unit any more (set *CA*).

   Let us consider the following example code:

   ```
   %eax := &struct_1
   param eax
   call some_func
   //
   // hereinafter only tags of struct_1 are being used, but
   not the structure as a unit
   //
   ```

   In this example, *some_func* function gets a pointer to *struct_1* structure as a parameter. *some_func* works with the tags of *struct_1* (e.g. modifies the values). Later on, the routine performs some actions depending on the values of structure tags. It means that after having called *some_func* the structure is no longer used as a unit, but its tags are used separately, so in fact they correspond to ordinary variables. Such tags (elements) constitute the *CA* set. Consequently, the *CA* set should be determined and assigned for each point of the routine. Like for *CO*, elements of *CA* can also be moved to a fake local context.

5) Separation of abstract memory regions that constitute indivisible memory elements, but that are used separately within one or more basic blocks (set *CB*). These variables can be moved to the fake local context only within the boundaries of the basic block in which they are used. Subsequently, the values of these variables are to be restored.

6) Separation of abstract memory regions with known values (the set *V*). The set *V* can be left empty if there are no such variables in the routine.

The main point of the input data analysis is to ensure separation of different types of data for obfuscating method. The phases of obfuscation are the following:

1) Dynamic calculation of constants and some certainly known values.

2) Selection of a number of fake instructions per original.

3) Generation of fake instructions.

4) Generation of dead code.

5) Meshing of control flow transitions.

6) Partitioning of basic blocks.

Let us consider the above steps in details.

3.1   Dynamic Calculation of Constants and Known Values

A reverse engineer may use signature search to make certain conclusions about the functional logic of the routine being analyzed. This step is necessary as an obstacle to a signature search.

If a routine contains constants that are specific for implementation of some cryptographic algorithm [18], then its analysis becomes much easier, at least by the usage of a signature search. However, if there are memory regions, which content is exactly known in some fixed points of the routine code, one can increase the complexity metrics of a routine by introducing dynamic calculation of such values (to escape from constants). Thus, reverse engineering becomes more complicated.

It is important to mention that in order to exclude the possibility of routine optimization and to make it optimizer-resistant, constants should not be fixed in code, but instead they are to be calculated during runtime. To provide parameters for runtime calculation, we use global (with respect to the routine) initializing values, which are a special type of fake arguments intended to conceal the constants. The main goal here is to complicate the signature search in the entangled code. For even greater complexity, not only the routine constants are to be concealed, but a number of other well-known constants, e.g. the magic initialization constants of MD5 and SHA-1 [18]: *0x98badcfe, 0x10325476*, etc.

Calculation algorithms can vary from a trivial addition to a sophisticated cryptographic algorithm. The choice of a particular algorithm depends on the desired execution speed (i.e. acceptable execution slowdown) of the obfuscated routine. Furthermore, it should be noted that using very complex and slow algorithms is not a good choice, since the growth of constants calculation complexity is not directly proportional to the growth of deobfuscation resistance (i.e. resistance to reverse-engineering).

3.2   Selection of Fake Instructions per Original

The fake per original (FPO) number is a global user-defined parameter. The obfuscation algorithm generates a FPO number for each original instruction; this number indicates the desired approximate number of fake instructions per single original. The greater the numerical value of this parameter, the more difficult it is to deobfuscate the routine and to determine its original operational logic.

To provide higher deobfuscation resistance, any fake variable must not be disposed until it has been used at least once. Therefore, original instructions must interact with fake context, as well as fake instructions must interact with original context. Let us denote by M_{W_ORIG} and M_{W_FAKE} the sets of memory regions that original and fake instructions write to; M_{R_ORIG} and M_{R_FAKE} will stand for the sets of memory regions that original and fake instructions read from. So we get the formal description:

$$M_{W\_ORIG} \cap M_{W\_FAKE} \neq \Theta$$

$$M_{R\_ORIG} \cap M_{R\_FAKE} \neq \Theta \qquad\qquad (1)$$

However, with increase of FPO, execution slowdown is increased as well. It is important to mention that a number of fake instructions is not necessarily exactly equal to FPO, so that FPO serves as some approximate value. Sometimes, the number of fake instructions can exceed FPO by several instructions in order to comply with the aforesaid.

3.3   Generation of Fake Instructions

When a fake instruction is generated, the obfuscating algorithm takes one of the following actions:

1) Write the instruction to any free abstract cell from the set *D*.

    The term "free" in this context means that the cell is not filled with data. It should be noted that a fake abstract memory cell can have 4 states: *FILLED*, *FREE*, *USED*, and *NOT_INITIALIZED*. If a cell is in a *FILLED* state, then we can write a new value to this cell only if its old (original) value is used for the new value calculation. If a cell is in *FREE* state, then we can write a new value to it without any restrictions. Ia a cell is in *USED* state, then it contains some value, which is necessary to perform some further operation. For example, such cell can contain a value used for dynamic calculation of some constants. The value written to a *FREE* cell can be read from any other abstract cell. If we read from a fake abstract memory cell, which is in *FILLED* state, then after having read the value we procced to changing the state of such cell to *FREE*. The cell with *NOT_INITIALIZED* state cannot be read. However, we can write some new value to such cell only if a new value

is not computed as a result of some operation with the previous (unknown) value of this cell.

2) Write the instruction to any abstract cell from sets *C'*, *CA'* or *CB'*.

The *C'*, *CA'* and *CB'* sets are such memory cells, the content of which has been moved to a fake local context. Consequently, *C'*, *CA'* and *CB'* are subsets of *C*, *CA* and *CB,* respectively. Writing the generated instruction to these abstract cells should be done as described in Point 1.

3) Write the instruction by performing arithmetical operations. The sets are chosen as described in Point 2, but arithmetic operations are performed with original values of the cells.

4) Write the instruction to the fake global context. The sets and operations are chosen as described in Points 1-3, but the instruction is interacting with the fake global context.

5) Generate control flow transitions (jumps) to dead code.

Such jumps should look "plausible" and should not differ from the original control flow transitions (CFTs). This means that conditions, at which the transitions take place, must seem feasible. It definitely requires the usage of the fake global context.

3.4   Generation of Dead Code

Dead code is a piece of code that is never executed. The task of an obfuscating algorithm is to ensure that neither a reverse-engineer (analyst) nor an automatic deobfuscation tool can prove that the specific piece of code is never executed (dead). It is obvious that such code should not differ a lot from the original instruction set or data. Its main purpose is to increase the complexity metrics of obfuscated code. Moreover, by injecting dead code we can counteract the alias analysis, analysis of values of the abstract memory cells, and consequently the determination of individual variables, structures, arrays, etc.

In general, alias analysis determines whether separate memory references point to the same area of memory. This allows the compiler to determine what variables in the program will be affected by the given statement. Thus, we can significantly reduce the effect of optimizing transformations with respect to obfuscated routine through compliance with the conservativity principle of preliminary code analyzing algorithms. For example, we can counteract the alias analysis by using the following techniques:

1) Initialization of abstract cells (including those used in original code) with pointers to other abstract cells.

2) Creation of loop structures in a dead code, in which previously initialized (see point 1) abstract cells are used.

In other aspects, the algorithms used in the dead code should be similar to those used in the original executable code.

3.5   Meshing of Control Flow Transition Blocks

A branch is a piece of code in a computer program which is conditionally executed depending on how the flow of control is altered at the branching point. Explicit branches in high-level programming languages (e.g. C/C++) usually take the form of various conditional statements that encapsulate the branches of code that should be executed (or not) upon some condition; low-level instructions that define corresponding branches of code are called jump instructions. A three-address code has support for both conditional and unconditional jumps, which are essentially *goto* statements.

In general, jump instructions have unconditional and conditional forms where the latter may be fulfilled or not, depending on some conditions. The truthiness of this condition is typically evaluated and temporarily stored by some previous instruction, but not necessarily the one immediately before. Usually, this temporary information is stored in a flag register.

3.5.1   Unconditional Jumps

Fig. 2 shows an example of a meshed CFT block that represents an unconditional jump. *Blocks 1, 2, 3, 4, 5,* and *6* are basic blocks. Considering the memory map, *Block 3* directly follows *Block 4*, while other blocks are contained separately. On this figure, full straight lines denote unconditional jumps in direction of the arrow. Dashed lines denote true jumps, i.e. the CFTs that actually take place in the original routine. Dotted lines denote transitions that do not take place in the routine.

On Fig. 2, we see three different types of jumps denoted by straight, dashed, and dotted lines, respectively. However, since CFT conditions are dynamically calculated using a global context, for a reverse engineer the transitions are equally likely, and therefore the reverse engineer cannot distinguish them. Moreover, *Blocks 2, 5* and *6* may contain either dead code, or a piece of code that is actually executed. It should be noted in particular that *Blocks 3* and *4* form a single memory region, so *Block 3* can be totally fake. In this case, it does not matter whether the control flow is transferred to *Block 3* or to *Block 4*.
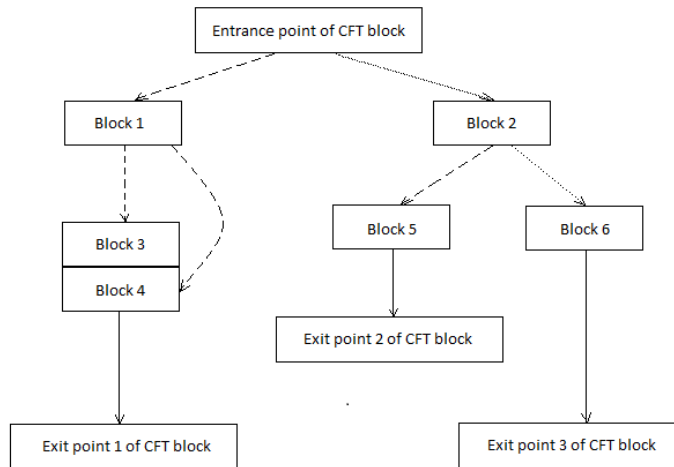
Figure 2
Meshed unconditional jump

3.5.2   Conditional Jumps

As we know, there are six logical conditions (comparisons): greater, less, equal, not equal, greater or equal, less or equal. We can make comparisons between variables, while the value of one of them is fixed. For example, *if (a > 3) goto L*. We will call this kind of comparison a *constant comparison*.

Comparisons can also be made between variables, while the values of both of them are unknown. For example, *if (a > b) goto L*. We will call this kind of comparison a *variable comparison*.

In computer programs, both types of logical comparisons are widely used, and if reverse engineered, can contain sensitive information for better understanding of a logical and functional structure of a program. Consequently, logical comparisons must be taken into account during obfuscation process.

A three-address code instruction *if (a > 3) goto L* can be represented as shown on the listing below:

```
if (a<0) goto L1;
; Garbage code
if (a<2) goto L2;
; Garbage code
if (a>6) goto L3;
; Garbage code
if (a<4) goto L4;
; Garbage code
if (a>=4) goto L5
```

Labels *L1*, *L2*, and *L4* denote the code that will be executed if the original condition *a > 3* is not satisfied. Labels *L3*, and *L5* denote the code that will be executed if the condition is satisfied.
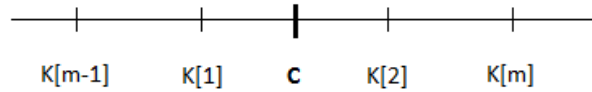

Figure 3
Number scale with generated constants

The proposed method of obfuscating constant comparisons is described below. At first, integers are selected on a number scale in such a way that among these integers there must be numbers greater and less than the original constant *C*. The distance between any adjacent selected integers is equal to two. Fig. 3 represents a number scale with original constant *C* and generated constants *K[i]*, where *i=[1..m]*.

Propositions:

$$a > C \equiv a \geq C+1 \tag{2}$$

$$a < C \equiv a \leq C+1 \tag{3}$$

$$a \geq C \equiv a > C-1 \tag{4}$$

$$a \leq C \equiv a < C+1 \tag{5}$$

$$a > C+i \Rightarrow a > C \tag{6}$$

$$a < C-i \Rightarrow a < C \tag{7}$$

$$a = C+i \Rightarrow a \geq C+i \tag{8}$$

$$a = C-i \Rightarrow a \leq C-i \tag{9}$$

$$a = C \equiv (a > C-1) \& (a < C+1) \tag{10}$$

$$a \neq C \equiv (a \geq C+1) \& (a \leq C-1) \tag{11}$$

For better understanding of meshing method, let us consider the logical comparison *a > C* as an example. Here follows an algorithm for meshing this logical comparison:

1) Choose *m* and generate *K[i]*, where *i=[1..m]*.

2) Select *i*.

3) Select one operation out of five: *a>K[i], a<K[i], a=K[i], a≥K[i], a≤K[i]*.

4) For selected operation do the following:

   a) If *a>K[i]* or *a≥K[i]* was selected:

For *K[i]>C* we should generate a CFT to code that should have been executed when original condition *a>C* is true, and then label *i* as "used", so that it will not be used at the next iteration. Otherwise we can have an ambiguous situation when *a* can be at the same time greater than, less than, or equal to *C*. In case of ambiguous situation, we should mark *i* as usable with all operations except for *a>K[i]* and *a≥K[i]*. In this case garbage code is to be generated, and further condition testing is requited. For the nearest right-hand adjacent constant *K[2]*, the following condition is to be generated: *a≥K[2]*.

b) if *a<K[i]* or *a≤K[i]* was selected:

If *K[i]<C*, then we should generate a CFT to the piece of code that should have been executed when original condition *a>C* is false, and then label *i* as "used". If *K[i]>C*, then we have an ambiguous situation.

5) Iterate steps 2-4 in a loop until all *i*-s are labeled as "used".

For *a≥C* the steps are similar to those described above with the only difference: if conditions *a>K[i]* (*a≥K[i]*) for *K[i]>C*, and *a<K[i]* (*a≤K[i]*) for *K[i]<C* are not satisfied, then consequently *a=C* and hence the condition *a≥C* is satisfied.

Similarly we can write this algorithm for *a<C* and *a≤C*.

Proposition.

A set of conditions, generated by the above meshing algorithm, coincides with the original condition.

Proof.

Let us prove this proposition with respect to condition *a>C*. The proof for other conditions is similar.

For all *K[i]>C*, the algorithm generates *a>K[i]* (or *a≥K[i]*). If the conditions are satisfied, the control flow is transferred to in the same block as if the original condition *a>C* were satisfied. Since difference between two neighboring *K[i]*-s is equal to *2*, we get: $|K[1]-C|=|K[2]-C|=1$. Thus, it follows:

$$a > C \equiv \forall K[i] > C : (a \geq K[2] \,\&\, (a > K[i], i = [1..m]) \tag{12}$$

For all *K[i]<C*, the algorithm generates *a<K[i]* (or *a≤K[i]*). If the conditions are satisfied, the control flow is transferred to in the same block as if the original condition *a>C* were not satisfied. Since $|K[1]-C|=|K[2]-C|=1$, we get:

$$\neg(a > C) \equiv \forall K[i] < C : (a \leq C \,\&\, (a < K[i], i = [1..m]) \tag{13}$$

∎

The proposed algorithm can be supplemented by a code duplication technology. In fact, if one creates multiple polymorphic duplicates of code that is executed if the condition *a>C* is true, and the control flow is transferred to different duplicates in

generated branches, then obfuscated code significantly better resists automatic deobfuscation tools.

For hiding constants, as well as for increasing the complexity metrics of a routine, we can use a method based on the following identity:

$$a > C \equiv a*k > C*k \qquad (14)$$

The * operation here is not a multiplication, but denotes any operation that satisfies (14). For example,

$$a > C \equiv a+k > C+k \qquad (15)$$

It must be specially noted that this method should be used with great caution, because it can possibly lead to an overflow error, and consequently the identity (14) will not hold true anymore. In this case we can use another operation, such as the following:

$$a > C \equiv a-k > C-k \qquad (16)$$

Herewith, the overflow error which arose in (15), will not arise in (16).

This method is suitable not only for constant comparison, but for variable comparison as well.

3.6   Partitioning of Basic Blocks

Basic blocks obtained at routine transformations are ranked by the following algorithm:

1) Determine the maximum number of functions that a basic block (BB) can be partitioned to: *MAX_NL-1*. Here *MAX_NL* stands for a maximum nesting level; this serves as an external parameter of the obfuscation algorithm. The greater the *MAX_NL* is, the more functions will be obtained from the BB.

| NL = 0 |
| NL = 1 |
| NL = 2 |
| . . . |
| NL = MAX_NL |
| . . . |
| NL = 2 |
| NL = 1 |
| NL = 0 |

Figure 4
Assigning nesting levels to groups of instructions in a basic block

2) Nesting levels are assigned to groups of three-address code instructions within the BB as shown on Fig. 4.

3) Zero-nested instructions *NL=0* remain at their original positions. These instructions are supplemented with additional instructions that assign arguments (parameters) for the subsequent nested functions.

4) Instructions of the first nesting level *NL=1* are transferred to a separate function. The operands of these instructions are replaced with the appropriate function arguments (parameters). As in the previous step, the function instructions are supplemented with instructions that assign arguments (parameters) for the subsequent nested functions.

5) Step 4 is repeated for all nesting levels sequentially until *NL=MAX_NL*.

Summary and Conclusions

In comparison with other known algorithms of obfuscation (Table 1), the intermediale-level obfuscator presented in this paper looks very promising. Its low complexity results from easiness of analysis of three-address code and simplicity of implementation of entangling transformations on the intermediate level. Other areas of comparison include the following:

1) Portability is an indicator of transferability of an implemented algorithm from one machine to another.

2) Flexibility is an indicator of the possibility of usage of the algorithm in different development environment or programming language.

3) Scalability is an indicator of degree of controllability of obfuscation process by the user.

Table 1
Comparison of different obfuscation techniques

	Collberg	*Wang*	*TAC IL Obfuscator*
Portability	no	yes	yes
Flexibility	medium	medium	high
Scalability	high	low	high
Complexity	high	medium	low

Collberg's algorithm [19] cannot be named as portable, because it was designed directly for use with the Java Virtual Machine. Intermediate level obfuscation algorithm is most flexible, since it is most isolated from high level programming constructs. Wang's algorithm [20] shows low scalability, because it uses very specific constructs. However, Wang's algorithm has a very high resilience, but specific opaque constructs are not protected at all, which is a significant drawback. The problem we have faced with Collberg's algorithm was that a high number of parameters controlling the algorithm makes empirical testing almost impossible. The intermediate level obfuscator, however, has just several parameters, and consequently we can control its output.

It should be specially noted that the presented method preserves its optimization-resistance only if there is no essential difference between routines, implementing original and fake operational logic. For example, if the instructions of the original routines used floating-point type operands, and fake instructions only work with integer numbers, then in such case the separation of original and fake instructions can be done automatically in polynomial time.

The great advantage of the presented intermediate level obfuscation method is that it can be applied to partitioned routines. Even if there is no possibility to add a fake global context to the original routine as a whole, it can always be done with respect to the partitioned routines with nesting level greater than zero. For that, we can use parameters of nested functions to pass pointers to buffers containing a fake local context. The resulting nested routines can be "scattered" in different parts of the application in an arbitrary manner.

Neither Collberg nor Wang provide a method for multiple obfuscation; that is, after having been obfuscated once, the program cannot be obfuscated again. The presented algorithm allows to obfuscate already obfuscated programs, or to obfuscate the selected routines of a program. By that, we obtain a multistage obfuscation technique.

In general, the presented method can be used to protect software from unauthorized analysis and modification, and consequently to prevent its reverse engineering. The algorithm based on this method is completely automatic and can therefore be used as a part of a software protection utility. The main advantage of this method compared to its counterparts is its platform independence. Doing obfuscation at intermediate level allows us to use the same software module at different hardware platforms.

This paper has set the ground for a new understanding of obfuscation. This research, furthermore, besides advancing academic research has major practical implications in software development, in counteracting software piracy, and in information protection. Intermediate level obfuscation raises the barriers to someone decompiling and stealing your code, and by that discourages casual attacks and makes one's intellectual property less likely to be stolen.

Future research of the authors will include, but will not be limited to, working out methods of translation from machine code into an intermediate representation and back. Such translation mechanisms must be implemented using machine-level obfuscation techniques, which would further increase the security of the program.

Acknowledgement

References

[1] Popa, M. (2011) Techniques of Program Code Obfuscation for Secure Software. *Journal of Mobile, Embedded and Distributed Systems*, Vol. 3(4)

[2] Barak B., Goldreich O., Impagliazzo R., Rudich S., Sahai A., Vadhan S., and Yang K. (2001) On the (im)possibility of Obfuscating Programs. In Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology, CRYPTO'01 (London, UK) pp. 1-18, Springer-Verlag

[3] Hessler A., Kakumaru T., Perrey H., Westhoff D. (2012) Data Obfuscation with Network Coding. *Computer Communications*, Vol. 35(1), pp. 48-61

[4] Bitansky N., Paneth O. (2012) From the Impossibility of Obfuscation to a New Non-Black-Box Simulation Technique. In Proceedings of IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS), pp. 223-232

[5] Balakrishnan A., Schulze C. (2005) Code Obfuscation Literature Survey. CS701 Construction of Compilers, Computer Sciences Dept., University of Wisconsin

[6] Ceccato M., Di Penta M., Nagra J., Falcarin P., Ricca F., Torchiano M, Tonella P. (2009) The Effectiveness of Source Code Obfuscation: An Experimental Assessment. In Proceedings of the 17th International Conference on Program Comprehension (Vancouver, Canada), pp. 178-187

[7] Fang H., Wu Y., Wang S., Huang Y. (2011) Multi-Stage Binary Code Obfuscation using Improved Virtual Machine. In ISC (X. Lai, J. Zhou, and H. Li, eds.), *Lecture Notes in Computer Science*, Vol. 7001, pp. 168-181, Springer Verlag

[8] Barak B. (2004) Non-Black-Box Techniques in Cryptography. PhD thesis, Department of Computer Science and Applied Mathematics, Weizmann Institute of Science

[9] Goldwasser S., Kalai Y. T. (2005) On the Impossibility of Obfuscation with Auxiliary Input. In Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, pp. 553-562

[10] Lynn B., Prabhakaran M., Sahai A. (2004) Positive Results and Techniques for Obfuscation. *Advances in Cryptology*. *Lecture Notes in Computer Science*, Vol. 3027, Springer Verlag, pp. 20-39

[11] Canetti R., Goldwasser S. (1997) Towards Realizing Random Oracles: Hash Functions that Hide All Partial Information. *Advances in Cryptology.*

Lecture Notes in Computer Science, Vol. 1294, Springer Verlag, pp. 455-469

[12]   Wee H. (2005) On Obfuscating Point Functions. In Proceedings of the 37th Annual ACM Symposium on Theory of Computing, pp. 523-532

[13]   Dodis Y., Smith A. (2005) Correcting Errors without Leaking Partial Information. In Proceedings of the 37th Annual ACM Symposium on Theory of Computing, pp. 654-663

[14]   Dunaev D., Lengyel L. (2014) Formal Considerations and a Practical Approach to Intermediate-Level Obfuscation. *WSEAS Transactions on Information Science and Applications*, Volume 11, pp. 32-41

[15]   Dunaev D., Lengyel L. (2012) Overview of an Obfuscation Algorithm. In Proceedings of the International Conference on Computer Science and Information Technologies, CSIT'2012 (Lvov, Ukraine), pp. 36-38

[16]   Grune D., Langendoen K. G., Jacobs C. J., Bal H. E. (2001) Modern Compiler Design. *Worldwide Series in Computer Science*, Chichester, New York, Weinheim: J. Wiley and sons

[17]   Dunaev D., Lengyel L. (2013) Aspects of Intermediate Level Obfuscation. In Proceedings of the 3rd Eastern European Regional Conference on the Engineering of Computer Based Systems, ECBS'2013 (Budapest, Hungary) pp. 138-143

[18]   National Institute of Standards and Technology (2002) "Secure Hash Standard", Fips 180-2, Federal Information Processing Standard, publication 180-2, tech. rep., Department of Commerce

[19]   Collberg C., Thomborson C., Low D. (1997) A Taxonomy of Obfuscating Transformations. Technical Report 148, Department of Computer Science, University of Auckland

[20]   Wang C., Hill J., Knight J., Davidson J. (2000) Software Tamper Resistance: Obstructing Static Analysis of Programs. Technical Report CS-2000-12, 2000

Thermal Insulation of the Clothing 2nd Royal Hungarian Army in Winter Campaign in the Light of Thermal Manikin Measurements

Zoltán Magyar[1], Tamás Révai[2]

[1]Budapest University of Technology and Economics
E-mail: magyar@egt.bme.hu

[2]University National of Public Service, Budapest
E-mail: drrt@t-online.hu

Abstract: This paper proposes to give a brief summary of thermal comfort of soldiers of the 2nd Royal Hungarian Army in Winter Campaign, at the River Don in 1943, and call attention to the importance of the proper clothing and the thermal sensation measured by thermal manikin. The soldiers in the proper sense of heat insulating garments affected no cold, and the effects of climatic factors were also reduced. The 2nd Hungarian Army in the Winter Campaign and military organization has been assigned to very difficult major tasks. Neither the desired equipment nor the clothing arrived as needed due to the overload of railways. The large number of freeze damage is also due to the clothing of soldiers, according also to that our results did not meet the criteria of the extreme variability of the weather conditions and the thermal comfort.

Keywords: clothing; thermal comfort; thermal insulation; thermal manikin

1   Introduction

The 70 years of the breakthrough at Don is one of the greatest tragedies in Hungarian history. More than 100 thousands of soldiers died, disappeared or were imprisoned. In addition, the Hungarians did not have cold-weather equipment, which were designed by the Germans, who promised weapons too, which finally never arrived [1]

It is important to identify the way clothing contributed to thermal comfort or discomfort, so our group on the base of the measurement, we defined the heat submission of the thermal manikin dressed in the uniform of the soldier serving at the River Don. The goals of the research projects were mainly carried out at the Department of Building Service Engineering at University of Pécs. We defined it the single body parts and the whole man's heat submission beside air velocity

changing in case of a different operative temperature. We compared the measured values, the reference with a value, and the heat submission of the naked human body. We proved that the heat insulating ability of the clothing worn at the bend of River Don can be measured and the worn clothing did not meet the requirements of thermal comfort [2].

Winter clothing of the 2nd Royal Hungarian Army [3]:

- **snow suit (jacket and trousers) and snow tires:** in snowy weather it served the adaptation to the environment

- **panties knitted or woven**: heated the lower leg and the body.

- **trusses**: served for warming the kidneys, the stomach, it is required to wear on the shirt.

- **knitted or woven sleeve:** kept warm the upper body and the arms.

- **leg warmers:** warmed the legs, and it was necessary to be long enough so that it covered the knees.

- **ankle warmers**: kept warm the ankles, it was unnecessary to wear them with boot.

- **knee warmers:** were helpful to keep the knees warm, especially for the ones, whose knees were directly in contact with the cold, such as riders, cyclists or motorbike users. It was always worn outside trousers [3].

- **wool footcloth:** was a good piece for keeping the foot and the ankle warm.

- **boots:** helped protecting the feet.

- **felt boots:** soldiers were **a**llowed to wear them only in dry weather.

- **warming worth:** was helpful keeping the palm warm.

- **snow caps:** were used to prevent the face, but not the head and the neck [3].

- **fur vest:** it served to keep the upper body warm, and it was recommended to wear it over the tunic, but under the jacket.

- **fur gloves:** the mitten fur glove was very helpful in keeping the hands warm.

- **cloth quilt:** was keeping the entire body warm., expect for the face, hands and feet.

- **fur hat/cap**: it was helpful for keeping the head warm, except for the face [3].

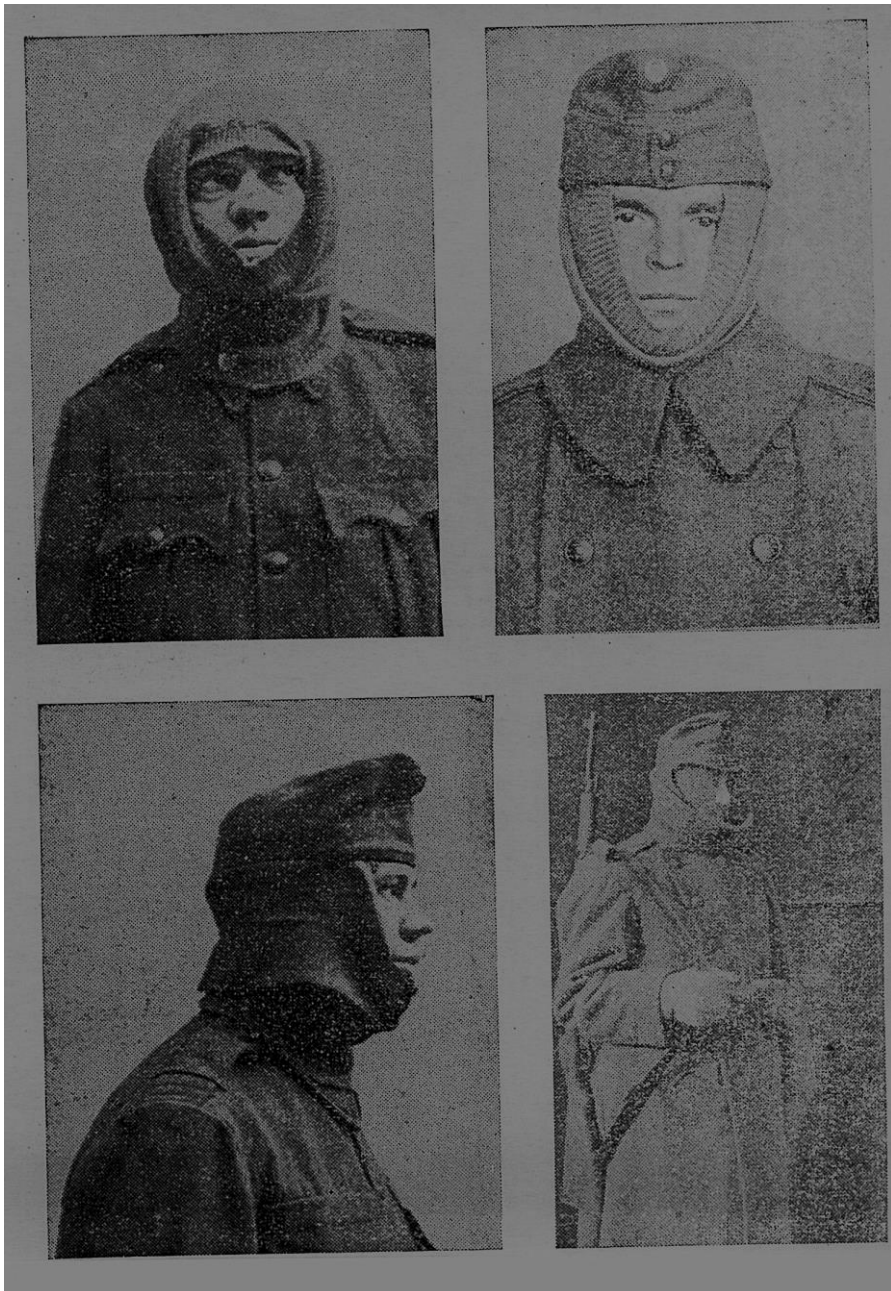- **Field cap:** was worn usually on the snow cap [3].

Figure 1
Field cap is worn on the snow cap [3], with copyright permission

Figure 2
Snow cap, which keeps the head, the ears and the neck warm, but not the face [3], with copyright
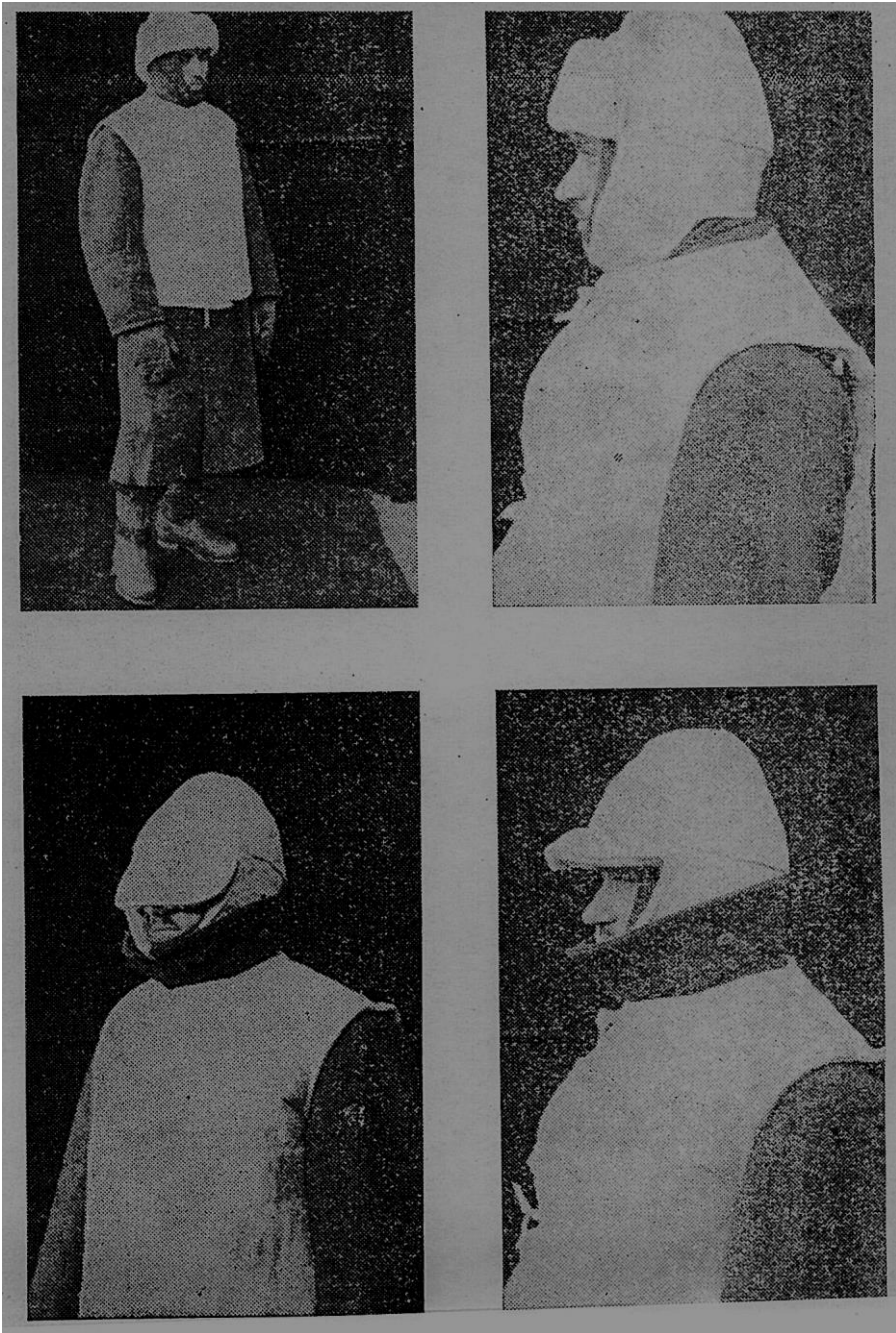permission

Figure 3
Fur hat, which keeps the head warm, but not the face [3], with copyright permission

2   Thermal Manikin

The clothing of the 2nd Royal Hungarian Army in winter campaign was tested on thermal manikin. The thermal manikin were made in the 1980's, in Hungarian Institute for Building Science, and it made by Swedish experts. Later, this thermal manikin were in the Technical University of Budapest, Department of Building Service Engineering, then from 2010 for research purposes it has been in University of Pécs, Pollack Mihály Technical Faculty, Department of Building Service Engineering, where the manikin got new, advanced datalogger and processing software [4]. The thermal manikin is a model with a body, a control unit and a data logger. The whole body is consisted of 16 body parts (Figure 4). The thermal manikin is a gauge system with big complexity, which was made of an average adult plastic puppet tallying with a man's body sizes. We kept a constant temperature on the body surface heated by electrical power. The electrical power of the 16 parts of the body is measured and logged. During the measurement we measured the body heat loss each parts of the manikin in the clothing under various ambient temperatures [4].

Figure 4
Different body part of the thermal manikin

The medical scientific literature gives the temperature of the human skin from 1919 [5] until nowadays [6]. It is an ongoing research in the medical science [7].

Figure 5 shows the set temperatures of the different body part based on the general medical practice. We developed new software and the set temperature can be changed.

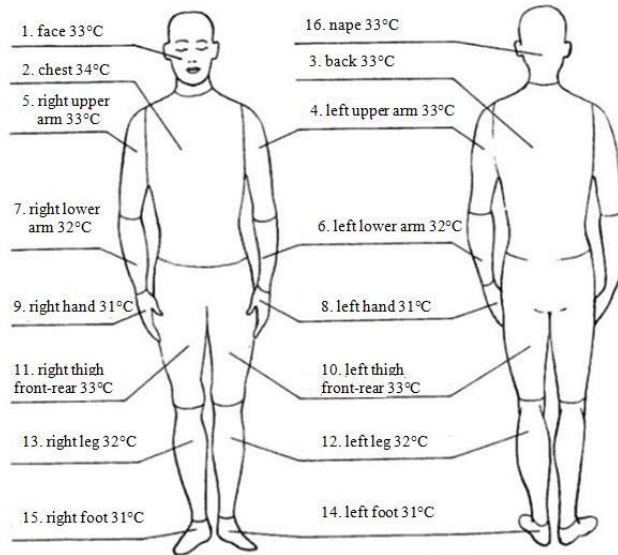The heat loss and heat sense of the people are influenced by the thermal insulation of clothing.

Figure 5
Surface temperature of thermal manikin [8]

3 Thermal Insulation of Clothing

The heat loss and the heat sense of the people are influenced by the thermal insulation of clothing. The convective and radiative heat transfer can be measured on the whole body surface with the thermal manikin. From the surface temperature and heating power data that measured at the body parts of the thermal manikin, by summation of body part area weighted data can be calculated the heat loss of the whole body and the thermal insulation of the clothing. The total insulation, that is thermal insulation of clothing and boundary air layer around clothing, can be calculated as follows [9], [10], [13]:

$$I_T = I_{cl} + \frac{I_a}{f_{cl}} \tag{1}$$

where:

I_T – total thermal insulation of clothing and boundary air layer, m^2K/W

I_{cl} – thermal insulation of clothing, m^2K/W

I_a – thermal insulation of boundary air layer, m^2K/W

f_{cl} – clothing area factor, that is the ratio of the outer surface area of the clothed body to the surface area of the nude body, -

To calculate the total thermal insulation we measured the heat loss of the thermal manikin dressed in the examined clothing. To calculate the boundary air layer around the body we measured the heat loss of the nude thermal manikin.

The used parallel calculation method [9], [11], [14] determines the total thermal insulation as an area-weighted average of the local insulations.

$$I_{T,\ parallel} = \frac{\left[\left(\sum_i f_i \cdot T_i\right) - T_a\right] \cdot A}{\sum_i H_i} \qquad \frac{m^2 \cdot K}{W} \tag{2}$$

We have to calculate the thermal insulation of boundary air layer around the nude body (I_a) similarly to the method of calculation of total thermal insulation.

The clothing area factor – the value of f_{cl} – can be determined by measurement, but approximate, indirect calculation is also possible, which correlation can be used [12]:

$$f_{cl} = 1 + 0{,}28 \cdot I_{cl} \tag{3}$$

In the practice the clo unit is used for the thermal insulation of clothing with the definition:

$$I_{cl} = I_{cl}\,[m^2 K/W]\,/\,0{,}155 \quad [clo] \tag{4}$$

4   Measurements and Results

The thermal manikin used the clothing of the 2nd Royal Hungarian Army in different outdoor temperatures (Figure 6).

Figure 6
Thermal manikin wearing the clothing of the 2nd Royal Hungarian Army

Figure 7 shows the specific heat loss of the body parts at the clothing of the 2nd Royal Hungarian Army was measured under different outdoor temperatures. The body part without clothing (such as the face) causes the high heat loss. The heat loss of legs and foots are relatively high.
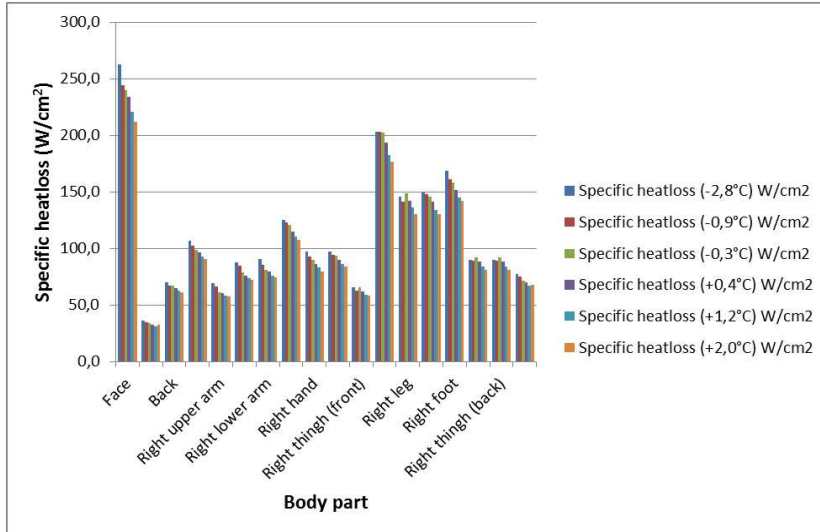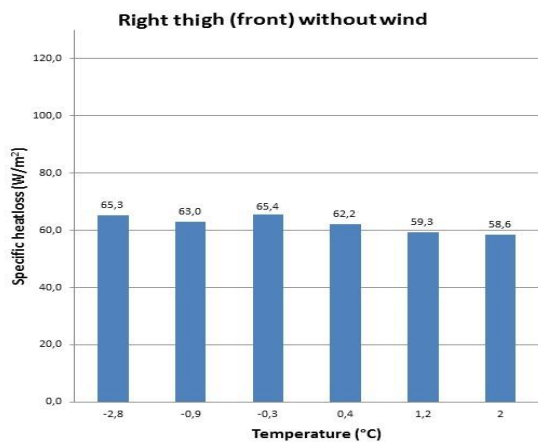

Figure 7
Specific heat loss of body parts in the clothing of the 2nd Royal Hungarian Army under different outdoor temperature

As an example we show the specific heat loss of the right thigh in the clothing of the 2nd Royal Hungarian Army under different outdoor temperatures (Figure 8).

Figure 8
Specific heat loss of the right thigh in the clothing of the 2nd Royal Hungarian Army under different outdoor temperatures without wind

The effect of the wind on the heat loss of the human body is significant. A relatively strong wind (15 km/h) was used below, while the -2.5 ^{0}C heat loss of the whole body nearly doubled its volume (Figure 9).
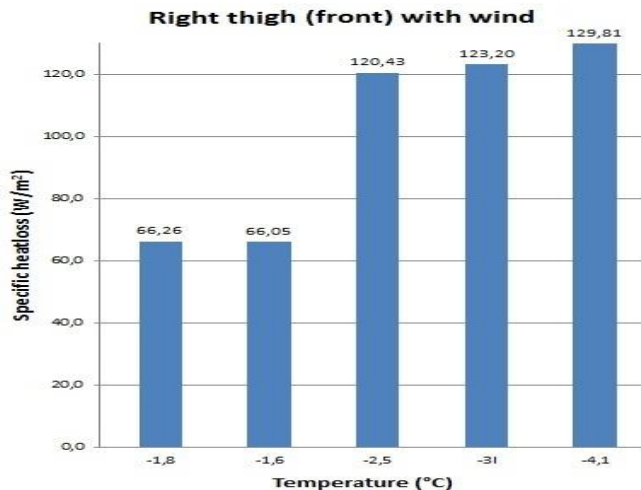

Figure 9

Specific heat loss of the right thigh in the clothing of the 2nd Royal Hungarian Army under different outdoor temperatures with wind

Conclusion

The calculation of the thermal insulation of the clothing was based on the equation (1),(2),(3) and (4), the result is 2,57 clo. The thermal insulation of the clothing of the 2nd Royal Hungarian Army is too low for the given outdoor temperature.

The measurements with thermal manikin justified, that the thermal insulation of the clothing of the 2nd Royal Hungarian Army in Winter Campaign, at the River Don in 1943 is measurable and the clothing worn did not meet the expectations.

The 2nd Hungarian Army in the Winter Campaign and military organization has been assigned to very difficult major tasks. The large number of freeze damage is also due to the clothing of soldiers, according also to that our results did not meet the criteria of the extreme variability of the weather conditions and the thermal comfort.

References

[1]   M. Horváth: A 2. magyar hadsereg megsemmisülése a Donnál. H. I. Bp., Zrínyi Kiadó, 1959, pp. 248-250

[2]   Z. Magyar, T. Revai: Behaveral Comfort and Nutritional Problems of the Hungarian Soldiers in World War II. Pakistan Journal of Nutrition, Vol. 10, No. 10, 2011, pp. 996-999

[3]   Chief of Staff of 2nd Royal Hungarian Army: Winter Campaign (notes) 2. booklet. Archives of Military History. The financial services of the Army in the winter, Vol. 58, No. 128, 1944, pp. 1029-3

[4]   Banhidi, L – Magyar, Z – Revai, T: Investigate History with the Thermal Manikin, Hungarian HVAC, 2010/11. pp. 6-10

[5]   F. G. Benedist, W. R. Miles, A. Johnson.: The Temperature of the Human Skin, Proc. Natl. Acad. Sci. USA. 1919 June 5 (6), pp. 218-222

[6]   Y. Liu, L. Wang, J. Liu, Y. Di: A Study of Human Skin and Surface Temperature in Stable and Unstable Thermal Environments, Journal of Thermal Biology, Elsever, Vol. 33, Issue 7, October 2013, pp. 440-448

[7]   M. Ponec: Characterization of Reconstructed Skin Models, Skin PharmacolAppl. Skin Physiol. 2002. 15. Suppl. pp. 4-17

[8]   Z. Magyar: Applicability of the Thermal Manikin for Thermal Comfort Investigations, PhD Thesis, Szent István University, Gödöllő, 2011

[9]   ISO/FDIS 9920:2007, Ergonomics of the Thermal Environment – Estimation of Thermal Insulation and Water Vapor Resistance of a Clothing Ensemble

[10]   ISO 15831: 2004, Clothing – Physiological Effects – Measurement of Thermal Insulation by Means of a Thermal Manikin

[11]   H. Attonen, J. Niskanen, H. Meinander, V. Bartels, K. Kuklane, R. E. Reinertsen, S. Varieras, K. Kuklane, R. E. Reinertsen, S. Varieras, K. Soltynski: Thermal Manikin Measurements – Exact or Not?, International Journal of Occupational Safety and Ergonomics (JOSE) 2004, 10 (3). pp. 291-300

[12]   E. McCullogh, B. Jones, J. Huck: A Comprehensive Data Base for Estimating Clothing Insulation. ASHRAE Transaction, USA, 1985. 91 (2) pp. 29-47

[13]   H. Meinander, H. Anttonon, V. Bartels, I. Holmér: Overview of the SUBZERO Project, 2nd European Conference on Protective Clothing and NOKOBETEF 7, Montreux, Switzerland 2003, pp. 9-11

[14]   H. Anttonon: Subzero Project: Manikin Measurements – Exact or Not, 2nd European Conference on Protective Clothing and NOKOBETEF 7, Montreux, Switzerland 2003, pp. 16-21

Economic Optimum of Thermal Insulating Layer for External Wall of Brick

Jozsef Nyers

Obuda University Budapest, Becsi ut 96, 1034 Budapest, Hungary
Subotica Tech, Marko Oreskovic 16, 24000 Subotica, Serbia
e-mail : jnyers@vts.su.ac.rs

Slavica Tomić

Faculty of Economics Subotica, Segedinski put 9-11,24000 Subotica, Serbia
e-mail : tomics@ef.uns.ac.rs

Arpad Nyers

Subotica Tech, Marko Oreskovic 16, 24000 Subotica, Serbia
e-mail : nyarp@vts.su.ac.rs

Abstract: This paper analyzes the technical and economic optimum thickness of a thermal insulation layer of an external wall. The observed wall was made of brick and the used thermal insulation material was polystyrene. The heat transfer through the wall takes place in the stationary regime. The mathematical model consists of algebraic equations for investment, exploitation and saving. The graph-analytical method was applied to solve the mathematical model. The applied optimization criterion is the minimum payback period of the investment. The numerical results obtained by the simulation are presented graphically. The optimum thickness of the thermal insulation layer can be seen from the diagram. Based on current prices in Serbia 2013 the techno-economical optimum thickness of thermal insulation layer (polystyrene) is 9 cm. The minimum payback period is 1.96 years.

Keywords: optimum; mathematical model; thermal insulation; criterion

Nomenclature

q heat flux per unit area$[W/m^2]$

Q heat per unit area per year$[Wh/m^2/year]$

ΔQ heat difference per unit area per year$[Wh/m^2/year]$

α convective heat transfer coefficient$[W/m^2/°C]$

λ conductive heat transfer coefficient$[W/m/°C]$
k overall heat transfer coefficient$[W/m^2/°C]$
t temperature$[°C]$
Δt temperature difference$[°C]$
τ time period per year [h/year]
δ thickness of thermal isolator layer

Subscripts and superscripts

i   input
o   output
m   middle

1   Introduction

Serbia now has the lowest level of energy efficiency in Europe. It is located on the bottom of the list among the countries that use energy rationally. Buildings in Serbia are huge energy consumers. The average consumption of final energy in buildings of the European Union in thermal purposes is 138 kWh/m[2].In most developed countries the tendency is to reach a value below 70 kWh/m[2]. In Serbia, it is about 200 kWh/m[2]. Therefore, the reduction of energy consumption and improvement of energy efficiency is necessary.

The effectiveness of this study is an energy-economic category. Achieving maximum results with minimal investment is a general principle of economic efficiency. It is necessary to increase efficiency and reduce costs, reduce the harmful impact on the environment. The solution reduces the use of natural resources, less waste, the air is less polluted. In this way, the demanding task of formula 3 E - energy, economy, ecology can be realized.

Respecting the principles of energy efficiency has become a liability, not the individual's choice. The introduction of the directive on the energy performance of buildings (EPBD - Energy Performance of Buildings Directive), the European Union is trying to reduce the amount of energy used in buildings.

Directive EPBD includes the adoption of appropriate regulations, incorporated in the legislation of Serbia. The study of energy efficiency dating September 2012 has become a legal obligation. This means that all houses built according to this law, when used, will consume less energy. But new buildings make up a small percentage of the total amount of constructed buildings.

From the aspect of energy saving, it is important to carry out the reconstruction of the existing buildings. 50% of these buildings was built before the existence of any regulations on thermal protection.

There are numerous studies which were performed on the definition of the optimum thermal insulation thickness[3]. Some of them are based on the degree-days method.This is calculated as the difference between the base temperature and the mean outdoor air temperature. Betul Bektas, Ayca Aytac Gulten and Teoman Aksoy are investigating the effects of the wall type and degree day values on the optimum insulation thicknesses for different insulation materials. The calculation was carried out for four different cities in different climate zones in Turkey. The cost of fuel decreases with the increase in insulation thickness. The total cost is the sum of the cost of fuel and insulation material. The optimum insulation thickness is determined as the minimum of the total cost curve.[4] A number of other authors also used the same "economic energy" method, including Bolatturk A,.K. Çomaklı, B. Yüksel,. J. Yu, C. Yang, L. Tian, D. Liao [5], [6], [7]. Several studies used the degree time method with financial components-market discount rate (for the value of money), the inflation rate (for the energy cost), to calculate the net energy savings [4], [8]. [9]. Dynamic transient models based on numerical methods were also applied in the literature [10], [11].

From the point of view the economic-energy efficiency of buildings, it is primary to use the thermal insulation layer to cover all external surfaces. The thermal insulation layer greatly reduces building heat loss. The reduction of losses depends on the thickness of the thermal insulation layer.

Increasing the thermal insulation layer increases the investment costs, but also reduces costs in the exploitation. Costs for investment and exploitation have opposite tendencies. Thus it follows, there is the techno-economic optimum of thermal insulation layer.

The optimum can be found by applying appropriate mathematical models and an efficient mathematical method. From the standpoint of analysis, the important issue is the way results are presented. One of the convenient forms of representation is the graphic display of results. The graphics show the solution visually and trends of change solutions.

This paper analyzes the economic optimum thickness of the thermal insulation layer of the external wall of the brick. The analysis provides a mathematical model attached. The model is composed of functions to describe the cost of investment, exploitation and saving function. In the functions, the values are expressed in Euros. The independent variables are the thickness of thermal insulation layers and the payback period of the investment.

In order to solve the model's equations the graph-analytical method is applied.The pure analytical method is applicable but very complicated. Solutions are shown in the accompanying graphs for the cost of investment, exploitation and savings over the exploration and the payback period of the investment. The graphs show the optimum thickness of the thermal insulation layer. The optimum is obtained by equating functions for investment and saving functions during operation. The optimal solution thickness of thermal insulation layer is obtained only after the

determination and replacement of the optimum for the minimum payback period in the equation.

Based on the application of the mathematical model and graph-analytical mathematical procedure it can be concluded that in Serbia, in 2013, the optimum thickness of the thermal insulation layer is 9 cm if the minimum payback period of the investment is 1.98 years and if the heat source used is electric energy with the price of 0,08 eu / kWh.

2   The Physical Model

The observed physical system for techno-economic optimization consists of an outer wall with and without thermal insulation. The static part of the observed wall was made of 25-cm-thick bricks and thermal insulation of the polystyrene (styrofoam).

From the standpoint of thermal calculation transmission heat losses observed over the outer wall, there are two layers that produce heat-resistance, thermal insulation and brick. The transmission heat loss depends on the thickness and physical characteristics of the layers.

The ventilation heat loss is not taken into account because it does not affect the techno-economic optimum of the thermal insulation layer's thickness. These losses depend on the size of the gap sealing windows and external doors.
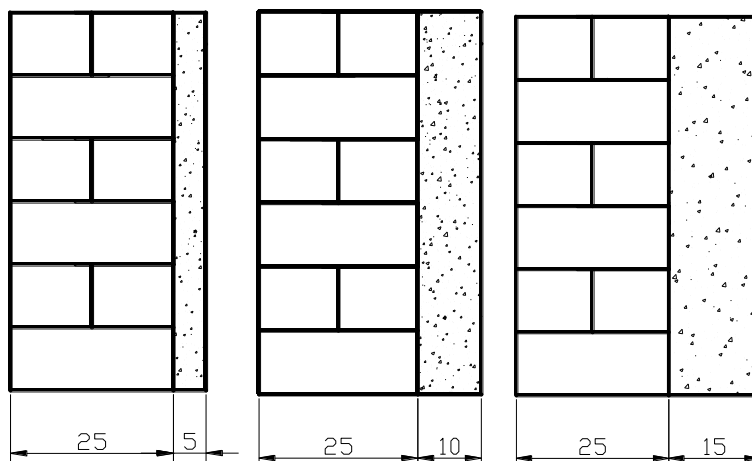

Figure 1

The cross section of the external walls of 25-cm-thick bricks with thermal insulation

layer thickness of = 5 cm, 10 cm, 15 cm

3 The Energy Part of the Mathematical Model

3.1 Transmission Heat Losses of External Walls

3.1.1 The Mean Heat Flux through the External Walls per Unit Area for the Heating Season

$$q_m = k \cdot \Delta t_m [W/m^2] \tag{1}$$

where are:

- overall heat transfer coefficient$[W/m^2 /^o C]$

$$k = \frac{1}{\frac{1}{\alpha i} + \Sigma \frac{\delta}{\lambda} + \frac{1}{\alpha o}} [W/m^2 /^o C] \tag{2}$$

- medium temperature difference between the temperature of the internal t_i and the mean outside air temperature t_{mo} for the heating season

$$\Delta t_m = t_i - t_{mo} [^o C] \tag{3}$$

3.1.2 The Amount of Heat of the Heating Season per Unit Area

$$Q = q_m \cdot \tau \tag{4}$$

$$Q = k \cdot \Delta t_m \cdot \tau \ [Wh/m^2/year] \tag{5}$$

where the following are denoted:

- duration of the heating season in hour τ $[h/year]$

3.1.3 Amount of Heating Energy for the Heating Season per Area of External Wall and for Various Thermal Insulation Layer Thickness δ

for $\delta = 0$ cm $Q_{(0)} = k_{(0)} \cdot \Delta t_m \cdot \tau$ (6)

for $\delta = 5$ cm $Q_{(5)} = k_{(5)} \cdot \Delta t_m \cdot \tau$ (7)

for $\delta = 10$ cm $Q_{(10)} = k_{(10)} \cdot \Delta t_m \cdot \tau$ (8)

for $\delta = 15$ cm $Q_{(15)} = k_{(15)} \cdot \Delta t_m \cdot \tau$ (9)

In general, the amount of heating energy for heating season per area as a function of thickness of insulating layer δ

$$Q_{(\delta)} = k_{(\delta)} \cdot \Delta t_m \cdot \tau \qquad \delta = 0, 5, 10, 15 cm \tag{9}$$

3.1.4 Amount of Saved Heat between the Thermal Insulated and Thermal Un-insulated External Wall of the Heating Season per Unit Area

$$\Delta Q_{(5)} = Q_{(0)} - Q_{(5)} \tag{10}$$

$$\Delta Q_{(5)} = k_{(0)} \cdot \Delta t_m \cdot \tau - k_{(5)} \cdot \Delta t_m \cdot \tau = (k_{(0)} - k_{(5)}) \cdot \Delta t_m \cdot \tau \tag{11}$$

$$\Delta Q_{(5)} = (k_{(0)} - k_{(5)}) \cdot \Delta t_m \cdot \tau \tag{12}$$

In general, the amount of heat

$$\Delta Q_{(\delta)} = (k_{(0)} - k_{(\delta)}) \cdot \Delta t_m \cdot \tau \; [Wh/m^2/year] \qquad \delta = 0, 5, 10, 15cm .. \tag{13}.$$

3.1.5 Amount of Saved Heat for the Heating Season per Unit Area in Continuous Form

$$\Delta Q_{(\delta)} = \Delta k_{(\delta)} \cdot \Delta t_m \cdot \tau \; [Wh/m^2/year] \tag{14}$$

where:

- The difference of heat transfer coefficient

$$\Delta k_{(\delta)} = \cfrac{1}{\cfrac{1}{\alpha_i} + \Sigma \cfrac{\delta}{\lambda} + \cfrac{\delta = 0}{\lambda} + \cfrac{1}{\alpha_o}} - \cfrac{1}{\cfrac{1}{\alpha_i} + \Sigma \cfrac{\delta}{\lambda} + \cfrac{\delta > 0 ..... n}{\lambda} + \cfrac{1}{\alpha_o}} \quad \delta > 0 .... n$$

$$[W/m^2 /^o C] \tag{15}$$

4   Economic Part of the Mathematical Model

4.1   The Cost of Heating in Euros as a Function of Thermal Insulation per Unit Area for the Heating Season

$$f_h = \sum_{0}^{n} Q_{(\delta)} \cdot e \; [eu/m^2/year] \tag{16}$$

4.2   Economic Savings in Euros as a Function of Thermal Insulation per Unit Area for the Heating Season

$$f_s = \sum_{0}^{n} \Delta Q_{(\delta)} \cdot e \; [eu/m^2/ year] \tag{17}$$

where: e [eu/kW] - unit price of energy for heating

5 Investment in Thermal Insulation

5.1 The Function of Investment Depending on the Thickness of the Thermal Insulation Layer(δ)

$$f_i = C_{isol} (\delta) + C_{tiple} (\delta) + C_{net} + C_{glue} + C_{pay}[eu/m^2] \tag{18}$$

The price of thermal insulation and the price of thermal insulation dowels depend on the thickness of the thermal insulating layer. Other terms are independent of the layer thickness.

5.1.1 The Constant Part of the Function of Investment

$$C= C_{net} + C_{glue} + C_{pay} \ [eu/m^2] \tag{19}$$

where:

- Price of fiber glass net per m^2

$$C_{net} \ [eu/m^2] \tag{20}$$

- Price of glue for polystyrene per m^2

$$C_{glue}= C_{glue,kg} \ [eu/kg] \ x \ n \ [kg/m^2] \tag{21}$$

- price of glue per kilogram

$$C_{glue,kg}. \ [eu/kg] \tag{22}$$

- consumption of glue per m^2

$$n \ [kg/m^2] \tag{23}$$

- Price of labor per m^2

$$C_{pay} \ [eu/m^2] \tag{24}$$

5.2 Investment Function in the Final Form

$$f_i = C_{isol} (\delta) + C_{tiple} (\delta) + C \ [eu/m^2] \tag{25}$$

where:

- Price of thermal insulation per m^2 depending on the thickness of insulating layer

$$C_{isol} (\delta) = C_{isol,m} \ [eu/m^3] \ x \ C_m \ [m^3/m^2m]x \ \delta \ [m] \tag{26}$$

- $C_{isol,m} \ [eu/m^3]$ Price of thermal insulation
- $C_m \ [m^3/m^2m]$ Number of m^2 in m^3 cubic meter depending on δ of thickness
- $\delta \ [m]$ Thickness of thermal insulation
- C_{tiple} Price of dowels per m^2 depending on the thickness of the insulating layer

6   The Simulation

6.1   About the Simulation

The simulation to determine the optimal thickness of thermal insulating layer is made based on the techno-economic situation in Serbia in 2013. The static part of the outer wall is made of 25-cm-thick bricks and the material of thermal insulation is polystyrene. The energy source of heat is electrical energy. The optimization criterion is the minimum payback period of investment.

6.2   Data for the Simulation

- thickness of the brick wall $\delta = 0.25$ m,

- thickness of the polystyrene to the thermal isolation $\delta = 1; 2; 3;.....20$ cm,

- coefficient of heat conduction for the brick wall $\lambda = 0.5$ W/m/^{o}C,

- coefficient of heat conduction of polystyrene $\lambda = 0.05$ W/m/^{o}C,

- convection heat transfer coefficient for the indoor air $\alpha_i = 8$ W/m^2/^{o}C,

- convection heat transfer coefficient for the outside air $\alpha_o = 20$ W/m^2/^{o}C,

- internal temperature $t_i = 20$ ^{o}C,

- medium outdoor temperature for heating season $t_m = 4$ ^{o}C,

- the duration of the heating season 200 days/year x 24 h/day = 4 800 h/year,

- price of polystyrene 40 eu/m^3 or 4 eu/m^2 for the thickness of 0.1 m, 2 eu/m^2 for, the thickness of 0.05 m, 6 eu/m^2 for the thickness 0.15 m,

- price of glue 5 eu/25 kg or 0.2 eu/kg,

- consumption of glue 7 kg/m^2,

- price dowels for polystyrene of 5 cm, 10 cm, 15 cm,

- price of labor for the installation of thermal insulating layer 6 eu/m^2.

6.3   Functions of Simulation

6.3.1   The Function of Economic Savings with Data

$$f_s = \Delta k_{(\delta)} \cdot \tau \cdot e \ [Wh/m^2/god] \tag{27}$$

- The difference of the overall coefficient of heat transfer through the wall

$$\Delta k_{(\delta)} = \frac{1}{\dfrac{1}{20} + \dfrac{0.25}{0.5} + \dfrac{1}{8}} - \frac{1}{\dfrac{1}{20} + \dfrac{0.25}{0.5} + \dfrac{\delta}{0.05} + \dfrac{1}{8}} \qquad (28)$$

- The mean temperature difference between the indoor air temperature and the mean temperature of the outside air during the heating season can be seen below

$$\Delta t_m = t_i - t_o = 20 - 4 = 16\ ^oC$$

$$(29)$$

- Period of the heating season

$$\tau = 200\,day\,/\,year \cdot 24\,h\,/\,year = 4800\ \ [h\,/\,year]$$

$$(30)$$

- mean unit price of electric energy in Serbia in 2013

$$e = 0.8\ [\ eu/kWh\ ] \qquad (31)$$

6.3.2   The Function of Investments in the Thermal Insulation, with Data

$$f_i = C_{isol}\,(\delta) + C_{tiple}\,(\delta) + C_{net} + C_{glue} + C_{pay}[eu/m^2]$$

$$f_i = 40\ eu/m^3\ x\ \delta\ m + 0 + 2\ [eu/m^2] + 1.4\ eu/m^2 + 6\ [eu/m^2][eu/m^2] \qquad (32)$$

where is the following are included:

- Price of fiberglass net per m^2 $C_{net} = 2$ [eu/m^2]

- Price of glue per m^2 $C_{glue} = 5$ eu/25 kg * 7 kg/m^2 = 1.4 [eu/m^2]

- Price of labor per m^2 $C_{pay} = 6$ [eu/m^2]

- Price of thermal insulation per m^3 $C_{isol}\,(\delta) = 40$ [eu/m^3]

- Price of dowels can be neglected due to the very low values. $C_{tiple}\,(\delta) = 0$

7   Mathematical Optimization Procedure

The mathematical model consists of two equations which are related through the thickness of thermal insulating layer. The equation for investment is linear while for the savings it is exponential. The economic optimum thickness of the insulating layer can be obtained using the analytical or graphical procedure.

7.1   Analytical Optimization Procedure

The analytical solution of the optimal thermal insulating layer in terms of the economy is obtained by equating the functions of investment and savings function.

$$f_i = C_{isol}\,(\delta) + C_{tiple}\,(\delta) + C_{net} + C_{glue} + C_{pay} \tag{33}$$

$$f_s = \Delta k(\delta) \cdot \tau \cdot e \tag{34}$$

$$fi = fs \tag{35}$$

After the replacement, the equation of investment is obtained

$$f = C_{isol}\,(\delta) + C_{tiple}\,(\delta) + C_{net} + C_{glue} + C_{pay} = 0 \tag{36}$$

The resulting equation (36) contains the three independent variables: The thickness of the insulating layer, the certain time period, the unit price of energy. The equation can be solved by one or the other of the third variable.

The economic optimum thickness of the thermal insulating layer is obtained by the minimum payback period. The analytical determination of the minimum payback period is possible by means of the first partial differencing of the implicit function (37), by the thickness of the thermal insulating layer. The first differential is equal to zero.

$$\partial f(\rho, \tau)/\partial \rho = 0 \tag{38}$$

From the equation (38) the minimum payback period is expressed. The result of the minimum payback period is replaced in the equation (36) and after an adjustment, the optimum thickness of thermal insulating layer may be determined.

From the above it can be seen that the analytical process is quite complicated and the obtained solution is analytic. For the analysis of the solutions in a wide range of changes in thickness of thermal insulating layers must work in graphics.

It follows from the foregoing that the complete process optimization works better in numerical-graphical form.

7.2   Numerical - Graphical Optimization Procedure

The numerical - graphical optimization procedure is much more appropriate than the analytical procedure. The solution is obtained in numerical form and after that, the solution is displayed in graphical form in the coordinate system. The analysis is visual, the graphics show the tendency of change in the investment, the savings, the cost of heating as a function of the wide range of changes of the thermal insulating layer's thickness.

The chart shows the optimum thickness of the thermal insulating layer and the minimum payback period of the investment.

The realization of the simulation is as follows: an algorithm was created for the numerical solution of the mathematical model and the algorithm is implemented in one of the numerical and mathematical package. The packages are supplied with graphics support for displaying numerical results.

In this case the algorithm for the numerical solution of function (36) is implemented in the mathematical package Matlab.

8   Results

Using simulation the obtained numerical results are presented graphically in Figures 2 and 3.
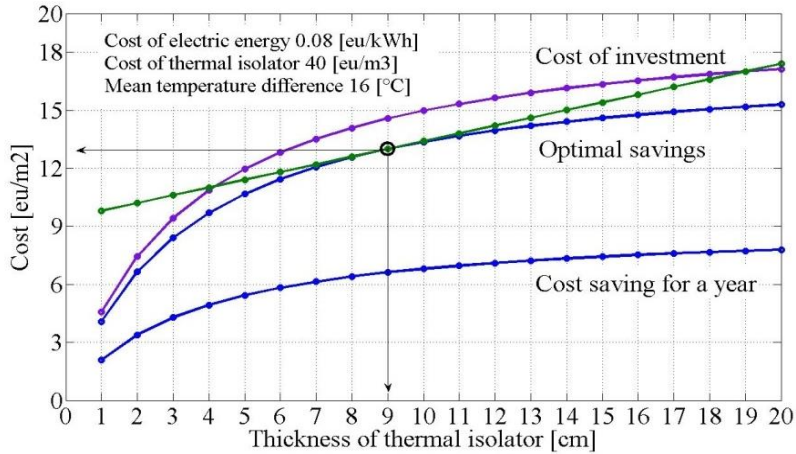

Figure 2

The function of investment, green straight line. Savings function is exponential blue-line and the investment return period of 1 year, from 1.98 years and 2.2 years.
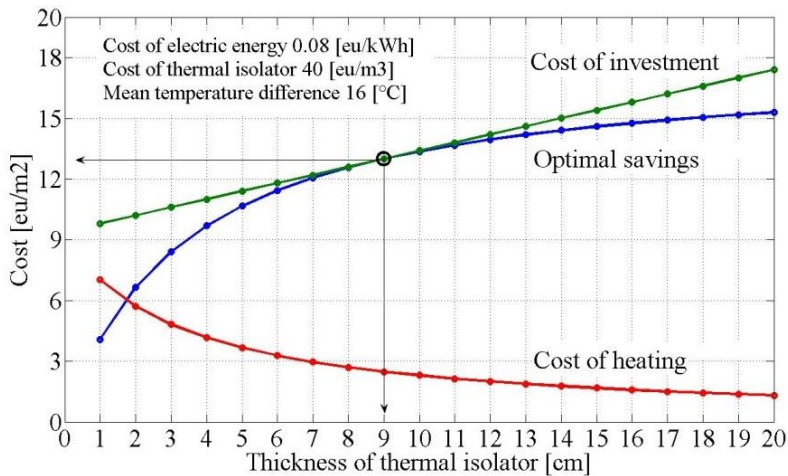

Figure 3

The optimum thickness of the thermal insulating layer in a joint investment function point, straight line and savings function exponential-line for a minimum payback period. The hyperbolic red line represents the cost of exploitation

Discussion

1   Discussion of the Mathematical Model

The mathematical model consists of the energy and economic part.

The energy of the model includes the equation of the heat loss through the wall and the equation of saving heat by using the thermal insulating layer. These algebraic equations are two dependent and two independent variables. The dependent variables are the amount of heat and the amount of heat saving, while the independent variables are the thickness of the thermal insulating layer, and the time of heating.

The economic part of the model includes a function of the cost of heating, the function of financial savings and the function of the balance between the costs and savings for a certain time period. These algebraic equations are two dependent and three independent variables. The dependent variables are the cost heating and financial savings and independent variables are the thickness of the thermal insulating layer, the observed time and energy prices.

In this paper, for a given techno-economic structure, the conditions of optimality are: 1. the investment equal to savings. 2. the payback period is minimal.

2   Discussion of Numerical Procedure

During the mathematical description of the defined problems, the authors tended to have a mathematical model that accurately describe the behavior of analyzed physical system.

The value-price of the model's elements is taken from the practice based on the data obtained from the market.

In order to solve the mathematical model the graphical-numerical procedure was chosen. The advantage of the procedure is the representation of the numerical results from the simulation in graphical form. Graphs are highly suitable for the analysis of the problem, because they provide a very wide range of changes of the independent variables and analyze the behavior of the observed dependent variables. The algorithms of mathematical models are realized in the mathematical-graphical package Matlab.

3   Discussion of Results Obtained by Simulation

The simulation was performed in order to obtain the optimum of thickness of the thermal insulation layer for the defined technical and economic conditions in Serbia in 2013. From a technical aspect, the examined outside wall is covered with thermal insulation of polystyrene, while from an economic point of view the fixed and variable costs are considered. All economic data are valid for the year 2013.

Figure 2. above shows the function of investment and savings functions for the reference period of 1 year; 1.96 years and 2.2 years. It can be noted that there is only one common point or value of function of investment and savings for the

payback period of 1.96 years. The common point defines the optimal value of the thickness of polystyrene which is 9 cm if the payback period is 1.96 years for the defined technical and economic conditions in Serbia.

The time of 1.96 years is the minimum payback period for the defined technical and economic conditions in Serbia in 2013. Only the minimal payback period provides the common solution of investment function and savings function. For the observed period of 1 year there is no solution, the functions pass each other, thus the requirement of a minimum period of payback period is not fulfilled.

For the period of 2.2 years the saving function intersects the line of investments in two points. The two solutions are obtained for the thickness of polystyrene. What is common in these two solutions is that their pay back period is the same for 2.2 years not least for 1.96 years.

The minimum payback period is obtained analytically, or graph-analytically. Analytically it is the first derivative of the function (36) which is equal to zero. Graph-analytically the minimum is obtained by moving closer the graphics savings while not touching the line of investment.

Figure 3 shows the functions of investment, savings and the cost of heating per square meter of external wall insulation with polystyrene. Graphs apply to the technical and economic conditions for Serbia in 2013. The common point of the lines of investment and savings defined the optimum thickness of polystyrene for a minimum payback period.

In the graph of the cost functions it can be seen that the optimum thickness of 9 cm polystyrene cost of heating per m^2 of external thermal insulated brick wall is 2.5 eu/m^2, the noninsulated wall is 9 eu/m^2. The saving is 6.5 eu/m^2 while the payback period is 1.96 years in Serbia in 2013, if the heat source used is electric energy with a price of 0.08 eu / kWh.

References

[1]   Asiye Aslan, Bedri Yuksel, Tugrul Akyol: Energy Analysis of Different Types of Buildings in Gonen Geothermal District Heating System. Applied Thermal Engineering, Vol. 31, Issues 14-15, October 2011, pp. 2726-2734

[2]   Betul Bektas Ekici, Ayca Aytac Gulten, U. Teoman Aksoy: A Study on the Optimum Insulation Thicknesses of Various Types of External Walls with Respect to Different Materials, Fuels and Climate Zones in Turkey. Applied Energy, Vol. 92, April 2012, pp. 211-217

[3]   Ray Galvin: Thermal Upgrades of Existing Homes in Germany: The Building Code, Subsidies, and Economic Efficiency. Energy and Buildings, Vol. 42, Issue 6, June 2010, pp. 834-844

[4]   Omer Kaynakli: A Review of the Economical and Optimum Thermal Insulation Thickness for Building Applications. Renewable and Sustainable Energy Reviews, Vol. 16, Issue 1, January 2012, pp. 415-425

[5] R. Pacheco, J. Ordonez, G. Martinez: Energy Efficient Design of Building: A Review Renewable and Sustainable Energy. Reviews, Vol. 16, Issue 6, August 2012, pp. 3559-3573

[6] Soteris A. Kalogirou, George Florides, Savvas Tassou: Energy Analysis of Buildings Employing Thermal Mass in Cyprus. Renewable Energy, Vol. 27, Issue 3, November 2002, pp. 353-36

[7] Meral Ozel: Cost Analysis for Optimum Thicknesses and Environmental Impacts of Different Insulation Materials. Energy and Buildings,Vol. 49, June 2012, pp. 552-559

[8] Surapong Chiraratananon, Vu Duc Hien: Thermal Performance and Cost Effectiveness of Massive Walls under Thai Climate, Energy and Buildings, Vol. 43, Issue 7, July 2011, pp. 1655-166

[9] Meral Ozel: Effect of Wall Orientation on the Optimum Insulation Thickness by Using a Dynamic Method. Applied Energy, Vol. 88, Issue 7, July 2011, pp. 2429-243

[10] Bolatturk A: Determination of Optimum Insulation Thickness for Building Walls with Respect to Various Fuels and Climate Zones in Turkey (2006) Applied Thermal Engineering, 26(11-12), pp. 1301-1309

[11] K. Çomaklı, B. Yüksel: Optimum Insulation Thickness of External Walls for Energy Saving, Applied Thermal Engineering, 23 (2003) pp. 473-479

[12] J. Yu, C. Yang, L. Tian, D. Liao: A Study on Optimum Insulation Thicknesses of External Walls in Hot Summer and Cold Winter Zone of China. Applied Energy, 86 Issue 11, Nov (2009) pp. 2520-2529

[13] A. Uçar, F. Balo: "Determination of the Energy Savings and the Optimum Insulation Thickness in the Four Different Insulated Exterior Walls". Renew Energy, 35 (2010), pp. 88-94

[14] Imrich Bartal, Hc László Bánhidi, László Garbai: "Analysis of the Static Thermal Comfort Equation" Energy and Buildings Vol. 49 (2012) pp. 188-191

[15] Garbai L., Jasper A.: "A matematikai rendszerelmélet feldolgozása és alkalmazása épületgépészeti optimalizációs feladatok megoldására"; Magyar Épületgépészet LX. évfolyam, 2011/3. szám, pp. 3-6

[16] Nyers J., Nyers A.: "COP of Heating-Cooling Systemwith Heat Pump" IEEE International Symposium CFP 1188N-PRT EXPRES 2011. Proceedings, pp. 17-21, Subotica, Serbia. 11-12 03. 2011

[17] László Garbai, Szabolcs Méhes: "The Amount of Extractable Heat with Single U-Tube in the Function of Time" Periodica Polytechnica, 52/2 (2008) pp. 49-56, DOI: 10.3311/pp.me.2008-2.02

A Novel Approach for Controlling the Electric Drives Used in Electric Trains

Mohamadali A. Vali, Jafar Monfared, Mehdi Amiri Dehcheshmeh and Hadi Givi

Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
mohamadalivali@aut.ac.ir, monfared@aut.ac.ir, mehdi750@aut.ac.ir, hadi.givi@aut.ac.ir

Abstract: Nowadays, the DC motors have been replaced by induction motors in most of the industrial applications. Substantially, Induction motors are single speed, but they can produce high power density in a wide range of speed by applying new drive controllers based on constant V/f, DTC and vector control theory. Electric motors control is mainly accomplished by power electronic drives based on control schemes like constant v/f. However, the challenge of torque generation with fast time response and high accuracy was met by using the controllers based on vector control and DTC. In applications like electric train, generation of torque with rapid time response is so vital. As an example, in Anti skid-Anti slid system, the generated torque should vary between zero and the desirable value quickly to prevent the electric train from slip and the wheels from skidding on the rail. However, the accuracy of the torque magnitude due to its high value is not an important issue and most of the time, errors up to 3 percents in the generated torque is acceptable.

In this paper, the control systems used in constant v/f drives are modified in order to generate torque with fast time response and acceptable accuracy. Control systems stability is an essential concern in electric drives and is arisen from the complexities in the controller's process and the time delays. In the proposed system, by simplification of the drive controllers, the instability problem of the control units is eliminated which in turn reduces the calculation time delays and results in torque generation with appropriate time response. Furthermore, the overall cost of the drive system is reduced considerably.

Keywords: electric drive control; electric train; induction motor; Anti skid-Anti slid system

1 Introduction

Nowadays, railway transportation systems have received a great deal of attention. Ease of displacement, capability of extensive transit of goods in a short time and environmentally friendly feature are all among the advantages that attracts the

attention of the governments and people to the electric trains. Nevertheless, the drive systems face to some limitations and challenges when they are applied to electric trains. For example, the motor torque must be generated according to the status of the wheel with respect to the rail so as to prepare the passengers comfort and to prevent the wheels from sliding on the rails in different situations like braking, primary accelerating and moving through slopes. In comparison to induction motors, DC motors were used in applications with a wide range of speed due to their simpler controllability. However, DC motors are bulk and massive and require various maintenance and services. On the other hand, their commutator and brush system is so sensitive and vulnerable. Furthermore, their manufacturing costs are considerable. Advances in power electronic and drive technologies led to the speed control of the induction motors in a wide range with high values of efficiency and power factor [11]. Although the first vector control based drive was installed in the Japan railway system in 1995 and became a conventional scheme, it faced to the problem of generating constant torque with high value while this problem is not concerned in the *v/f* drives [1]. As the effect of the stator voltage drop is ignored in *v/f* drives, complicated controllers are required to minimize the steady state speed error especially in low speeds and high values of load [2]. Using these controllers leads to the instability of the control system due to the time delays in the voltage drop compensation system. Therefore, PI controllers are required which in turn result in the severe dependency of the system to the machine parameters [3]. One of the problems in designing control systems for motors is the nonlinear relationship between speed and torque which is arisen from the dependency of these two quantities to the motor status including frequency and slip. In [4], using a new approximation of these two quantities, A.M-Garcia tries to improve the motor control but the requirement to calculation of the air gap power in his proposed linear expression increases the controller's sensitivity which in turn leads to the error increase.

In this paper, a constant *v/f* based controller is designed using the linear relation between the torque and speed. This controller receives the reference values of the speed and torque and locates the drive at the desirable operating point with the shortest time response. Due to the elimination of differentiation and integration systems from the voltage compensator, the time delay problem in compensating voltage generation is solved which in turn leads to the drive overall stability. On the basis of the mentioned conditions, the *v/f* method for torque generation and speed control in an electric train is analyzed in this paper. Finally, the validity of the proposed method is verified by simulation of the drive system implemented on a 200 kVA induction motor [6, 7].

2   Speed Control by Changing the Voltage and Frequency

According to the following expression, the synchronous speed is directly related to the power supply frequency:

$$\omega_{ms} = \frac{4\pi}{p} \tag{1}$$

Thus, changing the supply frequency leads to the change in the motor synchronous speed. On the other hand, the induced voltage is proportional to the product of the frequency and the air gap flux. Therefore, if the frequency is reduced without any change in the supply voltage, the air gap flux will increase. Induction motors are designed to operate in the vicinity of the knee point of the magnetic saturation curve. Thus, the flux increase may result in the motor saturation. On the other hand, reduction of the motor flux is not desirable as it reduces the motor capacity. Therefore, the frequency and the supply voltage are changed simultaneously to prepare the motor with a constant flux. Increasing the frequency to the levels more than the rated value is accomplished in a constant voltage. This is due to the stator insulation limitations. Suppose that the variable "a" is defined as:

$$a = \frac{f}{f_{rated}} \tag{2}$$

Where f and f_{rated} represent the operating and the rated frequencies respectively.

Obviously, for the operation in a constant flux, the magnetizing current "I_m" should be kept in a certain constant value. Thus:

$$I_m = \frac{E_{rated}}{X_m} = \frac{E_{rated}}{f_{rated}} \frac{1}{2\pi L_m} \tag{3}$$

It is evident that for each new operating point, the new values of "E" and "f" should satisfy the above expression to ensure a new constant "I_m". Therefore, the machine flux remains constant.

In this paper, one approximation is utilized that is $V = E$. This approximation is achieved by ignoring the voltage drop in the machine stator and obviously, it leads to desirable results in high speeds (high operating frequencies). In low speeds, this voltage drop can be compensated by adding this voltage to the machine input and considering the machine saturation effect. For the machine performance under and over the rated speed, the following expressions are presented where "ω_{ms}" represents the rated synchronous speed:

$$a = \frac{f}{f_{rated}} \tag{4.a}$$

$$s = \frac{a\omega_{ms} - \omega_m}{a\omega_{ms}} \tag{4.b}$$

$$T = \frac{3}{\omega_{ms}} \left[\frac{V_{rated}^2 \left(\frac{R_r'}{as} \right)}{\left(\frac{R_r'}{as} \right)^2 + \left(X_s + X_r' \right)^2} \right] \quad \text{if } a < 1 \tag{5.a}$$

$$T = \frac{3}{\omega_{ms}} \left[\frac{V_{rated}^2 \left(\frac{R_r'}{as} \right)}{\left(\frac{R_r'}{s} + R_s \right)^2 + a^2 \left(X_s + X_r' \right)^2} \right] \quad \text{if } a > 1 \tag{5.b}$$

If the flux of the operating point is kept constant in a certain value by using a special strategy, a torque-speed characteristic like Figure 1 is achieved:
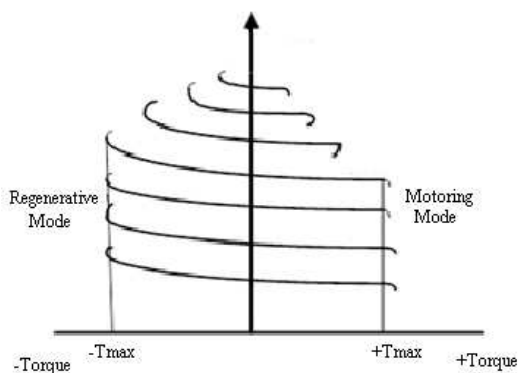
Figure 1
Torque-speed curve for the rated flux

But using the approximation $V = E$, results in the characteristic presented in Figure 2.
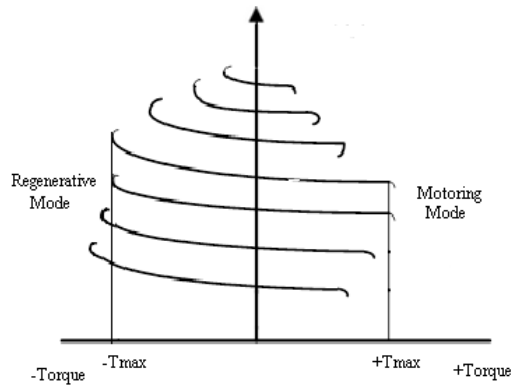
Figure 2
Torque-speed curve for "v/f = constant"

As seen in the above curves, in low operating frequencies (a<<1), the motor characteristic has a smaller break down torque. This characteristic can be improved by compensation of the stator voltage drop in the machine input and consideration of the machine saturation.

3 The State of the Voltage Source Inverter (VSI) Operation

The configuration of a 6-pulse inverter is illustrated in Figure 3:
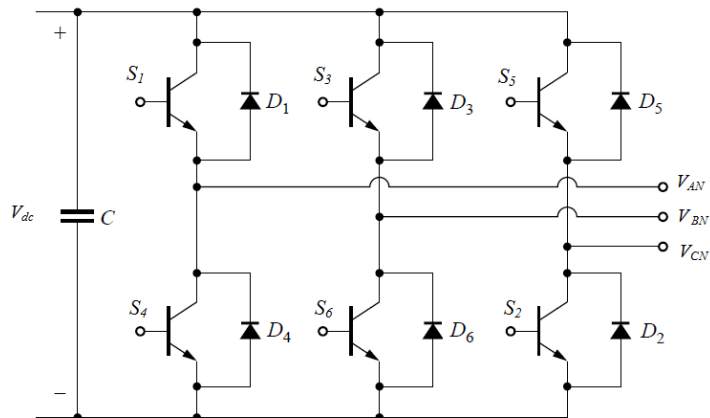
Figure 3
Configuration of a 6-pulse inverter

Suppose that the switching control signals applied to the switches are in the form presented in table I:

Table 1

Switching control signals of the 6-pulse inverter

Time (ωt)	Switches(ON)
$0 < \omega t < \pi/3$	1,5,6
$\pi/3 < \omega t < 2\pi/3$	1,2,6
$2\pi/3 < \omega t < 3\pi/3$	1,2,3
$3\pi/3 < \omega t < 4\pi/3$	2,3,4
$4\pi/3 < \omega t < 5\pi/3$	3,4,5
$5\pi/3 < \omega t < 6\pi/3$	6,4,5

Using the switching signals presented in table I, the output voltage V_{AN} will be in the form illustrated in Figure 4.
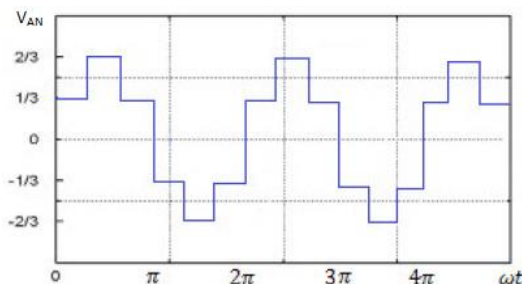
Figure 4

The inverter output voltage

It is obvious that the voltages V_{BN} and V_{CN} produced at the inverter output are the same as V_{AN} but with the phase shifts equal to 120 and -120 degrees with respect to V_{AN} respectively. Considering the Fourier series of the voltages V_{AB} and V_{AN}, the RMS value of the main harmonic and the RMS value of the phase voltage are presented below:

$$V_1 = \frac{\sqrt{2}}{\pi} V_d \tag{6.a}$$

$$V = \frac{\sqrt{2}}{3} V_d \tag{6.b}$$

Therefore, the value of the sinusoidal voltage (main harmonic) applied to the motor is controlled by regulating the voltage V_d. The frequency of the inverter output voltage can be controlled by changing the switching intervals.

4   The State of Controlling the Drive Speed at the Speeds Lower and Higher than the Rated Speed

In drives utilized for electric train, the reference torque and speed are given to the drive and the goal is to control the drive in a way that it follows the reference torque and speed with an acceptable transient duration. To attain this purpose, two strategies can be chosen.

In the first strategy, the values of the reference torque and speed are given to the drive. The motor speed in sensed in each instant using a sensor and it is used for calculating the values of the slip and operating frequency in order to lead the motor to the desirable values of torque and speed. Obviously, this system is so accurate but deals with complexities in determination of the slip and operating frequency in each instant. On the other hand, since these values are attained using two nonlinear control systems, this process is time-consuming and may result in the instability of the system.

In the second strategy proposed in this paper, by considering the torque-speed characteristic obtained in the previous parts and using proper approximations, it is shown that using an estimator system for the operating frequency, the machine follows the values of the reference torque and speed with an acceptable accuracy in the steady state.

5   The State of Estimating the Machine Operating Frequency

By attaining the torque-speed characteristic of the machine in the rated frequency, it is deduced that in the frequencies lower than the rated value, the torque-speed characteristic has the same slope but it is just shifted with respect to the characteristic in rated frequency and is appeared with a new synchronous speed equal to $a\omega_{ms}$ (ω_{ms} is the rated synchronous speed). In the speeds higher than the rated value, the characteristics slope is the same as the characteristic in the rated frequency but with a new break down (pull out) torque. This new break down torque can be determined by considering a constant power for the machine. Different torque-speed characteristics can be obtained by changing the value of "a". So a wide range of speeds and torques can be gained using these characteristics. By specifying the region of the reference torque and speed, namely by determining the operating frequency corresponded to the values of the reference torque and speed, the machine could be controlled in a way to follow the reference torque and speed in the steady state operation. In Figure 5, by regulating the reference torque and speed in a way to be located in the boundary area that has been shown in below figure, it is possible to control the motor so as to reach the desirable torque and speed with an excellent approximation in the steady state.
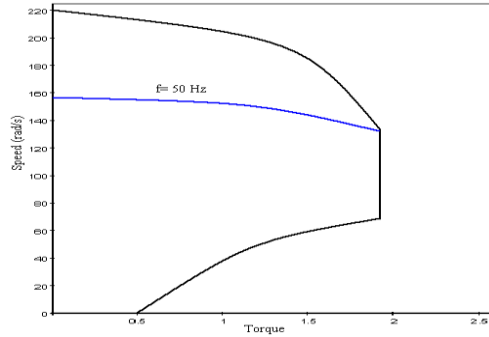
Figure 5

The operating region for following the reference torque and speed

It should be mentioned that by compensating the stator voltage drop in low operating frequencies, the drive operating region can be extended to the torques with values near the break down torque. Using the linear approximation, the torque – speed expression is attained for the speeds lower than the rated value in the following form:

$$\left|\omega - \omega_{ms}\right| = -\frac{\omega_{ms} - \omega_0}{T_{\max}} \tag{7}$$

6　Performance of the Drive under the Rated Speed

The slope of the other curves under the rated speed is assumed to be equal to (7). Therefore, for the drive performance under the rated speed, the following expression is given:

$$\omega - a\omega_{ms} = mT \tag{8}$$

This expression results in:

$$a = \frac{\omega - mT}{\omega_{ms}} \tag{9}$$

In the above expression: $T = T_{ref}$ and $\omega = \omega_{ref}$. The value of "m" is attained using the torque-speed characteristic of the machine. So, by utilizing the values of T_{ref} and ω_{ref}, "a" can be calculated using equation (2). As the drive is performing under the rated speed, the reference torque and speed are followed by using a voltage equal to "a" times of the rated voltage ($V=a*V_{rated}$).

7   Performance of the Drive over the Rated Speed

At the speeds higher than the rated value, T_{max} is not constant and based on the operating speed, T_{max} decreases. Thus, the following expression can be considered as the linear approximation for this operating region:

$$|\omega - \omega_{ms}| = \frac{\omega_{ms} - \omega_0}{\dfrac{\omega_{ms}}{\omega}T_{max}}T \tag{10}$$

As a result:

$$\omega - a\omega_{ms} = \frac{\omega}{\omega_{ms}}mT \tag{11}$$

The same as the operation under the rated speed, $T = T_{ref}$ and $\omega = \omega_{ref}$. "m" and "ω_{ms}" are gained from the normal operating characteristic of the machine. Therefore, "a" is calculated as:

$$a = \frac{\omega - \dfrac{\omega}{\omega_{ms}}mT}{\omega_{ms}} \tag{12}$$

Using this value of "a" and this time with the rated voltage, it is possible to follow the reference torque and speed. Drive system block diagram is shown in Figure 5:
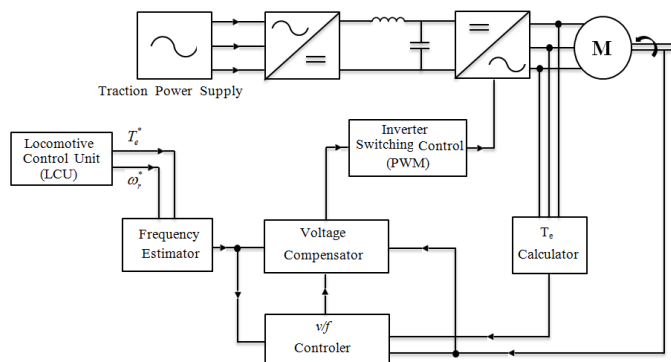
Figure 6
Drive system block diagram

8   Drive System Simulation Results

The discussed drive is tested on a 200 kVA motor in order to evaluate the correct operation and calculations of the proposed method. It will be shown that the proposed drive can follow the reference torque and speed so accurately. The characteristics of the simulated motor are presented in table 2.

Table 2
Characteristics of the simulated motor

Quantity	Value
Rated Power	160 kW
Rated Voltage	400 V
Rated Current	300 A
Frequency	50 Hz
Rated Torque	1000 N.m
Motor and its load inertia	5 kg.m^2

There is a comparison between two states, when the system operation is considered with voltage compensator and another state that the system operates without voltage compensator block. The nominal torque with consideration of the drive characteristic is 1000 N.m, and maximum torque is considered approximately three times more than nominal torque. Intrinsically, drive confronts with unsteady condition when operates at low speeds with high load torque, and by addition a new block named compensator block, the drive overcomes this situation and can produce high electromagnetic torque in low speed. The results of comparison between the states will be shown in the following:

A. T=3000 N.m; ω =25 $rad/\sec$
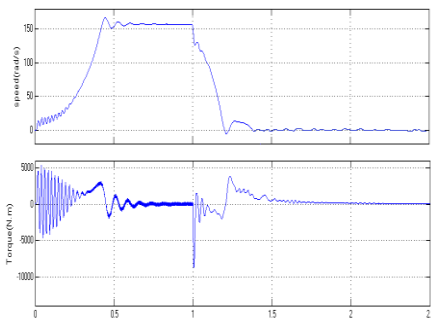
Figure (7.a)	Figure(7.b)
Drive speed and torque without voltage compensator	Drive speed and torque with voltage compensator

B. T=2000 N.m; ω =15 $rad/\sec$

Figure(8.a)
Drive speed and torque without voltage compensator

Figure(8.b)
Drive speed and torque with voltage compensator

C. T=1500 N.m; ω =12 $rad/\sec$
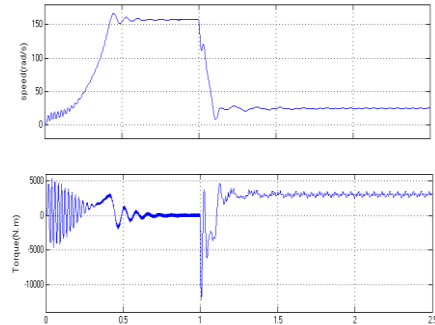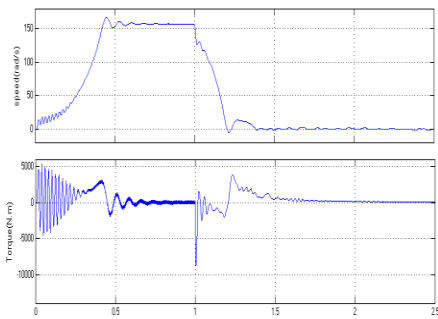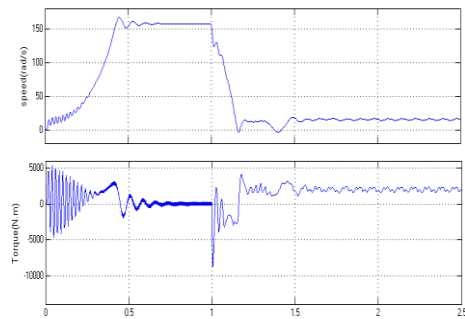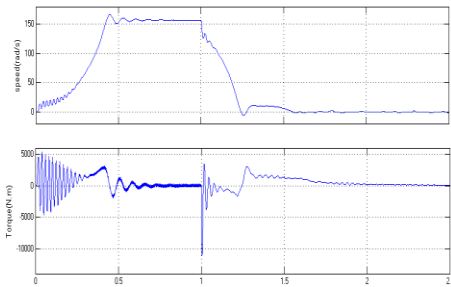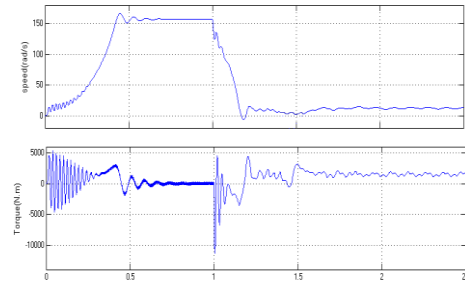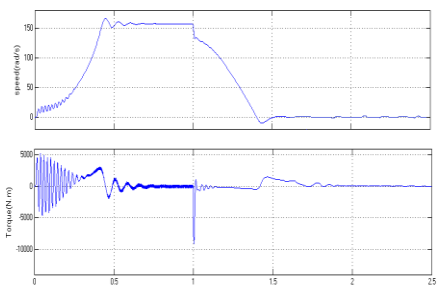

Figure(9.a)
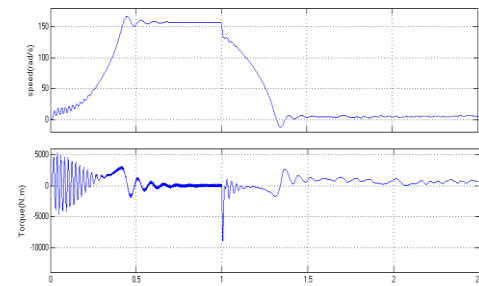Drive speed and torque without voltage compensator

Figure(9.b)
Drive speed and torque with voltage compensator

D. T=1000 N.m; ω =5 $rad/\sec$

Figure(10.a)
Drive speed and torque without voltage compensator

Figure(10.b)
Drive speed and torque with voltage compensator

As seen, in all mentioned conditions, when compensator block is not used, the motor is not able to follow the operating point (reference speed and torque). Then the drive system performance under practical conditions has been studied.

The simulation results are as following:

First, it is supposed that the motor is started and at $t = 1$ sec, a torque is applied to the motor and this torque reaches to its rated value after 0.2 second and driver is going to produce this torque in the motor output and also put the train in a desirable speed (reference speed).

The simulation results for different values of the applied torque and the reference speed are presented below.
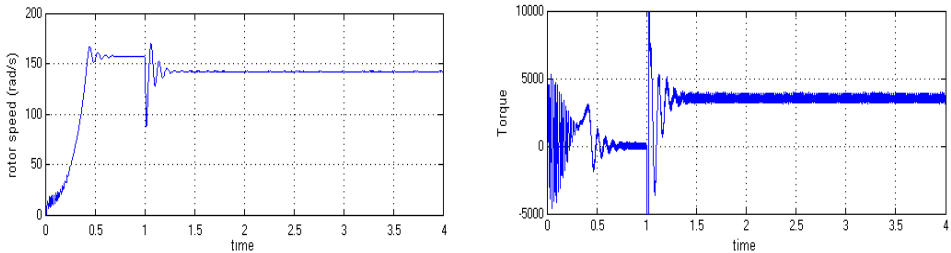
A. T=3500 N.m; $\omega_{ref} = 140$ $rad / \sec$

Figure (11.a)
Drive system performance

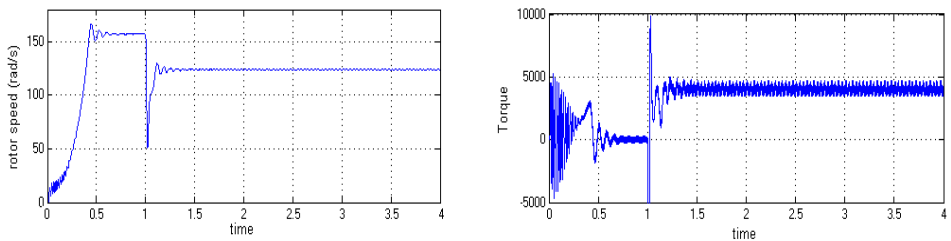B. T=4000 N.m; $\omega_{ref} = 120$ $rad / \sec$

Figure (11.b)
Drive system performance

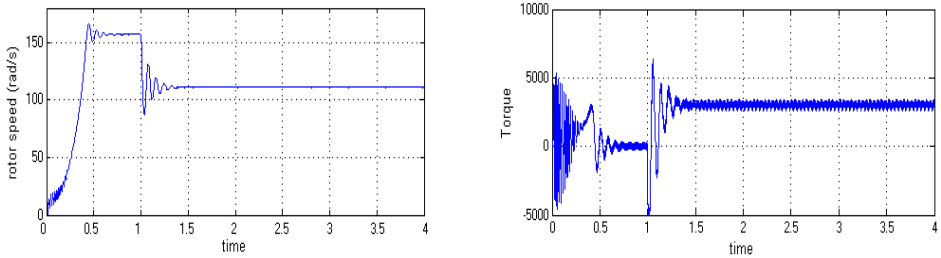C. T=3000 N.m; ω_{ref} =110 rad / sec

Figure (11.c)
Drive system performance

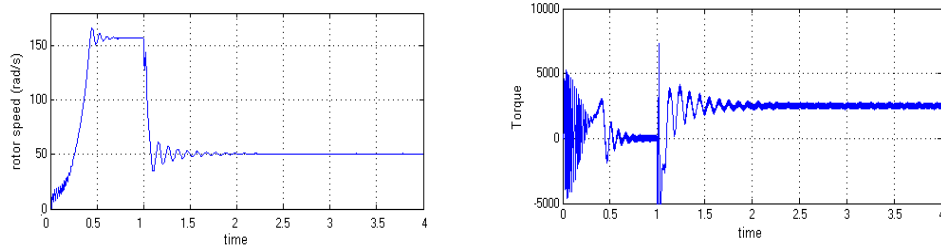D. T=2500 N.m; ω_{ref} =50 rad / sec

Figure (11.d)
Drive system performance

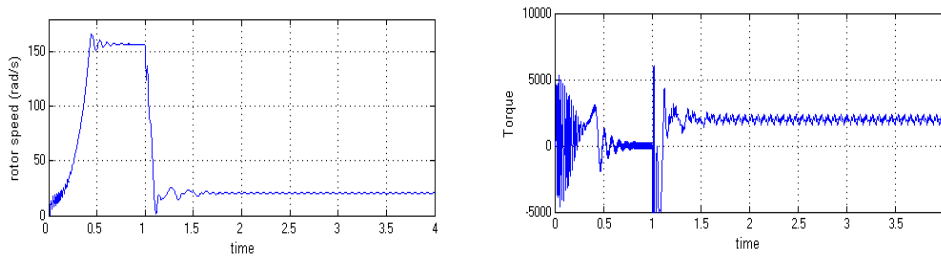E. T=2000 N.m; ω_{ref} =20 rad / sec

Figure (11.e)
Drive system performance

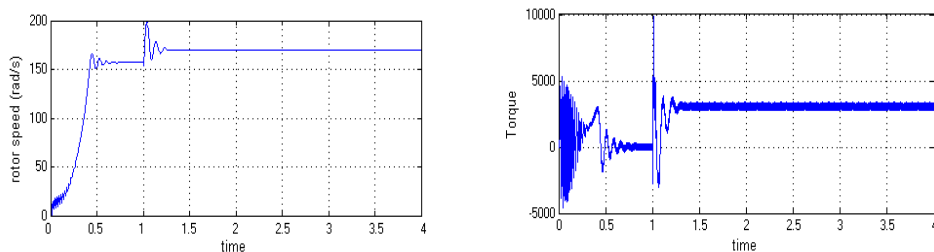F. T=3000 N.m; ω_{ref} =170 $rad/\sec$

Figure (11.f)
Drive system performance
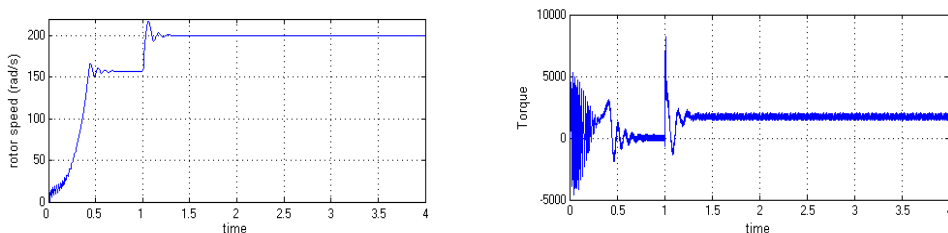
G. T=1700 N.m; ω_{ref} =200 $rad/\sec$

Figure (11.g)
Drive system performance

H. Test with variations of the reference speed, analysis of accelerating, braking and the steady state in a route:

In this test, it is assumed that the motor is under a constant load with the value of 2000 N.m; Our purpose is to analyze the braking and accelerating states under this load torque. First it is supposed that the motor is rotating at a speed equal to 180 rad/sec which is higher than the rated value. Then the motor receives a command for decreasing its speed to 130 rad/sec and after becoming stable at this speed, again the motor receives a command for decreasing its speed to 90 rad/sec. So the motor adjusts itself to this new speed. After positioning at the steady state, a command for increasing the speed to 110 rad/sec is received and after passing the transient state, the motor adjusts itself to this command. The rotor speed and the torque diagrams for this test are represented below.
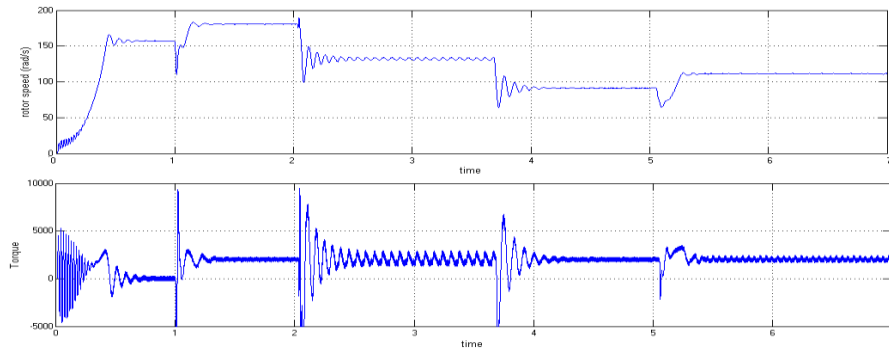
Figure (11.h)
Drive system performance

The tests (A), (B) and (C) are performed under the speeds lower than the rated speed, while the tests (D) and (E) are done at low speeds and the tests (F) and (G) are performed at the speeds which are higher than the rated value.

The test (H) verifies the correct operation of the motor at the compound cycles and confirms the motor ability for braking, accelerating and reaching to the steady state with minimum error.

Conclusion

In the conventional v/f systems, the control system considers the slip and in each instant, solves the motor equations to follow the reference values of the torque and speed but in the method proposed in this paper, the normal torque – speed characteristic of the motor is utilized and by designing the Frequency Estimator unit, this purpose is attained simply and by a high stability. As confirmed in the tests, the proposed drive system can easily control the speed of the electric train in a wide range. As seen, this system is able to prepare the controller or the train driver with their desirable torque and speed simply. On the other hand, using the simplifications discussed above, this system can be easily implemented and manufactured, so the drive system manufacturing cost is reduced considerably. Furthermore, as the control system for determination of slip and operating frequency is not utilized in the form of an instantaneous calculator, the system stability will increase and the control equipments for instantaneous calculation of the slip and operating frequency and also expensive sensors are not required anymore.

References

[1]   M. Yano, M. Iwahori,''Transition from Slip Frequency Control to Vector Control for Induction Motor Drives of Traction Applications in Japan,''IEEE, 2003

[2]   K. Koga, Y. Fukuoka,''Constitution of v/f Control for Reducting the Steady State Speed Error to Zero in Induction Motor Drive System'', IEEE, 1990

[3]   A. Munoz-Garcia, T. A. Lipo, D. W. Novotny. "A New Induction Motor V/f Control Method Capable of High-Performance Regulation at Low Speeds," Industry Applications, IEEE Transactions on, Vol. 34(4), pp. 813-821, July 1998

[4]   A. M. Garcia, T. A. Lipo, D. W. Novotny,''A New Induction Motor *V/f* Control Method Capableof High-Performance Regulation at Low Speeds'' IEEE Trans Ind. Appl., Vol. 34, No. 4, July/August 1998

[5]   P. L. Alger, Induction Machines, 2nd ed. New York: Gordon and Breach, 1970

[6]   Jee-Hoon Jung, Gang-Youl Jeong, and Bong-Hwan Kwon, "Stability Improvement of V/f-controlled Induction Motor Drive Systems by a Dynamic Current Compensator," Industrial Electronics, IEEE Transactions on, vol. 51(4), pp. 930-933, Aug. 2004

[7]   K. Suzuki, S. Saito, T. Kudor, A. Tanaka, and Y. Andoh, "Stability Improvement of V/F Controlled Large Capacity Voltage-Source Inverter Fed Induction Motor," Industry Applications Conference, 41st IAS Annual Meeting. Conference Record of IEEE, Vol. 1, pp. 90-95, Oct. 2006

[8]   R. A. Hamilton and G. R. Lezan, "Thyristor Adjustable Frequency Power Supplies for Hot Strip Mill Run-Out Tables," IEEE Trans. Ind. Gen. Applicat., Vol. IGA-3, pp. 168-175, Mar./Apr. 1967

[9]   F. Blaschke, 'The Principle of Field Orientation as Applied to the New Transvector Closed-Loop Control System for Rotating-Field Machines," Siemens Rev., Vol. 34, pp. 217-220, 1972

[10]   Zhiwen Ma, Trillion Zheng, and Fei Lin, "Stability Improvement of V/Hz Controlled PWM Inverter-fed Induction Motors Drives," Industrial Electronics and Applications, 1st IEEE Conference on, pp. 1-4, Man 2006

[11]   Cornell E. P., Lipo T. A., "Modelling and Design of Controlled Current Induction Motor Drive Systems", IEEE Trans Ind. Appl. 1977

Indentation Size Effect in Autopolymerized and Microwave Post Treated Poly(methyl methacrylate) Denture Reline Resins

Sebastian Balos[1], Leposava Sidjanin[1], Branka Pilic[2]

[1] Department of Production Engineering, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
sebab@uns.ac.rs; lepas@uns.ac.rs

[2] Department of Material Engineering, Faculty of Technology, University of Novi Sad, Bulevar Cara Lazara 1, Novi Sad, Serbia
brapi@uns.ac.rs

Abstract: Various hardness test methods have been used to determine the polymer materials resistance to local plastic deformation in scientific community. The most commonly used method is Vickers microindentation. However, it is of crucial importance to fully understand the influence of the indentation load, since its value influences the obtained result. In this paper, two commercial PMMA dental reline resins in untreated and microwave treated condition have been tested. Vickers microindentation with different loads has been used to assess the materials microhardness: 30, 50, 70, 200, 300 and 500 gf. One way ANOVA statistical analysis followed by Tukey's test was used to determine the statistical differences between various groups. Finally, three models that quantitatively describe load-dependence of the measured Vickers hardness values were used: Meyer's law, PSR and modified PSR model. It was found that the optimal load for determining PMMA dental reline resins in both untreated and microwave treated conditions is 300 gf. This value may be regarded as loading independent hardness, or H_{LIH}. At lower loads, a more or less pronounced indentation size effect was noticed, while at higher loads, forked crack development leads to an unreliable indentation diagonal measurement and therefore an unreliable microhardness result. The most adequate load-dependence model was found to be modified PSR, that takes into consideration the surface stresses induced by specimen preparation by grinding.

Keywords: Vickers microhardness; indentation size effect; poly(methyl methacrylate); microwave irradiation

1   Introduction

Microhardness has been widely used as an indicator of the materials mechanical properties, crucial for determining the materials performance. The most common hardness tests comprise of indentation techniques, which provide both materials resistance to local plastic deformation and a correlation to flexural and compressive strengths [1]. Although Vickers hardness is considered as the most widely used method for testing of denture materials of acrylic type, other methods have been used in scientific community. The Knoop hardness has been successfully used as an alternative to Vickers test, with an advantage of a less likely cracking in brittle materials. However, this method penetrates less into the specimen surface and thus, it is more sensitive to surface texture. Furthermore, the successful and accurate Knoop test indentations require larger specimens to be used [1]. Other researchers even used methods that are predominantly oriented towards hardness testing of metallic materials. Azzarri et.al. [2] used Rockwell P method (1/4 inch steel ball indenter and a load of 150 kgf), while Idol and Lehman [3] reported the application of Brinell hardness test on acrylic and other polymer materials.

For testing of acrylic resins, the lack of an agreement on the test method and testing parameters such as load prevents straightforward comparison of the results obtained in different studies. Another problem in obtaining comparable results is the Indentation size effect (ISE) which represents a phenomenon that may be briefly described as an indentation-depth-dependent hardness [4]. By the application of different loads, different hardness values are obtained, usually a lower indentation depth results in an increased hardness [5]. The ISE was observed in ceramic, metallic and polymer materials. While well understood and investigated for metals, which is not the case with polymers [6]. In metals, the ISE effect may be related to plastic deformation and dislocation movement and the notion of geometrically necessary dislocations increasing flow stress and hardness values [7, 8]. At indentation depths smaller than 0.2 μm, the roughness of the surface and other surface effects may influence the deformation mechanisms and hardness obtained [9, 10]. However, the ISE effect observed in polymers cannot be explained by dislocation theory. In turn, different theories developed to explain the ISE effect in polymers. Some researchers attributed the ISE to structural differences in depth [11, 12], while others linked ISE to the elastic strain and elastic deformation energy that is proportionally higher in polymers compared to plastic strain and deformation energy in metals [13, 14], related to polymer nematic-like molecular structure. In this model, polymer chains possess finite stiffness, which can be considered as a system consisting of interacting rodlike segments [15].

The aim of this study was to determine the ISE in Poly(methyl methacrylate) (PMMA), which are the most widely used types of materials for providing better retention of removable protheses in cases of alveolar resorption, as well as for

denture reparation in case of crack or fracture. Denture reline resins have a lower mechanical properties compared to denture resins, since the components (powder and liquid) are manually mixed together for a limited period of time, when polymerization process is started. Such material possesses a higher amount of monomer, which acts as empty space, or a microvoid, which, under load may propagate and cause fracture. Furthermore, the ISE of microwave treated dental reline resins, which is a common method for improving mechanical properties by decreasing the amount of monomer, were tested [16, 17].

2   Materials and Methods

Two autopolymerizing denture materials; Simgal-R (Galenika, Serbia) and Akrilat-R (ADA Dental Products/Dentaurum, Serbia/Germany) were tested in this study. Materials consisted of powder and liquid, which, when mixed, the polymerization process is initiated. Simgal-R and Akrilat-R powder consist of PMMA, benzoyl peroxide and inorganic pigments, while the liquid component contains the methylmetacrylate monomer and tertiary amine. Samples were prepared as advised by the manufacturer, with powder to liquid ratio of 2:1 in weight and subsequently cast in elastomer molds (Wirosil, Bego, Germany). After polymerization, a set of SiC papers (150, 400 and 1200 grit) were used to get the desired shape, dimensions and surface quality of the samples. Samples were of the cylindrical shape, 2 mm thick and 25 mm diameter. Two samples of each material were tested: control sample and a sample microwave after-treated with a power of 550 W during 4 min. Microwave aftertreatment was conducted in a microwave device with a turntable and output power of 800 W (Elin MW8020MG, Austria).

Microhardness measurements were made with two Vickers microhardness testing machines, used in two loading ranges: low (Zwick Z323, Germany) and high (Reichert Me-F, Austria). Low loads applied were 30, 50 and 70 gf, while high loads were 200, 300 and 500 gf. Vickers hardness number (VHN) was determined as an average of five indentations. VHN microhardness values were obtained by using a common formula:

$$HV = 1.8544 F / d^2 \tag{1}$$

Where 1.8544 is a constant geometrical factor for the Vickers pyramid, F is load [kgf] and d is indentation average diagonal [mm]. The data was compared by one way analysis of variance (ANOVA) followed by Tukey's test with the significance value set at 0.05.

3   Results

The results obtained by Vickers microhardness testing, as well as standard deviations and ANOVA statistical analysis are shown in Tables 1 and 2. Only the results obtained with 30, 50, 70, 200 and 300 gf are shown, due to forked crack development at the indentation tip that occurs when a load of 500 gf or more is applied for all tested materials, Fig. 1.

Figure 1

Vickers microhardness indentations obtained by the application of different loadings, from left to right: 500 gf, 300 gf, 200 gf and 70 gf. Indentation with 500 g load results in a cracking at indentation edge

In Table 1, mean microhardness results and standard deviations are shown. It can be seen that, as indentation load is increased, Vickers microhardness values decrease. This trend is obvious for both tested materials, Simgal-R and Akrilat-R. Furthermore, the same trend can be detected for microwave treated samples, but their microhardness values were higher, when the load independent hardness (H_{LIH}) is reached at applying 300 g load. By applying ANOVA one-way analysis, it can be seen that for untreated samples made of Simgal-R, all results obtained with 30, 50, 70 and 200 gf load are significantly different from the result obtained with 300 gf load. However, treated Simgal-R microhardness obtained with 70 and 200 g are not significantly different from the result obtained with 300 gf. Vickers microhardness testing of untreated and treated Akrilat-R revealed that the application of 70 and 200 gf load does not result in a significantly different results compared to the results obtained with 300 gf load. On the other hand, for values obtained with 500 gf load, all values have shown a statistically significant difference compared to the value obtained with 300 gf.

Table 1

Mean microhardness values and standard deviations given in parantheses

Load [gf]	Simgal-R untreated	Simgal-R treated	Akrilat-R untreated	Akrilat-R treated
30	22.25 (0.60)	22.18 (0.38)	22.25 (0.36)	26.81 (0.46)
50	21.95 (0.34)	20.92 (0.33)	21.22 (0.60)	24.50 (0.73)
70	19.66 (0.47)	19.96 (0.36)	20.28 (0.63)	23.22 (0.84)
200	18.16 (0.65)	19.88 (0.18)	18.92 (0.20)	21.51 (0.92)
300	17.27*(0.39)	19.70*(0.22)	18.21*(0.29)	20.43*(0.56)
500	20.65 (0.65)	22.38 (0.61)	22.54 (0.58)	23.85 (0.53)

∗ Values than may be considered as load independant hardness (H_{LIH})

Table 2

Statistical difference in relation to the result obtained with a load of 300 g of tested materials in relation to indentation load

Load [gf]	Simgal-R untreated	Simgal-R treated	Akrilat-R untreated	Akrilat-R treated
30	P<0.05	P<0.05	P<0.05	P<0.05
50	P<0.05	P<0.05	P<0.05	P<0.05
70	P<0.05	NS*	P<0.05	P<0.05
200	P<0.05	NS*	NS*	NS*
300	-	-	-	-
500	P<0.05	P<0.05	P<0.05	P<0.05

* No significance

For more convenience, indentation size effect is shown in Figs. 2 and 3 in the form of a diagram. A trend showing the drop of microhardness by applying a higher indentation load is shown. Furthermore, Simgal-R test results have shown that at relatively low loads, results of actually softer material (untreated Simgal R) may show higher microhardness compared to an actually harder material (treated Simgal-R), where $H_{(LIH\text{-}untreated\ Simgal)} < H_{(LIH\text{-}treated\ Simgal)}$.
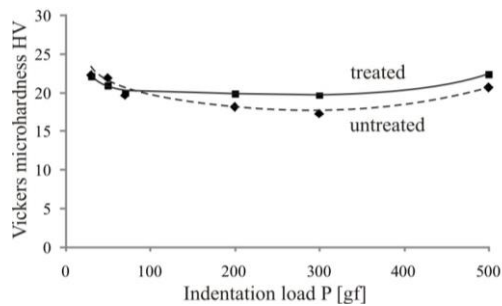

Figure 2

Vickers microhardness in relation to indentation load for Simgal-R in untreated and treated condition
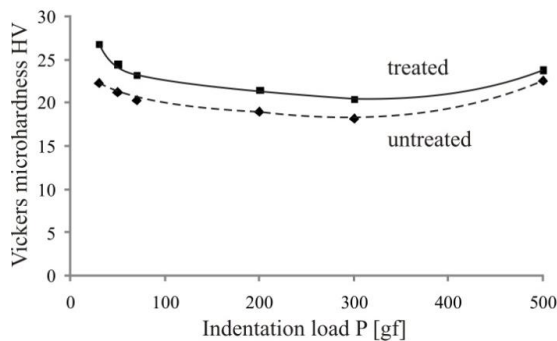

Figure 3

Vickers microhardness in relation to indentation load for Akrilat-R in untreated and treated condition

4 Discussion

Vickers hardness and microhardness consists of forcing a diamond pyramidal indenter into the prepared surface of the material being tested. After indenter removal, indentation diagonals are being measured by application of light microscope. Hardness value is calculated, basically as dividing indentation force by indentation area obtained from the indentation diagonals. In this study, four materials (PMMA denture reline resins) were tested, all of them showing an ISE, as well as the cracking phenomenon that occurs at higher indentation loads near the indentation edge.

By increasing indentation load, the apparent hardness of the material decreases, asymptotically reaching an apparently constant value. This value is the materials true hardness, or load independent hardness, H_{LIH}. Only this value may be used to assess and compare microhardnesses of two tested materials. If hardness obtained with a lower indentation load is used, relations between the obtained results may be misleading, or even worse, quite opposite to the relations between true hardness values (H_{LIH}). A typical example are the results obtained for Simgal-R in untreated and microwave treated conditions, where H_{LIH} of the treated sample is higher than that of the untreated, while at lower indentation loads of 30 and 50 gf, the apparent hardness of the untreated sample is higher compared to the treated one. This may be explained by the samples inhomogeneous structure, obtained after manual mixing and subsequent polymerization that resulted in 6.5% of unconverted monomer [16]. After microwave post – treatment, unconverted monomer in Simgal-R slightly decreased to 4.8% [16], which may have left sufficient degree of inhomogenity that influenced the microhardness results obtained with a low indentation loads. In this respect, polymer microhardness testing may be related to the hardness testing of cast iron, where Brinell method is applied. This method comprises of the indentation of a steel ball having a sufficiently large diameter of 5 or 10 mm to eliminate the influence of the low-hardness graphite present in the structure which is randomly distributed. By applying such a large indenter, a mean hardness is obtained, since a sufficient number of graphite particles is present in the indentation.

Another limiting factor is the cracking of the relatively brittle polymer such as PMMA when 500 gf or more indentation load is applied. Cracking occurs at indentation edge in form of forked cracks that limit the length of the diagonal, resulting in a considerable diagonal length difference. This effect influences the reliability of the obtained results, which, limits the indentation load to less than 500 gf.

When ISE and cracking effects are taken into consideration, Vickers microhardness optimal load for PMMA testing is 300 gf, the highest that does not lead to cracking in the specimen. Lower loads compared to 300 gf may in some experiments result in statistically insignificant difference compared to the results obtained with 300 gf, however, loads under 200 gf should be avoided.

The load-dependence of the measured Vickers hardness values can also be described quantitatively through the application of the classical Meyer's law:

$$P=Ad^n \tag{2}$$

where P is the indentation load and d is the resulting indentation size, that is, diagonal. The parameter A and n are values that can be derived directly from the curve fitting of the experimental data [18]. The Meyer's law parameters determined by the regression analyses are summarized in Table 3 and presented in Fig. 4, with 500 gf loading results omitted. The ISE is commonly related to the deviation of the n-value from two. For virtually all materials the power law exponent n is experimentally observed to be between 1 and 2, which indicated that lower indentation test loads result in higher apparent microhardness. ISE is more pronounced in specimens having the power law exponent n closer to 1 and vice versa. The results of n-exponent presented in Table 3 indicate that the most significant ISE is obtained in the case of untreated Simgal-R specimen ($n=1.796$), while the most non-significant ISE was obtained in the case of treated Simgal-R specimen ($n=1.914$). The correlation factors R^2 are consistent at 0.999 for all tested specimens.

Table 3

Regression analysis resuts of the experimental data according to Mayer's law

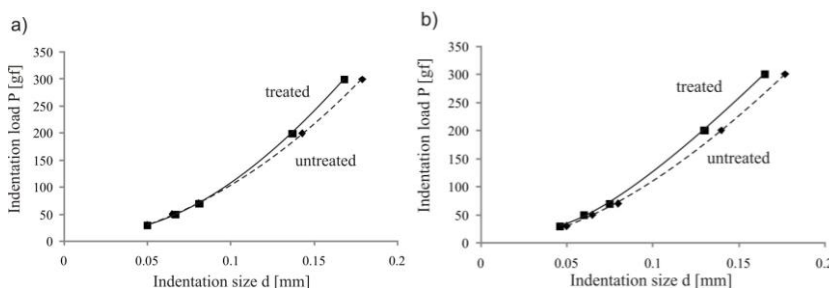|	A	$\log A$	n	Correlation factor (R^2)
Simgal-R untreated	8964.0	3.952	1.796	0.999
Simgal-R treated	6579.1	3.818	1.914	0.999
Akrilat-R untreated	7815.2	3.893	1.820	0.999
Akrilat-R treated	7080.1	3.850	1.805	0.999

Figure 4

Correlation between P and d according to Meyer's law (full line treated, dashed line untreated specimens): a) Simgal-R; b) Akrilat-R

An alternative analysis of ISE to the Meyer's law is proportional specimen resistance (PSR) model based on the following equation:

$$P=a_1d+a_2d^2 \tag{3}$$

where a_1 and a_2 are experimental constants. Eq. (2) describes the observed ISE by testing linearity between *P/d* and *d*. The results are summarized in Table 4 and Fig. 5. From Fig. 5 it is evident that in each system the data points show linearity with correlation factor (R^2) between 0.996 and 0.998.
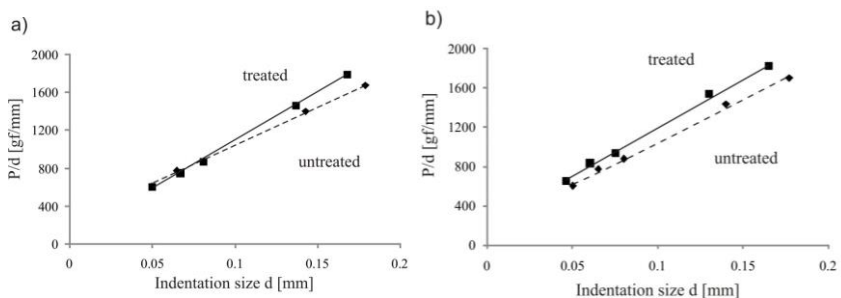

Figure 5

Correlation between P/d and d according to PSR model (full line treated, dashed line untreated specimens): a) Simgal-R; b) Akrilat-R

Table 4

Regression analysis resuts of the experimental data according to PSR model

	a_1	a_2	Correlation factor (R^2)
Simgal-R untreated	67.89	10172.31	0.998
Simgal-R treated	205.41	8264.21	0.998
Akrilat-R untreated	215.62	9868.14	0.996
Akrilat-R treated	189.60	8631.22	0.997

The modified PSR model proposed by Gong and Li [18] may be mathematically described as:

$$P = P_0 + a_1 d + a_2 d^2 \tag{4}$$

where P_0 is experimental constant, while a_1 and a_2 have the same physical meaning as in the Eq. (3). This model was proposed by Gong and Li, who found that the surface of the specimen is not in stress free state, but rather exposed to the stress induced by, in this case, grinding, necessary for conducting microhardness test. The fit values of all parameters included in Eq. (4) are given in Table 5, while their graphical representation is presented in Fig. 6. The regression analysis returns correlation coefficients between 0.999 and 1. From Table 5, it can be seen that parameters P_0 and a_1 may have positive or negative values. In accordance to the work by Gong and co-workers [18], this phenomenon may be reasonably explained by the porosity of the materials examined. Indeed, the materials used were mixed manually, by a technician [19], where some porosity is very difficult to avoid, partially due to the limited amount of time before polymerization commences.

Table 5

Regression analysis resuts of the experimental data according to modified PSR model

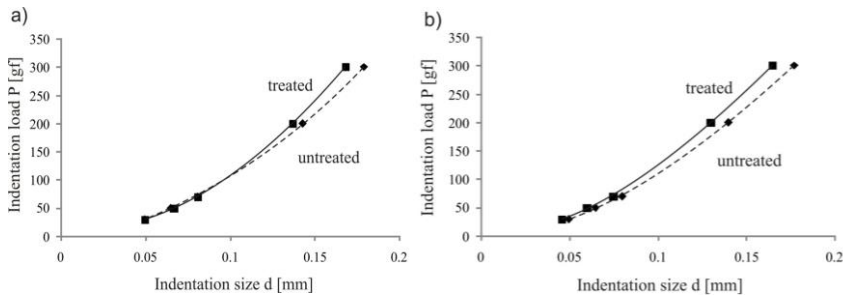|	P_0	a_1	a_2	Correlation factor (R^2)
Simgal-R untreated	8.45	-128.71	11108.11	1
Simgal-R treated	-5.54	330.40	7700.24	0.999
Akrilat-R untreated	-14.32	563.09	8178.31	0.999
Akrilat-R treated	-14.91	525.92	7107.10	0.999

Figure 6

Correlation between P and d according to modified PSR model (full line treated, dashed line untreated specimens): a) Simgal-R; b) Akrilat-R

Conclusion

In accordance to the presented results, some conclusions can be drawn:

- For testing of PMMA denture reline resins, untreated or microwave treated, optimal load is 300 gf.

- The loads lower than 300 gf result in an inconsistent statistical significance that may or may not give an adequate result. Loads lower than 200 gf are inadequate. Loads lower than 200 gf are inadequate, since the diagonal of the indentation is not sufficient to avoid materials imperfections such as unconverted monomer, which may influence the obtained result.

- At loads higher than 300 gf where forked cracks occur, the result can be regarded as unreliable and should be avoided.

- Before microhardness testing, a careful optimization of indentation load is needed to reveal the materials true hardness, or load independent hardness H_{LIH}. To determine this, conducting a pre-experiment is needed.

References

[1]   O. Yoldas, T. Akova and H. Uysal: Influence of Different Indentation Load and Dwell Time on Knoop Microhardness Tests for Composite Materials, Polymer Testing, 23 (2004) pp. 343-346

[2]   MJ. Azzarri, MS. Cortizoa and JL. Alessandrin: Effect of the Curing Conditions on the Properties of an Acrylic Denture Base Resin Microwave-Polymerised, Journal of Dentistry, 31 (2003) pp. 463-468

[3]   JD. Idol and RL. Lehman: The CRC Handbook of Mechanical Engineering. CRC Press. 2004

[4]   CS. Han and S. Nikolov: Indentation Size Effects in Polymers and Related Rotation Gradients, Journal of Materials Research, 22 (2007) pp. 1662-1672

[5]   G. Zamfirova and A. Dimitrova: Some Methodological Contributions to the Vickers Microhardness Technique, Polymer Testing, 19 (2000) pp. 533-542

[6]   VS. Tatiraju, CS. Han and S. Nikolov: Size Dependent Hardness of Polyamide/Imide, Open Mechanics Journal, 2 (2008) pp. 89-92

[7]   Q. Ma and DR. Clarke: Size Dependent Hardness of Silver Single Crystals, Journal of Materials Research, 10 (1995) pp. 853-863

[8]   WD. Nix and H. Gao: Indentation Size Effects in Crystalline Materials: A Law for Strain Gradient Plasticity. Journal of Mechanics and Physics of Solids, 46 (1998) pp. 411-425

[9]   JG. Swadener, EP. George and GM. Pharr: The Correlation of the Indentation Size Effect Measured with Indenters of Various Shapes, Journal of Mechanics and Physics of Solids 50 (2002) pp. 681-694

[10]   CS. Han, A. Hartmaier, H. Gao and Y. Huang: Discrete Dislocation Dynamics Simulations of Surface Induced Size Effects in Plasticity, Materials Science and Engineering, 415 (2006) pp. 225-223

[11]   BJ. Briscoe, L. Fiori and E. Pelillo: Nanoindentation of Polymeric Surfaces. Journal of Physics, 31 (1998) pp. 2395-2405

[12]   X. Li and B. Bhushan: Continuous Stiffness Measurement and Creep Behavior of Composite Magnetic Tapes, Thin Solid Films 377-378 (2000) pp. 401-406

[13]   DCC. Lam, F. Yang, ACM. Chong, J. Wang and P. Tong: Experiments and Theory in Strain Gradient Elasticity, Journal of Mechanics and Physics of Solids, 51 (2003) pp. 1477-1508

[14]   AW. McFarland and JS. Colton: Role of Material Microstructure in Plate Stiffness with Relevance to Microcantilever Sensors, Journal of Micromechanics and Microengineering 15 (2005) pp. 1060-1067

[15]   S. Nikolov, CS. Han and D. Raabe: On the Origin and Modeling of Size Effects in Small-Strain Elasticity of Solid Polymers. International Journal of Solids and Structures, 44 (2007) pp. 1582-1592

[16]   S. Balos, T. Balos, L. Sidjanin, D. Markovic, B. Pilic and J. Pavlicevic: Flexural and Impact Strength of Microwave Treated Autopolymerized Poly(methyl-methacrylate), Materiale Plastice 46 (2009) pp. 261-265

[17]   S. Balos, T. Balos, L Sidjanin, D. Markovic and B. Pilic: Study of PMMA Biopolymer Properties Treated by Microwave Energy, Materiale Plastice 48 (2011) pp. 127-131

[18]   J. Gong and Y. Li: An Energy-Balance Analysisfor the Size Effect in Low-Load Hardness Testing, Journal of Materials Science 35 (2000) pp. 209-213

[19]   T. Šuh, D. Mitić, D. Lebl-Antonić, A. Lebl: Determination of the Necessary Number of Technicians on the Faculty, Acta Polytechnica Hungarica 11 (2014) pp. 21-36