

MATHEMATICAL MODELS and COMPUTATIONAL METHODS

2nd Edition

**Proceedings of the International Conference on Applied Mathematics,
Computational Science & Engineering (AMCSE 2015)**

**Proceedings of the International Conference on Mathematical Models
and Methods in Applied Sciences (MMMAS 2015)**

**Proceedings of the International Conference on Economics and Applied
Statistics (EAS 2015)**

**Agios Nikolaos, Crete, Greece
October 17-19, 2015**

MATHEMATICAL MODELS and COMPUTATIONAL METHODS

**Proceedings of the International Conference on Applied Mathematics,
Computational Science & Engineering (AMCSE 2015)**

**Proceedings of the International Conference on Mathematical Models
and Methods in Applied Sciences (MMMAS 2015)**

**Proceedings of the International Conference on Economics and Applied
Statistics (EAS 2015)**

**Agios Nikolaos, Crete, Greece
October 17-19, 2015**

Copyright © 2015, by the editors

All the copyright of the present book belongs to the editors. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the editors.

All papers of the present volume were peer reviewed by no less than two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

Series: Mathematics and Computers in Science and Engineering Series | 55

ISSN: 2227-4588

ISBN: 978-1-61804-350-4

MATHEMATICAL MODELS and COMPUTATIONAL METHODS

**Proceedings of the International Conference on Applied Mathematics,
Computational Science & Engineering (AMCSE 2015)**

**Proceedings of the International Conference on Mathematical Models
and Methods in Applied Sciences (MMMAS 2015)**

**Proceedings of the International Conference on Economics and Applied
Statistics (EAS 2015)**

**Agios Nikolaos, Crete, Greece
October 17-19, 2015**

Organizing Committee

Editor:

Prof. Imre J. Rudas, Obuda University, Hungary

Organizing Committee:

Prof. Nikos Mastorakis, Technical University of Sofia, Bulgaria

Prof. Aida Bulucea, University of Craiova, Craiova, Romania

Prof. Kleanthis Psarris, The City University of New York, USA

Prof. Joseph Quartieri, University of Salerno, Italy

Dr. Claudio Guarnaccia, University of Salerno, Italy

Steering Committee:

Prof. George Vachtsevanos, Georgia Institute of Technology, USA

Prof. Valeri Mladenov, Technical University of Sofia, Bulgaria

Prof. Imre Rudas, Obuda University, Budapest, Hungary

Prof. Olga Martin, Politehnica University of Bucharest, Romania

Prof. Georgi Tsenov, Technical University of Sofia, Bulgaria

Prof. Panos M. Pardalos, University of Florida, USA

International Scientific Committee:

Prof. Martin Bohner, Missouri University of Science and Technology, USA

Prof. Dashan Fan, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Prof. Luis Castro, University of Aveiro, Aveiro, Portugal

Prof. Metin Demiralp, Istanbul Technical University, Istanbul, Turkey

Prof. Kamisetty Rao, IEEE Fellow, Univ. of Texas at Arlington, USA

Prof. Alberto Fiorenza, Universita' di Napoli "Federico II", Napoli (Naples), Italy

Prof. Patricia J. Y. Wong, Nanyang Technological University, Singapore

Prof. Salvatore A. Marano, Universita degli Studi di Catania, Catania, Italy

Prof. Martin Schechter, University of California, Irvine, USA

Prof. Ivan G. Avramidi, New Mexico Tech, Socorro, New Mexico, USA

Prof. Michel Chipot, University of Zurich, Zurich, Switzerland

Prof. Narsingh Deo, IEEE Fellow, ACM Fellow, University of Central Florida, USA

Prof. Xiaodong Yan, University of Connecticut, Connecticut USA

Prof. Ravi P. Agarwal, Texas A&M University - Kingsville, Kingsville, TX, USA

Prof. Yushun Wang, Nanjing Normal university, Nanjing, China

Prof. Ferhan M. Atici, Western Kentucky University, Bowling Green, KY 42101, USA

Prof. Anastassios Venetsanopoulos, IEEE Fellow, University of Toronto, Canada

Prof. Ravi P. Agarwal, Texas A&M University - Kingsville, Kingsville, TX, USA

Prof. Feliz Minhos, Universidade de Evora, Evora, Portugal

Prof. Mihai Mihailescu, University of Craiova, Craiova, Romania

Prof. Aggelos Katsaggelos, IEEE Fellow, Northwestern University, USA

Prof. Abraham Bers, IEEE Fellow, MIT, USA

Prof. Lucas Jodar, Universitat Politecnica de Valencia, Valencia, Spain

Prof. Jim Zhu, Western Michigan University, Kalamazoo, MI, USA

Prof. Andrei Korobeinikov, Centre de Recerca Matematica, Barcelona, Spain

Prof. Josef Diblik, Brno University of Technology, Brno, Czech Republic

Prof. Jianqing Chen, Fujian Normal University, Fuzhou, Fujian, China

Prof. Naseer Shahzad, King Abdulaziz University, Jeddah, Saudi Arabia

Prof. Sining Zheng, Dalian University of Technology, Dalian, China

Prof. Leszek Gasinski, Uniwersytet Jagielloński, Krakowie, Poland

Prof. Satit Saejung, Khon Kaen University, Muang District, Khon Kaen, Thailand

Prof. Ferhan M. Atici, Department of Mathematics, Western Kentucky University, USA

Prof. Meirong Zhang, Tsinghua University, Beijing, China

Prof. Lucio Boccardo, Universita degli Studi di Roma "La Sapienza", Roma, Italy
Prof. Tiecheng Xia, Department of Mathematics, Shanghai University, China
Prof. Lucas Jodar, Universitat Politecnica de Valencia, Valencia, Spain
Prof. Noemi Wolanski, Universidad de Buenos Aires, Buenos Aires, Argentina
Prof. Zhenya Yan, Chinese Academy of Sciences, Beijing, China
Prof. Shanhe Wu, Longyan University, Longyan, Fujian, China
Prof. Natig M. Atakishiyev, National Autonomous University of Mexico, Mexico
Prof. Jianming Zhan, Hubei University for Nationalities, Enshi, Hubei Province, China
Prof. Narcisa C. Apreutesei, Technical University of Iasi, Iasi, Romania
Prof. Detlev Buchholz, Universitaet Goettingen, Goettingen, Germany
Prof. Patricia J. Y. Wong, Nanyang Technological University, Singapore
Prof. Juan J. Trujillo, Universidad de La Laguna, La Laguna, Tenerife, Spain
Prof. Juan Carlos Cortes Lopez, Universidad Politecnica de Valencia, Spain
Prof. Wei-Shih Du, National Kaohsiung Normal University, Kaohsiung City, Taiwan
Prof. Chun-Gang Zhu, Dalian University of Technology, Dalian, China
Prof. Abdelghani Bellouquid, University Cadi Ayyad, Morocco
Prof. Jinde Cao, Southeast University/ King Abdulaziz University, China
Prof. Dumitru Baleanu, Cankaya University, Ankara, Turkey
Prof. Kailash C. Patidar, University of the Western Cape, Cape Town, South Africa

Additional Reviewers

Abelha Antonio	Universidade do Minho, Portugal
Angel F. Tenorio	Universidad Pablo de Olavide, Spain
Miguel Carriegos	Universidad de Leon, Spain
Alejandro Fuentes-Penna	Universidad Autónoma del Estado de Hidalgo, Mexico
Bazil Taha Ahmed	Universidad Autonoma de Madrid, Spain
M. Javed Khan	Tuskegee University, AL, USA
Moran Wang	Tsinghua University, China
Kazuhiko Natori	Toho University, Japan
Zhong-Jie Han	Tianjin University, China
Genqi Xu	Tianjin University, China
James Vance	The University of Virginia's College at Wise, VA, USA
Jose Flores	The University of South Dakota, SD, USA
Valeri Mladenov	Technical University of Sofia, Bulgaria
Minhui Yan	Shanghai Maritime University, China
Francesco Zirilli	Sapienza Universita di Roma, Italy
Tetsuya Shimamura	Saitama University, Japan
Andrey Dmitriev	Russian Academy of Sciences, Russia
Dmitrijs Serdjuks	Riga Technical University, Latvia
George Barreto	Pontificia Universidad Javeriana, Colombia
Francesco Rotondo	Polytechnic of Bari University, Italy
Masaji Tanaka	Okayama University of Science, Japan
Imre Rudas	Obuda University, Budapest, Hungary
Ole Christian Boe	Norwegian Military Academy, Norway
Stavros Ponis	National Technical University of Athens, Greece
Frederic Kuznik	National Institute of Applied Sciences, Lyon, France
Jon Burley	Michigan State University, MI, USA
Eleazar Jimenez Serrano	Kyushu University, Japan
Hessam Ghasemnejad	Kingston University London, UK
Konstantin Volkov	Kingston University London, UK
Takuya Yamano	Kanagawa University, Japan
João Bastos	Instituto Superior de Engenharia do Porto, Portugal
José Carlos Metrôlho	Instituto Politecnico de Castelo Branco, Portugal
Philippe Dondon	Institut polytechnique de Bordeaux, France
Xiang Bai	Huazhong University of Science and Technology, China
Tetsuya Yoshida	Hokkaido University, Japan
Shinji Osada	Gifu University School of Medicine, Japan
Kei Eguchi	Fukuoka Institute of Technology, Japan
Yamagishi Hiromitsu	Ehime University, Japan
Santoso Wibowo	CQ University, Australia
Sorinel Oprisan	College of Charleston, CA, USA
Deolinda Rasteiro	Coimbra Institute of Engineering, Portugal
Matthias Buyle	Artesis Hogeschool Antwerpen, Belgium

Table of Contents

Plenary Lecture 1: A Comparison of Evolutionary Algorithms to Construct Phylogenetic Trees and Language Families	13
<i>Peter Z. Revesz</i>	
Plenary Lecture 2: Knowledge Processing through Parameter Identification and Computer Aided Scale Up/Down in Engineering	14
<i>Fragiskos Batzias</i>	
Plenary Lecture 3: Approach to Electric Power Equipment Modelling through Sustainability Key Concepts	16
<i>Cornelia Aida Bulucea</i>	
Plenary Lecture 4: Short Term and Asymptotic Properties of Minimal-Exploration Sequential Allocation Rules	18
<i>Michael N. Katehakis</i>	
Plenary Lecture 5: Artificial Intelligence Technology in Health Informatics	19
<i>Abdel-Badeeh M. Salem</i>	
A Computational Model of the Spread of Ancient Human Populations Based on Mitochondrial DNA Samples	21
<i>Peter Z. Revesz</i>	
Evaluation of the Electromagnetic Properties of a Concrete Sample by Radar Measurements	26
<i>M. Albrand, G. Klysz, Y. Grisel, X. Ferrieres</i>	
Creating a Quasi-Continuum between Repeatability and Reproducibility in Statistical Experimental Design	31
<i>Fragiskos A. Batzias</i>	
An Incremental Phylogenetic Tree Algorithm Based on Repeated Insertions of Species	40
<i>Peter Z. Revesz, Zhiqiang Li</i>	
Randomized Global Optimization for Robust Pose Estimation of Multiple Targets in Image Sequences	45
<i>Johannes Brünger, Imke Traulsen, Reinhard Koch</i>	
A Computational Translation of the Phaistos Disk	53
<i>Peter Z. Revesz</i>	

Knowledge Base Modelling for Lignocellulosic Materials Optimal Processing by Means of Fuzzy Fault Tree Analysis	58
<i>Dimitrios Batzias, Dimitrios Sidiras, Christina Siontorou, Fragiskos Batzias, Michael Tsapatsis, Ivo Safarik</i>	
Curve Restoration with Implementations Based on Probability Distribution Functions	69
<i>Dariusz J. Jakóbczak</i>	
Mutations of Adjacent Amino Acid Pairs are not Always Independent	75
<i>Jyotsna Rmanan, Peter Revesz</i>	
On the Tradeoff between Biomass Exploitation and Exploration within an Interdisciplinary Research Project for Developing a New Product	80
<i>Dimitrios Batzias, Dimitrios Sidiras, Christina Siontorou, Leonidas Kamarinopoulos, Yiannis Pollalis, Fragiskos Batzias</i>	
A-Maze-D: Advanced Maze Development Kit Using Constraint Databases	89
<i>Shruti Daggumati, Peter Z. Revesz, Corey Svehla</i>	
International Trade of Environmental Goods: Is Trade Liberalization Fostering the Mexican Environmental Industry?	95
<i>Petr Sauer, René Fernando Lara Cervantes</i>	
A Computational Study of the Evolution of Cretan and Related Scripts	101
<i>Peter Z. Revesz</i>	
Computer Controlled Low Cost System for Professional 360° Photography	106
<i>Krzysztof Szklanny, Alicja Wieczorkowska</i>	
Prediction of Surface Roughness in CNC Milling of Al7075 Alloy: A Case Study of Using 8mm Slot Mill Cutter	110
<i>J. Kechagias, P. Kyratsis, K. Kitsakis, N. Mastorakis</i>	
Why an Economy Needs More than One Currency: the Scientific Evidence	115
<i>Bernard Lietaer</i>	
Aphastory for Google Glass	124
<i>Krzysztof Szklanny, Marcin Wichrowski, Alicja Wieczorkowska</i>	
Design of an Intelligent System for Disasters Management	128
<i>Theodora Dumitrescu, Razvan Popescu, Radu Dobrescu</i>	
Using Chroma Subsampling in Lossy Compression	134
<i>P. Pokorny</i>	

Application of Ostwald Ripening to a Prediction of Grain Mean Radius in Grain Coarsening and Coalescence	138
<i>Aliki D. Muradova</i>	
An Intelligent Diagnosis System Based on Extreme Learning Machine for Diabetes Diseases	143
<i>R. Coteli</i>	
Simulation Studies for Coalescence of Carbon Nanotubes from Graphene Using Controlled Methods	147
<i>D. Fülep, I. Zsoldos, I. László</i>	
An Investigation of Dimensional Accuracy of Multi-Jet Modeling Parts	151
<i>K. Kitsakis, Z. Moza, V. Iakovakis, N. Mastorakis, J. Kechagias</i>	
Human-Centered Architecture of a Medical Cyber-Physical System	158
<i>Razvan Popescu, Theodora Dumitrescu, Radu Dobrescu</i>	
Semi-Automated Object Identification and Features Extraction of Underground Faults from Tectonic Maps	162
<i>Antonios J. Konstantaras, Nikolaos S. Petrakis, Theofanis S. Frantzeskakis, Emmanouil N. Antonidakis</i>	
Advantages and Disadvantages of Family Entrepreneurship and How to Prevent Distress: Evidence from the Czech Republic	166
<i>Ondřej Machek, Petra Votavová</i>	
Design Sequences over the Finite Field of Order Four with High Linear Complexity and Arbitrary Even Period	171
<i>Vladimir Edemskiy</i>	
Study on Starting the High Power Induction Motors with Wounded Rotor	174
<i>Ion Vlad, Sorin Enache, Monica A. Enache, Ionut D. Smarandache</i>	
A Printed Circuit Board Exposure Device Design for Course Projects Using LEDs	179
<i>J. Chatzakis</i>	
Educational App for Android Mobile Devices	184
<i>Krzysztof Szklanny, Marcin Wichrowski, Alicja Wieczorkowska</i>	
The Hybrid Methods are 60 Years in the Scientific Works	189
<i>G. Mehdiyeva, V. Ibrahimov, M. Imanova</i>	
Hybrid Control System for the Hyper Redundant Arm HHR in Creeping Movements	194
<i>Ionel Cristian Vladu, Viorel Stoian, Ileana Vladu</i>	

Implementation of Type Wheels to Design Caterpillar Prototype Chair with Automated Features for People with Disabilities in Paraplegia	200
<i>Maribel Aguilar Echeverri, Arnold Romero Chaparro, Milena Alejandra Pirajan Zamora, Juan Diego López Vargas</i>	
Metaheuristics Based on the Variables Integration Method Applied to Reactive Power Compensation in Multi-Objective Optimization	207
<i>Iliana González Palau, Secundino Marrero Ramírez, Arístides Legrá Lobaina, Daniel Mendiola Ellis</i>	
Edge Detection Based Nearest Neighbor Linear Cellular Automata Rules	213
<i>Nima Aberomand</i>	
Automatically Diagnosis of Suspicious Lesions in Mammograms	216
<i>A. Elmoufidi, K. El Fahssi, S. Jai-Andaloussi, A. Sekkaki, G. Quellec, M. Lamard, G. Cazuguel</i>	
Exploitation of Chaotic and Synchronization Properties of Logistic Maps for Application in Wireless Communication	221
<i>Bijoy Kamal Bhattacharyya, Hemanta Kumar Sarmah, Kandarpa Kumar Sarma, Nikos Mastorakis</i>	
Reliability Polynomials: Obtaining and Usage	226
<i>Alexey S. Rodionov, Olga Rodionova</i>	
Mesh Refinement with Finite Elements and Artificial Neural Networks	230
<i>Fatima Belhabib, Mohamed Ettaouil</i>	
Determination the Coefficient of Regenerative Losses in Stirling	240
<i>Traian Florea, Catalin Oita, Traian Vasile Florea</i>	
Authors Index	250

Plenary Lecture 1

A Comparison of Evolutionary Algorithms to Construct Phylogenetic Trees and Language Families



Professor Peter Z. Revesz

Department of Computer Science and Engineering
University of Nebraska-Lincoln
USA

E-mail: revesz@cse.unl.edu

Abstract: Computer algorithms for the reconstruction of phylogenetic trees based on genome data have greatly facilitated the study of biological evolution. However, the existing phylogenetic tree algorithms often give implausible and sometimes clearly incorrect reconstructions. We present some novel phylogenetic tree algorithms that give biologically more acceptable reconstructions. We also describe how the phylogenetic tree algorithms can be adapted to the study of other types of evolution. In particular, we study the evolution of languages and reconstruct language evolutionary trees. We discuss the similarities and differences in trying to reconstruct phylogenetic trees and language families.

Brief Biography of the Speaker: Peter Z. Revesz holds a Ph.D. degree in Computer Science from Brown University. He was a postdoctoral fellow at the University of Toronto before joining the University of Nebraska-Lincoln, where he is a professor in the Department of Computer Science and Engineering. Dr. Revesz is an expert in databases, data mining, big data analytics and bioinformatics. He is the author of *Introduction to Databases: From Biological to Spatio-Temporal* (Springer, 2010) and *Introduction to Constraint Databases* (Springer, 2002). Dr. Revesz held visiting appointments at the IBM T. J. Watson Research Center, INRIA, the Max Planck Institute for Computer Science, the University of Athens, the University of Hasselt, the U.S. Air Force Office of Scientific Research and the U.S. Department of State. He is a recipient of an AAAS Science & Technology Policy Fellowship, a J. William Fulbright Scholarship, an Alexander von Humboldt Research Fellowship, a Jefferson Science Fellowship, a National Science Foundation CAREER award, and a “Faculty International Scholar of the Year” award by Phi Beta Delta, the Honor Society for International Scholars.

Plenary Lecture 2

Knowledge Processing through Parameter Identification and Computer Aided Scale Up/Down in Engineering



Professor Fragiskos Batzias

Laboratory of Simulation of Industrial Processes
Department of Industrial Management and Technology
School of Maritime and Industry
University of Piraeus
Greece
E-mail: fbatzi@unipi.gr

Abstract: Knowledge Processing is a modern domain, forming part of (but not limited to) the Computer Science and Information Technology discipline. On the other hand, 'Parameter Identification' is a method for transforming implicit to explicit knowledge by top-down penetration from surface/empirical to deeper/scientific phenomenological levels. Herein, we present a methodological framework under the form of an algorithmic procedure for optimizing the 'depth' of this penetration by means of techno-economic criteria, since depth increase implies non-linear (due to the validity of the 'Law of Diminishing Returns') increase of Research and Development (R&D) cost. Implementation of this methodology is presented in several topics of Environmental and Chemical Engineering, from the point of view of interdisciplinary R&D. The dimension of such an interdisciplinarity should be emphasized when a Research Programme (especially if concerning environmental or energy issues) is submitted by a consortium to EU or member State authorities for financial support. Finally, the problems, appearing when heterogeneous data/information processing takes place within a Knowledge Base, are analyzed/discussed and certain solutions, through simulation and Model Based Reasoning (MBR), are suggested.

Brief Biography of the Speaker: Prof. Fragiskos Batzias holds a 5years Diploma and a PhD degree in Chemical Engineering, and a BSc in Economics. He has also studied Mathematics and Philosophy. He designed/developed the Laboratory of Simulation of Industrial Processes and the Research Group on Systems Analysis at the Department of Industrial Management and Technology of the University of Piraeus, Greece. He is teaching at the postgraduate courses (i) Systems of Energy Management and Protection of the Environment, running by the University of Piraeus, and (ii) Techno-Economic Systems, running by the Electr. & Comp. Eng. Dept. of the Nat. Tech. Univ. of Athens in cooperation with the University of Athens and the University of Piraeus. His research interests are in chemical engineering systems analysis and knowledge based decision making. He has >100 publications in highly ranked journals and conference proceedings, including 29 research monographs in collective volumes, with 652 citations and an h-index of 13 (Scopus). He has participated (and chaired after invitation from the organizers) in

prestigious international conferences, such as those organized periodically by the IEEE, the European Federation of Chemical Engineering (EFCE), the DECHEMA, CHISA, WSEAS Organizations. He organizes the annual Symposium on Industrial and Environmental Case Studies running successfully since 2004 within the International Conference of Computational Methods in Sciences and Engineering (ICCMSE).

Plenary Lecture 3

Approach to Electric Power Equipment Modelling through Sustainability Key Concepts



Professor Cornelia Aida Bulucea

Faculty of Electrical Engineering

University of Craiova

ROMANIA

E-mail: abulucea@em.ucv.ro

Abstract: Since electrical power is used all over the world, the standards of life and development of civilization are often interpreted in correlation with the use of electricity. Nonetheless, concerns and questions have been raised regarding how to achieve a sustainable industrial metabolism. Integrating technical and ecological aspects should represent a significant challenge to humanity within the present industrial world. In line with this idea, sustainability concepts linked to mathematical models can improve understanding of the efficiencies of electric power equipment and systems and guide improvement efforts. Over the last few decades, international legislation have required environmental impact assessment be carried out for all phases of electric power equipment life, according to Life Cycle Assessment tool, which includes the production phase, operation phase and end-of-life phase. Modelling of all these life stages of electric power equipment might offer solutions for further improvement potential, focusing on patterns that reduce the electricity losses during the use phase, and on alternative technologies for reducing human health and environmental impacts. This lecture addresses some aspects illustrating energy conversion processes during the operation of power transformers and induction motors, as modeling examples of sustainable electric equipment. Taking a holistic view, this study focuses on highlighting that industrial ecology permits an alternate view of anthropogenic applications, related both to technical and environmental reference systems. Modelling of an electric power transformer in the use phase, and of an induction motor operating within an electrically driven system according to an industrial ecosystem pattern enhances thinking that anthropogenic activities can and should be viewed in concert with the entire system on Earth.

Brief Biography of the Speaker: Cornelia Aida Bulucea is currently an Associate Professor in Electrotechnics, Electrical Machines and Environmental Electric Equipment in the Faculty of Electrical Engineering, University of Craiova, Romania. She is graduate from the Faculty of Electrical Engineering Craiova and she received the Ph.D degree from Bucharest Polytechnic Institute. In Publishing House she is author of four books in electrical engineering area. Research work is focused on improved solutions for electrical networks on basis of new electric equipment, and environmental impact assessment of electric transportation systems. She has extensive experience in both experimental and theoretical research work, certified by over 70 journal and conference research papers and 15 research projects from industry. Due to WSEAS

recognition as huge scientific Forum she participated over time in nineteen WSEAS International Conferences, presenting papers and chairing sessions. She was Plenary Speaker in the 13th International Conference on Electric Power Systems, High Voltages, Electric Machines (POWER'13), Chania, Crete Island, Greece, August 27-29, 2013, in the 5th IASME/WSEAS International Conference on ENERGY&ENVIRONMENT (EE'10), held by the University of Cambridge, UK, February 23-25, 2010, in the 4th IASME/WSEAS International Conference on ENERGY&ENVIRONMENT (EE'09), held by the University of Cambridge, Cambridge UK, February 24-26, 2009, in the 8th WSEAS International Conference on POWER SYSTEMS (PS'08), held by the University of Cantabria, Santander, Spain, September 23-25, 2008. She is very proud by her over 30 papers published in the WSEAS Conferences Books and in the WSEAS TRANSACTIONS ON ENVIRONMENT AND DEVELOPMENT, WSEAS TRANSACTIONS ON POWER SYSTEMS, WSEAS TRANSACTIONS ON CIRCUITS AND SYSTEMS and WSEAS TRANSACTIONS ON ADVANCES IN ENGINEERING EDUCATION.

Plenary Lecture 4

Short Term and Asymptotic Properties of Minimal-Exploration Sequential Allocation Rules



Professor Michael N. Katehakis

Rutgers University

NJ, USA

E-mail: mkatehakis@gmail.com

Abstract: Consider the problem of sampling sequentially from a finite number of $N \geq 2$ populations or ‘bandits’, where each population i is specified by a sequence of random variables $\{X_{ki}\}_{k \geq 1}$, X_{ki} representing the reward received the k th time population i is sampled. For each i , the $\{X_{ki}\}_{k \geq 1}$ are taken to be i.i.d. random variables with finite mean. For any slowly increasing function g , subject to mild regularity constraints, we construct two policies (the g -Forcing, and the g -Inflated Sample Mean) that achieve a measure of regret of order $O(g(n))$ almost surely as $n \rightarrow \infty$. Additionally, asymptotic probability one bounds on the remainder term are established. In our constructions, the function g effectively controls the ‘exploration’ of the classical ‘exploration/exploitation’ tradeoff.

When additional parametric assumptions can be made, one can construct policies that are asymptotically optimal in the sense of achieving the lower bound on the logarithmic rate of increase of the regret of Burnetas and Katehakis (1996). We present such asymptotically optimal policies for the cases in which $\{X_{ki}\}$ are: a) Normal with unknown means and unknown variances and b) Uniform with unknown supports.

Brief Biography of the Speaker: Dr. Katehakis is a Professor in the Management Science and Information Systems Department at Rutgers University and chair of the Department. He holds a courtesy appointment in Rutgers’ New Brunswick Department of Mathematics Graduate Faculty, and he is a member of DIMACS the Center for Discrete Mathematics and Theoretical Computer Science, he is a Primary Investigator of CDDA the Rutgers Center for Dynamic Data Analytics, and a member of RUTCOR, the Rutgers Center for Operations Research.

Much of his work has been on the interaction between optimization and statistical inference.

Professor Katehakis joined the Rutgers University faculty after receiving his doctorate in Operations Research at Columbia University under the supervision of Cyrus Derman, and after being a faculty member at SUNY Stony Brook and at the Technical University of Crete. In addition, professor Katehakis was a member of the technical staff at the Operations Research Center of Bell - Laboratories, West Long Branch and a consultant at Brookhaven National Laboratory and he has held visiting appointments and taught at Columbia University, Stanford University and the National and Kapodistrian University of Athens, Greece.

Dr. Michael N. Katehakis is a Fellow of the Institute for Operations Research and the Management Sciences (INFORMS), an Elected Member of the International Statistical Institute (ISI) and a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE).

Dr. Michael N. Katehakis is the President of the College of Service Operations, Production and Operations Management Society (POMS).

Plenary Lecture 5

Artificial Intelligence Technology in Health Informatics



Prof. Abdel-Badeeh M. Salem

Faculty of Computer and Information sciences

Ain Shams University

Cairo, Egypt

E-mail: absalem@cis.asu.edu.eg

Abstract: Artificial intelligence (AI) is science and technology and is based on many disciplines such as: computer science, philosophy, psychology, mathematics, biology, linguistics, knowledge computing and engineering. AI has been mainly studied as computer based technologies. Various intelligent methodologies, computational techniques and knowledge-based systems have been developed for automated reasoning and learning. AI technologies are robust, can be successfully applied to complex problems, efficiently adaptive, and usually have a parallel computational architecture. For those reasons they have been proved to be effective and efficient in developing intelligent systems for many tasks in health sciences.

The aim of this talk is to make an overview of some of AI techniques and approaches and their applications in medical informatics and health care. The talk covers the following applications: (a) expert systems using the case-based reasoning approach for cancer and heart diagnosis, (b) ontological engineering approach for breast cancer knowledge management, and (c) mining patient data using rough sets theory to determine thrombosis disease.

Brief Biography of the Speaker: Prof. Dr. Abdel-Badeeh M Salem is a professor emeritus of Computer Science since September 2007 till now. He was a former Vice Dean of the Faculty of Computer and Information Sciences at Ain Shams University, Cairo-Egypt (1996-2007). He was a professor of Computer Science at Faculty of Science, Ain Shams University from 1989 to 1996. He was a Director of Scientific Computing Center at Ain Shams University (1984-1990). His research includes intelligent computing, expert systems, medical informatics, and intelligent e-learning technologies. He has published around 350 papers in refereed journals and conference proceedings in these areas. He has been involved in more than 400 conferences and workshops as an Int. Program Committee , organizer and Session Chair. He is author and co-author of 15 Books in English and Arabic Languages.

He was one of the founders of the following events, First Egyptian Workshop on Expert Systems 1987, Int. Cairo Conference on Artificial Intelligence Applications in 1992 and Int. Conf. on Intelligent Computing and Information Systems 2002, and one of the main sustainers of annual Int. Romanian Internet Learning Workshop Project (RILW), 1997. In addition he was Secretary of Egyptian Computer Society (1984-1990), Member of National Committee in Informatics – Academy of Scientific Research and Technology (1992-200), Member of Egyptian Committee in the Inter-Governmental Informatics Program, IIP-UNISCO, Paris (1988-1990) and Coordinator of

the Annual International Conference for Statistics, Scientific Computing, and Social and Demographic Research (1983-1990). In addition he was a partner of a MEDCAMPUS Projects on Methodologies and Technologies for Distance Education in Mediterranean (1993-1995)

He is a member of the Editorial Board of the following Journals: Int. Journal of Computing and Information Sciences(IJCIS), Canada; Egyptian Computer Science Journal, EC Newsletter, Education in Computing and Computers in Education, Italy; Scientific Journal of Studia Universitatis Babes-Bolyai, Series Informatica, Cluj – Napoca, Romania; International Journal of intelligent computing in medical sciences and image processing (IC- MED), Japan; Egyptian Journal for Specialized Studies, Faculty of Specific Education,Ain Shams University,Egypt; Int. Journal of Intelligent Computing & Information Science”,IJICIS, Egypt; Enformatika Transactions on Engineering, Computing and Technology, World Enformatika Society, Turkey; and Int. Journal of Soft Computing Approaches (IJSCA), Eurojournals.

He is a member of Int. Scientific Societies: American Association of Artificial Intelligence (AAAI), USA; British Computer Society, Expert Systems Specialist Group (SGES), Int. Neural Network Society (INNS), USA; Association for the Advancement of Computing Education (AACE), USA; Int. Society for Computers and their Applications ((ISCA), NC, USA, Dec. 95); Int. Society for Telemedicine & eHealth ISfTeH,, Switzerland; Member of Int. Federation for Information Processing (IFIP) Technical Committee WG 12.5, Knowledge-Oriented Development of Applications, Austria (2000 till now), Member of Int. Association for Science and Technology for Development (IASTED), TC on AI and Expert Systems, Int. Association for Science and Technology for Development, Canada, (2000 till now).

A Computational Model of the Spread of Ancient Human Populations Based on Mitochondrial DNA Samples

Peter Z. Revesz

Abstract— The extraction of mitochondrial DNA (mtDNA) from ancient human population samples provides important data for the reconstruction of population influences, spread and evolution from the Neolithic to the present. This paper presents a mtDNA-based similarity measure between pairs of human populations and a computational model for the evolution of human populations. In a computational experiment, the paper studies the mtDNA information from five Neolithic and Bronze Age populations, namely the Andronovo, the Bell Beaker, the Minoan, the Rössen and the Únětice populations. In the past these populations were identified as separate cultural groups based on geographic location, age and the use of, decoration or shape of cultural artifacts.

Keywords—Evolution, Mitochondrial DNA, Population Genetics, Similarity Measure, Phylogenetic Tree.

I. INTRODUCTION

Recent advances in biotechnology enable the extraction of ancient mitochondrial DNA (mtDNA) from human bones going back thousands of years. These advances already facilitated several studies of the origin and spread of various mtDNA types, called haplogroups. However, most human populations are highly heterogeneous in terms of their mtDNA haplogroup compositions. Hence even with the newly available mtDNA information, it is not obvious how human populations spread geographically over time. In particular, there are two main challenges for such studies.

The first challenge in studying the relationships among human populations is to develop an easy-to-compute and flexible similarity measure between pairs of human populations based on mtDNA samples from those two populations. Flexibility in this case means that the similarity measure has to accommodate mtDNA haplogroups that are defined to an arbitrary depth or level. For example, we need to be able to compare a relatively short haplogroup description, such as H5 with a long haplogroup description, such as H1a5b2. We define in Equation (2) below for any pair of populations an overall similarity measure that is both easy-to-compute and flexible.

Once a pairwise overall similarity measure is defined, it is possible to build a similarity matrix for all the populations for which mtDNA sample data is available. The second challenge

is making valid inferences from the similarity matrix regarding the mutual interaction and spread of human populations. In the area of phylogenetics, which is the study of biological phyla, similarity matrices are used to derive a hypothetical evolutionary tree of the phyla [1]-[4]. However, the algorithms that build hypothetical evolutionary trees, such as Neighbor Joining [9], UPGMA [11] and the common mutations similarity matrix (CMSM) algorithm [5] may not be applicable to the study of human populations for several reasons. First, the time scale of phyla evolution is vast compared to the time scale of the development of human populations. The evolution of biological phyla may take millions of years [10], [12], while ancient human mtDNA samples do not go back more than about ten thousand years. Second, while biological phyla diverge from each other in genetic isolation, when human populations come in contact with each other, they tend to merge their genetic pool. Therefore, the set of mtDNAs in a human population may come from several different ancestor human populations that were each more homogeneous in their mtDNA compositions. In general, if P_1 and P_2 are two human populations with set of mtDNAs S_1 and S_2 , respectively, such that the condition

$$S_1 \subseteq S_2 \quad (1)$$

holds, then P_1 can be assumed to be an ancestor of P_2 . However, the reverse is not true. In other words, P_1 may be an ancestor of P_2 but the above condition may not hold because either not all mtDNAs were transferred from P_1 to P_2 or some of the transferred mtDNAs have evolved to a different form.

This paper is organized as follows. Section II presents a computational model of the overall similarity between two populations based on mitochondrial DNA haplogroup samples from the two populations. Section III describes experimental results based on five different ancient Neolithic and Bronze Age European populations. These populations are identified by and associated with different cultural artifacts and were not considered related. However, the material culture can change over time to a degree that the relationships among various cultures become unrecognizable. Our experimental study reveals which populations are closer or more distantly related with each other. Finally, Section IV gives some conclusions and directions for future work.

Peter Z. Revesz is with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA (revesz@cse.unl.edu).

II. A COMPUTATIONAL MODEL

The degree of relatedness between two individuals can be estimated based on a comparison of their mtDNA haplogroups. In this paper, we use the mtDNA haplogroup classification provided by PhyloTree.org at <http://www.phylotree.org>.

We say that a *level 1 relationship* exists between two individuals if they belong to the same haplogroup (single capital letter) but do not share further classifications. We say that a *level 2 relationship* exists between two individuals if they belong to the same sub-haplogroup (capital letter and number) but do not share further classifications. In general, we say that a *level n relationship* exists between two individuals if their haplogroup classifications share the first n elements. For example, H1a2 and H1a5b have a level 3 relationship because they share H1a, that is, three elements, namely the haplogroup H, the sub-haplogroup H1 and the sub-sub-haplogroup H1a.

Note that the largest shared level is a unique number for any pair of haplogroups. This allows us to define the function

$$\text{Level}: s_1 \times s_2 \rightarrow N$$

which takes as input two haplogroups s_1 and s_2 and returns the maximum level numbering of the relationship that exists between them. For example,

$$\text{Level}(H1a2, H1a5b) = 3.$$

We also define the *weight function*

$$W: N \rightarrow N$$

which takes as input a level number and returns a weight value. For example, $W(3)$ returns the weight of level 3 relationships. The weight is intended to describe the degree of unusualness of the existence of a relationship. Normally we would expect the weights to increase exponentially in value because the mtDNA haplogroup tree has many branches at all levels.

We define the *overall similarity* between two populations P_1 and P_2 with associated mtDNA samples S_1 and S_2 , respectively, by the following equation:

$$\text{sim}(P_1, P_2) = \frac{\sum_{a \in S_1, b \in S_2} W(\text{Level}(a, b))}{n \times m} \quad (2)$$

where n and m are the number of samples in the two populations. Here P_1 and P_2 are bags (instead of sets) and can contain repetitions. Equation (2) says that the similarity of two populations equals to the weighted sum of the relationships between pairs of individuals from the two populations divided by the total number of possible pairs. Overall similarity measures closely related to Equation (2) were previously studied also in arbitration theory [7], [8] and cancer research [6]. Equation (2) defines a symmetric relation. Hence

$$\text{sim}(P_1, P_2) = \text{sim}(P_2, P_1) \quad (3)$$

Equation (2) could be further refined if we would know precisely the probabilities of each haplogroup because then we could select the weigh function to return for each level a value that is in inverse proportion to the probability that two random haplogroup samples have the given level of relationship.

Although Equation (2) could be improved with more statistical information, it is a good first approximation of the overall similarity between two populations. For simplicity, in this paper we assume that the weight function contains the following:

$$\begin{aligned} W(1) &= 0 \\ W(2) &= 0 \\ W(3) &= 1 \\ W(4) &= 5 \\ W(5) &= 25. \end{aligned}$$

III. EXPERIMENTAL RESULTS

A. The mtDNA Database

We obtained mtDNA data from five ancient populations from the website <http://suyun.info/index.php?p=ancientdna> which lists the source and age of the samples and classifies them according to cultural groupings. From that database, we selected the following five ancient populations:

1. **Andronovo culture:** The Andronovo culture, which is noted for the domestication of horses and burial in kurgans, flourished in the steppe region to the north and the east of the Caspian Sea in today's Kazakhstan and Russia [13]. The database contains nine Andronovo mtDNA samples dated 1800 – 1400 BC.

$$\text{Andronovo} = \{H6, K2b, T1a, T2a1b1, U2e, U4, U4, U5a1, Z1\}$$

2. **Bell Beaker culture:** The Bell Beaker culture is a prehistoric Western European culture that was named after its characteristic bell-shaped pottery [14]. Some megalithic structures, for example, Stonehenge is associated with the Bell Beaker culture [14]. The database contains eighteen Bell Beaker mtDNA samples dated 2600 – 2050 BC.

$$\text{Bell_Beaker} = \{H, H, H1, H1e7, H3, H3b, H4a1, H5a3, H13a1a2c, I1a1, J, K1, T1a, U2e, U4, U5a1, U5a2a, W5a\}.$$

3. **Minoan culture:** The Minoan culture flourished on Crete, Santorini and some other Aegean islands [15]. The Minoan culture is noted for building the ancient palace of Knossos that is associated with the mythical labyrinth where King Minos supposedly hid the Minotaur [15]. The database contains 34 Minoan mtDNA samples dated 2400 – 1700 BC.

Minoan = {H, H, H, H, H, H, H5, H7, H13a1a, HV, HV, HV, I5, I5, I5, J2, K, K, K, K, K, K, R0, T, T1, T2, T2, T2, T3, T5, U, U5a, W, X}.

In some cases, the haplogroup classification can be refined based on the <http://www.phylotree.org> website that gives an mtDNA classification tree based on the known mutations that characterizes each branch. The PhyloTree.org classification tree also changed slightly since the Minoan study was done. For example, in the latest version (February 19, 2014) the classifications T3 and T5 are now placed within the T2 branch. Using the updated classifications, the Minoan group can be refined as follows, where the updated values are highlighted in blue:

Minoan₂ = {H, H, H, H, H, H, H5a1b, H7, H13a1a, HV, HV, HV, I5, I5, I5, J2, K, K, K, K, K, K, R0, T2, T1a, T2, T2, T2, T2, T2e, U, U5a1f1/U5a2e, W, X}.

Note that the H5a1b identification is possible because of the mutation 11719A. Note also that U5a can be expanded to either U5a1f1 or U5a2e because both of these contain the 16311C mutation.

- Rössen culture:** The Rössen culture is a Neolithic Central European culture that built settlements consisting of trapezoidal or boat-shaped long houses [16]. The database contains ten mtDNA samples dated 4625 – 4250 BC.

Rössen = {H1, H5b, H16, H89, HV0, K, N1a1a, T2, T2e, X2j}

- Únětice culture:** The Únětice culture is a Bronze Age culture with sites known from Central Germany, the Czech Republic and Slovakia [17]. The Únětice culture is noted for the Nebra Sky disk and other metal artifacts [17]. The database contains twenty mtDNA samples dated 2200 – 1800 BC.

Únětice = {H11a, H2a1a3, H82a, H4a1a1a5, H3, H7h, I, I1, T1, T2, T2, T2b, U, U2, U5a1, U5a1a, U5b, W, X}

B. Computation of a Similarity Matrix

Using Equation (2), we computed the similarity matrix for the five ancient populations as shown in Table 1. Note that the similarity matrix is symmetric because of Equation (3). Each non-diagonal entry of the similarity matrix contains the overall similarity value between two different populations described in the corresponding row and column.

Table 1 Similarity matrix among five different ancient populations.

	Andro	Bell-B	Minoan	Rössen	Únětice
Andro.		.0432	.0196	0	.0556
Bell-B.	.0432		.0523	0	.0417
Minoan	.0196	.0523		.0029	.0074
Rössen	0	0	.0029		0
Únětice	.0556	.0417	.0074	0	

Table 2 shows pairs of mtDNA samples that indicate a level 3 or higher distant relationship between the Andronovo and the Bell Beaker populations.

Table 2 Andronovo and Bell Beaker relationships.

Andronovo	Bell Beaker	Relationship Level
T1a	T1a	3
U2e	U2e	3
U5a1	U5a1	4

Hence by Equation (2) the overall similarity between the Andronovo and the Bell Beaker populations can be calculated to be:

$$sim(Andronovo, Bell_Beaker) = \frac{1 + 1 + 5}{9 \times 18} = 0.0432$$

Similarly, when we compare the Bell Beaker and the Minoan samples, Table 3 shows the pairs that indicate a level 3 or higher relationship.

Table 3 Bell Beaker and Minoan relationships.

Bell Beaker	Minoan ₂	Relationship Level
H5a3	H5a1b	3
H13a1a2c	H13a1a	5
T1a	T1a	3
U5a1	U5a1f1	4

Hence the overall similarity between the Bell Beaker and the Minoan₂ populations is:

$$sim(Bell_Beaker, Minoan_2) = \frac{1 + 25 + 1 + 5}{18 \times 34} = 0.0523$$

As another example, in comparing the Minoan₂ and the Rössen populations only one pair indicates a level 3 or higher relationship as shown in Table 4.

Table 4 Bell Beaker and Rössen relationships.

Minoan ₂	Rössen	Relationship Level
T2e	T2e	3

Hence the overall similarity between the Minoan₂ and the Rössen populations can be calculate to be:

$$\text{sim}(\text{Minoan}_2, \text{Rössen}) = \frac{1}{34 \times 10} = 0.0029$$

Likewise, we can calculate the following similarity values:

$$\text{sim}(\text{Andronovo}, \text{Minoan}) = \frac{1 + 5}{9 \times 34} = 0.0196$$

$$\text{sim}(\text{Andronovo}, \text{Únětice}) = \frac{5 + 5}{9 \times 20} = 0.0556$$

$$\text{sim}(\text{Bell_Beaker}, \text{Únětice}) = \frac{5 + 5 + 5}{18 \times 20} = 0.0417$$

$$\text{sim}(\text{Minoan}, \text{Únětice}) = \frac{5}{34 \times 20} = 0.0074$$

Finally, between the Andronovo and the Rössen, the Bell Beaker and the Rössen, and the Rössen and the Únětice populations, no pair of samples shows a level 3 or higher relationship. Hence

$$\begin{aligned} \text{sim}(\text{Andronovo}, \text{Rössen}) &= 0 \\ \text{sim}(\text{Bell_Beaker}, \text{Rössen}) &= 0 \\ \text{sim}(\text{Rössen}, \text{Únětice}) &= 0 \end{aligned}$$

C. Discussion of the Results

In general, the higher is the similarity value between two populations, the more closely related those two populations are. According to that intuition, the highest similarity (0.0556) is between the Andronovo and the Únětice populations as shown Table 1. There is an almost equally high similarity (0.0523) between the Bell Beaker and the Minoan populations. The results also reveal that the Rössen population is only related with the Minoan populations with a relatively small similarity (0.0029).

Perhaps a deeper insight can be gained from the data if we also consider which haplogroups are the major links (level 3 or higher relationships) between each pair of populations.

Table 5 shows that the Minoan and the Rössen populations are related via the T2e haplogroup, while the Minoan and the Bell Beakers populations are related via H5a, H13a, T1a and U51a haplogroups. The major links between the Andronovo and the Minoan populations are via T1a and U5a1 haplogroups, while the U5a1 haplogroup is the only major link between the Minoan and the Únětice populations.

It needs to be mentioned that in the current database many of the ancient mtDNA samples are only fragments instead of complete mtDNAs. Hopefully, the haplogroup classifications may be further refined with improved testing methods in the future. The results may change slightly as some haplogroup classifications are extended from two to three or more letters. Nevertheless, it seems extremely unlikely that the refinement of some of the mtDNA classifications would change the current clustering picture instead of further strengthening the already existing groupings.

Table 5 The level 3 or higher haplogroup relationships among the five different ancient populations. Only the entries in the upper triangular part of the matrix are shown because the matrix is symmetric.

	Andro	Bell-B	Minoan	Rössen	Únětice
Andro.		T1a U2e U5a1	T1a U5a1		U5a1
Bell-B.			H5a H13a1a T1a U5a1		H4a1 U5a1
Minoan				T2e	U5a1
Rössen					
Únětice					

The experimental results suggest either T2e gene flows between or a common origin of the Minoan and the Rössen populations. Similarly, the results suggest either H5a and H13a1a gene flows between or a common origin of the Bell Beaker and the Minoan populations. The origin and the dispersal of the T1a and U5a1 haplogroups are less clear because they are shared more widely among the five studied populations.

Unfortunately, the mtDNA data does not allow drawing conclusions regarding the language associated with each of the five sample populations in this study. However, either gene flows or common origin between pairs of populations raises the chance of similarity of language. Hence some language similarity is plausible between Minoan and Rössen and between Bell Beaker and Minoan.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we defined a measure for the overall similarity between two populations with mtDNA haplogroup samples. Our study is not merely the study of the dispersal of various mtDNA haplogroups but the dispersal of various populations that are already heterogeneous in terms of their mtDNA compositions.

Our mtDNA haplogroup-based population similarity measure could be extended easily to a Y-DNA haplogroup-based population similarity measure. It would be interesting to perform a similar analysis on Y-DNA data for the populations studied in this paper and compare the similarity matrices generated by the mtDNA and the Y-DNA haplogroup-based data. However, ancient Y-DNA data is much harder to obtain than ancient mtDNA data with current technology. Hence such a Y-DNA study may have to wait until further DNA extraction technology improvements.

Another way to extend the research is to study a larger number of populations. We intend to study more ancient populations as well as currently living populations to gain more insight into the origin and dispersal of various populations. The five populations were all ancient Neolithic or Bronze Age European cultures. Considering populations that encompass a broader time scale and a larger geographic area may give a deeper insight into human pre-history.

REFERENCES

- [1] D. Baum and S. Smith, *Tree Thinking: An Introduction to Phylogenetic Biology*, Roberts and Company Publishers, 2012.
- [2] B. G. Hall, *Phylogenetic Trees Made Easy: A How to Manual*, 4th edition, Sinauer Associates, 2011.
- [3] P. Lerney, M. Salemi, and A.-M. Vandamme, editors. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edition, Cambridge University Press, 2009.
- [4] P. Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, Springer, New York, 2010.
- [5] P. Z. Revesz, "An algorithm for constructing hypothetical evolutionary trees using common mutations similarity matrices," *Proc. 4th ACM International Conference on Bioinformatics and Computational Biology*, ACM Press, Bethesda, MD, USA, September 2013, pp. 731-734.
- [6] P. Z. Revesz and C. J.-L. Assi, "Data mining the functional characterizations of proteins to predict their cancer relatedness," *International Journal of Biology and Biomedical Engineering*, 7 (1), 2013, pp. 7-14.
- [7] P. Z. Revesz, "On the semantics of arbitration," *International Journal of Algebra and Computation*, 7 (2), 1997, pp. 133-160.
- [8] P. Z. Revesz, "Arbitration solutions to bargaining and game theory problems," *Annales Universitatis Scientiarum Budapestinensis, Sect. Comp.*, 43, 2014, pp. 21-38.
- [9] N. Saitou and M. Nei, "The neighbor-joining method: A new method for reconstructing phylogenetic trees," *Molecular Biological Evolution*, 4, 1987, pp. 406-425.
- [10] M. Shortridge, T. Triplet, P. Z. Revesz, M. Griep, and R. Powers, "Bacterial protein structures reveal phylum dependent divergence," *Computational Biology and Chemistry*, 35 (1), 2011, pp. 24-33.
- [11] R. R. Sokal, and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Science Bulletin*, 38, 1958, pp. 1409-1438.
- [12] T. Triplet, M. Shortridge, M. Griep, J. Stark, R. Powers, and P. Z. Revesz, "PROFESS: A protein function, evolution, structure and sequence database," *Database -- The Journal of Biological Databases and Curation*, 2010, Available: <http://database.oxfordjournals.org/content/2010/baq011.full.pdf+html>
- [13] Wikipedia, "Andronovo culture," downloaded August 19, 2015. Available: https://en.wikipedia.org/wiki/Andronovo_culture
- [14] Wikipedia, "Beaker culture," downloaded August 19, 2015. Available: https://en.wikipedia.org/wiki/Beaker_culture
- [15] Wikipedia, "Minoan civilization," downloaded August 19, 2015. Available: https://en.wikipedia.org/wiki/Minoan_civilization
- [16] Wikipedia, "Rössen culture," downloaded August 19, 2015. Available: https://en.wikipedia.org/wiki/Rössen_culture
- [17] Wikipedia, "Unetice culture," downloaded August 19, 2015. Available: https://en.wikipedia.org/wiki/Unetice_culture

Peter Z. Revesz holds a Ph.D. degree in Computer Science from Brown University. He was a postdoctoral fellow at the University of Toronto before joining the University of Nebraska-Lincoln, where he is a professor in the Department of Computer Science and Engineering. Dr. Revesz is an expert in bioinformatics, databases, data mining, and data analytics. He is the author of *Introduction to Databases: From Biological to Spatio-Temporal* (Springer, 2010) and *Introduction to Constraint Databases* (Springer, 2002). Dr. Revesz held visiting appointments at the IBM T. J. Watson Research Center, INRIA, the Max Planck Institute for Computer Science, the University of Athens, the University of Hasselt, the U.S. Air Force Office of Scientific Research and the U.S. Department of State. He is a recipient of an AAAS Science & Technology Policy Fellowship, a J. William Fulbright Scholarship, an Alexander von Humboldt Research Fellowship, a Jefferson Science Fellowship, a National Science Foundation CAREER award, and a "Faculty International Scholar of the Year" award by *Phi Beta Delta*, the Honor Society for International Scholars.

Evaluation of the electromagnetic properties of a concrete sample by radar measurements

M. Albrand, Université de Toulouse, UPS, INSA, LMDC (Laboratoire Matériaux et Durabilité des Constructions),
135, Avenue de Rangueil, F-31 077 Toulouse Cedex 04, France

Email albrand@insa-toulouse.fr

G. Klysz, Université de Toulouse, UPS, INSA, LMDC (Laboratoire Matériaux et Durabilité des Constructions),
135, Avenue de Rangueil, F-31 077 Toulouse Cedex 04, France

Email: klysz@insa-toulouse.fr

Y. Grisel, Université de Pau et des pays de l'Adour, Avenue de l'Université, 64012 Pau, France

Email: yann.grisel@univ-pau.fr

X. Ferrieres, ONERA French Aerospace Lab., 2 avenue Edouard Belin, 31055 Toulouse, France

Email: xavier.ferrieres@onera.fr

Abstract—In this paper is proposed an inverse formulation to evaluate the dielectric values of a concrete sample by using radar measurements in the context of buildings security. To solve the problem, an optimisation strategy based upon a Gauss-Newton method is presented and, finally, it is applied on several configurations using experimental measurements.

I. INTRODUCTION

The detection of corrosion inside old buildings is an important topic to ensure the safety of people using these structures. To make large scale investigations without affecting the integrity of the structures, the radar measurement, based on the propagation of electromagnetic waves, seems a useful tool. Indeed, the identification of some pathologies of concrete, essentially due to moisture content [1], becomes equivalent to the knowledge of the mapping of the concrete's conductivity and permittivity. In this paper, we present an electromagnetic study to evaluate the dielectric constants of a concrete sample, by using a specific measurement procedure. In a first part, we present the formulation of the inverse problem and the strategy proposed to find the dielectric values of a concrete sample. In a second part, we apply our optimization process on homogeneous concrete samples and we give some comparisons between the values found by simulation and the values obtained experimentally. Finally, in a third part, we adapt our method to find the dielectric values of a heterogeneous concret material in the case of properties varying linearly with respect to depth.

II. FORMULATION OF THE INVERSE PROBLEM

To detect possible pathologies in the concrete, it is interesting to determine the dielectric values of the concrete by using electric field measurements taken on its surface. By using these data obtained in time domain for several points [1], we can define an inverse problem to find them. In the general case, when ε and σ are functions of the space, inverse problem is

formulated by:

$$\min_{\varepsilon, \sigma} \|E(\varepsilon, \sigma) - E_{mes}\|^2 \quad (1)$$

under the constraint defined by the Maxwell equations given by:

$$\begin{cases} \varepsilon_0 \varepsilon \partial_t E + \sigma E = \nabla \times H \\ \mu \partial_t H = \nabla \times E \end{cases} \quad (2)$$

In the cost function E_{mes} defines the measured values and E the computed values. ε represents the relative permittivity of the material and σ its conductivity. ε and σ are functions of the space point. In this paper, we are interested in two cases. First ε and σ are constant. Second ε and σ linearly dependent of the depth z and constant with respect to the axis x and y .

To solve the inverse problem we use a Gauss-newton optimisation method combined with a FDTD simulation method to treat the Maxwell equations. The process of optimisation can be described by the following sequence, where ε and σ are constant or vector quantities:

- 1) choose an initial point $(\varepsilon_0, \sigma_0)$;
- 2) compute the electric field E by using the Maxwell equations on all the computational domain ;
- 3) evaluate the jacobian matrices $J = \left(\frac{\partial E}{\partial \varepsilon}, \frac{\partial E}{\partial \sigma} \right)$;
- 4) evaluate the new solution $(\varepsilon^k, \sigma^k)$ with the previous solution $(\varepsilon^{k-1}, \sigma^{k-1})$ by using:

$$\begin{pmatrix} \varepsilon^k \\ \sigma^k \end{pmatrix} = \left(\frac{\partial E^{k-1}}{\partial \varepsilon} \quad \frac{\partial E^{k-1}}{\partial \sigma} \right)^+ (E^{k-1} - E_{mes}) \quad (3)$$

where A^+ denotes the pseudo-inverse of the matrix A .

- 5) evaluate the cost function and return to the step 2 if the value does not satisfy a stopping criterion.

In this process, we need to evaluate at each step the value of the gradients of the cost function. In this study, we used a Broyden formula [6] to evaluate the gradients at the step k by using their values at the step $k-1$ of the process :

$$J_{k+1} = J_k + \frac{\Delta F_k \cdot \Delta x_k^T - J_k \Delta x_k \cdot \Delta x_k^T}{\Delta x_k^T \cdot \Delta x_k} \quad (4)$$

with $\Delta F_k = F_c(\varepsilon_{k+1}, \sigma_{k+1}) - F_c(\varepsilon_k, \sigma_k)$ and $\Delta x_k = \begin{pmatrix} \varepsilon_{k+1} - \varepsilon_k \\ \sigma_{k+1} - \sigma_k \end{pmatrix}$, where F_c and J_k respectively defines the cost function and the gradients at the step k of the reconstruction process.

This descent algorithm makes the solution strongly dependent on the choice of the initial point.

III. APPLICATION TO THE CASE OF HOMOGENOUS MATERIAL

In this part we want to apply our method to retrieve the dielectric values of an homogeneous material. First we are going to see how to adapt and improve the algorithm presented above in the case of an homogeneous material and then we will present some numerical results.

A. Formulation of the problem

In the case of an homogeneous material, the problem (1) consists in minimizing a function of two variables, σ and ε . We want to know how the convergence of the algorithm presented in section (II) depends on the choice of the initial point. To see that, we plot the variations of the cost function

first by fixing σ and making ε vary and then by fixing ε and making σ vary. The figure (1) shows the behaviour of these cost functions on a particular sample. This behaviour is very similar on other samples.

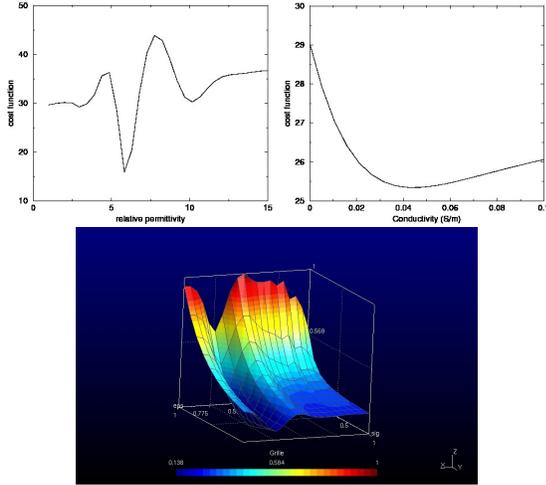


Fig. 1: variations of the cost function versus relative permittivity and conductivity.

On these figures, we can see that the variation of the cost function $f_c(\varepsilon, \sigma) = \|E(\varepsilon, \sigma) - E_{mes}\|^2$ is very different whether it is ε or σ that varies. When σ varies and ε is fixed – whatever the value of ε is – the reconstruction process converges to the minimum of the cost function $f_c(\sigma)$. On the contrary, when σ is fixed and ε varies, the obtained cost function $f_c(\varepsilon)$ presents many local minima. Then, for this parameter, it is very important to take an initial value which

corresponds to the attractive domain of the global minimum. The last curve on this figure presents the variation of the cost function by considering both variables. We notice on this curve the quasi independence between the two variables. This remark is very important because it permits to consider separately the two variables in the optimisation process.

By using the last remark, we propose a two step strategy to find the dielectric values:

- 1) we set a constant value σ_0 in the possibilities interval and we evaluate the cost function for different values of ε on the interval of the possibilities, to obtain an initial value of ε in the attractive domain of the global minimum.
- 2) the value of ε determined at the previous step and the value σ_0 are then used to initialize the Gauss-Newton process

This strategy permits to obtain a good initial point for the Gauss-Newton optimization. However, it requires a lot of evaluations of the electromagnetic fields by solving the Maxwell equations. Using the FDTD simulation method on a 3D model and a 3D grid of the physical domain, every evaluation is very expensive in terms of CPU-time. This leads to consider first a 2D model of the physics and to solve (1) by computing $E(\varepsilon, \sigma)$ with this simplified model and the algorithm defined above. Then we can take the 2D solution as an initial point for the 3D optimization.

B. Numerical results

To validate our approach, we have applied it on experimental measurements. The data are carried out by using a SIR-2000 radar system, equipped with two 1.5 GHz coupled antennas both from GSSI. The GSSI-5100 antenna is a double coupled antenna made up of two dipolar elements, one used as the transmitter and the other as the receiver. The measurements considered in this paper, are obtained by using two GSSI antennas. On the first we use only the transmitting part. The other is the receiving antenna. Several points of measurement are made by fixing the emitter antenna and doing a translation on the concrete surface for the receiver antenna.

The figure (2) shows the experimental test process on a sample of concrete in laboratory. We can see on this figure the two boxes antennas and the measurement at a point posted up the screen of the notebook.

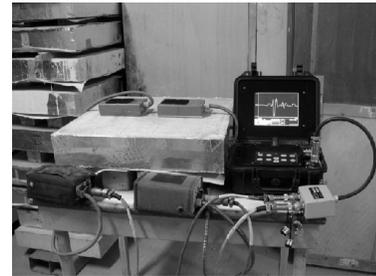


Fig. 2: Experimental device on a sample of concrete.

By using this experimental setup, nine concrete samples were designed and prepared in such a way as to avoid any bias in the measurements from border effects or uncontrolled reflection. Two compositions of concrete, B-25 and B-40, were chosen for the concrete slabs in order to obtain concretes with different porosities and different aggregate sizes. The calculation of the dimensions of these slabs based on the duration of the radar pulse led us to choose $60 \times 60 \times 12 \text{ cm}$. The slabs were covered with aluminium foil on all their faces except the top face, where the measurements were made. This allowed preventing the water evaporation of the slab and also ensured perfect reflection of the radar waves on those faces. This configuration also allowed us to neglect the possible perturbations due to the boundary conditions of the computational domain taken into account in the numerical approach. In order to study the effects of the water content on the radar waves, we used concrete slabs having different water contents with an uniform distribution of water in the slabs.

The first step in the characterisation of the concrete dielectric values by an inverse problem consists in getting good simulations of the measurement process, and then, a good model for the device. In our case it is very difficult to obtain this model. In part, this is due to the fact that the device is a commercial instrument and the values of its components are not available. In particular, the source device and the dielectric materials around the antennas inside the box are unknown. Then, in a first step, we have developed a numerical model which is not exactly representative of the device, but which gives the best fitting between measurements and simulations.

The method proposed in this paper has been applied to different samples of concrete, by using experimental measurements. The table I gives the results obtained for each sample. The values obtained are compared with values provided by previous studies [1], [7], [8]. The figures (3) and (4) show respectively the quite good agreement between relative permittivity and conductivity obtained by solving the inverse problem or by direct measurement. The solving process of the inverse problem is very easy to perform and does not require a long time. Indeed, the other processes, essentially experimental, need a lot of time to correctly adjust the different measurements.

sample	ϵ computed	measured ϵ	σ computed (S/m)	σ measured (S/m)
B40-1	6.41	5.15	3.e-2	2.8e-2
B40-2	7.01	6.38	5.e-2	5.e-2
B40-3	7.45	6.57	6.e-2	5.5e-2
B40-4	7.97	7.97	8.e-2	7.8e-2
B25-1	5.9	4.94	2.e-2	2.5e-2
B25-2	6.72	5.71	4.e-2	4.1e-2
B25-3	6.84	6.59	5.e-2	5.2e-2
B25-4	7.99	7.88	9.e-2	7.6e-2
B25-5	8.3	8.39	10.e-2	8.5e-2

TABLE I: simulations versus measurements for relative permittivity and conductivity.

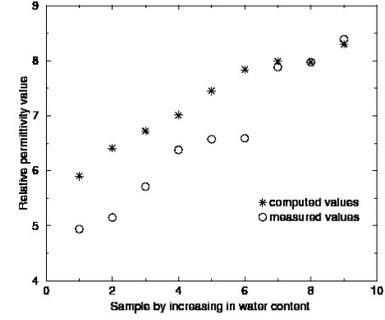


Fig. 3: Relative permittivity variations according to the water content.

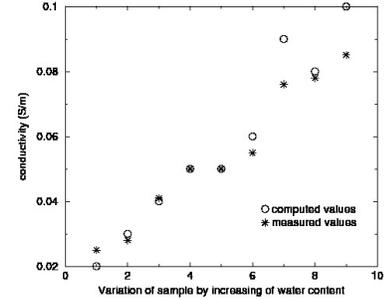


Fig. 4: Conductivity variations according to the water content.

IV. APPLICATION TO THE CASE OF AN HETEROGENEOUS MATERIAL

In the context of building security, in order to have information on the state and aging of a structure, it is interesting to have a more accurate information about the cartography of the moisture content in the concrete rather than a simple mean (the material considered as an homogeneous medium). Particularly meaningful is the evolution in depth of the moisture content, which can cause a corrosion of the armature. In this paper we are going to study the case of a material presenting a linear distribution in depth of its dielectric properties. That is ϵ and σ defined by:

$$\epsilon = a_\epsilon \frac{z - z_0}{z_1 - z_0} + b_\epsilon \text{ and } \sigma = a_\sigma \frac{z - z_0}{z_1 - z_0} + b_\sigma \quad (5)$$

where $z_0 = 0$ is the level of the sample's surface and z_1 is the material's thickness. a_ϵ and b_ϵ are real values so that $\epsilon(z) > 1$. a_σ and b_σ are real values so that $\sigma(z) > 0$. The unknowns of the problem are then a_ϵ , b_ϵ , a_σ and b_σ .

First we are going to see how to adapt the previous algorithm to this new problem. Then we shall present some preliminary numerical results obtained with simulated measurement.

A. Formulation of the problem

In the model that we now consider, the material is defined by four values, a_ϵ , b_ϵ , a_σ and b_σ , as defined in (5). The problem

(1) can be then rewritten as follows:

$$\min_{a_\varepsilon, b_\varepsilon, a_\sigma, b_\sigma} \|E(a_\varepsilon, b_\varepsilon, a_\sigma, b_\sigma) - E_{mes}\|^2 \quad (6)$$

The convergence of the algorithm presented in section II to solve the inverse problem is strongly dependent on the initial values of the unknowns. To find a good point to start our descent algorithm, a first idea is to take the solution point obtained when we consider the material as an homogeneous medium. However numerical experiments show that this approach is not efficient. On the basis of the homogeneous case and in the view of the cost function $f_c(a_\varepsilon, b_\varepsilon, a_\sigma, b_\sigma) = \|E(a_\varepsilon, b_\varepsilon, a_\sigma, b_\sigma) - E_{mes}\|^2$ we observe that the two couples of variables $(a_\varepsilon, b_\varepsilon)$ and a_σ, b_σ seem to be independent. We can see this particular property on the figure (5) where we draw the variations of f_c for different fixed values of a_σ and b_σ . For the first subfigure (top left) we fixed $(a_\sigma, b_\sigma) = (0.01, 0.01)$, for the second (top right) we fixed $(a_\sigma, b_\sigma) = (0.05, 0.05)$, for the third (bottom left) we fixed $(a_\sigma, b_\sigma) = (0.1, 0)$ and for the last subfigure we fixed $(a_\sigma, b_\sigma) = (0.01, 0)$. The shape of the surface as well as the position of the minimum is independent of the choice of (a_σ, b_σ) .

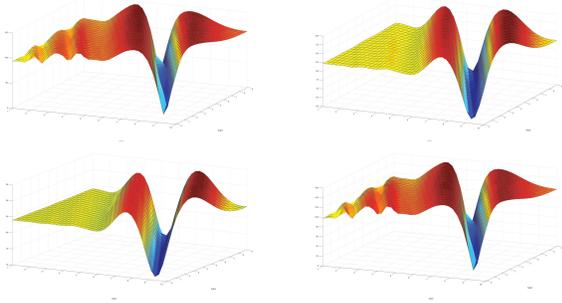


Fig. 5: $f_c(a_\varepsilon, b_\varepsilon)$ for different fixed values of a_σ and b_σ

Next, on figure (6) we draw the variations of f_c for different fixed values of a_ε and b_ε . For the first subfigure we fixed $((a_\varepsilon, b_\varepsilon) = (7, 5))$ and for the second we fixed $((a_\varepsilon, b_\varepsilon) = (5, 10))$. It appears that f_c is convex relatively to a_σ and b_σ .

Consequently, the idea proposed to choose an initial point is to search a couple $(a_\varepsilon, b_\varepsilon)$ by a global optimization method minimizing f_c for an arbitrarily fixed couple (a_σ, b_σ) .

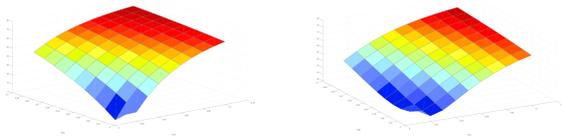


Fig. 6: $f_c(a_\sigma, b_\sigma)$ for different fixed values of a_ε and b_ε

B. Numerical results

Currently, some samples of concrete are in preparation. After the pouring of the concrete, three months are necessary

to stabilize the moisture content inside a slab. Those new test specimens are $50 \times 24.5 \times 12$ cm blocs, some containing a round steel bar and some not. To set up a moisture gradient in depth z independent of x and y , we made the lateral faces impermeable and fixed the relative humidity and the temperature of the upper and lower faces at two different values. When the equilibrium is reached, the moisture content is a linear function of the depth. In a first approximation we can consider that ε and σ are linearly linked with the moisture content. Before we can apply our process to retrieve the dielectric parameters from measurements, we use as E_{mes} a simulated field. The material that we took as a test material is defined as follows: $a_\varepsilon = 7$, $b_\varepsilon = 5$, $a_\sigma = 0.01$ and $b_\sigma = 0.01$. To validate our inverse method, by choosing an initial point for several fixed couples (a_σ, b_σ) as previously described and using the algorithm presented in section II, we evaluated an optimal solution and compared it to the true solution. Table (II) presents for the different fixed couples (a_σ, b_σ) , the initial point and the optimal solution obtained. We note on this table that the evaluated solution is the expected solution which validate our inverse process. Figure (7) shows, for each initial point, the evolution of the cost function and then, the convergence of our optimization process.

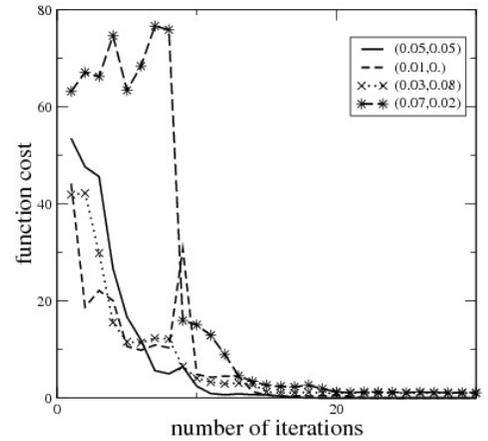


Fig. 7: Evolution of the cost function for several couples (a_{sigma}, b_{sigma}) .

$(a_\varepsilon, b_\varepsilon, a_\sigma, b_\sigma)$	Initial point $(a_\varepsilon, b_\varepsilon, a_\sigma, b_\sigma)$	optimal solution
(0.05, 0.05)	(5, 10, 0.05, 0.05)	(6.999, 5., 0.01, 0.00999)
(0.03, 0.08)	(5, 10, 0.03, 0.08)	(7, 4.999, 0.01, 0.00998)
(0.07, 0.02)	(5, 10, 0.07, 0.02)	(7, 5, 0.00998, 0.01)
(0.01, 0.)	(5, 10, 0.01, 0)	(7, 4.999, 0.01, 0.00998)

TABLE II: Convergence of the descent algorithm according to several chosen fixed couples (a_σ, b_σ) .

V. CONCLUSION

In this paper, we have presented an inverse problem for the characterization of concrete electromagnetic properties, in

a context of sustainability of concrete structures. The aim is to obtain the dielectric values of this material by putting a source on its surface and making measurements of the scattered fields at several points of its surface. First, we have proposed a process to rebuild the permittivity and conductivity values of an homogeneous concrete sample. Some results were given that validated the proposed approach. In a second step, we adapted this process to rebuild gradients in depth of dielectric values inside reinforced concrete. The aim is achieve a better estimation of quality of concrete close to the reinforcement in order to evaluate the risk of corrosion. The first result obtained in the 2D case on simulated measurements shows that it is possible to find the solution if you take a good initial point for the optimization process. To find this point a strategy based on global optimization has been proposed. In our future works, we will introduce measurement data on concrete samples in our inverse process.

VI. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the French Agency of Research for their financial support of the CONTINUS project.

REFERENCES

- [1] G. Klysz, J.-P. Balayssac, X. Ferrieres, *Numerical FDTD model of a GPR coupled antenna for dielectric properties of concrete evaluation: validation and parametric study*, NDT & E International (Non-Destructive Testing and Evaluation), Volume 41, Issue 8, December 2008, pp. 621-631.
- [2] G. Klysz, X. Ferrieres, J.P. Balayssac, S. Laurens, *Simulation of direct wave propagation by numerical FDTD for a GPR coupled antenna*, NDT& E International 39 (2006), 338-347
- [3] R. Holland and L. Simpson, *Finite difference analysis of EMP coupling to thin struts and wire*, IEEE Trans. on EMC, vol.23, pp.88-97, May 91.
- [4] K.S. Yee, *Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media*, IEEE Trans. Antennas Prop., Vol.14, No.3, pp. 302-307, May 1966
- [5] J.-P. Berenger, *A Perfectly Matched Layer for the Absorption of Electromagnetic Waves*, J. Comput. Phys. 114, 185-200, 1994.
- [6] V. Eyert, *A comparative study on methods for convergence acceleration of iterative vector sequences*, J. Comput. Phys. 124, 271-285, (1996)
- [7] G. Klysz, *Caractérisation du béton d'enrobage par technique radar : exploitation de l'onde directe émetteur-récepteur*, Thèse de Doctorat de l'Université Paul Sabatier, Toulouse, 2004
- [8] M. N. Soutsos, J. H. Bungey, S. G. Millard, M. R. Shaw, A. Patterson *Dielectric properties of concrete and their influence on radar testing*, NDT & E International, volume 34, Issue 6, September 2001, pp. 419-425.

Creating a Quasi-Continuum between Repeatability and Reproducibility in Statistical Experimental Design

Fragiskos A. Batzias

Abstract— The aim of this work is to create a quasi-continuum between repeatability and reproducibility conditions, so that both may take place in the same laboratory within a unified experimental design framework. Initially, an independent/explanatory variable I is defined, named Reproducibility Index, standing for Inter-Laboratory Study (ILS) expenditure expressed as a percentage of total expenditure for testing materials and processes. Each optimal value, I_{opt} , can be determined as the equilibrium point in the tradeoff between the partial costs C_1 and C_2 corresponding to experimental design (including the execution of the suggested ILS tests) and downward technical/economic/environmental impact, respectively. Since cost minimization is an incomplete optimality criterion, we have developed an algorithmic procedure, including 24 activity stages and 7 decision nodes, in order to enable the simulation of reproducibility conditions in the laboratory where repeatability conditions have already been established. This maximizes utility by allowing either for scaling up/down of the process under consideration or for the examination of the material to be tested. Subsequently, implementation of certain stages of this algorithmic procedure is presented, concerning (i) continuous flow stirred tank reactor and plug flow reactor (used also as environmental models by simulating lakes and rivers, respectively), as regards process scale up/down, and (ii) anodized aluminum, as regards materials examination. In all cases, the results are discussed, while the ASTM Standards are referenced as a methodological framework.

Keywords— experimental design, metrology standardization, modelling evaluation, optimization, repeatability/reproducibility.

I INTRODUCTORY ANALYSIS

Laboratory measurement carried out on specimens, supposed to be identical, do not necessarily give the same results, due to unpredictable errors that may occur during each examination. These errors depend mainly on the following six factors: (i) the variation of constituents within the (supposed identical) specimens to be tested, (ii) the experience/attention of the human operators in duty, (iii) the kind/type of equipment used, (iv) the condition (including maintenance and calibration) of this equipment, (v) the natural environment (e.g., light intensity, other radiation activity, temperature, humidity), and (vi) the anthropogenic environment, including pollution as principal component.

The closeness of any result to the ‘true’ or the accepted

reference value is expressed by the term of ‘accuracy’, which is expressed (in its turn) in terms of the corresponding bias and precision, the latter being evaluated through repeatability and reproducibility. Under repeatability conditions, the factors listed above are kept or remain reasonably constant and usually contribute only minimally to the variability; under reproducibility conditions, the factors are generally different (i.e., they change among laboratories) and frequently contribute significantly to the variability of examination results; thus, repeatability and reproducibility are two practical extremes of precision [1-5].

These distinct extremes define a domain full of discontinuities as regards Man, Machine, Materials, Methods, and Measurement (5M), constituting the main categories of causal factors representing dependence in an Ishikawa ‘fishbone’ structure [6,7]. The aim of the present work is to create a quasi-continuum between Repeatability and Reproducibility, and contribute to optimization of the variables/ parameters/coefficients (VPCs) involved.

For illustrating the respective procedure we consider minimization of the total cost $C(I) = C_1(I) + C_2(I)$ as the optimization criterion, where C_1 is the partial direct cost due to performing experimental design for additional measurements and statistical processing of the corresponding results, and C_2 is the partial indirect or uncertainty cost, due to downward technical/economic/environmental impact (i.e., related to the end-use or the subsequent processing to the end/intermediate product under examination), especially when strict specifications should be followed, while the independent/explanatory variable I is the Reproducibility Index, standing for Inter-Laboratory Study (ILS) expenditure expressed as a percentage of total expenditure for testing materials and processes. Evidently, C_1 is an increasing function of I with an increasing rate (i.e., $dC_1/dI > 0$, $d^2C_1/dI^2 > 0$), because of the validity of the Law of Diminishing marginal or differential Returns (LDR). On the other hand, C_2 is a decreasing function of I , with an increasing algebraic or a decreasing absolute rate (i.e., $dC_2/dI < 0$, $d^2C_2/dI^2 > 0$ or $d |dC_2/dI| /dI < 0$), since the higher the I -values the lower the partial cost C_2 , because certainty increases. The optimal value I_{opt} is the abscissa of C_{min} -point, where $(C_1+C_2)_{min}$ implies $MC_1=MC_2$, meaning equality of the marginal or differential costs (i.e., $dC_1/dI = |dC_2/dI|$) at the equilibrium point of the tradeoff between C_1 and C_2 .

F. A. Batzias is with the Dep. Industrial Management & Technology, Univ. Piraeus, 80 Karaoli & Dimitriou, GR18534 Piraeus, Greece (corresponding author phone: +30-210-4142360; fax: +30-210-4142392; e-mail: fbatzi@unipi.gr).

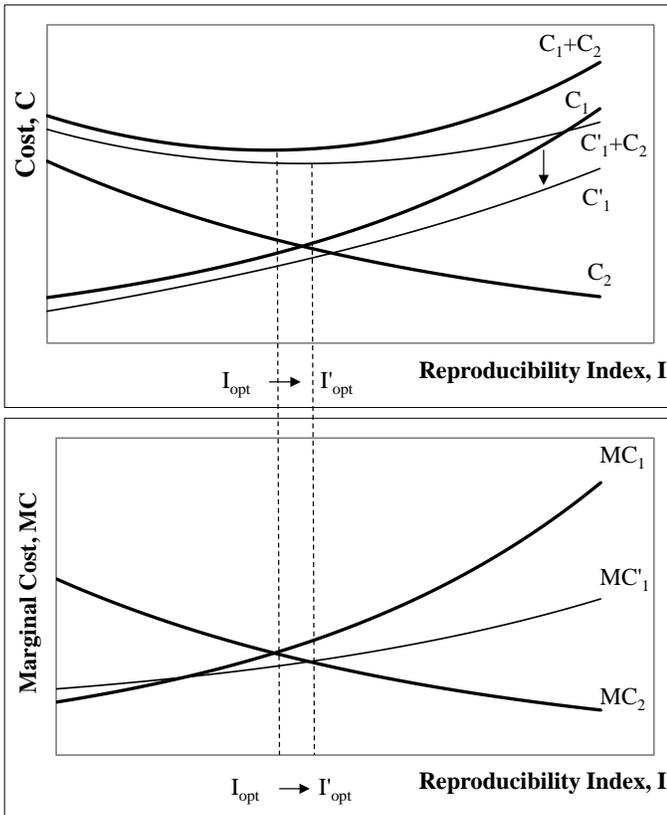


Fig. 1a. Dependence of partial direct cost C_1 (for performing experimental design) and partial indirect or uncertainty cost C_2 on I ; in case of introducing computer aided statistical processing to improve experimental design within the network of cooperating laboratories, the optimal value I_{opt} is shifting to I'_{opt} , where $I'_{opt} > I_{opt}$.

In case of introducing computer aided statistical processing for better experimental design within the network of cooperating laboratories, the C_1 -curve will move downwards to the new position C'_1 becoming more flat, since the cost decrease is expected to be more expressed in the region of higher I -values, where the partial cost is already high due to the validity of the LDR; consequently, the I_{opt} value is shifting to I'_{opt} , where $I'_{opt} > I_{opt}$, while C_{min} will move to C'_{min} , where $C'_{min} < C_{min}$, as shown in Fig. 1a. In case of specifications relaxation (as regards downward processing and/or marketability of the final product), the C_2 -curve will move downwards to the new position C'_2 becoming steeper, since the cost decrease is expected to be more expressed in the region of higher I -values, where the partial cost is already high due to the validity of the LDR; consequently, the I_{opt} value is shifting to I'_{opt} , where $I'_{opt} > I_{opt}$, while C_{min} will move to C'_{min} , where $C'_{min} < C_{min}$, as shown in Fig. 1b.

Similar movements to the ones described above of the partial cost curves C_1 and C_2 may be realized in the medium/long run as a result of 'learning by doing'. It is worthwhile noting that the vectors $(I'_{opt} - I_{opt})$ and $(I'_{opt} - I_{opt})$ have the same direction, indicating/supporting the extendability increase of the co-operating laboratories network, while the minimum total cost decreases in both cases.

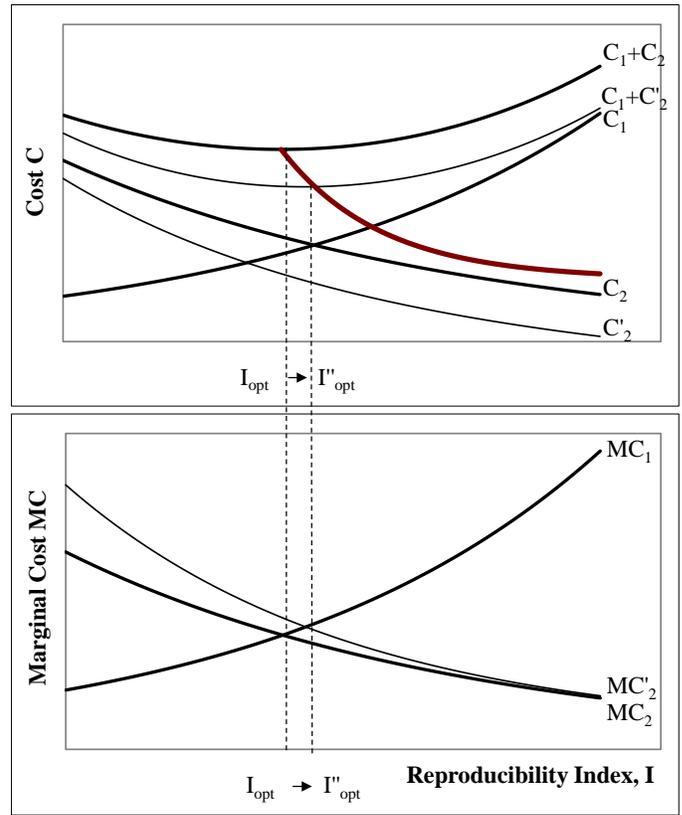


Fig. 1b. Dependence of partial direct cost C_1 (for performing experimental design) and partial indirect or uncertainty cost C_2 on I ; in case of specifications relaxation (as regards downward processing and/or marketability of the final product), the optimal value I_{opt} is shifting to I''_{opt} . The thick curve in the upper diagram is the locus of C_{min} points as a function of I_{opt} .

The conceptual optimization method (through a respective simple tradeoff) presented above is a rather static one in the sense that only certain partial costs are taken into account while total utility is not considered. Such a consideration might give a dynamic character to the optimization procedure allowing for either (i) scale up/down of the process under examination or/and (ii) information acquisition at higher granularity level as regards the properties of the material to be tested.

II METHODOLOGY

The dynamic character of the optimization procedure, mentioned in the last paragraph above has been incorporated into the following algorithm, presented as a flowchart in Fig. 2, where interconnections between activity stages and decision nodes are also shown.

1. Description of the testing method required to either examine a material or understand a phenomenon or monitor/control a process.
2. Determination of the metrological target and the consequent confidence intervals of the involved variables/parameters/coefficients (VPCs) of the involved models.
3. Determination of required precision as conceived/specified by means of repeatability and

- reproducibility conditions.
- 4. Denotation of simple one-laboratory repeatability (SOR) conditions, where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.
- 5. Denotation of horizontal multi-laboratory reproducibility (HMR) conditions, where test results are obtained with the same method on identical test items, by simulating the different laboratories with different operators using different equipment, although performing within the laboratory where SOR conditions have already been established.
- 6. Denotation of vertical simulatative reproducibility (VCR, in parallel with deterministic or probabilistic/possibilistic modelling) leading to either scale-up from laboratory to pilot-plant scale or scale-down, when a fault appears during the larger scale operation and remedial path is suggested/followed.
- 7. Reconsideration of stages 4-6 and determination of optimal VPCs values.
- 8. Examination/testing at prototype scale.
- 9. Examination/testing at industrial scale.
- 10. Determination of parameters creating a quasi-continuum between SOR and simulated HMR performing in the same laboratory.
- 11. Extension of this determination along the scale up/down.
- 12. Localization of faults.
- 13. Reconsideration of VPCs identification.
- 14. Corrective action according to these suggestions.
- 15. Determination of the relations connecting the VPCs identified so far, within the framework of a corresponding Ontology.
- 16. Bridging of any discontinuity gaps by considering additional VPCs, even at a superficial phenomenological level.
- 17. Trial operation of the process or usage of the material at the larger scale under normal and extreme (i.e., expected frequently and rarely, respectively) conditions.
- 18. Investigation for revealing the possible causes of these faults.
- 19. Synthesis of a fault tree, setting each observed fault as top event, following a top-down/deductive method.
- 20. Selection of experts to confirm/enrich/modify the dendritic structure and assign significance indices (following a bottom-up/inductive method) on the causal connections in a fuzzy version to count for uncertainty.
- 21. Determination/identification (after defuzzification) of the event-paths most responsible for the existence of each top event.

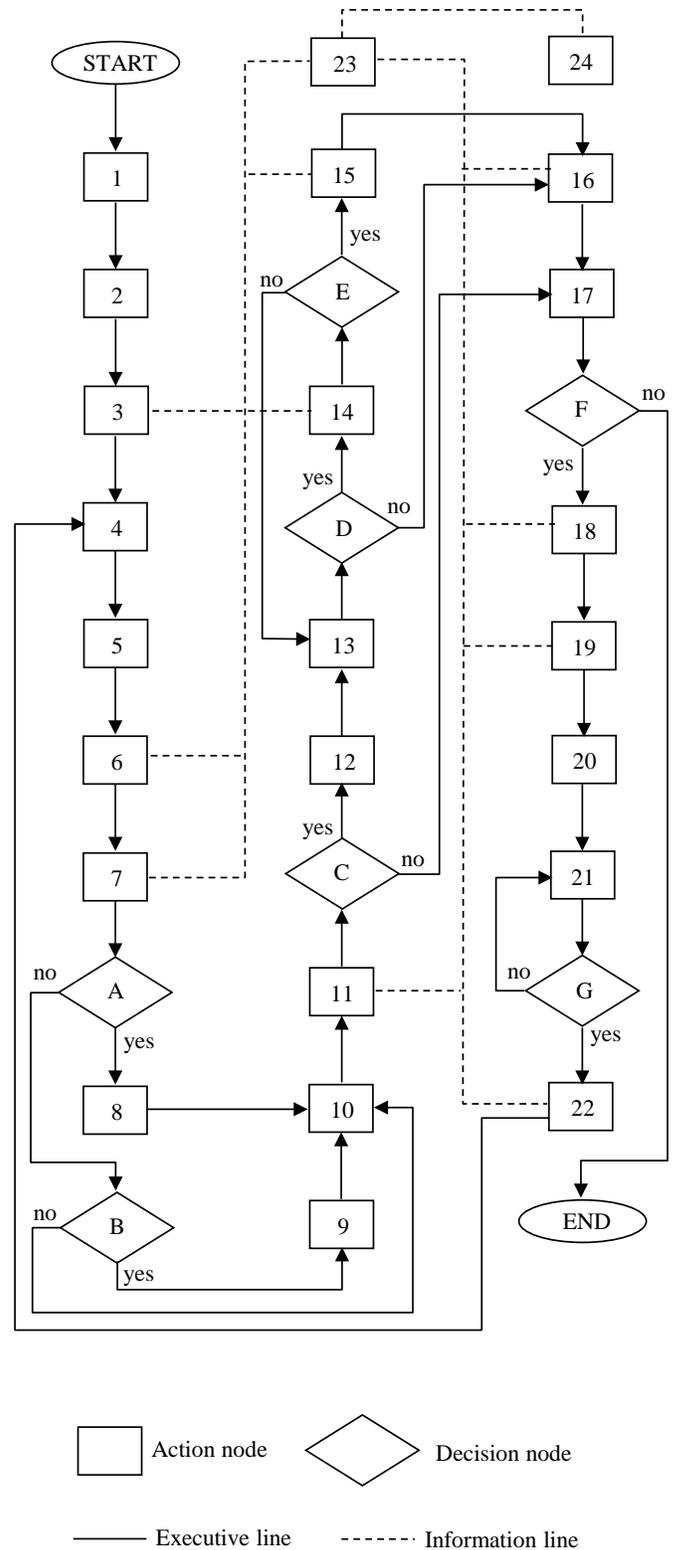


Fig. 2. Flowchart of the algorithmic procedure developed for (i) creating a continuum between repeatability and reproducibility, and (ii) extending this methodological framework to cover process scale up/down and material properties examination.

22. Design/development of procedures to obtain the required additional knowledge.
23. Knowledge processing within the internal KB.
24. Searching in external KBs by means of an Intelligent Agent (IA), according to [8].

- A. Are laboratory facilities available for testing at prototype level?
- B. Are laboratory facilities available for testing at industrial level?
- C. Are there discontinuities in these domains?
- D. Are there any suggestions for VPCs splitting/merging/modification made as a result of this reconsideration?
- E. Is the corrective action satisfactory?
- F. Are further faults observed or even suspected to appear later on?
- G. Is this identification satisfactory?

III IMPLEMENTATION

The difference between reproducibility and repeatability conditions may due to the latent/hidden change of a parameter value that occurs either in the laboratory environment (e.g., temperature, humidity, light intensity) or within an apparatus (like chemical reactor, adsorption column, sedimentation or flotation vessel) used for testing. Such a parameter value variation, within a preset range, is reasonable for a laboratory planned to serve a specific purpose at minimum cost (i.e., with devices/equipment, as simple as possible [9]). The change of this parameter value causes (in its turn) a corresponding change either (i) in the dependent variable of the model that simulates the phenomenon taking place in the process under examination or (ii) in the properties of the material to be tested. Obviously, the model continuity is closely related with the continuum we are trying to create/reveal and show forth herein. Consequently, the case examples, selected for implementing certain significant stages of the algorithmic procedure quoted in the Methodology Section, are emphatically expressing the continuity characteristic mentioned above.

The first case example is the mass balance (rate of accumulation = input – output – consumption) over a continuous flow stirred tank reactor (CFSTR) where a (bio)chemical reaction of n^{th} order takes place, giving the following model

$$VdS/dt = QS_0 - QS - kS^n V \quad (1)$$

where

S is the substance concentration in tank and discharge,
 S_0 is the substance concentration in input,
 Q is the input volumetric rate into and out of tank,
 V is the constant volume of liquid in the tank,
 k is the rate constant.

For $n=1$, this is a first-order linear differential equation, which can be rearranged to give the more familiar standard form $dS/dt + \alpha S = (Q/V)S_0$, where $\alpha = Q/V + k$, that has the general analytic solution,

$$S = (Q/V) \exp(-at) \int_0^t S_0 \exp(at) dt + S_i \exp(-at) \quad (2)$$

where S_i is the initial value of S at $t=0$.

For S_0 independent of t (i.e., constant input concentration)

$$S = S_0 Q / (aV) [1 - \exp(-at)] + S_i \exp(-at) \quad (3)$$

with the limiting value for $t \rightarrow \infty$

$$S = S_0 / (1 + kV/Q) \quad (4)$$

For steady state conditions (i.e., $dS/dt=0$), the last expression is also valid, although the previous Eqs. (2,3) are not valid, since they are based on the hidden assumption $dS/dt \neq 0$. Indeed, Eq. (1) gives $0 = QS_0 - QS - kSV$ for $n=1$, from which Eq. (4) is obtained, while other simple models can be found [10,11]. On the other hand, the respective kinetic equation for an ideally operating Plug Flow Reactor (PFR) is given by $S/S_0 = \exp(-kt)$, where t is the retention time in the reactor, a concept that has a rather statistical spatiotemporal meaning in the case of an ideal CFSTR and a rather clock time meaning in the case of an ideal PFR.

An ideal PFR exhibits higher efficiency E , given as $E = (S_0 - S)/S_0$ or $E = 1 - S/S_0$, in comparison with an ideal CFSTR of equal V , as we can show by expanding e^x in power series:

$\exp(x) = 1 + x + x^2/2! + x^3/3! + \dots$, where $x = kV/Q$; consequently, $\exp(x) > 1 + x$ or $1/(1+x) > \exp(-x)$ or $1/(1+kV/Q) > \exp(-kV/Q)$ or $E_{\text{PFR}} > E_{\text{CFSTR}}$.

In practice, a PFR is not operating ideally because of axial dispersion of the substance under (bio)chemical processing, tending to partially behave as a CFSTR, as such dispersion is increasing. For first order kinetics, the respective kinetic equation has the form

$$\frac{S}{S_0} = \frac{\exp[1/(2d)]}{(1+b)^2 \exp[b/(2d)] - (1-b)^2 \exp[-b/(2d)]} \quad (5)$$

where $b = (1 + 4ktd)$, dimensionless,

$d = D/(uL)$, the reactor dispersion number (dimensionless) characterizing the mixing condition in the reactor, u = velocity along the reactor length L . For $d \rightarrow 0$ or $d \rightarrow \infty$, the reactor simulates the PFR or the CFSTR, respectively. Consequently, the parameter d can be used as a basis for creating a quasi-continuum, especially in the case of scale-up, where reproducibility is playing a key-role under VCR conditions (see stage 6 of the flowchart in Fig. 2).

For other kinetic expressions (i.e., differential kinetic equations in the original/mechanismic form with $n \neq 1$) non-analytic solutions are obtained, consequently numerical analysis methods should be followed. Attempting to avoid such complicate modelling, we have devised a probabilistic/statistical method by weighting the two ideal models through a factor w , as follows

$$S/S_0 = w[\exp(-kt)] + (1-w)(1+kV/Q) \quad (6)$$

Evidently, the higher the w -value ($0 < w < 1$) the closer the reactor behavior to that of a PFR; consequently, the parameter w , having its value estimated by applying non-linear regression on data, plays (in a simpler way) the role of d in creating the intended quasi-continuum platform.

The subsequent case example refers to the application of the

combined Langmuir-Freundlich (LF) adsorption isotherm, which has the form

$$q=Q(K_a C)^n / [(K_a C)^n + 1] \quad (7)$$

where q is the amount of polluting species adsorbate at equilibrium, measured as mass per mass ratio, Q is the adsorption capacity of the system, estimated similarly, C is the aqueous phase concentration at equilibrium, measured as mass per volume, K is the affinity constant for adsorption, measured as volume per mass (therefore the product $K_a C$ is dimensionless, n is the index of adsorbent surface heterogeneity, representing the parameter that may be used to create the intended quasi-continuum by controlling the adsorbent production conditions, which are responsible for the adsorptive properties of the product: for $n \rightarrow 1$ or $n \rightarrow 0$ the expression (7) is reduced to the Langmuir or Freundlich isotherm, respectively, giving relatively satisfactory results, according to the Principle of Simplicity; otherwise, the complete/complex LF function should be used, representing the continuum under investigation.

Although, the parameter used to establish the intended quasi-continuum seems to be of pure physico-chemical nature, its formalism is closely related to the following weighted integral [12-14]

$$q = \int_{-\infty}^{+\infty} g(K_a) \frac{K_a C}{1 + K_a C} dK_a \quad (8)$$

where the statistical density function $g(K_a)$ stands for the individual site density values of elementary isotherms with the corresponding affinity constant K_a . Actually, $g(K_a)$ distribution reduces to the Dirac's delta function for Langmuir isotherm, while it resembles a log-normal distribution for Freundlich isotherm.

A similar situation is met in cases where a two-parameter statistical distribution is used (e.g., the Weibull one, like in [15]) and we convert it to the corresponding three-parameter distribution: in such cases, the additional parameter defines the intended continuum, since the condition change in simulating reproducibility is related as an independent/explanatory variable to this parameter value, by means of a continuous function.

A similar case example is the examination of the adsorption process in a fixed bed column packed with adsorbent, simulating a pollutant removal from aquatic solution. The kinetic study of this process can be performed by adopting the Bohart-Adams model [16],

$$\ln\left(\frac{C_i}{C} - 1\right) = \frac{KNx}{u} - KC_i t \quad (9)$$

in which C = effluent concentration; C_i = influent concentration; K = an adsorption rate coefficient; N = an adsorption capacity coefficient; x = bed depth; u = linear velocity; and t = time.

This model can be rearranged to give $C/C_i = [1 + \exp(\alpha - bt)]^{-1}$, where $a = KNx/u$ and $b = KC_i$.

Since this expression implies an S-shapes curve, which is symmetrical round the midpoint with coordinates $t = a/b$ and $C = C_i/2$, while this restriction does not reflect reality, a

relaxation is necessary by introducing an exponential parameter v to count for asymmetry:

$$C/C_i = [1 + \exp(\alpha - bt)]^{-v} \quad (10)$$

while other asymmetrical forms of the same category of sigmoid curves, having as common feature the reduction to symmetrical form for $v=1$, can be found in [17].

Clark [18] gave a physico-chemical meaning to v through the relation $v=1/(n-1)$, $n \neq 1$, where n is a measure of adsorbent surface heterogeneity, obtained from the Freundlich isotherm, (related also to the distribution of Gibbs free energy, which, in its turn, is a function of enthalpy, entropy and temperature). In relevant measurements performed in our Laboratory [19-21], we have estimated all parameter-values following a two-step procedure: firstly by applying the Freundlich isotherm and subsequently by applying the Bohart - Adams - Clark model with fixed the v -value. This approach contributes to parameter identification but may give an increased total standard error of estimate in comparison with a single step approach (i.e., without involving the Freundlich isotherm into the parameter values estimation procedure). Evidently, measurements under simulated/controlled reproducibility conditions may contribute decisively to (i) clarify the role of each pollutant in the adsorption from an aquatic mixture, (ii) reveal the mechanism of competitive adsorption, (iii) investigate interaction between pollutants, (iv) enhance the parameter identification approach.

If we are not certain about the validity of a model, as it is the case in VCR (see stage 6 in the Methodology Section and the flowchart in Fig. 2) when performing scale up/down, we can apply experimental design in combination with dimensional analysis (see [22]), and use at least one of the parameters inter-relating the dimensionless groups, in order to define the intended continuum.

If there is not a model providing VPCs appropriate to create the intended continuum, we can use directly the standard definition of reproducibility conditions and simulate these conditions within the same laboratory where measurements under repeatability conditions are already performed. The advantage of this simulation method is that we can change one factor each time we want to examine its impact on any dependent variable; moreover, we can (i) incorporate this procedure into an R&D project performing scale up/down, and (ii) decrease ILS cost, and obtain independence as regards outsourcing. The disadvantages concern (i) the danger to limit examination ranges, as result of bias, and (ii) the inability to perform arbitration closes, a condition usually met in contracts/agreements between providers and users. In practice, the second disadvantage is actually turned to advantage in the sense that such a simulation might be used for estimation *a priori* or checking *a posteriori* possible results obtained by an external laboratory involved in an arbitration procedure.

Subsequently, we present a case example based on thickness measurements of the anodic coating of aluminum (formed in the 15% sulphuric acid as electrolyte at direct current density 2000 Cb/dm² and 20°C and dissolved chemically within a bath made of d=1.75 phosphoric acid 35 ml/L and chromic acid 20 g/L, determined as CrO₃) performed by using the classic gravimetric method. The results under real

reproducibility conditions are presented in table I, together with the corresponding one-way (or single factor) ANOVA, where M_{ij} is the measurement j in sample S_i , M_i is the arithmetic mean in Sample i , V_i is the variance in Sample i , SV is the source of variation, SS is the sum of squares of residuals, either between the samples (BS) or within the samples (WS), df is the number of degrees of freedom, MS is the mean square of residuals, F_e is the estimated F-value, P_e is the estimated probability or significance level, F_c is the critical one-tailed F-value at significance level 0.05 or 5%. These results indicate that the null hypothesis (that there is no significant difference between intra- and inter-laboratory examination) is rejected. On the contrary, the results obtained under simulated reproducibility conditions (obtained within the same laboratory by changing only the operators measuring the samples S_3 and S_4) indicate that the null hypothesis is not rejected, implying the existence of a quasi-continuum between real repeatability and real reproducibility through simulation. Therefore, we can conclude about the influence each factor is exerting on the material or process under examination in the same mode we determine the impact on the depended variable by changing the value of an independent/explanatory variable within a function, even if this function has not been explicitly expressed. In case that such an expression has been achieved, the intended creation of the continuity domain is simply represented by the partial derivative of the function as regards the VPC changing its value (i.e., a concept that coincides with performing mono-parametric sensitivity analysis).

The ANOVA results may also contribute to the quantification of the metrological impact caused by changing the expected value of a factor from laboratory to laboratory or from condition to condition. In the case example analyzed above, the success of the gravimetric examination depends heavily on the effectiveness of the reagents mixture that will dissolve the anodic film if aluminum without attacking the basic metal.

Table I. One-way ANOVA for the results obtained under real reproducibility conditions (different laboratories, same method, different equipment and operators): the null hypothesis for statistical testing is rejected, since $F_e > F_c$, $P < 0.05$.

S_i	M_{i1}	M_{i2}	M_{i3}	M_i	$V_i 10^2$
S_1	10.3	10.0	10.2	10.17	2.33
S_2	10.1	10.5	10.3	10.30	4.00
S_3	9.4	9.3	9.8	9.50	7.00
S_4	9.1	8.9	9.2	9.07	2.33

SV	SS	df	MS	F_e	P	F_c
BS	3.016	3	1.005			
WS	0.313	8	0.039			
Total	3.329	11				
Rest				25.667	18610^{-6}	4.066

Table II. One-way ANOVA for the results obtained under simulated reproducibility conditions (same laboratory, same method, same equipment but different operators for measuring the samples S_3, S_4): the null hypothesis for statistical testing is not rejected, since $F_e < F_c$, $P > 0.05$.

S_i	M_{i1}	M_{i2}	M_{i3}	M_i	$V_i 10^2$
S_1	10.3	10.0	10.2	10.17	2.33
S_2	10.1	10.5	10.3	10.30	4.00
S_3	9.8	10.1	9.9	9.93	2.33
S_4	10.0	9.8	10.2	10.00	4.00

SV	SS	df	MS	F_e	P	F_c
BS	0.247	3	0.082			
WS	0.253	8	0.032			
Total	0.500	11				
Rest				2.596	0.125	4.066

For saving time (i.e., capital cost, since the equipment depreciation does not change significantly over time) a strong mixture used at relatively high temperature should apply. During the dissolution procedure, a whitish appearance on withdrawing from the bath indicates that the anodic coating removal is not complete and the procedure must be continued until the original metallic surface is revealed. The sample is weighted before and after dissolution; the difference m is used to provide the film thickness $T=m/(A\rho)$, where A is the apparent surface area and ρ is the density of the anodic film, estimated with an error of $\pm 2.5\%$.

If greater accuracy is required, the apparent density (including any air filled pores) can be estimated by carrying out a weight loss determination in a coating that has had its thickness measured by a more precise (and more expensive, without mentioning equipment availability) method [23], like (i) the microscopic section or (ii) the split-beam microscope or (iii) eddy current, or (iv) the beta backscattering method.

Apart from the appearance of the specimen after film removal, the completion of the stripping operation can be checked by repeating the immersion in phosphoric/chromic acid for a very short time interval followed by drying and re-weighting, stopping when no further loss of weight is observed. Nevertheless, on some alloys, particularly those containing zinc and/or copper, a constant weight is not achieved due to continuing slow attack by the acid on the basis metal. In such a case, the suggested immersion time should be restricted to the shortest time for the visible removal of the coating.

Obviously, estimation through visual observation (in combination with the measuring equipment used) is rather subjective, heavily depended on the operator's experience and his/her tacit knowledge about the behavior of the material under examination. Performing measurements under simulated reproducibility conditions in the same laboratory, we can categorize the personnel according to their (i) experience through 'learning by doing', (ii) explicit knowledge background, (iii) skillfulness, especially in copying with

difficult situations, (iv) typical qualifications, and (v) familiarity with the equipment currently in use. By progressively changing quantitatively and/or qualitatively each factor we create/direct the intended quasi-continuum while applying steady state conditions to keep the examined process under continuous control.

IV DISCUSSION

In an ontological approach, instead of using cost, we can use the broader concept of ‘utility’ U , as the dependent variable within a corresponding tradeoff in order to determine I_{opt} . Evidently, I_{opt} is determined as the abscissa of the $U_{max}=(U_1+U_2)_{max}$ point, where U_1 and U_2 represent ILS reliability/validity (or, equivalently, consistency/suitability) and expenditure for the material/process under examination, respectively. U_1 is an increasing function of I with a decreasing rate (i.e., $dU_1/dI>0, d^2U_1/dI^2<0$) because of the validity of the LDR, as regards reliability. On the contrary, U_2 is a decreasing function of I with a decreasing algebraic or an increasing absolute rate (i.e., $dU_2/dI<0, d^2U_2/dI^2<0$ or $d|dU_2/dI|/dI>0$), since the higher the I -values, the higher the corresponding expenditure and, consequently, the lower the utility. At U_{max} (i.e., at the equilibrium point of the tradeoff between U_1 and U_2), we have $MU_1=MU_2$, denoting equality of marginal or differential utilities $MU_1=dU_1/dI$ and $MU_2=|dU_2/dI|$.

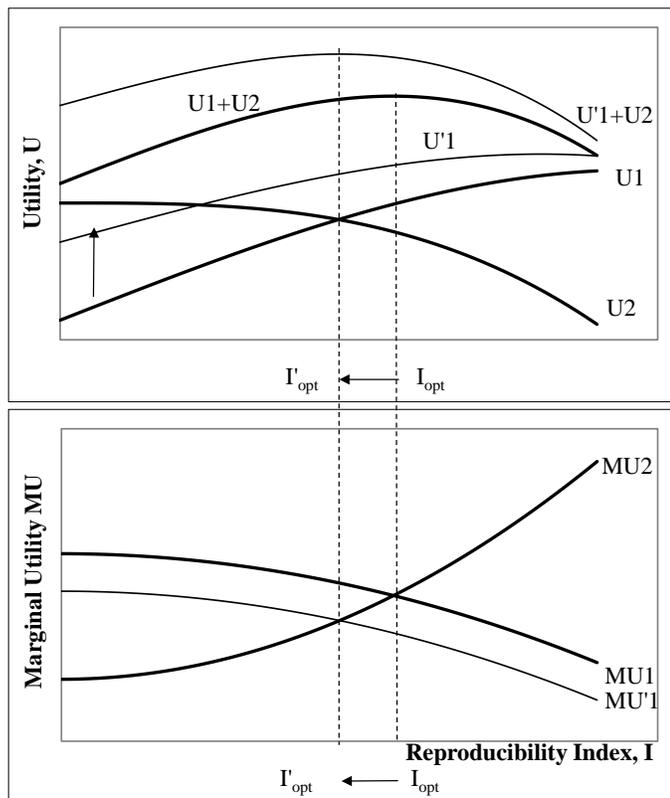


Fig. 3a. Dependence of the partial utilities U_1 and U_2 (representing ILS reliability/validity and expenditure for the material/process under examination, respectively) on I ; by introducing appropriate software based on the algorithmic procedures suggested herein, the optimal value I_{opt} is expected to decrease while U_{max} will increase.

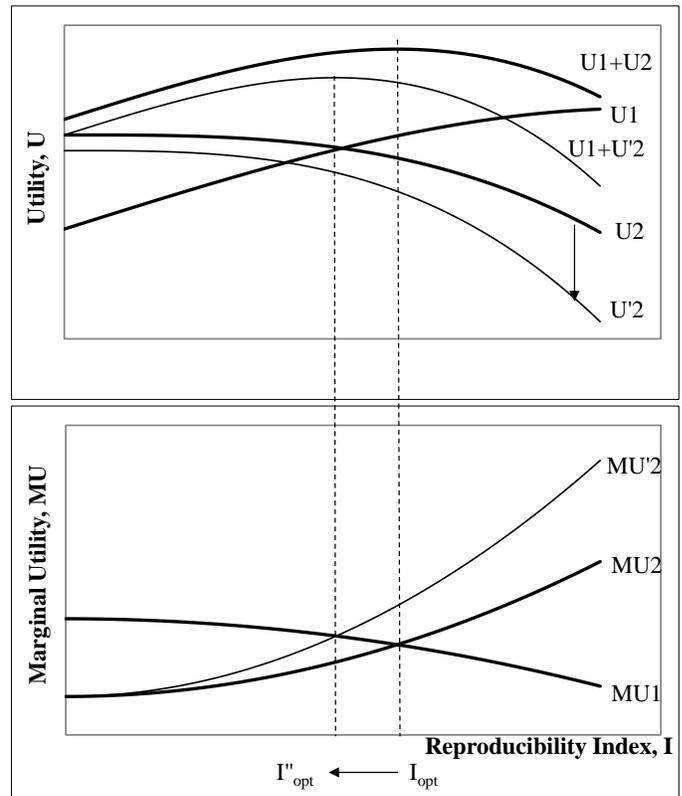


Fig. 3b. Dependence of the partial utilities U_1 and U_2 (representing ILS reliability/validity and expenditure for the material/process under examination, respectively) on I ; during a period of inflation the optimal value I_{opt} is expected to decrease and U_{max} will decrease too.

By introducing appropriate software based on the algorithmic procedures suggested herein, especially designed to enhance system reliability, the U_1 -curve moves upwards to its new position U_1' becoming more flat, since the fewer the inter-laboratory measurements (at the initial stage of the tradeoff design) the higher the imprecision; consequently, the I_{opt} -value is shifting to I'_{opt} , where $I'_{opt}<I_{opt}$, while U_{max} will move to U'_{max} , where $U'_{max}>U_{max}$, as shown in Fig. 3a. Moreover, during a period of inflation, the expenditure for outsourcing is increasing, resulting to the downward movement of the U_2 -curve to its new position U_2' becoming steeper, since the utility difference (in comparison with the corresponding initial curve) is higher in the region of high I -values because of the plethora of inter-laboratory measurements that correspond to this region; as a result, the I_{opt} -value is shifting to I''_{opt} , where $I''_{opt}<I_{opt}$, while U_{max} will move downwards.

It is worthwhile noting that the vectors $(I'_{opt}-I_{opt})$ and $(I''_{opt}-I_{opt})$ have the same direction, indicating/supporting the limitation of the co-operating laboratories network, in contrast to the conclusion reached by using special versions of direct and indirect partial costs in the Introductory Analysis Section. On the other hand, there is an uncertainty about the direction to which U_{max} will move on, depending on the number of measurements performed in the cooperating laboratories and the necessary information granularity level of the ILS results. Regularly, an ILS should include 30 or more laboratories [1],

a number which is also used as the minimum of measurements taken into account to confirm goodness of fitting of normal distribution to data, by means of the chi-squared test. Nevertheless, the achievement of this number is not usually feasible and a network with fewer laboratories (but not less than 6, meaning that an ILS should start with at least 8 laboratories in order to allow for attrition) may be used, provided that reasonable estimates concerning reproducibility are finally approximated.

It should be noticed that the relevant Ontology of standards is continually enriched, following a non-linear pattern, as shown in Fig. 4, where the number of revisions of the basic or 'root' standard E 691 is represented as a function of time.

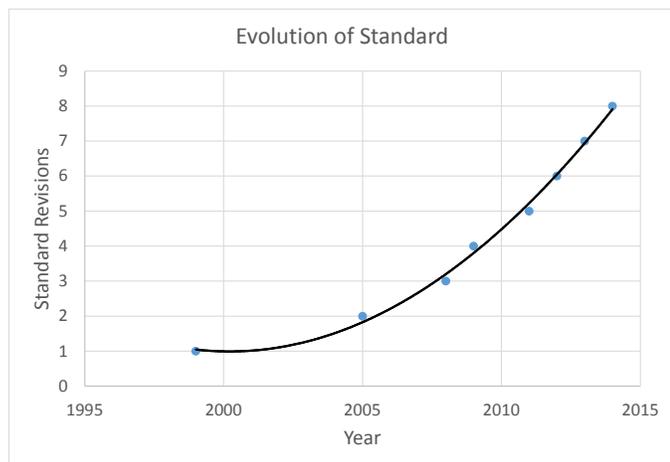


Fig. 4. The evolution of standard practice E691 'Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method'.

This quantitative result is accompanied by a qualitative one: the number of referenced documents of this Standard is steadily increasing, even when a referenced standard pre-existed in the relevant thesaurus, before being assigned/recognized as a referenced one (i.e., as an essential for understanding or a prerequisite for the realization of E 691); such examples are the E 456-06 and the E 1402-13 [24] standards, referring to Quality & Statistics terminology and Sampling, respectively, indicating the worldwide interest on establishing a standard controlled vocabulary, which is a basic characteristic of a formal Ontology. The creation of the quasi-continuum between repeatability and reproducibility hopefully (i) contributes to the enrichment of such a controlled vocabulary and (ii) enhances the density and consistency of the respective network of relevant concepts, as they are inter-related through explicit/implicit functions and/or tacit/subjective knowledge.

V CONCLUSIONS

The functionality of the methodological framework we have designed under the form of an algorithmic procedure has been proved in both cases: the development/control of an industrial process (possibly simulating also the behavior of waterbodies in the aquatic environment) and the examination of materials of practical interest. In the first case, where mathematical

modelling is feasible, the created continuum is based on parameters representing homogeneity, which decreases in scaling up. In the second case, where mathematical modelling is frequently not feasible at a surface phenomenological level, the created continuum is based on numerical grid construction making use of statistical experimental design (mainly in various forms of Analysis Of VAriance – ANOVA).

In the first case, numerical analysis methods may be indispensable to cope with practical problems usually met in the real world; e.g., in the plug flow reactor with axial dispersion we consider herein, there is analytical solution only for the first order chemical kinetics and only as regards efficiency, while even the process continuity parameter cannot be obtained explicitly. In the second case, precision, being evaluated through repeatability and reproducibility (according to the ASTM standard E 691) depends on the density of the grid mentioned above.

It is worthwhile noting that the creation of the continuum suggested herein contributes to resources (including time) saving, laboratories network minimization, and equipment availability maximization (i.e., economotechnical optimization), facilitating also the performance of relevant Research & Development (R&D) projects by means of (i) dynamic programming during scale up/down, and (ii) examining the quality of produced goods, according to specifications set *a priori*.

ACKNOWLEDGEMENTS

Financial support by the Research Centre of the University of Piraeus is kindly acknowledged.

REFERENCES

- [1] ASTM Standard: E691-14 Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method, Vol. 14.02
- [2] ASTM Standard: E177-14 Practice for Use of the Terms Precision and Bias in ASTM Test Methods, Vol. 14.02
- [3] ASTM Standard: E456-13ae2 Terminology Relating to Quality and Statistics, Vol. 14.02
- [4] ASTM Standard: E1169-14 Practice for Conducting Ruggedness Tests, Vol. 14.02
- [5] ASTM Standard: E2282-14 Guide for Defining the Test Result of a Test Method G. S. Bohart, E. Q. Adams, "Adsorption in columns", *J. Chem. Soc.*, Vol. 42, pp. 523–544, 1920, Vol. 14.02
- [6] M. Gierczak, "The Qualitative Risk Assessment of MINI, MIDI and MAXI Horizontal Directional Drilling Projects", *Tunneling and Underground Space Technology*, Vol. 44, September 2014, pp.148–156.
- [7] K. Ishikawa, "Guide to Quality Control", *Juse*, Tokyo, 1968.
- [8] F.A. Batzias, E.C. Marcoulaki, "Restructuring the keywords interface to enhance CAPE knowledge via an intelligent agent", *Computer Aided Chemical Engineering*, Vol. 10, pp. 829–834, 2002.
- [9] J. Ponce, J. Liceaga, I. Siller, E. Zamudio, "Humidity and Temperature Control of a Metrology Laboratory", in Proc. of Intern. Conf. on Energy, Environment, Development and Economics, also appearing in *Energy, Environment and Structural Engineering Series vol. 38*, Eds. A. Bulucea et al., Zakynthos, Greece, July 2015, pp 150-154.
- [10] F. Batzias, I. Salapa, C. Siontorou, "On the Tradeoff between Reliability and Uncertainty when Combining Bioreactors for Wastewater Treatment", in Proc. of 10th WSEAS Intern. Conf. on Environment, Ecosystems and Development, Eds. S. Oprisan et al., Montreux, Switzerland, Dec. 2012, pp 181-186.
- [11] M. Hadjigeorgiou, V. Gekas, "Applying Mathematical Analysis in Biosciences", in Proc. of Intern. Conf. on Bioscience and

- Bioinformatics, Eds. N. Mastorakis, V. Mladenov, M. Demiralp, Z. Bojkovic, Vouliagmeni, Athens, Greece, Dec. 2010, pp 131-135.
- [12] Limousin et al., "Sorption isotherms: A review on physical bases, modeling and measurement" *Applied Geochemistry*, 22 (2007), pp. 249–275.
- [13] Turiel et al., 2003 E. Turiel, C. Perez-Conde, A. Martin-Esteban "Assessment of the cross-reactivity and binding sites characterisation of a propazine-imprinted polymer using the Langmuir-Freundlich isotherm" *The Analyst*, 128 (2) (2003), pp. 137–141.
- [14] Umpleby et al., 2001 R.J. Umpleby, S.C. Baxter, Y. Chen, R.N. Shah, K.D. Shimizu, "Characterization of Molecularly Imprinted Polymers with the Langmuir-Freundlich Isotherm", *Analytical Chemistry*, 73 (19) (2001), pp. 4584–4591.
- [15] C. Guarnaccia, J. Quartieri, N. Mastorakis, C. Tepedino, "Analysis of Noise Level Exceedances by Exponential Rate Function in Non-Homogenous Poisson Model", in *Proc. of Intern. Conf. on Energy, Environment, Development and Economics, also appearing in Energy, Environment and Structural Engineering Series vol. 38*, Eds. A. Bulucea et al., Zakynthos, Greece, July 2015, pp 335-339.
- [16] G.S. Bohart, E.Q. Adams, "Adsorption in columns", *J. Chem. Soc.*, 42 (1920), pp. 523–544.
- [17] F. Batzias, Z. Res, "Decision Making", *Hellenic Open University, Advanced Quantitative Methods for Managers*, Vol. 3, Patras, Greece, 2005, pp. 311-317.
- [18] R. M. Clark, "Modeling TOC removal by GAC: the general logistic function", *J. Am. Water Works Assoc.*, vol. 79, no. 1, pp. 33–131, 1987.
- [19] D. Sidiras, F. Batzias, E. Schroeder, R. Ranjan, M. Tsapatsis, Dye adsorption on autohydrolyzed pine sawdust in batch and fixed-bed systems, *Chemical Engineering Journal*, Volume 171, Issue 3, 15 July 2011, Pages 883-896.
- [20] F.A. Batzias, D.K. Sidiras, Dye adsorption by prehydrolysed beech sawdust in batch and fixed-bed systems, *Bioresource Technology*, Volume 98, Issue 6, April 2007, Pages 1208-1217.
- [21] F. Batzias, D. Sidiras, E. Schroeder, C. Weber, Simulation of dye adsorption on hydrolyzed wheat straw in batch and fixed-bed systems, *Chemical Engineering Journal*, Volume 148, Issues 2–3, 15 May 2009, Pages 459-472.
- [22] D. Sidiras, A. Bountri, F. Batzias, "Modeling an Adsorption Column for Wastewater Treatment by Using Dimensionless Groups in Scale-Up Procedure", in *Proc. of Intern. Conf. 8th WSEAS on Environment, Ecosystems and Development*, Eds. N. Mastorakis, V. Mladenov, M. Demiralp, Z. Bojkovic, Vouliagmeni, Athens, Greece, Dec. 2010, pp 234-239.
- [23] V. F. Henley, "Anodic Oxidation of Aluminium & its Alloys", *Pergamon Press*, 1982, pp 100-107.
- [24] ASTM Standard: E1402-13 Standard Guide for Sampling Design, Vol. 14.02

An Incremental Phylogenetic Tree Algorithm Based on Repeated Insertions of Species

Peter Z. Revesz, Zhiqiang Li

Abstract—In this paper, we introduce a new phylogenetic tree algorithm that generates phylogenetic trees by repeatedly inserting species one-by-one. The incremental phylogenetic tree algorithm can work on proteins or DNA sequences. Computer experiments show that the new algorithm is better than the commonly used UPGMA and Neighbor Joining algorithms.

Keywords—Data structure, Distance matrix, Phylogenetic tree, Protein.

I. INTRODUCTION

CURRENT phylogenetic tree construction algorithms [1]-[3], [5], [9], [11] are not incremental and have to be rerun from the beginning whenever a new species is added to the database. Moreover, a rerun from the beginning is necessary even if the new species is aligned with the already used species. In this paper, we develop an incremental algorithm that inserts new species one-by-one into a growing phylogenetic tree.

Our inspiration for such an incremental phylogenetic algorithm is the way biologists usually classify any newly discovered species. Starting from the root node of the existing classification tree, the newly discovered species is compared with existing species and always an appropriate branch is chosen to go one level down in the classification hierarchy. Eventually we reach one of the existing species, which is the closest relative. It is next to that nearest relative where the new species is normally inserted.

Our aim is to develop a computer algorithm that uses the above paradigm but works with both DNA sequences and proteins. As the genomes of a growing number of species are sequenced and become part of DNA and protein databases [4], [12], molecular biology increasingly augments, although not completely replaces, morphological considerations.

Reliable phylogenetic tree constructions are needed for a diverse set of studies, including theoretical studies on the rate of evolution in various phyla [10] and applied studies aimed at developing medical diagnosis methods [6] and pharmaceutical development.

Our algorithm has two main benefits compared to previous algorithms:

Peter Z. Revesz is with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA (revesz@cse.unl.edu).

Zhiqiang Li is with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA (zli@cse.unl.edu).

- 1) *Faster* because it can be used incrementally if the new sequence is aligned with the other sequences.
- 2) Generates *more accurate* phylogenetic trees as indicated by the computer experiments presented in Section 4.

This paper is organized as follows. Section II presents some related work. Section III describes the incremental phylogenetic tree algorithm. Section IV presents some experimental results. Finally Section V gives some conclusions and directions for future work.

II. RELATED WORK

The *UPGMA* [11] and the *Neighbor Joining* [9] algorithms are commonly used and familiar to most users. The maximum likelihood method is also well known, although it seems less frequently used than UPGMA and Neighbor Joining in practice because it requires more computational time. All of these algorithms are reviewed in textbooks, such as [1]-[3].

Revesz [5] introduced the *Common Mutations Similarity Matrix* algorithm, which has $O(n^3)$ time complexity, where n is the number of sequences. We briefly review this algorithm as a related work, which will also be used in the experimental results section of this paper.

Table 1 below shows seven DNA sequences, $S_1 \dots S_7$, each with a length fifteen nucleotides displayed by groups of five nucleotides per column.

Table 1 Seven input DNA sequences and a common ancestor μ

S1	AGCTA	CTAGT	AATCA
S2	AGCTA	CGAGT	AATCA
S3	ATCCA	CTAGT	ACACT
S4	ATCCA	CTAGT	ATACT
S5	CGGTA	TTTGT	AAGCT
S6	CGGTT	CATCA	AATGC
S7	AGGTA	CTTGA	AATCC
μ	AGCTA	CTAGT	AATCT

Let $S_i[k]$ denote the k th nucleotide of S_i . The *Hamming distance* between two DNA sequences S_i and S_j each with length n , denoted $\delta(S_i, S_j)$, is defined as the number of corresponding nucleotide pairs that are different, that is, $\sum_{1 \leq k \leq n} S_i[k] \neq S_j[k]$. μ is the common ancestor of seven sequences.

Evolutionary tree construction algorithms generally start

from a *Hamming distance matrix* to recursively combine pairs of sequences (rows and columns) until only a single combined sequence remains. For example, the UPGMA (unweighted pair group method with arithmetic mean) [11] method would always search for the closest pairs to combine. When several pairs are equally distant, then an arbitrary choice is made. In this case, the closest pairs are S1 and S2 and S3 and S4 because $\delta(S1, S2) = 1$ and $\delta(S3, S4) = 1$. The Neighbor Joining [9] method is a more sophisticated and commonly used method that is also based on distance matrices.

Instead of distance matrices, Revesz [5] introduced a *common mutations similarity matrix (CMSM)*. The motivation behind looking for common mutations is that in practice rare but shared features, such as rare mutations, often provide useful markers of similarity among a set of closely related items. Moreover, if mutations are rare, then it may be more efficient to count their occurrences than finding the Hamming distances for long sequences. Assuming that the seven DNA sequences in Table 1 are related, we can find the most likely common ancestor sequence, denoted μ , as the mode of each column. If there is no most frequent nucleotide in a column, then we arbitrarily chose one of the most frequent nucleotides in it.

The Common Mutations Similarity Matrix (CMSM) algorithm records for each pair of sequences the mutations that they share in common with respect to a global average μ , which is taken as the most likely common ancestor sequence.

Example 1. Given seven nucleotide sequences in Table 1 below (rows S1 to S7 where the sequences are displayed in groups of five), the common ancestor sequence μ is calculated in [5] as the most frequent in each column.

Alternatively, if S1 to S7 are considered amino acid sequences where A, C, G and T now stand for the amino acids Alanine, Cysteine, Glycine and Threonine, respectively, then the common ancestor sequence μ can be defined as in each column as the amino acid x out of the set S of twenty amino acids used in most proteins such that x is overall closest to the set of amino acids in that column. We make this statement more precise below using as an example the PAM250 amino acid similarity matrix. Let

$$PAM250[AminoAcid1, AminoAcid2] = a \quad (1)$$

denotes that AminoAcid1 and AminoAcid2 have a similarity score of a . For example, $PAM250[A, G] = 1$ means that Alanine and Glycine are slightly similar to each other. Then for the i th column,

$$\mu[i] = x \in S \quad (2)$$

such that

$$\sum_{j=1}^7 PAM250[S_i[j], x] \quad (3)$$

is maximum.

For example, we can see that the value of $\mu[1]$ changed from A to C because C is the amino acid that is overall closest to the each of the amino acids in the first column.

Table 2 Common ancestor μ from the new algorithm

S1	AGCTA	CTAGT	AATCA
S2	AGCTA	CGAGT	AATCA
S3	ATCCA	CTAGT	AACT
S4	ATCCA	CTAGT	ATACT
S5	CGGTA	TTTGT	AAGCT
S6	CGGTT	CATCA	AATGC
S7	AGGTA	CTTGA	AATCC
μ	CGCCA	CTTGT	AATCC

It can be assumed that in each sequence S_i those amino acids (or nucleotides) that do not match the corresponding amino acid (or nucleotide) in μ were mutated at some point during evolution. Intuitively, the more common mutations two sequences S_i and S_j share, the closer they are likely to be in an evolutionary tree. For the above set of sequences, the common mutations similarity matrix is shown in Table 3:

Table 3 Initial CMSM matrix

	S1	S2	S3	S4	S5	S6	S7
S1	0	4	2	2	1	1	2
S2	4	0	2	2	1	1	2
S3	2	2	0	5	1	0	1
S4	2	2	5	0	1	0	1
S5	1	1	1	1	0	2	2
S6	1	1	0	0	2	0	3
S7	2	2	1	1	2	3	0

According to the common mutations similarity matrix, the closest pair of sequences is S3 and S4. Hence these will be merged. When we merge two sequences S_i and S_j , in the merged sequence the k th element will be equal to the amino acid (or nucleotide) in the two sequences if $S_i[k] = S_j[k]$ and will be equal to $\mu[k]$ otherwise. Hence the matrix of sequences will be updated as Table 4:

Table 4 The updated sequences

S1	AGCTA	CTAGT	AATCA
S2	AGCTA	CGAGT	AATCA
S34	ATCCA	CTAGT	AACT
S5	CGGTA	TTTGT	AAGCT
S6	CGGTT	CATCA	AATGC
S7	AGGTA	CTTGA	AATCC
μ	CGCCA	CTTGT	AATCC

For example, since $S3[12] = C \neq T = S4[12]$, by the above merging rule $S34[12] = \mu[12] = A$.

After the merge, the common mutations matrix needs to be recalculated. The merge does not change μ , but the entries in the common mutations similarity matrix that are related to the newly merged sequence S34 need to be calculated. The values for S3 and S4 should be deleted. In this case, Table 5 shows the updated common mutation matrix.

Table 5 The updated CMSM matrix

	S1	S2	S34	S5	S6	S7
S1	0	4	2	1	1	2
S2	4	0	2	1	1	2
S34	2	2	0	1	0	1
S5	1	1	1	0	2	2
S6	1	1	0	2	0	3
S7	2	2	1	2	3	0

Now the closest pair is S1 and S2 with a value of 4 common mutations. Hence those two will be merged next. The merging will continue until there is only one sequence left. The CMSM evolutionary tree algorithm can be summarized as follows.

Table 6 The CMSM algorithm

<p>ALGORITHM CMSM (S1...Sn, n)</p> <ol style="list-style-type: none"> 1 Form n clusters of sequences, each with a single sequence. 2 Find the putative common ancestor μ of the sequences. 3 Construct a graph T with a node for each n cluster and for μ. 4 While (there is more than one cluster) <ol style="list-style-type: none"> 5 Find the common mutations similarity matrix. 6 If (exist distinct Si and Sj with some common mutations) <ol style="list-style-type: none"> 7 Merge a closest distinct Si and Sj pair into a new cluster Sij and create a node for Sij. 8 Connect the nodes for Si and Sj with parent node Sij. 10 Else <ol style="list-style-type: none"> 11 Connect the remaining clusters' nodes to parent μ. 11 Return T. 12 Return T.
--

Note: Alternatively, instead of only recording the values, the actual set of common mutations can be put into each entry of the common mutations similarity matrix. Clearly, the cardinality of the sets in the second representation determines the numerical values in the first representation.

III. INCREMENTAL PHYLOGENETICS BY REPEATED INSERTIONS

A. A New Phylogenetic Tree Algorithm

Suppose that we have n number of amino acid sequences S1, . . . , Sn. The sequences and the number n are the inputs to the following algorithm that constructs an evolutionary tree by repeated addition of new species that are represented by the amino acid sequences. We call the new algorithm IPRI (incremental phylogenetic by repeated insertions).

In the algorithm, the closest pair can be found by minimum Hamming distance if the sequences are DNA or RNA strings. If the sequences are proteins, then the closest pair can be found by using a PAM or a BLOSUM substitution matrix. The running time is $O(n^2m)$ where m is the length of the sequences because there are n insertions, and each insertion requires n comparisons between two strings of length m.

Table 7 The IPRI algorithm

<p>ALGORITHM IPRI(S1...Sn, n)</p> <ol style="list-style-type: none"> 1 Create an independent node Nk for each sequence Sk. 2 Let $N = \{ Nk : 1 \leq k \leq n \}$ 3 Find the closest pair of nodes Ni and Nj. 4 Create a tree T with root R, left child Ni and right child Nj. 5 $N = N \setminus \{Ni, Nj\}$ 6 While (N is not empty) <ol style="list-style-type: none"> 7 Find the closest pair of nodes $Ni \in N$ and $Mj \in T$. 8 If (Mj is not the root of T) <ol style="list-style-type: none"> 9 P = parent of Mj. 10 Delete P as a parent of Mj. 11 Create a node R. 12 Make P the parent of R. 13 Make R the parent of Ni and Mj. 14 Else <ol style="list-style-type: none"> 15 Create a node R. 16 Make R the parent of Ni and Mj. 17 $N = N \setminus \{Ni\}$. 18 Return T.
--

IV. EXPERIMENTAL RESULTS

We compared the algorithms on simulated evolutionary data as follows. We assumed that the original protein consists of a chain of one thousand Alanine amino acids. We mutated this original string two ways to generate to children. Both children were generated by first randomly selecting one percent of the amino acids. Then we changed the selected amino acids to one of the twenty amino acids. That is, each of the selected amino acids had a five percent chance of remaining A and ninety five percent chance of changing into another amino acid, with five percent chance of changing into C, five percent chance of changing into D and so on.

Next both of the children were further mutated to generate four grandchildren of the original protein. Then we general additional levels of the tree so that after N levels we had 2^N leaves.

With the above process of evolutionary tree generation, two siblings can be expected to differ from each other on twenty amino acids. Two first cousins can be expected to differ from each other on forty amino acids. Two seconds can be expected to differ from each other on sixty amino acids, and so on.

We ran ten tests on evolutionary trees with height four (and sixteen leaves). We implemented the CMSM and the IPRA algorithms in MATLAB. We used ClustalW2's implementation of the UPGMA and NJ algorithms. We chose on the ClustalW2 website the default parameters, that is, a gap open penalty of 10, a gap extension penalty of 0.2, and a maximum gap distance of 5. The results can be summarized in the Table 8, where "Perfect" means that the reconstructed tree is the same as the original evolutionary tree. When a reconstructed tree had errors, we checked only how many of the sibling pairs (SPs) were correctly handled.

Table 8 Experimental comparisons of the algorithms

Test	CMSM	IPRA	UPGMA	NJ
1	Perfect	Perfect	8 SPs	8 SPs
2	Perfect	Perfect	7 SPs	7 SPs
3	Perfect	Perfect	7 SPs	7 SPs
4	Perfect	Perfect	6 SPs	7 SPs
5	Perfect	Perfect	7 SPs	7 SPs
6	Perfect	Perfect	7 SPs	7 SPs
7	Perfect	Perfect	8 SPs	8 SPs
8	Perfect	Perfect	8 SPs	8 SPs
9	Perfect	Perfect	6 SPs	6 SPs
10	Perfect	Perfect	7 SPs	7 SPs

As an example, Fig. 1 shows the output of the IPRA algorithm in case 4. As a comparison, Fig. 2 shows the output of the UPGMA algorithm in the same case.

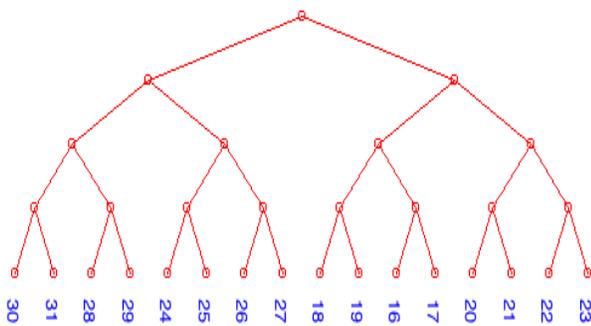


Fig. 1 Sample evolutionary tree reconstructed by the IPRA algorithm

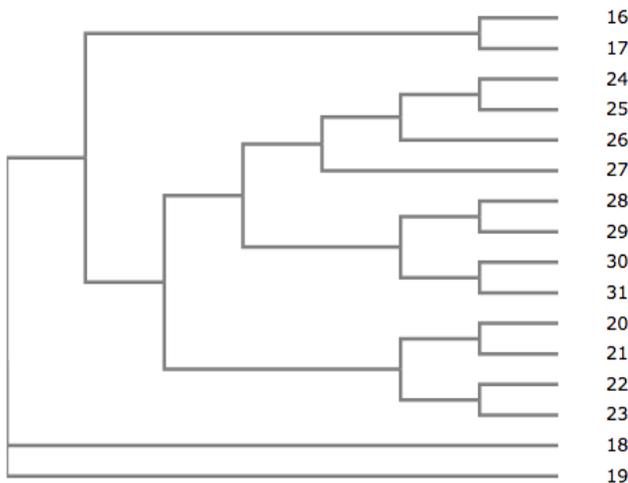


Fig. 2 Sample evolutionary tree reconstructed by the UPGMA algorithm

As can be seen from Figures 1 and 2, the IPRA algorithm has given back the original evolutionary tree. On the other hand, the UPGMA algorithm made a mistake in some of the sibling pairs. In particular, the leaves 26 and 27 and the leaves 18 and 19 are not paired correctly. In addition, there are more

mistakes in grouping together cousin leaves. For example, the sibling leaves 16 and 17 are paired correctly, but they are not grouped correctly with their cousin leaves 18 and 19.

V. CONCLUSIONS AND FUTURE WORK

The new incremental phylogenetic tree algorithm has a potential to improve the general of phylogenetic trees and our understanding of evolutionary history, as can be inferred based on molecular biology. Generally all phylogenetic tree algorithms improve with greater data size both in the number of species and in the length of the sequences. In the future, we plan to study additional protein families and their DNA and amino acid sequences. We also plan to develop computer animation software that shows the insertion of new species into the phylogenetic tree.

REFERENCES

- [1] D. Baum and S. Smith, *Tree Thinking: An Introduction to Phylogenetic Biology*, Roberts and Company Publishers, 2012.
- [2] B. G. Hall, *Phylogenetic Trees Made Easy: A How to Manual*, 4th edition, Sinauer Associates, 2011.
- [3] P. Lerney, M. Salemi, and A.-M Vandamme, editors. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edition, Cambridge University Press, 2009.
- [4] P. Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, Springer, New York, 2010.
- [5] P. Z. Revesz, "An algorithm for constructing hypothetical evolutionary trees using common mutations similarity matrices," *Proc. 4th ACM International Conference on Bioinformatics and Computational Biology*, ACM Press, Bethesda, MD, USA, September 2013, pp. 731-734.
- [6] P. Z. Revesz and C. J.-L. Assi, "Data mining the functional characterizations of proteins to predict their cancer relatedness," *International Journal of Biology and Biomedical Engineering*, 7 (1), 2013, pp. 7-14.
- [7] P. Z. Revesz and T. Triplet, "Classification integration and reclassification using constraint databases," *Artificial Intelligence in Medicine*, 49 (2), 2010, pp. 79-91.
- [8] P. Z. Revesz and T. Triplet, "Temporal data classification using linear classifiers," *Information Systems*, 36 (1), 2011, pp. 30-41.
- [9] N. Saitou and M. Nei, "The neighbor-joining method: A new method for reconstructing phylogenetic trees," *Molecular Biological Evolution*, 4, 1987, pp. 406-425.
- [10] M. Shortridge, T. Triplet, P. Z. Revesz, M. Griep, and R. Powers, "Bacterial protein structures reveal phylum dependent divergence," *Computational Biology and Chemistry*, 35 (1), 2011, pp. 24-33.
- [11] R. R. Sokal, and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Science Bulletin*, 38, 1958, pp. 1409-1438.
- [12] T. Triplet, M. Shortridge, M. Griep, J. Stark, R. Powers, and P. Z. Revesz, "PROFESS: A protein function, evolution, structure and sequence database," *Database -- The Journal of Biological Databases and Curation*, 2010, Available: <http://database.oxfordjournals.org/content/2010/baq011.full.pdf+html>

Peter Z. Revesz holds a Ph.D. degree in Computer Science from Brown University. He was a postdoctoral fellow at the University of Toronto before joining the University of Nebraska-Lincoln, where he is a professor in the Department of Computer Science and Engineering. Dr. Revesz is an expert in databases, data mining, big data analytics and bioinformatics. He is the author of *Introduction to Databases: From Biological to Spatio-Temporal* (Springer, 2010) and *Introduction to Constraint Databases* (Springer, 2002). Dr. Revesz held visiting appointments at the IBM T. J. Watson Research Center, INRIA, the Max Planck Institute for Computer Science, the University of Athens, the

University of Hasselt, the U.S. Air Force Office of Scientific Research and the U.S. Department of State. He is a recipient of an AAAS Science & Technology Policy Fellowship, a J. William Fulbright Scholarship, an Alexander von Humboldt Research Fellowship, a Jefferson Science Fellowship, a National Science Foundation CAREER award, and a “Faculty International Scholar of the Year” award by *Phi Beta Delta*, the Honor Society for International Scholars.

Zhiqiang Li got his B.S. and M.S. degrees from Xidian University, China, in 2009 and 2012, respectively. He is currently a Ph.D. student in the University of Nebraska-Lincoln. His research interests include bioinformatics, computer security, data mining and image processing.

Randomized global optimization for robust pose estimation of multiple targets in image sequences

Johannes Brünger, Imke Traulsen, and Reinhard Koch

Abstract—The problem of determining the position and pose of multiple targets in an image sequence can be treated as a mathematical optimization problem. Therefore we show a technique based on randomized black-box optimization to detect multiple targets with similar appearance in bird’s-eye view image sequences. An easy to adapt cost-map and parameter-constraints are used to design a fitness-function evaluated for each target separately. Especially for crowded targets in limited space like in livestock-environments this approach is able to demonstrate its strengths. The detection performance is evaluated on two hand-labeled data-sets of piglets, recorded for behaviour studies. Detection rates above 90% at runtimes of 220ms per processed frame prove the potential of the proposed technique.

Index Terms—multiple indistinguishable targets, pose estimation, randomized black-box optimization

I. INTRODUCTION

THE correct estimation of the pose and position of an object in an image is part of many computer vision related problems. One possible solution is to fit a parametrized model to the image data and measure its accuracy by a specific evaluation-function. This exactly meets the definition of a mathematical optimization problem, where a fitness-function has to be maximized depending on the model parameters.

Under simple conditions this fitness-function may be linear, but in general it is more complex or even unknown. Furthermore sensors like cameras do not reflect the environment perfectly due to noise or projection issues. Therefore the fitness-function has to be designed by hand to suit the current problem.

If the shape of the fitness-function is non-convex, discontinuous or even unknown the related optimization problems are referred to as black-box optimization problems. In this situation the only accessible information about the function are function values of evaluated search points in the parameter-space.

As more and more veterinary or biological behavior studies make use of automated visual monitoring [1], [2], [3], the need for robust object detection by computer vision techniques increases. The correct detection of heavily interacting indistinguishable targets in combination with noisy, poorly illuminated video footage is an obvious representative for the class of black-box optimization problems. Using

different measurements a multi dimensional fitness-function can be defined which only relies on information received from the camera image. In this work we show the application of the black-box optimization scheme "Covariance Matrix Adaptation Evolution Strategy" (CMA-ES) [4] for a robust detection of multiple indistinguishable targets in livestock environments. We particularly address the detection of piglets in image sequences recorded for behaviour studies. CMA-ES has proven its capabilities in several comparisons [5], [6] and has a broad acceptance as state-of-the-art algorithm for continuous optimisation.¹

We use a fixed number of simplified pig-models and try to find the optimal assignment of the models to the targets. Exploiting the continuity in the image sequence, the parameters of all models are refined in each frame separately by CMA-ES and a particularly designed fitness-function starting from the last valid position. Thereby all other targets around are considered to avoid overlap. This competitive approach ensures the representation of all targets and helps to mark out the individual targets if they are crowded together. To manage occlusion the models can be deactivated if their fitness-value drop. Likewise the models get recovered if they spot an uncovered region with high fitness-value.

In section II we briefly introduce related work showing the use of CMA-ES in several domains. Next we give a short summary of related work in behaviour studies. The proposed method is introduced in section III in detail. Last the used setup and the results are presented in section IV.

II. RELATED WORK

Mathematical optimization has many applications in computer science research. CMA-ES in particular with its capabilities in non-linear domains is often used in image registration tasks but also in engineering problems or classical parameter optimizations. In the medical field for instance, computer vision is used to guide the physician on interventions. In [7] the authors used CMA-ES to optimize transformation parameters in multislice-to-volume registration. Thereby the position of the patient is tracked to compensate motion during the operation. Another example for 2D-3D-registration in medical imaging using CMA-ES can be found in [8]. The comparison in [6] even attested CMA-ES the best results concerning registration rate and accuracy for a comparable task. In the field of geoenvironment Bouzarkouna et al. [9] used CMA-ES

J. Brünger and R. Koch are with the Institute of Computer Science, Christian-Albrechts-Universität zu Kiel (see <http://www.mip.informatik.uni-kiel.de>).

I. Traulsen is with the Institute of Animal Breeding and Husbandry, Christian-Albrechts-Universität zu Kiel.

Manuscript received September 15, 2015; revised ...

¹see COmparing Continuous Optimisers, <http://coco.gforge.inria.fr>

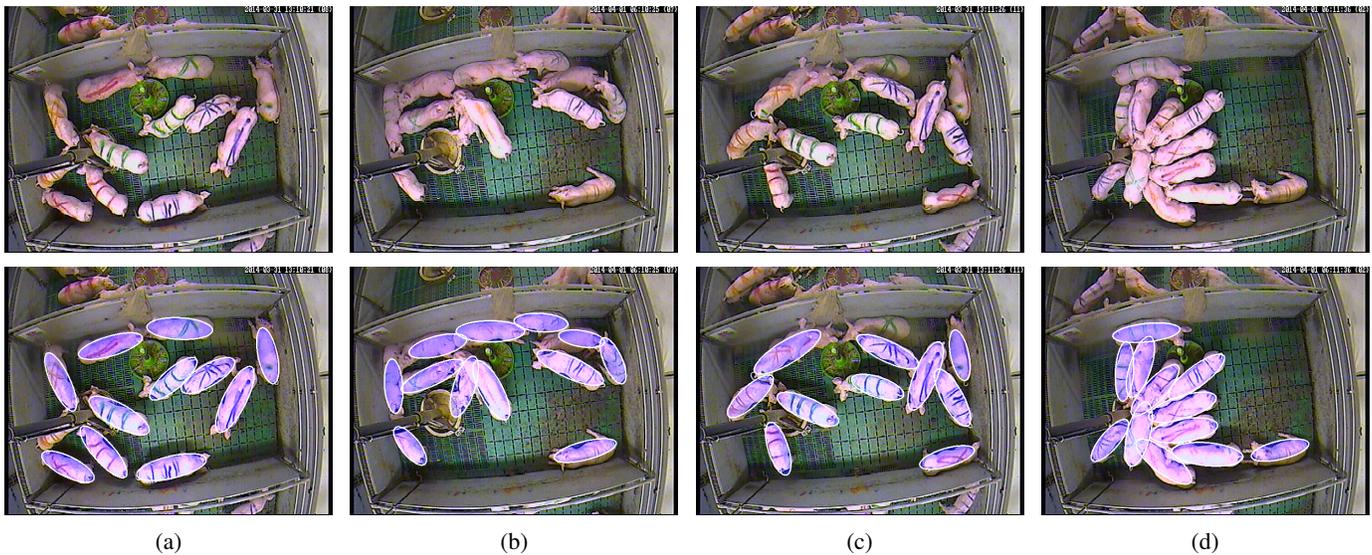


Fig. 1: First row: Different frames of our footage (the colored pattern of paint were used for ground truth only). Second row: The results of our algorithm for the same frames. Note the partially occlusion of individual pigs by structure or other pigs in (a) and (c).

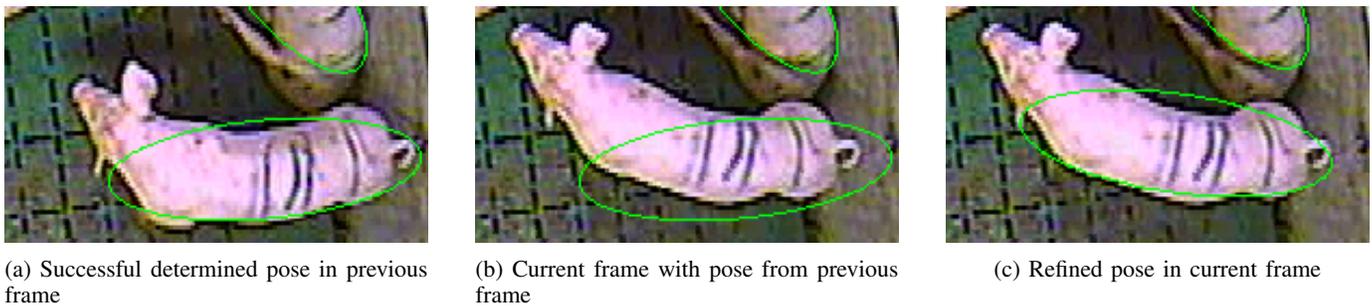


Fig. 2: Given an estimated pose from the previous frame, determining the pose for the current frame only needs refinement.

for optimized well placement, since such problem depends on a large number of not influenceable parameters. The same problem arises for the placement of sensors in a Wireless Sensor Networks (WSN). The sensors often have limited sensing and communication capabilities and need therefore to be placed in optimal positions to cover an area as large as possible. Hence their positions can be optimized with CMA-ES too, as showed by Akbarzadeh et al. [10].

As in medical imaging, classical computer vision tasks include the need for registration of image features to model parameters. Jordt and Koch [11], [12], [13] used CMA-ES to track deformation of objects in real time. They used a time-of-flight-camera to obtain depth information and an analysis-by-synthesis approach to optimize the model parameters. In [14] Zelenka detected the contours of gas bubbles in water and reconstructed their shape. This helped to track the individual bubbles over time and to measure the throughput of the emitting source. Also classical target tracking was implemented with the help of CMA-ES. In [15] the model parameters obtained by a particle filter are optimized with CMA-ES. The examples show that CMA-ES is a versatile technique that can be used in a variety of problems.

Over the last years the detection of animals with computer

vision techniques gained more attention in behaviour studies. Classical tracking often relies on the predicted motion or the unique appearance of the target. Therefore the special conditions in livestock or laboratory environments require new approaches as the movements of the animals are hardly predictable. If multiple animals are caged in limited space this causes group dynamics and if further their appearance is similar target switching can occur while they are interacting. Tillet [16], [17] showed that image processing can be used to fit a model to pigs and hence determine their position and orientation. Cangar et al. [18] used image processing to fit a basic cow model to pregnant cows and monitor their movements. Dollár et al. used in [19] regression-methods to progressively refine a loosely specified initial guess of a pose. As their system is learned from human annotated training data the application is not limited to a specific target. So one of their examples shows the detection of a mouse in a laboratory environment.

Depending on the number of monitored targets the motivation differs. If only one animal is present its activity can be classified. So Jhuang et al. [3] used background subtraction to detect a mouse in a home-cage. By extracting space-time motion and velocity-based features they were able to

phenotype its behaviour. Zurn et al. [20] used near infrared light to detect the position and activity of a rodent in dark and light cycles. Farah et al. [21] used a sliding window tracking approach to determine the motion pattern of a laboratory rat. They combined several features in a fitness-function to refine the detected pose.

For social behaviour studies more than one animal has to be observed. To study the effects of genetic mutations, drugs or environmental stimuli Giancardo et al. [22] used a thermal-camera to detect multiple mice. The animals were segmented with blob detection followed by an adapted watershed algorithm. Building up on the estimated positions they classified social behaviour among the animals. Khan et al. [23] tracked interacting ants with an MCMC-based particle filter. They exploited the characteristics of Markov random fields to model interacting targets efficiently.

The more animals are part of the monitored group, the more the distance between the animals decreases. This results in more group dynamics and more points of contact. Pistori et al. [24] used a particle filter to track multiple mice or larvae. With k-means clustering the segmented pixel were assigned to the individual targets. This made it possible to tell them apart, if they crowd together.

McFarlane and Schofield [25] used a Laplacian operator to separate piglets that were clustered tightly together. Then ellipses were used to model the piglets from the remaining blobs. To track 3 loose-housed pigs Arendt et al. [26] used a support-map for each pig. With a 5D-Gaussian-Distribution as model, they updated the support-map in each loop, defining which pixel are probably part of a specific pig. They took into account that the pigs are uniformly colored and therefore the distance in the color-space is small. In every frame they measured the Mahalanobis distance of each pixel in 5 dimensions looking for the best match to an existing pig support-map. Kashiha et al. [2] used background subtraction and ellipse fitting to determine the position of 10 pigs. They used the position information to monitor the water consumption of the pigs by counting the visits at the drink nipple. In a newer work Kashiha et al. [27] used pattern recognition to identify 10 pigs by markers on their back. The positions were determined by ellipse fitting again. They analyzed 4 sequences of 390 images each. In the 15600 positions they were able to identify the pigs with a rate of 88.7%. With the proposed ellipse fitting tightly clustered pigs are still problematic as they may get covered by a single ellipse. In our approach individual object-tracker compete against each other. This ensures the representation of each target. Furthermore we take the limited movements into account, instead of evaluating each frame independently.

III. METHOD

In image sequences the position and pose of the monitored target as well as the the describing parameters of its model often barely changes between two successive frames. This can be utilized to find the unknown parameters by refining the parameters of the successfully estimated pose in the previous frame (see Figure 2). Given an initial starting point for the first frame, the complete sequence can then be processed

automatically without any further user-input.

Let $T = \{t_0, \dots, t_{n-1}\}$ be the set of n targets in the processed image sequence with k frames. For each image $I_{j=0, \dots, k-1}$ the position and pose of each target t_i can be described by a vector of l parameters $\Theta_{i,j} \in \mathbb{R}^l$. If a fitness-function $F : \mathbb{R}^l \rightarrow \mathbb{R}$ is available, which evaluates an estimate of the parameter-vector based on the appearance of the target in the current image, the refinement can be seen as a classical optimization task. The optimizer can use the last valid position and pose parameters as a starting point and evaluate the fitness-function on the current image to find the set of parameters which best describe the current pose of the target.

Given a fitness-function $F(I_j, \Theta_{i,j-1})$ which takes the current image I_j and the vector of parameters $\Theta_{i,j-1}$ to describe the position and pose of target t_i in the previous image I_{j-1} the optimization task for a optimizer can be formulated as:

$$\Theta_{i,j}^* = \arg \max F(I_j, \Theta_{i,j-1}) \quad (1)$$

with $\Theta_{i,j}^*$ as the vector of parameters describing the position and pose of target t_i in image I_j best.

Designing a suitable fitness-function is part of the proposed solution and is described in Section III-B. Unfortunately this high dimensional fitness-function tend to be non-convex and discontinuous as it is only based on the image data which is noisy and may include compression artifacts. In addition constraints are needed to prevent the models from degeneration, which also can result in breaks or sharp edges in the function's shape. All this leads to a complex and unknown function where the only information about function values can be obtained by sampling the function at certain points. This situation is referred to as black-box optimization as no knowledge about the function can be used for proper optimization-parameter tuning.

A. Covariance Matrix Adaptation - Evolution Strategy

To find the global maximum in our optimization problem we use the Covariance Matrix Adaptation - Evolution Strategy (CMA-ES) [4] as it has shown great performance² in the domain of randomized black-box search techniques. Randomized black-box optimizer use an internally maintained probability distribution to randomly choose points where the given fitness-function is to be evaluated. The gathered fitness-function values are then used to update all the distribution-parameters trying to maximize the probability of finding the best solution (see Algorithm 1).

In the case of CMA-ES a multivariate normal distribution $\mathcal{N}(m, C)$ is used. $m \in \mathbb{R}^l$ is the mean of the parameter space and $C \in \mathbb{R}^{l \times l}$ respectively its covariance matrix. C is usually initialized as a diagonal matrix filled with the variances of the individual parameters.

Since CMA-ES follows an evolutionary strategy the $\lambda \in \mathbb{N}$ sampled points $\Theta_{i=0, \dots, \lambda-1}$ can be interpreted as a population. Starting with the initial distribution $\mathcal{N}(m^{(0)}, C^{(0)})$ in each iteration CMA-ES regenerates its population and forms a new generation $\Theta_{i=0, \dots, \lambda-1}^{(g+1)}$ by sampling from an updated

²see 2009 Black-Box Optimization Benchmarking Competition (BBOB)

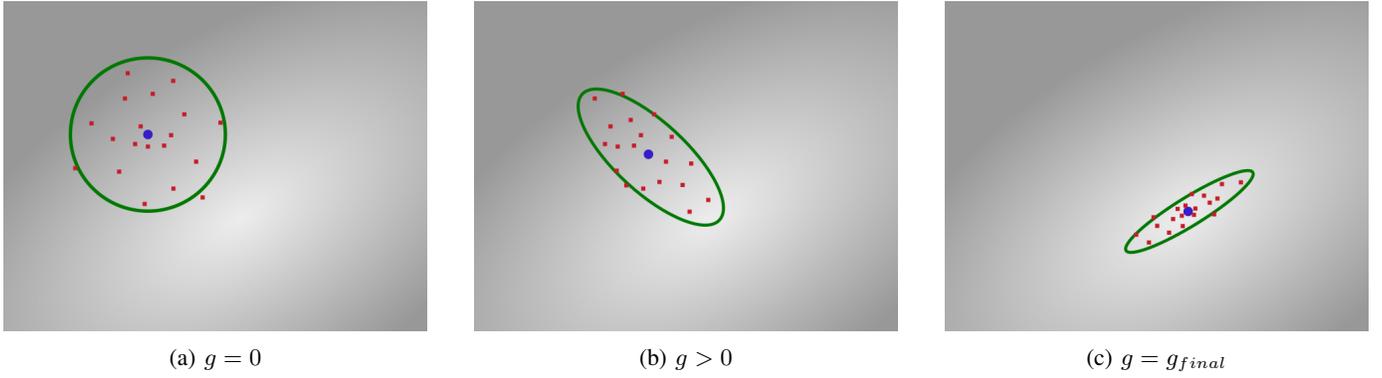


Fig. 3: Schematic visualization of the CMA-ES optimization process evolving over generation g with population-size $\lambda = 19$ and $l = 2$ parameters. The background shows the fitness function value (brighter = higher). Also shown are the covariance matrix C (green ellipse), the mean m (blue dot) and the sample points (population) according to C and m (red points).

Algorithm 1 Randomized black-box search

```

Initialize distribution
for generation  $0, 1, 2, \dots$  do
    Sample  $\lambda$  independent vectors of parameters  $\Theta$  from
    distribution
    Evaluate the samples on fitness-function  $F$  (see Fig. 4)
    Update distribution parameters
    break, if termination criterion met
end for
    
```

distribution $\mathcal{N}(m^{(g+1)}, C^{(g+1)})$, with $g \in \mathbb{N}$ depicting the generation number. Thus the evolution between g and $g + 1$ is done by updating the mean and the covariance-matrix. To move the mean the best $\mu \in \mathbb{N}_{\leq \lambda}$ samples of the current population are selected. Assuming the samples are sorted according to their cost $F(\Theta_0) \leq F(\Theta_1) \leq \dots \leq F(\Theta_{\lambda-1})$ the new mean is calculated by applying a weighted average with weights $w_{i=0, \dots, \lambda-1} \in [0, 1]$ and $\sum_{i=0}^{\lambda-1} w_i = 1$. To enforce the selection the last $\lambda - \mu$ weights are set to 0.

$$m^{(g+1)} = \sum_{i=0}^{\mu-1} w_i \Theta_i^{(g)} \quad (2)$$

To maximize the probability of finding the best parameter-vector the covariance-matrix is updated:

$$C^{(g+1)} = \frac{1}{\mu} \sum_{i=0}^{\mu-1} (\Theta_i^{(g)} - m^{(g)}) (\Theta_i^{(g)} - m^{(g)})^T \quad (3)$$

This effects the probability density function of the distribution and therefore the spread of the samples in the search-space (see Figure 3). For further optimized methods of the covariance-update we refer to the written tutorial of Nicolau Hansen³.

As long as no termination-criterion is met, the fitness-function is evaluated in each iteration for each individual of the population. The key termination-criteria are a stagnation in the history of the last fitness-function values and the maximum number of iterations (or generations of the

population). With a reasonable value for stagnation-limitation set, the maximum number of iterations is rarely reached and can therefore be kept untouched.

The remaining control-parameter is the population-size λ . The significant influence of the population-size on the global search performance is evaluated and confirmed in [28]. If the size of the population is too low, more iterations are needed to achieve a significant improvement of the distribution parameters. If it is too high, the count of fitness-function evaluations increases and therefore the runtime. In [29] the optimal population-size λ for l parameters is proposed as $\lambda = 4 + \lceil 3 \ln l \rceil$.

In our approach a separate CMA-ES-optimizer is initialized once with a manually labeled starting-position for each target in frame 1. In the subsequent frames each optimizer starts with the last known valid position of its target. To limit the search to the local neighbourhood the standard deviations of the parameters are set appropriately. Although the optimizers are evaluated independently, through the overlapping-prevention they compete for the segmented pixel. Therefore the optimizers are evaluated in the order of the result from the last frame to prevent good detections from being disturbed by inferior ones.

Optimizing the parameters of all targets at once has been tested too but gave inferior results.

B. Fitness-function

Besides the model-parameters to evaluate, the input of the fitness function consists of the last valid positions of all targets $\{\Theta_{0,j-1}, \dots, \Theta_{n-1,j-1}\}$ and the pixel-values of the current image I_j . To allow flexibility for a wide range of applications, only a binary segmentation of the targets is used. This segmentation can be provided by any detection-algorithm according to the processed scene. In addition to gray-scale thresholds, color based segmentation in RGB or HSV color-space are probably the easiest way to implement such a binary segmentation. But even more complex approaches like background-subtraction can be used. With this techniques the targets normally will form big clusters of segmented pixel

³<https://www.lri.fr/~hansen/cmatutorial.pdf>

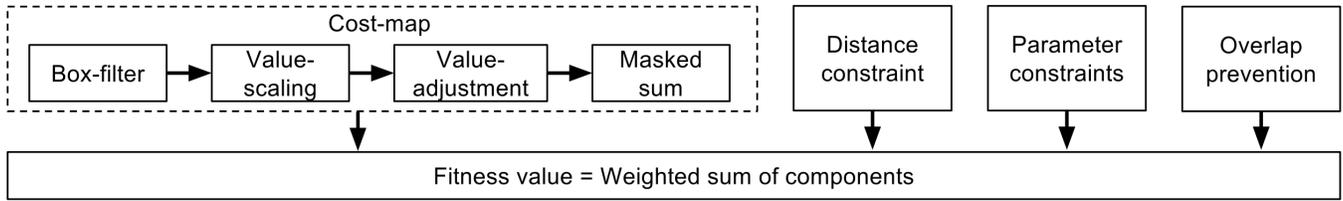


Fig. 4: Components of the fitness-function evaluation.

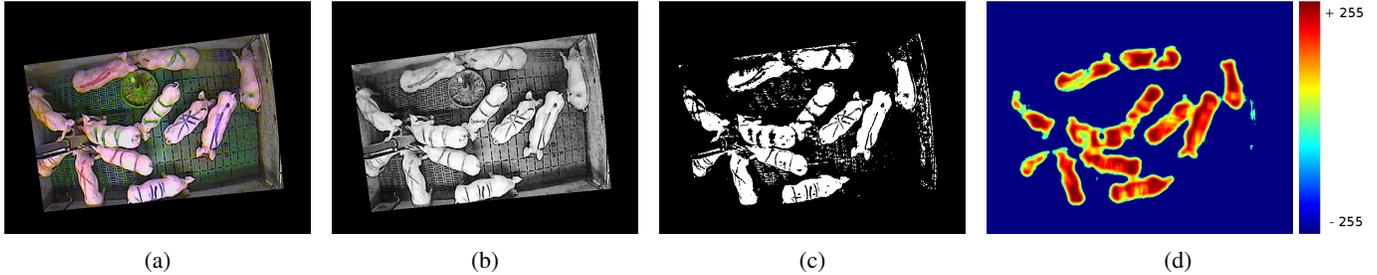


Fig. 5: The stages of segmentation: masking (a), histogram-equalization (b) binary threshold (c) and the resulting (colorized) cost-map (d).

except for disturbances and noise. To allocate these regions to the targets, a binary pixel-mask of the model is used to determine which pixel need to be evaluated. The mask is positioned, scaled and rotated according to the parameters proposed by the optimizer.

This segmentation-evaluation is the main part of the fitness-value. To refine the result, the positions of all other targets are included to avoid overlap and remove target ambiguities. Further improvements are achieved by constraints limiting the change of the model over time. To calculate the final fitness, the results of the overlap-calculation and the constraints are weighted and subtracted from the segmentation-evaluation (see Figure 4). The components of the fitness function are described in detail below.

Cost-map: To penalize a potential coverage of background-pixel we transform the segmentation to a cost-map where pixel are rated depending on their probability of belonging to a target or to the background (see Figure 5). To generate the cost-map we apply a box-filter with an $f \times f$ -window ($f \in \mathbb{N}$) on the binary segmentation. This changes the range of the segmentation-image to $[0, f^2]$ with high values on locations where many segmented pixel are clustered in the local neighbourhood. Next we normalize the values to a range of $[-c, c]$, $c \in \mathbb{N}$, so pixel with no or few segmented pixel in the neighbourhood get negative values, the clustered pixel get positive values. Finally all values of the cost-map covered by the mask are summed up which forces the optimizer to an optimal alignment of the model to the segmented pixel. To further optimize the detection results, the cost-map can be manipulated by setting all background pixel $[-c, -b]$ with $b \in \mathbb{N}, b < c$ to the maximum negative value $-c$. This penalizes these pixel even further and keeps the optimizer from jumping over the gap between two targets and gathering

pixel of two targets at the same time.

We found that emphasizing the center of the target gave better pose detection. So instead of using a simple binary mask, a weighting mask may be used to emphasize specific areas of the model.

Distance constraint: Since the targets hardly move between two frames of the sequence, guesses with a long distance to the last position are rated worse. So the Euclidean distance between the center of the current guess and the center of the last valid position is also weighted and subtracted.

Parameter constraints: At initialization the mean of certain parameters over all targets are computed. This gives an idea of the average dimension of the targets. Constraining the guessed parameters by this global mean prevents a degeneration of the proposed target-model. In addition a running mean of some parameters is useful. Deviations from these target-related means can be penalized to keep the changes of the parameters smooth.

Overlap prevention: To avoid overlap the fitness-function also evaluates the last valid position and pose of all other targets. If the current guess of the optimizer overlaps with other targets the return-value of the fitness-function is lowered to penalize the overlap.

Depending on the target's model the overlap-function has to be defined accordingly. We used a overlap-function which returns an percentage overlap of two given targets calculate pixel-wise. If the overlap of the current target with one of the other targets exceeds $o \in \mathbb{R}$ percent the current guess is invalidated. Likewise if the overlap of the current target with all other targets sums up to an value $> o$ the guess is invalidated. To speed up the overlap-check, only targets

in a defined radius around the current evaluated position are considered. If the percentage overlap is smaller than o , the number of overlapping pixel is multiplied by a weight and subtracted from the current fitness-value.

C. Recovery of lost targets

To handle occlusions the optimizers have a flag which marks their targets as active or not. If the fitness-value of the targets drops below a certain threshold it is marked as inactive. With this technique optimizers are automatically inactivated if the size of the segmentation of their target is reduced through occlusion.

To recover the reappearing targets the population-size λ of the optimizer with inactive targets is increased and the standard deviations of the x- and y-position are set to the half image-size. The starting-point for the recovery is chosen randomly. This allows the CMA-ES to spread its individuals over the whole image to find the lost target.

IV. EVALUATION

We tested the algorithm on videos recorded for behaviour studies. Our own data-set consists of two image sequences of piglets from a surveillance-camera. The sequences show a pen with 12 piglets on two consecutive days. All piglets were marked with pattern of paint for individual identification for ground truth labeling only. The patterns were not used for detection. There were no special preparations or additional lighting to enhance the quality of the footage. Moreover structure in the pen sometimes led to partial and total occlusion of single piglets. In addition sequence 2 includes the feeding-procedure where all pigs shove around the feeder. Both sequences have 1993 frames covering 15 minutes. The resolution is 720 x 540 pixel. Figure 1 shows different frames of the sequences.

As many others [25], [2], [27] we used simple ellipses as a model for the pigs. An ellipse in the 2D-plane can be fully described by only five parameters but approximates the body of a pig sufficiently. The five parameters are the x- and y-position of the centroid, the two axis lengths and the rotation of the ellipse. The pigs in our data-set have an average size of 136 x 45 px.

A. Parameters

We used simple binary threshold-segmentation on the histogram-equalized grayscale-image. To reduce disturbances the regarded pen was masked (see Figure 5).

For the generation of the cost-map we used a 19x19 box-filter. After the filtering the cost-map was scaled to [-255, 255] and all values in the range of [-255, -50] set to -255 (see Figure 5d).

As weighted binary model we used three nested ellipses. The inner one had a value of 2.0, the middle one a value of 1.0 and the outer one a value of 0.5.

The values of the cost-map were weighted with 0.001, overlapping pixel were weighted with 0.2.

As constraints we used the Euclidean distance of the guess to

the last valid position, weighted with 1.5. Derivations from the running mean of each target were squared and then weighted with 0.025 (width) and 1.0 (height). Derivations from the mean over all targets were squared and then weighted with 0.05 (width) and 0.5 (height).

In addition we refused all guesses whose ratio of width/height deviated from the same ratio over all targets at initialization with more than 0.6.

All weights were determined heuristically and need to be adjusted to the values of the cost-map. As threshold for inactive targets we used 50, so all targets with fitness-values < 50 were marked as inactive.

B. Detection-Performance

To measure the accuracy of our algorithm we segmented the ellipses for all 12 pigs in the first 500 frames of the two sequences by hand. For long term evaluation we labeled one pig over the complete first sequence. If pigs are obviously occluded by structures or by other pigs this was registered too. With this ground truth data we are able to give an accurate evaluation by comparing the manually labeled data with the ellipses denoted by our algorithm.

Like in [30] we used the Hausdorff distance to measure the accordance of ellipses. Let E and G be two ellipses, interpreted as a set of points $E = \{e_0, \dots, e_n\}$ and $G = \{g_0, \dots, g_m\}$. The Hausdorff distance H is then defined as:

$$H(G, E) = \max(\delta(G, E), \delta(E, G)) \quad (4)$$

with:

$$\delta(G, E) = \max_{e \in E, g \in G} (\|g - e\|) \quad (5)$$

where $\|\cdot\|$ is the Euclidean norm.

We denote the object which holds the parameters for one ellipse a tracker. Since the assignment of the trackers is not fixed to the pigs, two trackers may temporarily target parts of the same pig. Furthermore trackers may be marked as inactive if they fail to find a pig. To ensure a clear assignment of the trackers to the pigs the assignment-procedure for each frame was as follows:

- Evaluate the overlap of each tracker with all manually labeled pig-positions. If the Hausdorff distance is below a certain threshold it is stored as a possible match.
- Sort all possible matches according to the distance in ascending order.
- Go through the list of matches (starting at the best) and mark the dedicated pig as *found* and the dedicated tracker as *processed*.
- Proceed with all further possible matches, skipping all matches where the pigs are already *found* or where the trackers are already *processed*.

After all possible matches are examined, pigs that are not marked as *found* are counted as a miss. As mentioned earlier the manual labeled data includes also occlusion-information. So misses where pigs are obviously occluded can optionally be ignored.

As CMA-ES is a randomized procedure we ran each evaluation 50 times. Table I shows the average percentage of positions

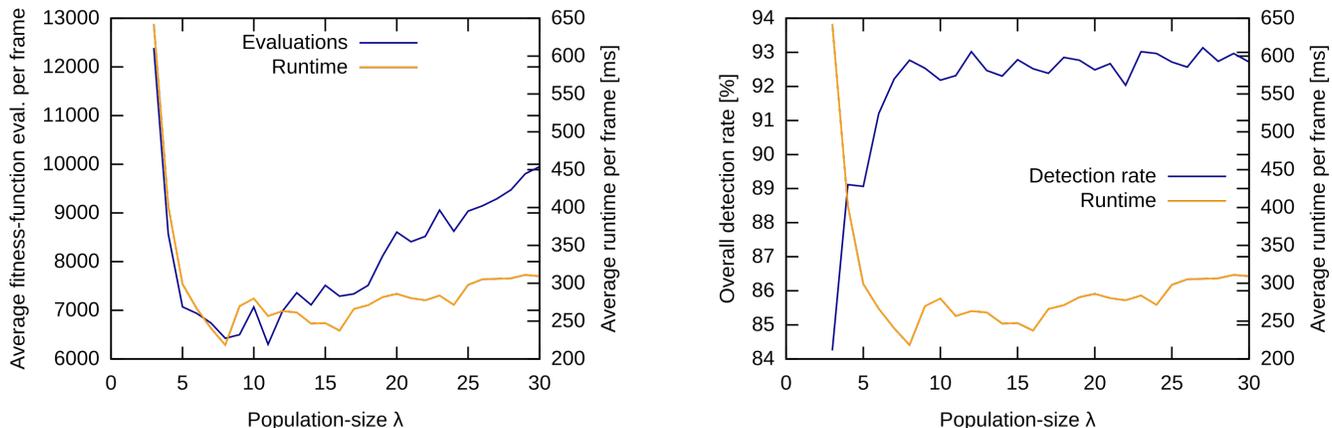


Fig. 6: Evaluation of the influence of the population-size for 500 frames of sequence 1.

our algorithm found for different thresholds with reference to the ground truth. The maximum standard deviation of 0.93% over the 50 runs confirms the stability of the proposed method.

Hausdorff threshold		45px	35px	25px
Sequence 1 (A)	%	94.65%	91.92%	81.10%
	\overline{hd}	14.02px	12.88px	9.57px
Sequence 1 (B)	%	91.86%	88.94%	77.84%
	\overline{hd}	13.64px	12.51px	9.27px
Sequence 1 (C)	%	97.02%	93.60%	85.91%
	\overline{hd}	12.66px	11.31px	9.06px
Sequence 2 (A)	%	94.16%	91.22%	81.21%
	\overline{hd}	14.40px	13.15px	10.01px
Sequence 2 (B)	%	90.74%	87.63%	77.34%
	\overline{hd}	13.93px	12.68px	9.67px

TABLE I: Percentage of correct positions with an Hausdorff distance below the given threshold. The average Hausdorff distance (\overline{hd}) for all correct positions is also given. All values are averaged over 50 runs. The population-size was set to $\lambda = 8$.

(A) depicts a setup of 12 pigs over 500 frames whereby misses of occluded pigs are ignored.

(B) depicts a setup of 12 pigs over 500 frames whereby misses of occluded pigs are not ignored.

(C) depicts a setup of 1 pig over 1993 frames whereby occlusions did not occur.

C. Runtime

As stated in Section III the population-size λ has significant influence on search performance and runtime. Therefore we evaluated the first 500 frames of the first sequence with different population-sizes on an Intel Core i7-4790 system. Figure 6 shows the clear correlation between population-size and the average number of fitness-function evaluations per frame.

As the individuals of the population are evaluated independently in each iteration, we used parallelization to speed this up. Since the used system was a quad-core CPU with Hyper-Threading, multiples of 8 threads could be handled optimally,

which can be clearly seen as local minima at $\lambda = 8, 16, 24$ in the runtime curve.

For our 5 parameters an optimal population-size of 8 was proposed in [29]. In Figure 6 one can see, this holds for our experiments as the detection rate reaches the desired high values from this point. With approx. 220 ms per frame the runtime is also minimal at this point.

As described in Section III-C the population-size for the individual optimizer is increased to recover lost targets. Since a lot of positions are tested in the recovery process where no target is present, the iteration count differs from the normal optimization process. Hence the runtime for individual frames varies depending on the number of lost targets. We evaluated the different parts separately to get an idea how much time the different parts needed. (see Table II).

In the pre-processing step the cost-map is calculated. This has static runtime depending on the image-size. The computing times for a normal optimization or a recovery depends on the number of iterations needed.

Population-size λ	Pre-proc. [ms]	∅ Optim. [ms]	∅ Rec. [ms]
3	30	38.32 (5294)	143.24 (706)
8	30	11.76 (5643)	71.06 (357)
20	30	17.43 (5635)	74.82 (365)

TABLE II: Average computing time for pre-processing, normal optimization and recovery over all 6000 Positions in Sequence 1. The values in brackets show the amount of normal optimizations resp. recoveries.

V. CONCLUSION

In this work we used CMA-ES to detect the pose and position of multiple piglets crowded together. Based on simple binary segmentation we presented a fitness-function to evaluate the pose detection based on the last valid position and the image data. With this the whole image sequence can be processed without user input, except the one-time initialization. An individual CMA-ES optimizer for each target ensures the representation and the best fit. Our approach achieved detection rates above 90% in around 14.000 evaluated

positions on two data sets.

We successfully adapted the proposed method to other videos of biological behaviour studies. More precisely we were able to detect ants and a single bee with just a few adjustments on the weights from section IV-A. In future work we want to find a way to make the algorithm self-adjusting on any kind of image data.

An other interesting extension of the presented work is the possibility to link the detections via tracklets and thereby track the animals over the whole sequence.

REFERENCES

- [1] D. Sergeant, R. Boyle, and M. Forbes, "Computer visual tracking of poultry," *Computers and Electronics in Agriculture*, vol. 21, no. 1, pp. 1 – 18, 1998. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169998000258>
- [2] M. Kashiha, C. Bahr, S. A. Haredasht, S. Ott, C. P. Moons, T. A. Niewold, F. O. Ödberg, and D. Berckmans, "The automatic monitoring of pigs water use by cameras," *Computers and Electronics in Agriculture*, vol. 90, no. 0, pp. 164 – 169, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169912002372>
- [3] H. Jhuang, E. Garrote, X. Yu, V. Khilnani, T. Poggio, A. D. Steele, and T. Serre, "Automated home-cage behavioural phenotyping of mice," *Nat Commun*, vol. 1, p. 68, Sep. 2010. [Online]. Available: <http://dx.doi.org/10.1038/ncomms1064>
- [4] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evol. Comput.*, vol. 9, no. 2, pp. 159–195, Jun. 2001. [Online]. Available: <http://dx.doi.org/10.1162/106365601750190398>
- [5] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Pošík, "Comparing results of 31 algorithms from the black-box optimization benchmarking bbob-2009," in *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, ser. GECCO '10. New York, NY, USA: ACM, 2010, pp. 1689–1696. [Online]. Available: <http://doi.acm.org/10.1145/1830761.1830790>
- [6] S. Winter, B. Brendel, and C. Igel, "Registration of bone structures in 3d ultrasound and {CT} data: Comparison of different optimization strategies," *International Congress Series*, vol. 1281, pp. 242 – 247, 2005, {CARS} 2005: Computer Assisted Radiology and Surgery Proceedings of the 19th International Congress and Exhibition. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0531513105006126>
- [7] H. Tadayyon, A. Lasso, A. Kaushal, P. Guion, and G. Fichtinger, "Target motion tracking in mri-guided transrectal robotic prostate biopsy," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 11, pp. 3135–3142, Nov 2011.
- [8] R. H. Gong and P. Abolmaesumi, "2d/3d registration with the cma-es method," pp. 69 181M–69 181M–9, 2008.
- [9] Z. Bouzarkouna, D. Ding, and A. Auger, "Well placement optimization with the covariance matrix adaptation evolution strategy and meta-models," *Computational Geosciences*, vol. 16, no. 1, pp. 75–92, 2012. [Online]. Available: <http://dx.doi.org/10.1007/s10596-011-9254-2>
- [10] V. Akbarzadeh, C. Gagne, M. Parizeau, and M. Mostafavi, "Black-box optimization of sensor placement with elevation maps and probabilistic sensing models," in *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, Sept 2011, pp. 89–94.
- [11] A. Jordt and R. Koch, "Fast tracking of deformable objects in depth and colour video," in *Proceedings of the British Machine Vision Conference, BMVC 2011*, S. McKenna, J. Hoey, and M. Trucco, Eds. British Machine Vision Association, 2011.
- [12] —, "Direct model-based tracking of 3d object deformations in depth and color video," *International Journal of Computer Vision*, 2013.
- [13] —, "Reconstruction of deformation from depth and color video with explicit noise models," pp. 128–146.
- [14] C. Zelenka, "Gas bubble shape measurement and analysis," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, X. Jiang, J. Hornegger, and R. Koch, Eds. Springer International Publishing, 2014, vol. 8753, pp. 743–749. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11752-2_63
- [15] A. Johansson and E. Lehmann, "Evolutionary optimization of dynamics models in sequential monte carlo target tracking," *Evolutionary Computation, IEEE Transactions on*, vol. 13, no. 4, pp. 879–894, Aug 2009.
- [16] R. Tillett, "Model-based image processing to locate pigs within images," *Computers and Electronics in Agriculture*, vol. 6, no. 1, pp. 51 – 61, 1991. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169991900222>
- [17] R. Tillett, C. Onyango, and J. Marchant, "Using model-based image processing to track animal movements," *Computers and Electronics in Agriculture*, vol. 17, no. 2, pp. 249 – 261, 1997, livestock Monitoring. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169996013087>
- [18] O. Cangar, T. Leroy, M. Guarino, E. Vranken, R. Fallon, J. Lenehan, J. Mee, and D. Berckmans, "Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis," *Comput. Electron. Agric.*, vol. 64, no. 1, pp. 53–60, Nov. 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.compag.2008.05.014>
- [19] P. Dollár, P. Welinder, and P. Perona, "Cascaded pose regression," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1078–1085.
- [20] J. Zurn, D. Hohmann, S. Dworkin, and Y. Motai, "A real-time rodent tracking system for both light and dark cycle behavior analysis," in *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on*, vol. 1, Jan 2005, pp. 87–92.
- [21] R. Farah, J. Langlois, and G.-A. Bilodeau, "Rat: Robust animal tracking," in *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, Sept 2011, pp. 65–70.
- [22] L. Giancardo, D. Sona, H. Huang, S. Sannino, F. Manag, D. Scheggia, F. Papaleo, and V. Murino, "Automatic visual tracking and social behaviour analysis with multiple mice," *PLoS ONE*, vol. 8, no. 9, p. e74557, 09 2013. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0074557>
- [23] Z. Khan, T. Balch, and F. Dellaert, "An mcmc-based particle filter for tracking multiple interacting targets," in *Computer Vision - ECCV 2004*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, vol. 3024, pp. 279–290. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-24673-2_23
- [24] H. Pistori, V. V. V. A. Odakura, J. B. O. Monteiro, W. N. Gonçalves, A. R. Roel, J. de Andrade Silva, and B. B. Machado, "Mice and larvae tracking using a particle filter with an auto-adjustable observation model," *Pattern Recognition Letters*, vol. 31, no. 4, pp. 337 – 346, 2010, 20th SIBGRAPI: Advances in Image Processing and Computer Vision. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865509001330>
- [25] N. McFarlane and C. Schofield, "Segmentation and tracking of piglets in images," *Machine Vision and Applications*, vol. 8, no. 3, pp. 187–193, 1995. [Online]. Available: <http://dx.doi.org/10.1007/BF01215814>
- [26] P. Ahrendt, T. Gregersen, and H. Karstoft, "Development of a real-time computer vision system for tracking loose-housed pigs," *Computers and Electronics in Agriculture*, vol. 76, no. 2, pp. 169 – 174, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169911000263>
- [27] M. Kashiha, C. Bahr, S. Ott, C. P. Moons, T. A. Niewold, F. Ödberg, and D. Berckmans, "Automatic identification of marked pigs in a pen using image pattern recognition," *Computers and Electronics in Agriculture*, vol. 93, no. 0, pp. 111 – 120, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S016816991300029X>
- [28] N. Hansen and S. Kern, "Evaluating the cma evolution strategy on multimodal test functions," in *Parallel Problem Solving from Nature - PPSN VIII*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, vol. 3242, pp. 282–291. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-30217-9_29
- [29] N. Hansen, "Benchmarking a bi-population cma-es on the bbob-2009 function testbed," in *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, ser. GECCO '09. New York, NY, USA: ACM, 2009, pp. 2389–2396. [Online]. Available: <http://doi.acm.org/10.1145/1570256.1570333>
- [30] L. Świrski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12. New York, NY, USA: ACM, 2012, pp. 173–176. [Online]. Available: <http://doi.acm.org/10.1145/2168556.2168585>

A Computational Translation of the Phaistos Disk

Peter Z. Revesz

Abstract— For over a century the text of the Phaistos Disk remained an enigma without a convincing translation. This paper presents a novel semi-automatic translation method that uses for the first time a recently discovered connection between the Phaistos Disk symbols and other ancient scripts, including the Old Hungarian alphabet. The connection between the Phaistos Disk script and the Old Hungarian alphabet suggested the possibility that the Phaistos Disk language may be related to Proto-Finno-Ugric, Proto-Ugric, or Proto-Hungarian. Using words and suffixes from those languages, it is possible to translate the Phaistos Disk text as an ancient sun hymn, possibly connected to a winter solstice ceremony.

Keywords—Acrophonic principle, Cretan Hieroglyph, Linear A, Linear B, Phaistos Disk, Proto-Finno-Ugric, Proto-Hungarian

I. INTRODUCTION

Luigi Pernier discovered a fired clay flat round object with an archaic form of writing at the Phaistos palace on the island of Crete in 1908. The object called the Phaistos Disk (also spelled Phaistos Disc) was the subject of several decipherment or translation attempts that did not yield any convincing results. For example, in their decipherment attempts, Faucounau [8] and Fisher [9] assume an archaic form of Greek, Aartun [1] assumes a Semitic language, Achterberg et al. [2] assume Luwian, Kovar [14] uses Proto-Slavic, Kvashilava [10] assumes Georgian, and Owens [15] assumes some Indo-European language. Duhoux [5] is a critique of previous decipherment attempts.

Not only does the language of the Phaistos Disk remain unknown, but even its authenticity was questioned by some researchers [5]. However, most researchers agree with Duhoux [4] that the Phaistos Disk is a Bronze Age Minoan artifact created between 1850 B.C. and 1600 B.C. on the island of Crete. The symbols on the disk have numerous connections to other native Cretan writings, which were first classified by Arthur Evans, the explorer of Knossos Palace, as Cretan Hieroglyph, the Linear A and the Linear B scripts [7]. In 1952 Michael Ventris gave a decipherment of Linear B as described in Chadwick [3]. Hooker [12] gives a good introduction to Linear B. The Cretan Hieroglyph [23] and the Linear A scripts are also not deciphered.

Most decipherment attempts relied heavily on the acrophonic principle, which is the taking of the first sound of a word referred to by an object. The acrophonic principle has

several problems. First, a symbol may be interpreted as denoting many different objects. Second, the depicted object could have many synonyms in the native language. Third, each of the synonym words may have gone through a linguistic development where the initial sound changed. The combination of these three problems almost guarantees that we can derive by the acrophonic principle numerous beginning sounds for each symbol

In this paper we give a translation of the Phaistos Disk. Unlike previous decipherment attempts, our decipherment relies only minimally on the acrophonic principle. Instead, we use the already established correspondences between Cretan writing symbols and other ancient scripts with known sound values [18]. These other ancient scripts include the Phoenician [21], the South Arabic [22], the Greek and the Old Hungarian (see Forrai [11] and Varga [20]) alphabets.

Our approach to the translation of the Phaistos Disk is guided by our previous study of biological evolution [16], [17], [19]. The sound changes within a word are similar to genetic mutations. While many mispronunciations of words are possible, certain mispronunciations are easier to produce than others spread more easily. Similarly, while many types of mutations could occur on a genome, only the beneficial mutations are likely to spread to successive generations of descendants.

This paper is organized as follows. Section II outlines a semi-automatic translation method. Section III describes a transliteration of the Phaistos Disk text. Section IV describes a basic dictionary of Proto-Finno-Ugric and Proto-Hungarian words and suffixes and their consonant base representation. Section V presents the translation of the Phaistos Disk using the dictionary. Finally Section VI gives some conclusions and directions for future work.

II. THE TRANSLATION METHOD

We outline below a five-step translation method for the Phaistos disk.

1. Transliterate the symbols on the Phaistos disk using the sound correspondences recently identified in [18]. Some symbols with unknown sound values are not transliterated but are denoted by numbers.
2. Set up a Proto-Finno-Ugric and Proto-Hungarian dictionary. The dictionary needs to include the most common and oldest prefixes and suffixes.
3. For each word in the dictionary find its consonant base.

Peter Z. Revesz is with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA (revesz@cse.unl.edu).

4. Find matches between the transliterated text and the words in the dictionary. In the match only the consonant bases are used and vowels are ignored. Each symbol with an unknown sound value can be matched to any consonant or vowel sound, but it has to take the same (or similar) sound value at each of its occurrence. Choose between alternatives based on grammatical correctness.
5. Form sentences and translate them into a modern form.

In the above method, Step 1 can be computerized because it is a simple symbol substitution problem. Step 4 can be also partially computerized because the matching between the Phaistos Disk text and the dictionary requires string searching. We can simply take each word in the dictionary and search for all of its occurrences in the text. We used a simple string searching computer program facilitate this matching process. However, the selection of the best-fit word to each location of the text requires some human judgment and grammatical understanding. Some grammatical knowledge can be built into computer software, like grammar checkers, but we did not have anything available for Proto-Finno-Ugric and Proto-Hungarian, which we presumed to be close relatives to the Phaistos Disk language because of the connection between the Phaistos Disk symbols and the letters of the Old Hungarian alphabet [18].

1  /m/	2  /k/	3  /v/	4  /v/	5  /ε/	6  /n/	7  /ts/	8  /m/	9  /s/	10  /t/
11  /i/	12  j*	13  /m/	14  /s/	15  /s/	16  /t/	17  /s/	18  /t/	19  /t/	20  /t/
21  u*	22  /g/	23  /d/	24  /z/	25  /o/	26  /p/	27  /j/	28  /l/	29  /j/	30  /b/
31  /n/	32  /l/	33  /h/	34  /m/	35  /n/	36  /r/	37  /k/	38  /f/	39  /z/	40  /u/
41  /j/	42  /j/	43  /j/	44  /j/	45  /cc/					

Fig. 1 Each element in this matrix lists from top to bottom the following corresponding triplet: A. Evan’s numbering, Phaistos disk symbol, and IPA sound symbol. These associations are from [18] except for symbol 28, which is new. Here j* means /j/, /jom/ or /jon/, and u* means /u/ or /uz/.

III. A TRANSLITERATION OF THE PHAISTOS DISK

Arthur Evans [7] gave an enumeration of the forty-five Phaistos Disk symbols, which are listed according to his order in Figure 1. Below his enumeration, we added the putative sound values based on a recent study of the similarities between the Phaistos Disk symbols and some ancient script symbols whose sound values are already known [18]. That study did not include the sound value of symbol 28. We assume that symbol 28 has a sound value of /l/ because it shows a leg, which would be in Proto-Finno-Ugric *lu, from which derives both Finnish luu (bone) and Hungarian láb (leg). In addition, [18] assumed /r/ for symbol 9 in analogy with Phoenician and South Arabic. However, the rhotacism /n/ → /r/ occurs in many languages such as Aramaic. Hence we assume that symbol 9 was /n/ originally.

Given the enumeration and the known sound values in Figure 1, side A of the Phaistos Disk can be converted into the following sequence when reading from the center to the edge of the disk:

f-3-ts	m-13	u*-k-n-f-j*-k
f-3-ts	n-t-d	m-13-j*-k
j*-p-n	t-17-s-v	f-s-l-m-f-j*-k
p-n-j*-k	m-l	s-d-ts-o-f-k
p-n-j*-k	d-h	u*-k-n-f-j*-k
p-n-j*-k	m-l	s-d-ts-o-f-k
i-z	f-d-l-j*-k	n-t-jj-j*-k
n-p-n	v-40-jj-m	ε-44-f
j*-v-cç-f	s-v-j*-k	m-j-j
v-cç-j	h-40-4-j*-k	s-m-13-j*-k
	j*-40-z	

Similarly, side B can be converted into the following:

v-cç	o-d-m-j	v-d-n-v-k	v-s-z-b-n
ε-v-r-j-g		z-s-d-v	
v-cç-v		n-s-v	o-d-m-f
ε-v-r-j-g		v-cç-j	13-ε-j
ε-v-r-j		m-f-n-k	h-z-l-n-v
m-h-j		s-m-16	n-20-z-z-j
m-f-o-f	40-r-p-k	n-40-z-v	o-42-k-g
s-m-13-v-s		h-z-m-13	
43-s-d-16		j*-20-z-h	f-o-g
5-d-k-k		n-v-cç-f	v-40-g-j*-k

In the above, we highlighted in gray the words or phrases that are repetitions of earlier words or phrases. These highlighted parts of the text do not need a separate translation. We also highlighted in bold the first letter of the words that have below them a slash mark on the disk. The slash seems to be added to the symbols by hand. When we break the text up into lines such that the new lines start with the bold letters, then we get in most lines three words or phrases, assuming that each block of the disk is a word or short phrase. The repetitions highlighted in gray tend to be exactly below their earlier occurrences. Some of the repetitions may be refrains. The structured gray repetitions and the relatively equal lengths of the lines suggest that the text may be either a poem or a song.

IV. A PROTO-FINNO-UGRIC AND PROTO-HUNGARIAN WORD AND SUFFIX DICTIONARY

We collected a set of Proto-Finno-Ugric, Proto-Ugric and Proto-Hungarian words using the etymological dictionary [24]. Each word was represented by a consonant base, which was obtained by omitting the vowels. We also use the convention of putting a star symbol before any word that is a hypothetical proto word in any language. When it is necessary, we indicate the source of each word by adding to the word the name of the language in a superscript. For example, the Proto-Ugric word *mäle (warm) was represented by the consonant base m-l. The following table gives some examples from the dictionary.

Base	Cognate or Proto Word	Hung. Word	Meaning
-ts		-ci	diminutive suffix
d-h		düh	anger
d-s	*ipse	édes	sweet, dear
d-tj̄		dicső	glorious
f-j	*päŋe	fej > fő	head > chief
h-l	*kale	háló	fishing net
h-j		hajó	ship
j-n	*jäŋe	jön	come
j-n-k		junk	1 st person plural verb suffix, present tense
j-v-t	*jome (good)	javít	improve, help
k-n-d		kende	ruler
k-n-z		kenéz	judge
m-l	*mäle	meleg	warm
m-l	*mele	mély	deep
m-n-d		mind	all
m-f	*mu	más	other
n-p		nap	sun
n-t	*jome	indít	make sm start > plead
n-t	ümetöl ^{Mansi}	ünöt int	protect caution
ŋ		anya	mother
p-d-l	*pentele	fedél	cover, protector
p-j	*päje	fejér, fény	light, white
p-ŋ			
f		és	and
f-t	šit ^{Mansi}	süt	bake, shine
s	*icā	ős	ancestor
s-l	*šalke	szál	rod, thread
t-v-s	tūje ^{Mansi} tulis ^{Permi}	tavas	spring
v-l-g	*βalke	villog virrad	to shine to rise (Sun)
v-t	*βete wit ^{Mansi}	víz	water
z		izzó	hot
-z	*-t	-z	verb forming suffix

Unfortunately, the symbols used to describe the word pronunciations in [24] do not follow the standard IPA (International Phonetic Alphabet). Hence we changed some of the vowel symbols in [24] to the corresponding IPA symbol. Since the consonant symbols in [24] correspond much better to the IPA, we left them intact in the table.

V. A TRANSLATION OF THE PHAISTOS DISK

A. Sound Changes

The Proto-Finno-Ugric language is assumed to have undergone several sound changes in reaching the Proto-Ugric and the Proto-Hungarian stage of language development. We presume that the language of the Phaistos Disk is close to Proto-Hungarian. Therefore, the language of the Phaistos Disk shares some of the common sound changes that have been identified between Proto-Finno-Ugric and Hungarian. Some of the common sound changes include the following:

$$p > f$$

$$\eta > j > i$$

$$\text{tj̄} > \text{ts}$$

$$z > ʒ$$

$$g > \text{ŋ}$$

$$t > \text{c}̣ \text{ or } t > z$$

Sometimes adjacent pairs of consonants change together in predictable way. For example, the following is a relatively common sound rule change:

$$*nt > nd > d$$

B. Matches between the Text and the Dictionary

The sound changes slightly influence the way we do string searches. We need to search the Phaistos Disk text to find both the exactly matches and the approximate matches where some of the consonant sound were changed using the sound change rules.

We also need to search for both root words and suffixes. The suffixes are restricted to the end of the blocks into which the Phaistos Disk is divided. The blocks are clearly indicated on the disk by the scribe. Each block can be assumed to be either a single word or a phrase. Words do not run across blocks.

In the following, we use the following color highlighting.

Yellow – root word.

Blue – suffix.

Gray – repetitions of earlier words or phrases.

Green – voice assimilation of the root due to suffix.

The next table shows the result of the string matches color-coded according to the above legend. We added some extra grammatical markers that are not listed in the above dictionary.

For example, in the middle of the third row the –a suffix describe a possessive relationship similar to the English possessive ‘s, but while the English language marks the possessor, the Hungarian language marks the possessed object. Hence the possessive phrase “fény tavasz-a” can be translated as “light’s spring.” That phrase is meaningful if we recall that the word “tavasz” derives from a rising of lake water levels [24]. The “light’s spring” may have referred to either any sunrise or a lengthening of the days after a winter solstice.

f-3-ts	m-13	u* k-n-f-j-j*-k
fö-3-ci	mind	az kenes-s-jün-k
chief 3	all	the ruler-1PL.POSS
f-3-ts	n-t-d	m-13-j*-k
chief 3	ünöt-öd	mind-jun-k
	protect-2SG	all 1PL
j* p-n	t-17-s-b	f-g-l m-f-j*-k
jő fény	tavasz-a	süssél másik
come light	spring-POSS	shine again
p-n-j*-k	m-l	s d-ts-o f-k
fény-jenek	meleg	s discsó szálak
shine-3PL	warm	and glorious ray-PL
p-n-j*-k	d-h	u*-k-n-f-j-j*-k
shine-they	anger	the ruler-1PL.POSS
p-n-j*-k	m-l	s-d-ts-o-f-k
i-3 f-d-l-j*-k	v-40-jj m	n-t-j-j*-k
izzó fedel-jünk	villog-ni	indít-jünk
hot cover-3PL.POSS	rise-INFIN	plead-1PL
n-p-n	s p-j*-k	ε-44-f
napfé-ny	ős anyá-nk	édes
sunlight	ancestor mother-3PL.POSS	dear
j*-v-cq-f	h-40-4-j*-k	m-j-j
javi-ss/javitys	hálóink / hajóink	mély
help	fishing net/ship-3PL.POSS	deep
v-cq-j	j*-40-z	s m-13-j*-k
vízi	jön-	
water-LOC	go-VERBALIZER	and all 1PL

Fig. 2 The above rough translation shows the result of matching the text with proto-words from the dictionary using the consonant bases with allowance for the sound rule changes described in the text and root changes due to voice assimilation. The root words are highlighted in yellow, the suffixes in blue, the voice assimilations in green, and repeated elements in gray.

From the context, there is a suspicion that symbol 3 may refer to some ancestor spirit or divinity, perhaps affectionately called with a diminutive suffix if the ending -ts is not part of the name. In fact, the text seems to be a hymn to a solar divinity to bring back (stronger) sunlight to the earth. This sun hymn may have been said at a winter solstice ceremony. Hence side A of the Phaistos Disk can be translated into the following text:

Chief god of all, our ruler.
 Chief god, you protect all of us.
 Come light’s spring, shine again
 Shine warm and glorious rays.
 Light up strong, our ruler.

Shine warm and glorious rays.
 For our hot cover, to rise we pray.
 Sunlight, our dear ancestor mother,
 Help our ships sailing on the seas
 And all of us.

We translated in a similar manner side B, which also seems to refer to a sun divinity, but because of space limitations we postpone presenting that translation in the forthcoming journal version of this paper.

VI. CONCLUSIONS AND FUTURE WORK

Based on recent advances in the comparative study of ancient scripts [18], we could start our translation by having a plausible sound value for the majority of the Phaistos Disk symbols. The sound values presented in [18] seem corroborated by being able to form words, phrases, and sentences with proper grammar after finding matches between the Phaistos Disk text and etymologically plausible proto words from the Proto-Finno-Ugric and the Proto-Hungarian languages [24]. Moreover, the translation yields a sun hymn, which seems to fit in Bronze Age cultural context. There are many Bronze Age cultures where the sun was worshipped. For example, there are Babylonian hymns to Shamash, their sun god, and in ancient Egypt, around 1350 B.C. Pharaoh Akhenaten also wrote several hymns to the sun. The relationship between the newly translated text and other ancient sun hymns may be an interesting direction for further study.

REFERENCES

- [1] K. Aartun, *Die minoische Schrift : Sprache und Texte vol. 1*, Wiesbaden, Harrassowitz, 1992.
- [2] W. Achterberg, J. Best, K. Enzler, L. Rietveld, F. Woudhuizen, *The Phaistos Disc: A Luwian Letter to Nestor*, Publications of the Henry Frankfort Foundation vol XIII, Dutch Archeological and Historical Society, Amsterdam 2004.
- [3] J. Chadwick, *The Decipherment of Linear B*, Cambridge University Press, 1958.
- [4] Y. Duhoux, *Le Disque de Phaestos*, Leuven, 1977.
- [5] Y. Duhoux, “How not to decipher the Phaistos Disc,” *American Journal of Archaeology*, 104 (3), 2000, pp. 597–600.
- [6] J. M. Eisenberg, “The Phaistos Disk: One hundred year old hoax?” *Minerva*, July/August 2008, pp. 9–24.
- [7] A. J. Evans, *Scripta Minoa: The Written Documents of Minoa Crete with Special Reference to the Archives of Knossos*, Volume II, Classic Books, 1909.
- [8] J. Faucounau, *Le Déchiffrement du Disque de Phaistos: Preuves et conséquences*. L’Harmattan, Paris/Montreal 1999.
- [9] S. R. Fisher, *Glyph-Breaker*, Springer, 1997.
- [10] G. Kvashilava, *On Reading Pictorial Signs of the Phaistos Disk and Related Scripts*, Ivane Javakhishvili Institute of History and Ethnology, Tbilisi, 2010.

- [11] S. Forrai, *The Old Hungarian Writing from Ancient Times to the Present*, (in Hungarian), Antológia Kiadó, 1994.
- [12] J. T. Hooker, *Linear B: An Introduction*, Bristol Classical Press, 1980.
- [13] G. Hosszú, *Heritage of Scribes: The Relation of Rovas Scripts to Eurasian Writing Systems*, Rovas Foundation Hungary, 2013.
- [14] J. Matejka, "Translation of the Phaistos Disc," WM Magazine, January 19, 2011. <http://www.wmmagazin.cz/view.php?cislocclanku=2011010004>
- [15] G. A. Owens, The Phaistos Disk and related inscriptions, 2008-2014. http://www.teicrete.gr/daidalika/pages/page.php?page=phaistos_disk
- [16] P. Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, Springer, New York, 2010.
- [17] P. Z. Revesz, "An algorithm for constructing hypothetical evolutionary trees using common mutations similarity matrices," *Proc. 4th ACM International Conference on Bioinformatics and Computational Biology*, ACM Press, Bethesda, MD, USA, September 2013, pp. 731-734.
- [18] P. Z. Revesz, "A computational study of the evolution of Cretan and related scripts," *Proc. International Conference on Applied Mathematics, Computational Science and Engineering*, October 2015.
- [19] M. Shortridge, T. Triplet, P. Z. Revesz, M. Griep, and R. Powers, "Bacterial protein structures reveal phylum dependent divergence," *Computational Biology and Chemistry*, 35 (1), 2011, pp. 24-33.
- [20] G. Varga, *Bronzkori Magyar Írásbeliség*, Írástörténeti Kutató Intézet publication, 1993.
- [21] Wikipedia, "Phoenician alphabet," downloaded July 6, 2015. Available: https://en.wikipedia.org/wiki/Phoenician_alphabet
- [22] Wikipedia, "South Arabian alphabet", downloaded July 5 2015. Available: https://en.wikipedia.org/wiki/South_Arabian_alphabet
- [23] J. G. Young, "The Cretan Hieroglyphic script: A review article," *Minos* 31-32 (1996-1997[1999]) 379-400.
- [24] G. Zaicz, chief editor, *Etimológiai Szótár: Magyar Szavak és Toldalékok Eredete*, (*Etymological Dictionary: Origin of Hungarian Words and Affixes*), Tinta Könyvkiadó, 2006.

Peter Z. Revesz holds a Ph.D. degree in Computer Science from Brown University. He was a postdoctoral fellow at the University of Toronto before joining the University of Nebraska-Lincoln, where he is a professor in the Department of Computer Science and Engineering. Dr. Revesz is an expert in databases, data mining, big data analytics and bioinformatics. He is the author of *Introduction to Databases: From Biological to Spatio-Temporal* (Springer, 2010) and *Introduction to Constraint Databases* (Springer, 2002). Dr. Revesz held visiting appointments at the IBM T. J. Watson Research Center, INRIA, the Max Planck Institute for Computer Science, the University of Athens, the University of Hasselt, the U.S. Air Force Office of Scientific Research and the U.S. Department of State. He is a recipient of an AAAS Science & Technology Policy Fellowship, a J. William Fulbright Scholarship, an Alexander von Humboldt Research Fellowship, a Jefferson Science Fellowship, a National Science Foundation CAREER award, and a "Faculty International Scholar of the Year" award by *Phi Beta Delta*, the Honor Society for International Scholars.

Knowledge Base Modelling for Lignocellulosic Materials Optimal Processing by Means of Fuzzy Fault Tree Analysis

Dimitrios Batzias, Dimitrios Sidiras, Christina Siontorou, Fragiskos Batzias,
Michael Tsapatsis, Ivo Safarik

Abstract— Biomass processing units (BPUs) as Small/Medium Enterprises (SMEs) of the industrial subsector of Waste Lignocellulosic Materials Exploitation (WLME) suffer of lack of means and knowledge to apply quality and process control. This work deals with modelling of a Knowledge Base (KB) for lignocellulosic materials optimal processing, established within a Technical Support Center (TSC), capable to provide the required knowledge to BPUs, by using as a knowledge vehicle the remedial advice offered by the TSC when defects are observed in the products (herein lignocellulosic adsorbents) without any evidence about the causing fault in the production line. For this purpose, a methodological framework has been developed under the form of an algorithmic procedure, including 34 activity stages and 11 decision nodes. Fault Tree Analysis (FTA) in its fuzzy version (to deal with uncertainty) was adopted as an inference engine for knowledge extraction and know-how transfer. Since WLME is usually connected with production units of the primary/secondary sector of the economy (i.e., agriculture and industry) and the product under investigation is a biomass based material of environmental importance, most optimization procedures are considered within an Industrial Ecology framework. An implementation referring to faulty biomass processing is presented, with emphasis on cause-of-fault investigation/identification and communication protocols functioning. Fuzzy multicriteria analysis (MCA) is also performed for ranking alternatives when they appear eventually or set *a priori* in the fault tree. A university spin-off is suggested as mostly suitable to

play the role of TSC, inasmuch as this view is further supported through experience accumulated into the Laboratory of Simulation of Industrial Processes in our Department at the University of Piraeus.

Keywords— biomass processing, fuzzy fault tree analysis, knowledge base, optimization.

I. INTRODUCTORY ANALYSIS

The main processes met in most biomass processing units (BPUs) providing intermediate/end products to the secondary sector of the economy are size reduction (cutting), milling, fluidization, heat and mass transfer, hydrolysis, bioconversion and surface treatment. Secondary/tertiary processes/operations (e.g., leaching, desorption, combustion) may take place for decontaminating/recycling wastes to satisfy environmental and/or market needs. The great majority of the firms involved in this business consists of Small/Medium Enterprises (SMEs). Even when specific biochemical reactions are involved, which give, by their nature, very precise results, the variance of control parameters, which is usually high under industrial conditions, implies faults at processing level and defects at product level. Some of these defects are not observable but will possibly influence the function of the corresponding product during its lifecycle. Most SMEs have not the proper equipment and the required knowhow in advanced quality control in order to (a) diagnose/identify/remedy a significant number of kinds of defects, (b) optimize production conditions and (c) establish/maintain a diagnostic Knowledge Base (KB) that will facilitate technology transfer/adaptation.

On the other hand, there are certain chemical/biochemical departments operating within large companies possibly engaged, directly or indirectly, with the processes mentioned above. These departments are not competing the SMEs, because their production is dedicated to serve the main purpose of the large enterprise to which they belong. They also (a) follow medium/high quality standards, (b) engage personnel with increased qualifications and (c) operate well equipped laboratories for quality control; nevertheless, a great deal of knowledge remains either tacit or dispersed in thousands of official routine papers, completely unstructured and hardly retrievable.

This research has been co-financed by the European Union (European Social Fund- ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS. Investing in knowledge society through the European Social Fund - Project: THALIS - University of Piraeus - Development of New Material from Waste Biomass for Hydrocarbons Adsorption in Aquatic Environments (MIS 377356).

F. A. Batzias is with the Dep. Industrial Management & Technology, Univ. Piraeus, 80 Karaoli & Dimitriou, GR18534 Piraeus, Greece (corresponding author phone: +30-210-4142360; fax: +30-210-4142392; e-mail: fbatzias@unipi.gr).

D. K. Sidiras, D. F. Batzias and C. G. Siontorou are with the Dep. Industrial Management & Technology, Univ. Piraeus, 80 Karaoli & Dimitriou, GR18534 Piraeus, Greece (e-mail: sidiras@unipi.gr, dbatzias@unipi.gr, csiontor@unipi.gr).

M. Tsapatsis is with the Department of Chemical Engineering & Material Sciences, University of Minnesota, 421 Washington Ave. SE, Minneapolis, USA (e-mail: tsapa001@umn.edu).

I. Safarik is with the Department of Nanobiotechnology, Institute of Nanobiology and Structural Biology, GCRC Academy of Sciences, Na Sadkach 7, 370 05 Ceske Budejovice, Czech Republic (e-mail: ivosaf@yahoo.com).

The aim of this work is to present the development of a methodology for creating/maintaining a diagnostic KB for knowledge acquisition/transformation to meet the quality needs of the lignocellulosics market, which can be used for (a) knowledge management either in the production Department of a Large Company (DLC) or in the operation of a Technical Support Center (TSC) and (b) technology transfer towards a SME in a mode of routine consultation on a quasi online / real-time basis. Such a TSC can be established either in a bottom-up way by the interested SMEs or in a top-down way by (i) a public Organization/Institution, (ii) an applied research center, (iii) a university, (iv) a DLC allowed to offer services to externals.

Since WLME is usually connected with production units of the primary/secondary sector of the economy (i.e., agriculture and industry) and the product under investigation is a biomass based material of environmental importance, most optimization procedures are considered within a framework of protection against pollution. As a paradigm, we can determine optimal pollutant concentration (in effluent wastewater) P_{opt} through a tradeoff equilibrium between environmental and economic cost, $C_1(P)$ and $C_2(P)$ respectively, by minimizing total cost C , as follows. The partial cost C_1 is an increasing function of P with an increasing rate (i.e. $dC_1/dP > 0, d^2C_1/dP^2 > 0$), since P -increments, ΔP , cause higher environmental damage (including impact on human health) in the region of higher P -values in comparison with equal ΔP taking place in the region of lower P -values.

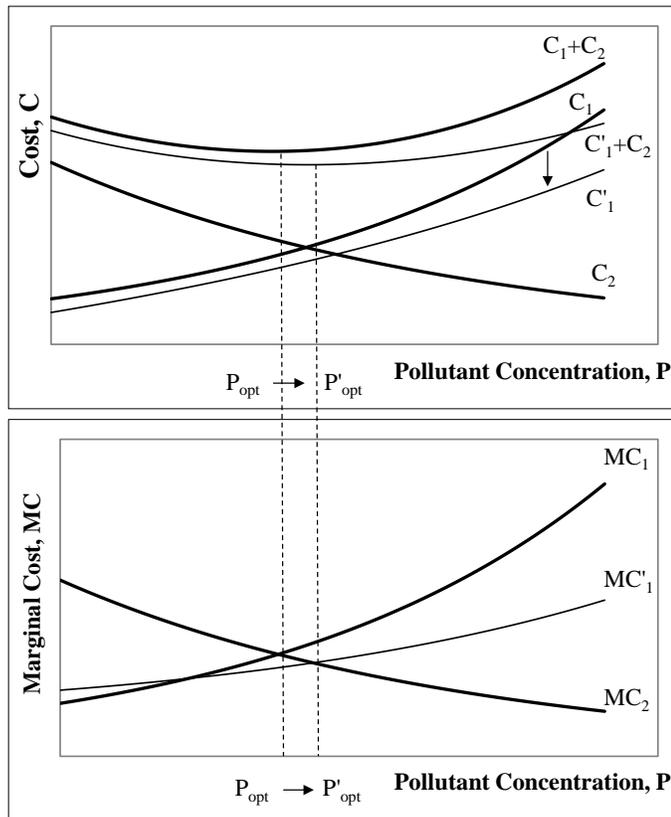


Fig. 1a. Optimization tradeoff and sensitivity analysis: dependence of environmental and economic cost, C_1 and C_2 , on pollutant concentration and shifting of P_{opt} to P'_{opt} , when C_1 -movement is considered under industrial ecology conditions.

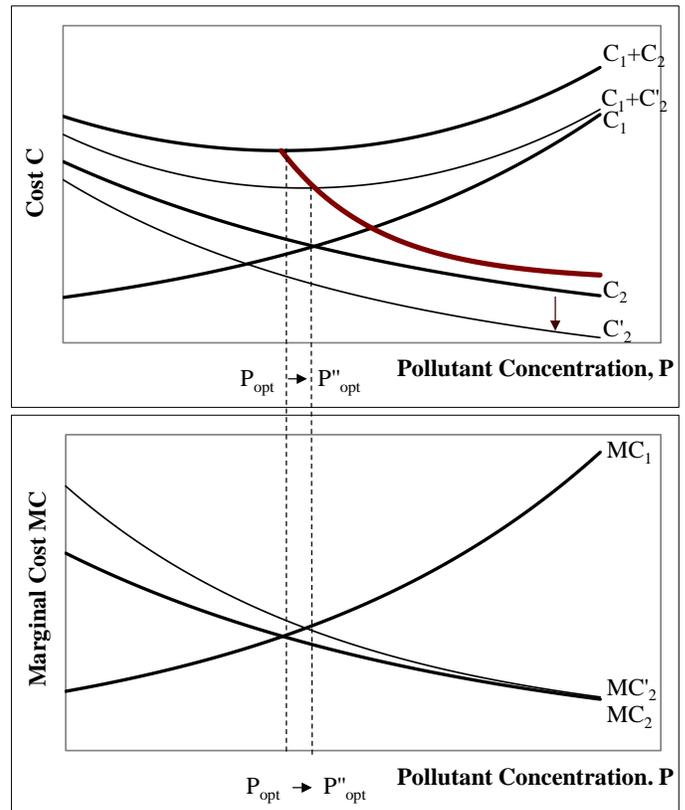


Fig. 1b. Optimization tradeoff and sensitivity analysis: dependence of environmental and economic cost, C_1 and C_2 , on pollutant concentration and shifting of P_{opt} to P''_{opt} , when C_2 -movement is considered under industrial ecology conditions; the thick line in the upper diagram represents the locus of C_{min} -points.

On the other hand, C_2 is a decreasing function of P with an increasing algebraic or a decreasing absolute rate (i.e., $dC_2/dP < 0, d^2C_2/dP^2 > 0$ or $d|dC_2/dP|/dP < 0$), because of the validity of the Law of Diminishing (marginal or differential) Returns (LDR), implying disproportionately higher economic cost for higher efficiency. Evidently, P_{opt} is the abscissa of C_{min} -point at $MC_1 = MC_2$, where $MC_1 = dC_1/dP$ and $MC_2 = |dC_2/dP|$, are the marginal partial costs at the tradeoff equilibrium point.

From the Industrial Ecology point of view (considered either in the narrow/topography or in the broader/sustainability sense, as quoted in [1], [2], respectively), the decrease of lignocellulosic waste disposed in landfills or abandoned/burned in farms implies a corresponding decrease of environmental damage caused by primary biomass producers or processors (estimated through 'supply side analysis'). As a result, the C_1 -curve is moving downwards to C_1' , becoming more flat, since there is a higher margin for environmental cost decrease in the region of higher P -values; consequently, P_{opt} is shifting to P'_{opt} , where $P'_{opt} > P_{opt}$, as shown in Fig. 1a. From the same point of view, the provision of inexpensive raw lignocellulosic material for producing innovative adsorbents will cause movement of the C_2 -curve to a lower position C_2' , giving a steeper form to it, since higher deviation (in relation with its initially estimated position) is expected in the region of lower P -values, where economic cost of wastewater treatment (for achieving higher efficiency in order to lower pollutant concentration in the effluent) is already high, according to the LDR; consequently P_{opt} is shifting to P''_{opt} ,

where $P''_{opt} > P_{opt}$, as shown in Fig. 1b. It is to be noted that the vectors $(P'_{opt} - P_{opt})$ and $(P''_{opt} - P_{opt})$ have the same direction, when an Industrial Ecology framework is adopted.

Finally, the optimization procedure described above is leading to lesser total cost and higher P_{opt} -values, which can be realized on condition that the new value is permitted by (i) the environmental legislation set by the State and/or the Local Authorities, and (ii) the rules agreed with the polluter (as a producer or an environmental agent) when the permission for the respective activities was granted. Nevertheless, a more global approach should be adopted, since P is not the only independent/explanatory variable to be optimized, while other criteria should also be taken into account.

II. METHODOLOGY

The methodology adopted herein, under the form of an algorithmic procedure, relies heavily on fuzzy reasoning within a top-down/bottom-up dialectic scheme of Fault Tree Analysis (FTA), in a fuzzy version (see [3], [4]) to deal with uncertainty, and includes the following 34 activity stages and 11 decision nodes (see Figure 2 for their interconnection):

1. Setup of the network including the TSC and the WLME SMEs under consideration.
2. Setup of the communication protocols.
3. Observation/measurements of/on low quality product (LQP), expressed in terms of the specifications set a priori, as regards the property under examination.
4. Examination of treatment/production conditions.
5. Remedial proposal, confirmation and corrective action.
6. Complete description of the fault in an internal report.
7. Preliminary identification of the fault via the DKB.
8. Recording of the production conditions, referring to the lot in which the defected output was observed.
9. Coding of message according to MF1 [variables/conditions/documents (standard photos, diagrams) reported under strict specifications]; this Message Format (MF) is obligatory in the response/remedial – proposal sent by the consulting/supporting TSC to the supported WLME SME.
10. Coding of message according to MF2 [Message Format allowing for exposing the problem in a narrative way (accompanied by available documentation, e.g. non-standard photos), because a strictly prescribed form may not apply in the situation under consideration].
11. Coding of message according to MF3 [Message Format, that simply accompany defected lignocellulosic output or guide specimens (set purposely in the same batch or treated in the same way) when they are dispatched to the consulting TSC].
12. Electronic transmission of message to TSC, according to the established communication protocols in stage 2.
13. Conventional dispatch of message to TSC (representative sample of the defected lignocellulosic material is included).
14. Measurements in the Quality Control Laboratory, according to mutually accepted standards and recommended practices.
15. Delivery decoding, and (if necessary) interpretation of message (to assign contextual meaning when description in natural language is used for giving complementary information), according to the principles of 2nd order Cybernetics.
16. Knowledge-based identification of fault/failure in a possibilistic or probabilistic mode [5], [6].
17. Computer aided localization of the corresponding fault tree where the identified defect is the top event [7], [8].
18. Successive causes path identification through the assignment of fuzzy significance values on events by an expert (or group of experts participating in a Delphi method to investigate the causal chain most likely responsible for the fault appearance), contributing to system reliability and risk analysis [9], [10].
19. DKB Creation/enrichment receiving information internally (see Figure 2) and externally via an Intelligent Agent (IA), like the one described in [11].
20. Experimental confirmation (after scaling down) of the corrective action.
21. Updating of the fuzzy rules used in the fault tree, if necessary [12]-[14].
22. Conditional Proposal for correction action to the production unit.
23. Unconditional Proposal for correction action to the production unit.
24. Coding of response (feedback) message, according to MF1, and electronic transmission to SME.
25. Decoding of the message by the BPU and application of the corrective proposal (given by the TSC), after finding the correspondence to production conditions (proper interpretation, according to 2nd order Cybernetics).
26. Registration of new information in the BPU's local KB and the TSC's central DKB.
27. Design of crucial experiments, putting emphasis on ruggedness testing (e.g., according to ASTM standard guide E 1169), in order to resolve such conflicts.
28. Performance of measurements according to experimental design and statistical processing of the results.
29. Investigation on the remaining causes of conflict after evaluating new information.
30. Monte Carlo simulation and new experimental design, possibly including supplementary VPCs retrieved from the internal KB, if necessary.
31. Recombination/revision of the IA key words/phrases selected and the ontological concepts/relations used to search in external KBs.
32. Activation of the revised IA and processing of results.
33. Suggested measurements carrying out and comparison of results with the ones obtained previously.
34. Knowledge integration and conclusive results.
 - A. Is there any parameter or control variable or a combination of them out of the recommended interval or combination of intervals?

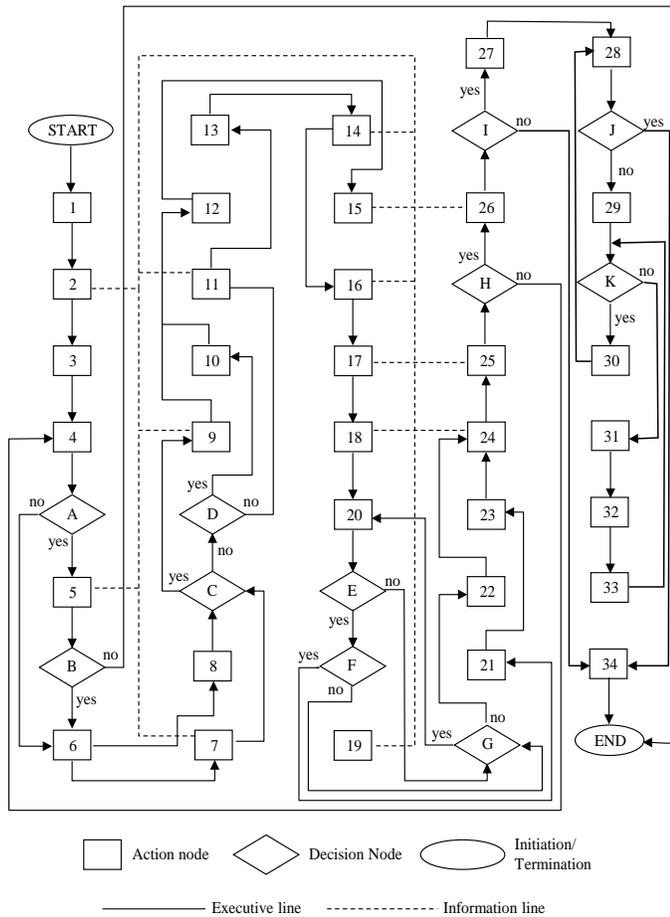


Fig. 2. The methodological framework, under the form of an algorithmic procedure, developed for interactive knowledge acquisition/processing within a KB and know-how transfer from a BPU to a TSC and vice versa.

- B. Is there any defected batch in the new lot?
- C. Is the preliminary identification complete?
- D. Is the unidentified fault describable under the specifications required by the means of MF2 terms?
- E. Is the confirmation of the path successful?
- F. Is the remedial proposal testing feasible?
- G. Is there another experimental test?
- H. Is the application successful?
- I. Is the new information (obtained in this way) implying any conflict with the knowledge or information already existing in the KBs?
- J. Has the conflict been resolved?
- K. Are the available results and the revised causal relations (forming an upgraded Ontology) adequate for a new experimental design?

Obviously, a Rules Based Reasoning (RBR) approach has been adopted as a background for the algorithmic procedure described above. Since, in practice, an interplay of rules and cases is unavoidable, all cases examined, either actually proved successful or unsuccessful, are stored to be used as raw material for Case Based Reasoning (CBR) when either an

unidentifiable defect appears in quality control performed in a BPU-lot or an holistic rearrangement of KB rules is needed in the TSC.

III. IMPLEMENTATION

The methodology described above has been successfully implemented in the case of knowledge acquisition/transformation to meet the quality needs of lignocellulosic adsorbents. Most of the necessary information was obtained from the Laboratory of Simulation of Industrial Process of the Department of Industrial Management at the University of Piraeus, Greece. The case example presented herein refers to biomass characterization described in a narrative way (see screenshot in Fig. 3), following the communication protocol MF2 (stage 10), since there is not corresponding standard terminology according to the respective fault tree (already in the DKB), and its vocabulary.

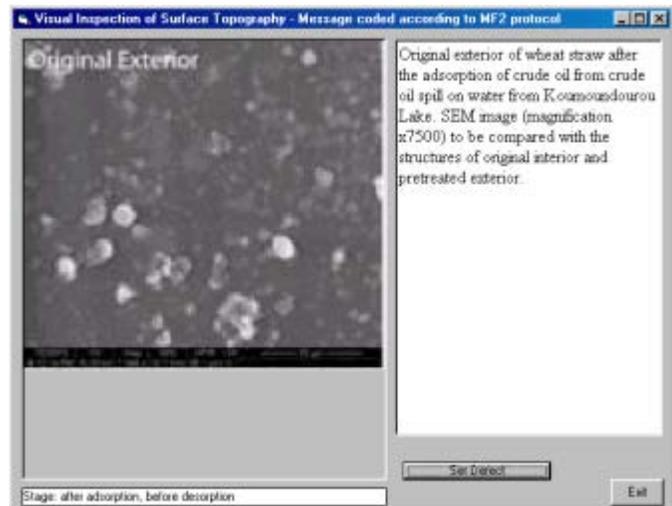


Fig. 3. Screenshot indicating electronic message from SME to TSC, coded according to MF2 protocol, referring to the heterogeneity of surface topography of the original exterior of wheat straw, after the adsorption of crude oil.

By simulating the TSC behavior, the system localizes, according to stage 17, as most relevant the defect 'low/uneven adsorptive capacity of modified biomass' presented as top event in Fig. 4. Applying the rules on input fuzzy data, we obtain the result on the top event in fuzzy form, which can be defuzzified to give a crisp number that is more familiar to the operator in the SME/BPU. In Fig. 5, an extract of a specimen run is shown, starting from the leaf 1.4.2.2 'high population of microorganisms in biomass before final storing' of the tree, which is experimentally found as one of the mostly responsible ultimate causes, and ending at the top event. The linguistic variables refer to each event, considered in the universe of discourse of the corresponding fault, and appear as indices (%), with terms L, M, H (Low, Medium, High, respectively), according to a three-division Likert scale. The rules are set in the usual IF-THEN form; e.g., IF 1.4.2.1 is M AND 1.4.2.2 is H THEN 1.4.2 is H.

The intermediate/final events in the tree under examination are described subsequently, while a sample of fuzzy input and intermediate/final output (first triad and second triad schemes, respectively, as shown in Fig. 4), represents FTA for the branch marked in gray (see Fig. 5).

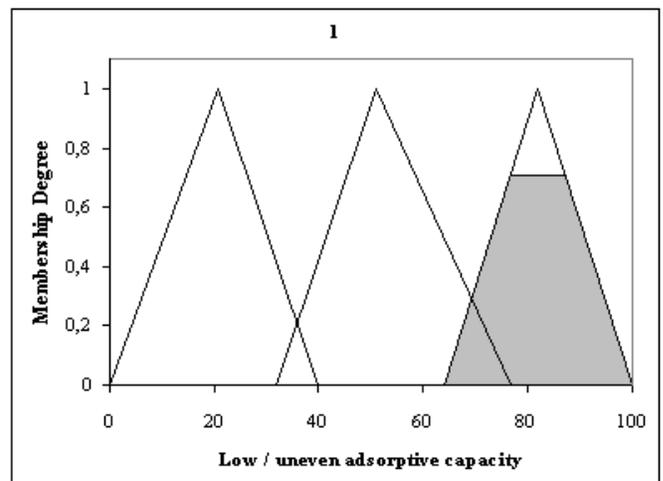
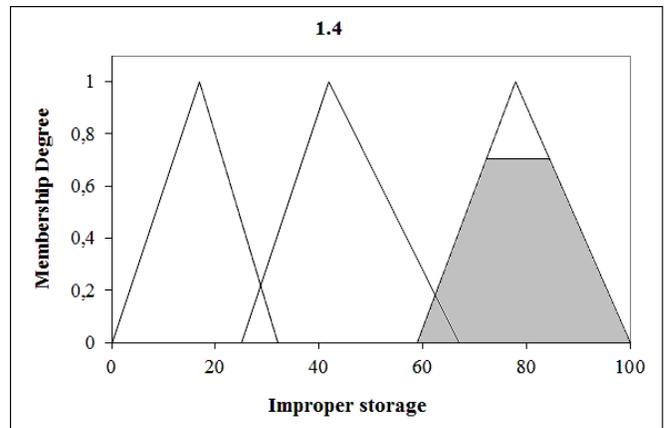
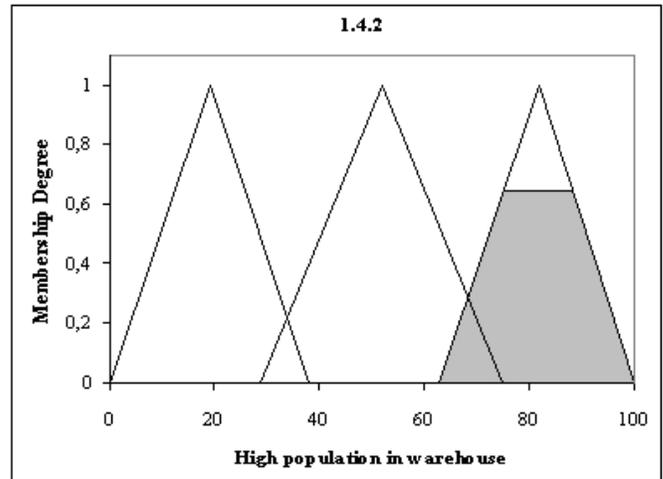
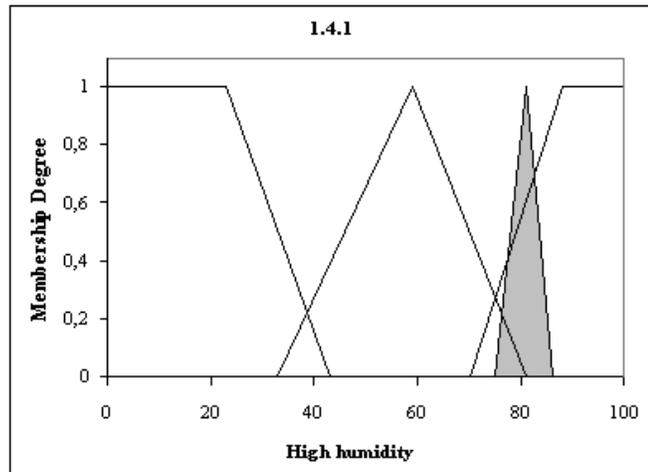
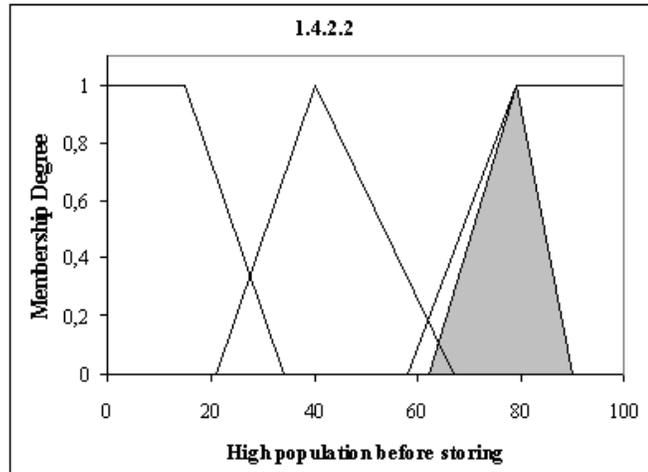
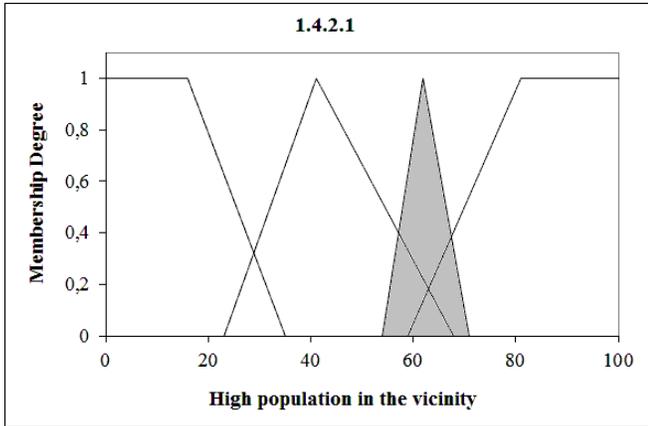


Fig. 4. A sample of fuzzy input and intermediate/final output (first triad and second triad of schemes, respectively), representing FTA for the right hand branch marked in gray, as shown in Fig. 5.

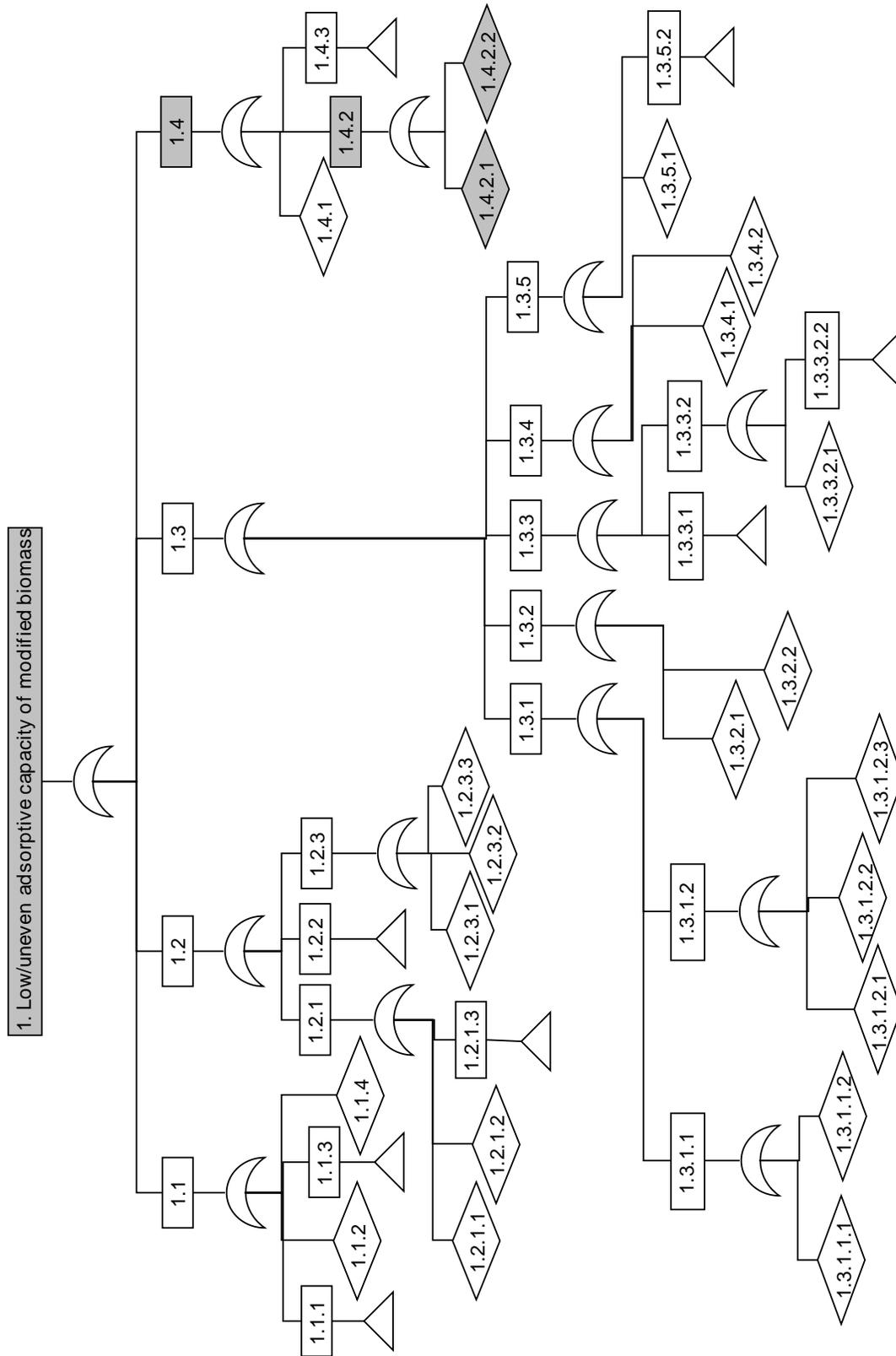


Fig. 5. Fault Tree Analysis (FTA), when the top event is "Low / uneven adsorptive capacity of modified biomass". The branch in gray represents a causal path, according to activity 18 described in the methodology section and depicted in the flowchart of Fig. 2 (see also Fig. 4). The symbols moon fraction (waxing/waning crescent), rectangular, rhombus (diamond), isosceles triangle stand for OR-gate (inclusive), top/intermediate event/cause, final event/cause, branch continuation in other subroutine.

- 1 Low / uneven adsorptive capacity of modified biomass
- 1.1 Raw biomass of high heterogeneity at cutting stage or waste delivery
- 1.2 Improper storage of cut/waste biomass in open air
- 1.3 Faulty modification
- 1.4 Improper storage conditions of packed/ensacked modified biomass (as final product) in warehouse
 - 1.1.1 Lignocellulosic original and/or waste materials have not been well seasoned.
 - 1.1.2 Improper mixing of biomass coming from quite different sources and exhibiting very different properties (porosity, constitution, including water).
 - 1.1.3 Biomass coming from plants suffering from a disease or injured by disease
 - 1.1.4 High concentration of insecticides/herbicides or other poisonous substances used for crop protection or wood preservation.
 - 1.2.1 Enzymatic hydrolysis.
 - 1.2.1.1 High temperature
 - 1.2.1.2 High humidity
 - 1.2.1.3 High population of microorganisms in the vicinity (quantitative modeling and qualitative approach, referring to kind of species).
 - 1.2.2 High deterioration of Lignocellulosic textures due to insect action.
 - 1.2.3 Chemical hydrolysis.
 - 1.2.3.1 Acidic hydrolysis
 - 1.2.3.2 Alkaline hydrolysis
 - 1.2.3.3 Autohydrolysis
 - 1.3.1 Low specific surface
 - 1.3.1.1 Low milling time
 - 1.3.1.1.1 Higher (than recommended through good practice methods for milling) cost decrease for electric energy saving.
 - 1.3.1.1.2 Underestimation of biomass properties (e.g., hardness, fragility, plasticity) as regards its resistance to size reduction.
 - 1.3.1.1.2 High agglomeration
 - 1.3.1.1.2.1 Very small biomass particles resulting as intermediate product in the process of size reduction.
 - 1.3.1.1.2.2 High temperature during milling
 - 1.3.1.1.2.3 High humidity in the mill
 - 1.3.2 Improper mean temperature during biomass modification.
 - 1.3.2.1 Temperature higher than the upper limit
 - 1.3.2.2 Temperature lower than the corresponding limit
 - 1.3.3 High standard deviation of temperature
 - 1.3.3.1 Wrong size/geometry of the reactor used for thermochemical modification
 - 1.3.3.2 Improper agitation.
 - 1.3.3.2.1 Wrong choice of agitation pattern.
 - 1.3.3.2.2 Wrong choice/application of simulation model in the scale up procedure performed for the optimization of agitation.
 - 1.3.4 Improper pressure during biomass modification.
 - 1.3.4.1 Temperature higher than the upper limit

- 1.3.4.2 Temperature lower than the corresponding limit
- 1.3.5 Impact of contaminant in the biomass mixture
 - 1.3.5.1 High concentration of contaminant
 - 1.3.5.2 Unacceptable kind of contaminant species.
- 1.4.1 High humidity in warehouse, decreasing available specific surface by partially filling the pores with water.
- 1.4.2 High population of microorganisms in warehouse, causing degradation of biomass texture, with dramatic impact on (i) porosity, and (ii) the surface concentration of chemical groups.
- 1.4.3 Presence of dangerous microorganism species
 - 1.4.2.1 High population in the vicinity
 - 1.4.2.2 High population in biomass before final storing (e.g., due to unsuitable intermediate storing or transportation under wrong conditions).

The impact of biomass properties (like porosity, specific surface, and acidic character of the surface, appeared in the fault tree as final/intermediate events under the code numbers 1.1.2, 1.2.3.1, 1.2.3.3, 1.3.1) and cost/availability on ranking raw material alternative sources is subsequently examined by means of fuzzy multicriteria analysis (MCA). A sample of fuzzy MCA for ranking two biomass and coal based adsorbents is presented below, by using technical literature [15]-[21] and human expertise, as information and comparative evaluation sources. The sum of weighted grades $a(i,j)$ for each alternative A_j gives the total criterion SWA_j for the final ranking. Since $SWA_2 > SWA_3 > SWA_1$. The preferred alternative is A_2 (pinewood biomass), which is a robust solution (as compared with A_1) when either $a(5,2)$ is changing from 6.58 to 2.11 or $a(5,1)$ is changing from 2.48 to 7.56, as proved by one-parameter sensitivity analysis shown in Fig. 6a and Fig. 6b, respectively; nevertheless, the two-parameter sensitivity analysis shown in Fig 6c reveals that when both $a(5,1)$ and $a(5,2)$ change simultaneously, the preference matrix solution is not robust, as compared only with the other biomass alternative A_1 , since this is not the case if the grade $a(5,3)$ of the alternative A_3 is taken into account, even if it is not changing too. The criteria are: meso-porosity, f_1 ; specific surface area, f_2 ; acidic character of surface, f_3 ; colonizability by bacteria, f_4 ; availability in the vicinity at reasonable cost, f_5 . The origin of the alternative adsorbents are: coconut biomass, A_1 ; pinewood, A_2 ; bituminous coal, A_3 .

Table I. The grades $a(i,j)$ ($i = 1, \dots, 5$ and $j = 1, 2, 3$) assigned to MCA preference matrix and the values f_i assigned to the elements of the weights vector, after defuzzification (consensus in input).

f_i	W_i	A_1	A_2	A_3	$W-A_1$	$W-A_2$	$W-A_3$
f_1	0,24	4,37	7,12	5,87	1,049	1,709	1,409
f_2	0,12	5,92	7,03	3,69	0,710	0,844	0,443
f_3	0,07	3,83	5,41	3,86	0,268	0,379	0,270
f_4	0,18	4,70	4,95	4,50	0,846	0,891	0,810
f_5	0,39	2,48	6,58	6,14	0,967	2,566	2,395
SWA_j	1,00				3,841	6,388	5,326

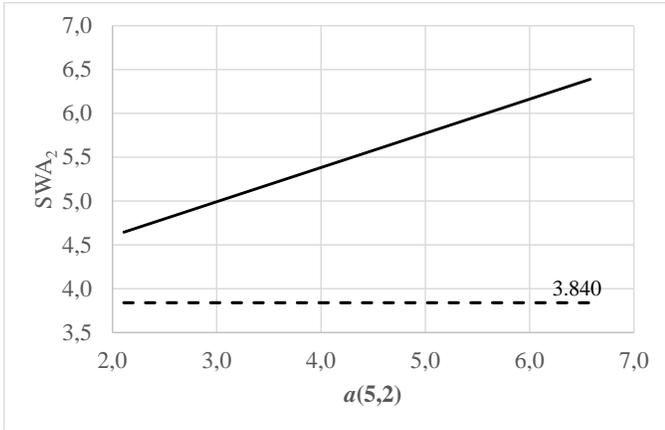


Fig. 6a One-parameter sensitivity analysis of the ranked first alternative A_2 , indicating robustness of the preference matrix solution, when the value of the $a(5,2)$ -element is changing.

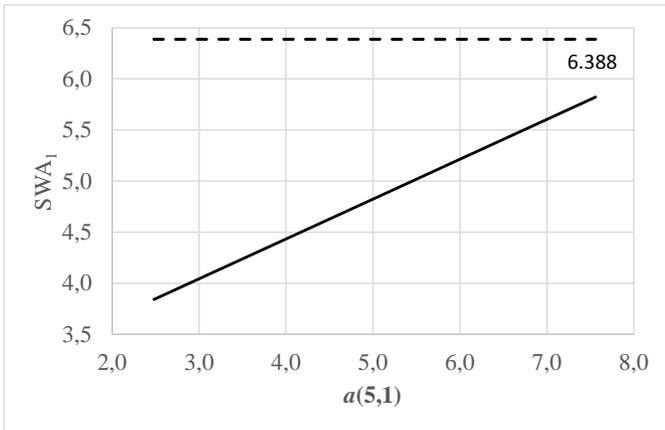


Fig. 6b One-parameter sensitivity analysis of the ranked first alternative A_2 , indicating robustness of the preference matrix solution, when the value of the $a(5,1)$ -element is changing.

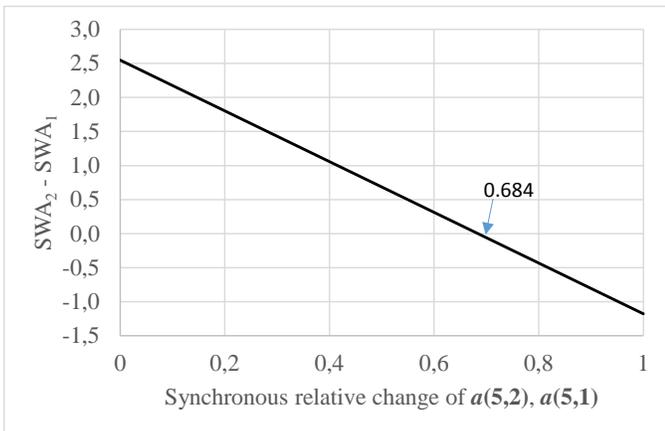


Fig. 6c Two-parameter sensitivity analysis, indicating a change of the preference order at point 0.684 in the normalized (0, 1)-interval.

IV. DISCUSSION

The main problem usually caused by a hidden assumption, as it can be recognized in the cases analyzed in the Implementation Section above, is that the

independent/explanatory variables should had been optimized and the relevant parameters should had been identified before synthesizing the fault tree and the preference matrix to perform FTA and MCA, respectively. Taking as a representative paradigm the optimization of the specific surface area (and the corresponding porosity characteristics, like the number of pores per apparent surface area and their mean diameter) of the adsorbent used in FTA and MCA (intermediate events 1.3.1.2.1/1.4.1/1.4.2 and criteria f_1, f_2, f_3 presented in Fig. 5 and Table 1, respectively we can proceed as follows:

Let $E(D) = E_1(D) + E_2(D)$ the objective function to be maximized in order to find the optimal mean pore diameter D_{opt} , where the pollutant total removal efficiency E has been split to partial efficiencies E_1 and E_2 . The former represents absorption capacity and is an increasing function of D with decreasing rate (i.e., $dE_1/dD > 0, d^2E_1/dD^2 < 0$), because of the gradually increasing loss of pore character for adequately higher D -values. The latter partial efficiency E_2 , represents adsorption capacity and is an increasing function of specific surface, therefore a decreasing function of D (implying fewer pores per apparent surface unit) in the meso-pore region, with decreasing algebraic or increasing absolute rate (i.e. $dE_2/dD < 0, d^2E_2/dD^2 < 0$ or $d|dE_2/dD|/dD > 0$). Evidently, D_{opt} is the abscissa of $E_{max} = (E_1+E_2)_{max}$ at $ME_1 = ME_2$ where $ME_1 = dE_1/dD$ and $ME_2 = |dE_2/dD|$, are the marginal values of E_1 and E_2 , respectively.

For improving the partial efficiency E_1 , when a stage of biodegradation follows, we can support proper diversification of bacterial population in sustainable mode (i.e., without violating natural paths in the food chain and prey-predator equilibrium points) in the local eco-system growing within the pores. In such a case, the E_1 -curve is expected to move upwards to E_1' , becoming steeper, since higher deviation (as compared with the initially estimated curve E_1) will take place in the region of higher D -values, where there is adequate room for the growth of the corresponding bio-societies; as a result, the D_{opt} is shifting to D'_{opt} , where $D'_{opt} > D_{opt}$, as shown in Fig. 7a. Similarly, we can improve the partial efficiency E_2 by enhancing surface hydrophobicity of the adsorbent to the extent that such a modification has not harmful impact to the development / growth of the bio-society finally established as described above. In such a case, the E_2 -curve is expected to move upwards to E_2' , becoming steeper, since higher deviation (as compared with the initially estimated curve E_2) will take place in the region of lower D -values, where specific surface obtains its higher values, influencing favorably the partial efficiency, E_2 ; as a result, the D_{opt} is shifting to D''_{opt} , where $D''_{opt} < D_{opt}$, as shown in Fig. 7b. If both actions (i.e., ecosystem differentiation and hydrophobicity enhancement) take place at the same time, E_{max} will certainly increase but D_{opt} is not certain whether increases or decreases, since the direction of vectors $(D'_{opt} - D_{opt})$ and $(D''_{opt} - D_{opt})$ is opposite to each other. The final direction is a result of synthesizing two different technologies with different cost applicability, and environmental impact.

From a conceptual point of view, we can represent knowledge acquisition/processing/ conversion, by using the Nonaka's four-stage SECI model [21]: Socialization (S), for converting freshly acquired tacit knowledge through

communication/sharing/diffusion at experiential level. Externalization (E), for articulating tacit into explicit knowledge either for hypotheses making, to be subsequently tested under real or simulated conditions, or for simple prototypes preparation, to be subsequently examined for further improvement. Combination (C), for converting simple/modular knowledge into more complex and systematic sets of explicit knowledge, including hypotheses testing and dissemination/diffusion at formal/ agreed level. Internalization (I), for embodying explicit into tacit knowledge by the individuals involved, mainly through learning by doing/observing and sharing formal knowledge.

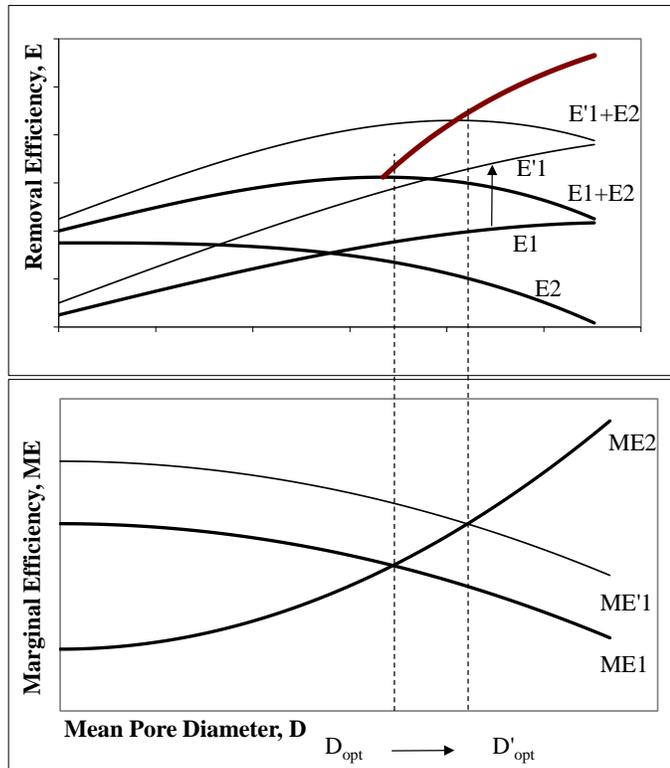


Fig. 7a. Dependence of partial pollutant removal efficiencies E_1 and E_2 , (standing for adsorption and absorption, respectively) on mean pore diameter D of the corresponding distribution on modified biomass based adsorbent/absorbent, and shifting of optimal value D_{opt} , when supporting proper diversification of bacterial population in the eco-system growing within the pores; the thick curve in the upper diagram represents the locus of E_{max} points.

Our modification of Nonaka’s model consists in coupling two knowledge subsystems, one for SME/BPU and another for the TSC. These subsystems are shown in Fig. 8, where (i) the asterisk signifies the defect observation (to be further examined by means of the KB) as information given explicitly, (ii) the symbols on the arrows stand for the SECI model, and (iii) the symbols E, T on each apex stand for explicit and tacit/implicit knowledge, respectively. The local diagnostic KB (DKB) of each SME/BPU, named DKB_s , and the central DKB which belongs to TSC, named DKB_t , are interconnected for information exchange.

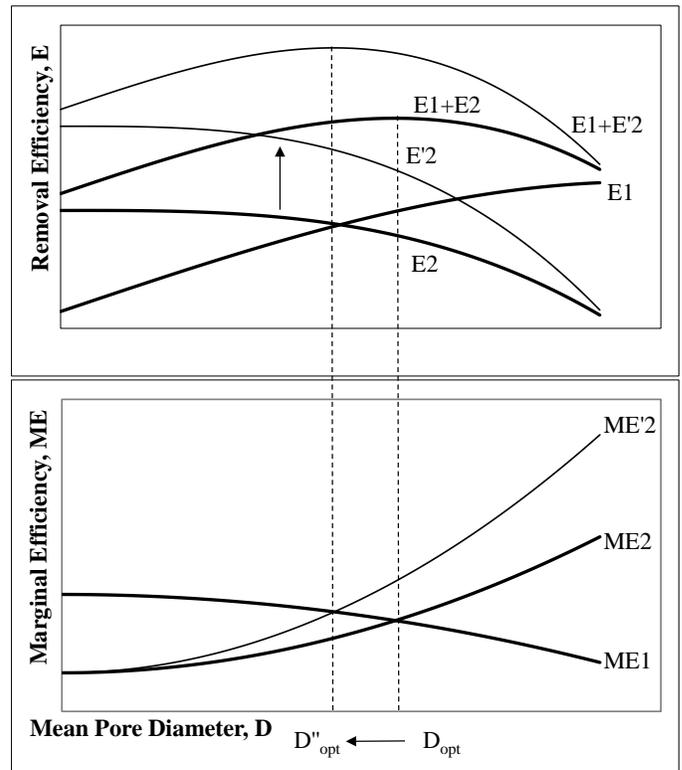
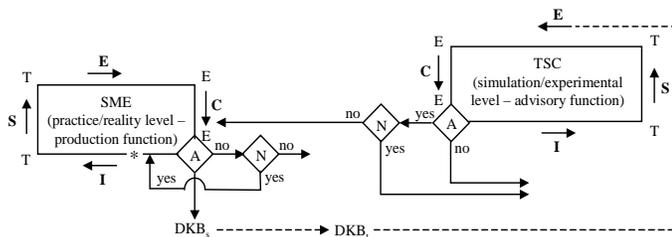


Fig. 7b. Dependence of partial pollutant removal efficiencies E_1 and E_2 , (standing for adsorption and absorption, respectively) on mean pore diameter D of the corresponding distribution on modified biomass based adsorbent/absorbent, and shifting of optimal value D_{opt} , when enhancing surface hydrophobicity of the adsorbent to the extent that such a modification has not harmful impact on bio-society.

In terms of Cybernetics, TSC is an observer/agent while SME is observed entity, both integrated within a system where interaction between them takes place. Since the ‘observer’ as a cybernetic subsystem tries to construct a model for another cybernetic subsystem (i.e. the ‘observed’), we have a “cybernetics of cybernetics”, i.e. a ‘meta’ or 2nd order Cybernetics. As a matter of fact, we have an analogue to the Principle of Uncertainty as extracted from quantum mechanics, which states that observer and observed cannot be separated, and the result of observation will depend on their interaction. This effect is not immediately obvious to the engineer in charge for quality management in the SME, since he knows thoroughly the internal structure of the production line to a high degree of accuracy as well as the points where a fault may appear; therefore, he tends to de-emphasize the system/model distinction, thinking as if the model is actually the subsystem under consideration; such an engineer behaving as a 1st order cyberneticist, tends to manage the subsystem under his supervision as if it were a passive, objectively given ‘thing’, that can be separately observed and independently managed. In the contrary, a 2nd order cyberneticist recognizes the same subsystem (herein, the SME) as an agent in its own right, interacting with another agent, herein the TSE), since limited knowledge or/and malfunctioning of the latter agent may lead to a vicious circle through this interaction (due to bilateral feedback mechanism, possibly as a result of

misinterpretation of the decoded proofs/descriptions quoted in stages 15, 25 of the algorithmic procedure shown in Fig. 2), a circulatory problem arises. A solution to this problem might be a permanent link of the TSC with a knowledge provider, functioning in a different mode; such a link is the relation of an Organization of public interest (established by the State, the Local Authorities or a Professional Union) with a spin-off having the Organization as a parent.



A : Has a satisfactory remedy been achieved?

N : Is continuation of cyclic SECI procedures in the same subsystem recommended?

Fig. 8. Interconnected subsystems forming a complex/modified Nonaka model, representing knowledge acquisition/processing/transformation; symbols in bold stand for processes quoted in the SECI model.

On the grounds of the above analysis, we can propose a university spin-off (USO), company as a feasible solution to the problem of establishing an independent/effective/ reliable and widely accepted TSC. The main advantage of such a company over a corporate spin-off (CSO) comes from the difference in intellectual or mental character of the people running these spin-off firms, who carry the spirit of the antecedent organization: while a private company often tries to keep research and technology within the firm selling solely products/services in a black-box form, a university usually allows or even encourages the transfer of results to outsiders in more or less transparent form. Moreover, because of the large dissimilarities between the parent organizations, it seems natural that the continued relation between the USO and the University, compared to the CSO and the private firm, is expected to be completely different. E.g., a CSO can easily be an associate customer or subcontractor to its parent firm (possibly developing a common attitude as regards the extent/degree up to which knowledge diffusion is allowable), while this is almost impossible for an USO. On the other hand, a CSO may develop a competitive character against its parent organization, consequently eliminating a path of knowledge transfer, a situation rather impossible in the case of a USO, since its activities are completely different in comparison to those of the parent University.

Suggestions for further research include (i) extension of FTA to consider human errors expected to occur in the interaction between industrial operator and thermochemical process equipment by analogy with the interaction between a computer user and the operating system, according to [22], and (ii) incorporation of the TSC's consultancy into the SME's managerial planning system, for the exploitation of special waste biomass adsorbents (e.g., see [23]), and (iii) investigation of the possibility of producing other biomass-based products (preferably chosen from the categories

supported by the EU, as described in [24]), in order to take advantage of the expected scale economies in the BPU, especially in the cases where renewable energy sources (RES) can be combined with biomass-based products or processes, as described in [25], [26].

V. CONCLUSIONS

A methodological framework, under the form of an algorithmic procedure, based on fuzzy fault tree analysis, has been developed for interactive knowledge acquisition/processing and know-how transfer from a biomass processing unit (BPU) to a technical support center (TSC) and vice versa. This procedure was simulated successfully in the case of a BPU using data provided by the Laboratory of Simulation of Industrial Processes of the Department of Industrial Management at the University of Piraeus, Greece. The functionality of this algorithmic procedure is further supported by performing fuzzy multicriteria analysis (MCA) for ranking alternatives, when they appear eventually or set a priori in the fault tree. The conceptual background of the methodological framework was shown to follow a complex/modified Nonaka SECI model while the interaction between the BPU and the corresponding TSC can be (i) studied through 2nd order cybernetics and (ii) satisfactorily realized by means of a university spin-off, which exhibits certain advantages over a corporate spin-off.

Last, optimization tradeoff techniques were used to find equilibrium points by maximizing total benefit or minimizing total cost, especially when environmental and economic cost are the partial conflict variables.

ACKNOWLEDGMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund. - Project: THALIS – University Of Piraeus – Development Of New Material From Waste Biomass For Hydrocarbons Adsorption In Aquatic Environments. MIS 377356.



REFERENCES

- [1] F. Batzias, A. Bountri, "Environmental Management within an Industrial Ecology Framework – The case of Waste Biomass", in proc. *WSEAS Intern. Conf. on Energy, Environment, Devices, Systems, Communications and Computers (EEDSCC '11)*, Eds. S. Chen, N. Mastorakis, F. Rivas-Echeverria, V. Mladenov, Venice, Italy, 2011, pp 209-214.
- [2] N. Bulz, A. Bogdan, S. Chelmu, A. Strateanu, "Integrated Project for the Exergy and Sustainable Development of the Agro-Biodiversity through an Interactive Modeling Analysis regarding the Synergy Rural Economy and Bioethics based upon Smart Growth, Eco-innovation and Large Scale Systems", in proc. *WSEAS Intern. Conf. on Energy, Environment, Devices, Systems, Communications and Computers*

- (EEDSCC '11), Eds. S. Chen, N. Mastorakis, F. Rivas-Echeverria, V. Mladenov, Venice, Italy, 2011, pp 96-102.
- [3] F. Batzias, "Improving Anaerobic Digestion of Waste Sludge by Means of Fuzzy Fault Tree Analysis", in Proc. of 10th WSEAS Intern. Conf. on Environment, Ecosystems and Development, Eds. S. Oprisan et al., Montreux, Switzerland, Dec. 2012, pp 187-192.
- [4] F. Batzias, A. Bountri, Ch. Siontorou, "Solving River Pollution Problems by Means of Fuzzy Fault Tree Analysis", in Proc. of Intern. Conf. 8th WSEAS on Environment, Ecosystems and Development, Eds. N. Mastorakis, V. Mladenov, M. Demiralp, Z. Bojkovic, Vouliagmeni, Athens, Greece, Dec. 2010, pp 228-233.
- [5] D. Yuhuaa, Y. Datao, "Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis", *J. Loss Prev. Proc. Ind.*, Vol. 18, 2005, pp. 83–88.
- [6] G.-S. Liang, M.-J. J. Wang, "Fuzzy Fault-Tree Analysis Using Failure Possibility", *Microelectron. Reliab.*, Vol. 33 (4), 1993, (Received for publication, Oct.1991), pp. 583-597.
- [7] A.F. Batzias, F.A. Batzias, "Fuzzy Fault Tree Analysis as a Means for Computer Aided Technology Transfer to Small/Medium Anodizers", *Computer Aided Chemical Engineering*, Vol. 18, pp. 577-582, 2004.
- [8] R. Ferdousa, F. Khana, B. Veitcha, P. R. Amyotte, "Methodology for computer aided fuzzy fault tree analysis", *Proc. Safety Env. Prot.*, Vol. 87, 2009, pp. 217–226.
- [9] M. Kumar, S.P. Yadav, "The weakest t-norm based intuitionistic fuzzy fault-tree analysis to evaluate system reliability", *ISA Transactions*, Vol. 51, 2012, pp. 531–538.
- [10] U. Sharma, "Use of recursive methods in fuzzy fault tree analysis: an aid to quantitative risk analysis", *Reliability Engineering and System Safety*, Vol. 41, 1993, pp. 231-237.
- [11] F.A. Batzias, E.C. Marcoulaki, "Restructuring the keywords interface to enhance CAPE knowledge via an intelligent agent", *Computer Aided Chemical Engineering*, Vol. 10, 2002, pp. 829–834.
- [12] D. Wang, P. Zhang, L. Chen, "Fuzzy fault tree analysis for fire and explosion of crude oil tanks", *J. Loss Prev. Proc. Ind.* Vol. 26, 2013, pp. 1390-1398.
- [13] S.M. Lavasania, A. Zendegani, M. Celik, "An extension to Fuzzy Fault Tree Analysis (FFTA) application in petrochemical process industry", *Process Safety and Environmental Protection*, Vol. 93, 2015, pp. 75–88.
- [14] H. Jian, C. Junying, L. Jiahong, Q. Dayong, "Risk Identification of Sudden Water Pollution on Fuzzy Fault Tree in Beibu-Gulf Economic Zone", *Procedia Environmental Sciences*, Vol. 10, 2011, pp. 2413-2419.
- [15] A. Mercier, C. Joulian, C. Michel, P. Auger, S. Coulon, L. Amalric, C. Morlay, F. Battaglia-Brunet, "Evaluation of three activated carbons for combined adsorption and biodegradation of PCBs in aquatic sediment", *Water Research*, Vol. 59, 2014, pp.304-315.
- [16] M. Caldeira, S.C. Heald, M.F. Carvalho, I. Vasconcelos, A.T. Bull, P.M. Castro, "4-Chlorophenol degradation by a bacterial consortium: development of a granular activated carbon biofilm reactor", *Appl. Microbiol. Biotechnol.*, Vol. 52 (5), 1999, pp. 722-729.
- [17] M.T. Jonker, A.A.Koelmans, "Extraction of polycyclic aromatic hydrocarbons from soot and sediment: solvent evaluation and implications for sorption mechanism", *Environ. Sci. Technol.*, Vol. 36 (19), 2002, pp. 4107-4113.
- [18] Y.Liu, S.F. Yang, Y.Li, H. Xu, L. Qin, J.H. Tay, "The influence of cell and substratum surface hydrophobicities on microbial attachment", *J. Biotechnol.*, Vol. 110 (3), 2004, pp. 251-256.
- [19] B. Jensen, T. Kuznetsova, B. Kvamme, A. Oterhals, "Molecular dynamics study of selective adsorption of PCB on activated carbon", *Fluid Phase Equilibria* Vol. 307 (1), 2011, pp. 58-65.
- [20] K. Amstaetter, E. Eek, G.Cornelissen, "Sorption of PAHs and PCBs to activated carbon: coal versus biomass-based quality", *Chemosphere* Vol. 87 (5), 2012, pp. 573-578.
- [21] I. Nonaka, R. Toyama, N. Konno, "SECI, Ba and leadership: a unified model of dynamic knowledge creation", *Long Range Planning*, Vol. 33, 2000, pp. 5-34.
- [22] M. Virvou, "Automatic reasoning and help about human errors in using an operating system", *Interacting with Computers*, Vol. 11, 1999, pp 545-573.
- [23] Y. B. Patil, "Development of a Low-Cost Industrial Waste Treatment Technology for Resource Conservation – An Urban Case Study with Gold-Cyanide Emanated from SMEs", *Procedia - Social and Behavioral Sciences*, Vol. 37, 2012, pp 379–388.
- [24] Ing. G. Grassi, "Modern bioenergy in the European Union", *Renewable Energy*, Vol. 16, Issues 1–4, January–April 1999, pp 985–990.
- [25] M.F. Milazzo, F. Spina, P. Primerano, J.C.J. Bart, "Soy biodiesel pathways: Global prospects", *Renewable and Sustainable Energy Reviews*, Vol 26, October 2013, pp 579–624.
- [26] D.K. Sidiras, "GIS based simulation of the biodiesel penetration in European Union markets: The case of Greece", *Biomass and Bioenergy*, Vol. 65, pp. 101-111, 2014.

Curve restoration with implementations based on probability distribution functions

Dariusz J. Jakóbczak

Abstract—Proposed method, called Probabilistic Nodes Combination (PNC), is the method of 2D curve modeling and handwriting identification by using the set of key points. Nodes are treated as characteristic points of signature or handwriting for modeling and writer recognition. Identification of handwritten letters or symbols need modeling and the model of each individual symbol or character is built by a choice of probability distribution function and nodes combination. PNC modeling via nodes combination and parameter γ as probability distribution function enables curve parameterization and interpolation for each specific letter or symbol. Two-dimensional curve is modeled and interpolated via nodes combination and different functions as continuous probability distribution functions: polynomial, sine, cosine, tangent, cotangent, logarithm, exponent, arc sin, arc cos, arc tan, arc cot or power function.

Keywords— handwriting identification, shape modeling, curve interpolation, PNC method, nodes combination, probabilistic modeling.

I. INTRODUCTION

Handwriting identification and writer verification are still the open questions in artificial intelligence and computer vision. Handwriting based author recognition offers a huge number of significant implementations which make it an important research area in pattern recognition [1]. There are so many possibilities and applications of the recognition algorithms that implemented methods have to be concerned on a single problem. Handwriting and signature identification represents such a significant problem. In the case of writer recognition, described in this paper, each person is represented by the set of modeled letters or symbols. The sketch of proposed method consists of three steps: first handwritten letter or symbol must be modeled by a curve, then compared with unknown letter and finally there is a decision of identification. Author recognition of handwriting and signature is based on the choice of key points and curve modeling. Reconstructed curve does not have to be smooth in the nodes because a writer does not think about smoothing during the handwriting. Curve interpolation in handwriting identification is not only a pure mathematical problem but important task in pattern recognition and artificial intelligence such as: biometric recognition [2-4], personalized handwriting recognition [5], automatic forensic document examination

[6,7], classification of ancient manuscripts [8]. Also writer recognition in monolingual handwritten texts is an extensive area of study and the methods independent from the language are well-seen. Proposed method represents language-independent and text-independent approach because it identifies the author via a single letter or symbol from the sample. This novel method is also applicable to short handwritten text.

Writer recognition methods in the recent years are going to various directions: writer recognition using multi-script handwritten texts [9], introduction of new features [10], combining different types of features [3], studying the sensitivity of character size on writer identification [11], investigating writer identification in multi-script environments [9], impact of ruling lines on writer identification [12], model perturbed handwriting [13], methods based on run-length features [14,3], the edge-direction and edge-hinge features [2], a combination of codebook and visual features extracted from chain code and polygonized representation of contours [15], the autoregressive coefficients [9], codebook and efficient code extraction methods [16], texture analysis with Gabor filters and extracting features [17], using Hidden Markov Model [18-20] or Gaussian Mixture Model [1]. But no method is dealing with writer identification via curve modeling or interpolation and points comparing as it is presented in this paper.

The author wants to approach a problem of curve interpolation [21-23] and shape modeling [24] by characteristic points in handwriting identification. Proposed method relies on nodes combination and functional modeling of curve points situated between the basic set of key points. The functions that are used in calculations represent whole family of elementary functions with inverse functions: polynomials, trigonometric, cyclometric, logarithmic, exponential and power function. These functions are treated as probability distribution functions in the range [0;1]. Nowadays methods apply mainly polynomial functions, for example Bernstein polynomials in Bezier curves, splines and NURBS [25]. But Bezier curves do not represent the interpolation method and cannot be used for example in signature and handwriting modeling with characteristic points (nodes). Numerical methods for data interpolation are based on polynomial or trigonometric functions, for example Lagrange, Newton, Aitken and Hermite methods. These methods have some weak sides [26] and are not sufficient for curve interpolation in the situations when the curve cannot be build by polynomials or trigonometric functions. Proposed 2D curve interpolation is the functional modeling via any elementary

Dariusz Jacek Jakóbczak, Department of Electronics and Computer Science, Technical University of Koszalin, Sniadeckich 2, 75-453 Koszalin, Poland
dariusz.jakobczak@tu.koszalin.pl

functions and it helps us to fit the curve during handwriting identification.

This paper presents novel Probabilistic Nodes Combination (PNC) method of curve interpolation and takes up PNC method of two-dimensional curve modeling via the examples using the family of Hurwitz-Radon matrices (MHR method) [27], but not only (other nodes combinations). The method of PNC requires minimal assumptions: the only information about a curve is the set of at least two nodes. Proposed PNC method is applied in handwriting identification via different coefficients: polynomial, sinusoidal, cosinusoidal, tangent, cotangent, logarithmic, exponential, arc sin, arc cos, arc tan, arc cot or power. Function for PNC calculations is chosen individually at each modeling and it represents probability distribution function of parameter $\alpha \in [0;1]$ for every point situated between two successive interpolation knots. PNC method uses nodes of the curve $p_i = (x_i, y_i) \in \mathbf{R}^2$, $i = 1, 2, \dots, n$:

1. PNC needs 2 knots or more ($n \geq 2$);
2. If first node and last node are the same ($p_1 = p_n$), then curve is closed (contour);
3. For more precise modeling knots ought to be settled at key points of the curve, for example local minimum or maximum and at least one node between two successive local extrema.

Condition 3 means for example the highest point of the curve in a particular orientation, convexity changing or curvature extrema. The goal of this paper is to answer the question: how to model a handwritten letter or symbol by a set of knots [28]?

II. CURVE INTERPOLATION

The method of PNC is computing points between two successive nodes of the curve: calculated points are interpolated and parameterized for real number $\alpha \in [0;1]$ in the range of two successive nodes. PNC method uses the combinations of nodes $p_1=(x_1, y_1)$, $p_2=(x_2, y_2), \dots, p_n=(x_n, y_n)$ as $h(p_1, p_2, \dots, p_m)$ and $m = 1, 2, \dots, n$ to interpolate second coordinate y for first coordinate $c = \alpha \cdot x_i + (1-\alpha) \cdot x_{i+1}$, $i = 1, 2, \dots, n-1$:

$$y(c) = \gamma \cdot y_i + (1-\gamma)y_{i+1} + \gamma(1-\gamma) \cdot h(p_1, p_2, \dots, p_m), \quad (1)$$

$\alpha \in [0;1]$, $\gamma = F(\alpha) \in [0;1]$.

Here are the examples of h computed for MHR method [29]:

$$h(p_1, p_2) = \frac{y_1}{x_1} x_2 + \frac{y_2}{x_2} x_1 \quad (2)$$

or

$$h(p_1, p_2, p_3, p_4) = \frac{1}{x_1^2 + x_3^2} (x_1 x_2 y_1 + x_2 x_3 y_3 + x_3 x_4 y_1 - x_1 x_4 y_3) + \frac{1}{x_2^2 + x_4^2} (x_1 x_2 y_2 + x_1 x_4 y_4 + x_3 x_4 y_2 - x_2 x_3 y_4)$$

The examples of other nodes combinations:

$$h(p_1, p_2) = \frac{y_1 x_2}{x_1 y_2} + \frac{y_2 x_1}{x_2 y_1}$$

or

$$h(p_1, p_2) = \frac{y_1 x_2}{y_2} + \frac{y_2 x_1}{y_1}$$

or

$$h(p_1, p_2) = x_1 y_1 + x_2 y_2$$

or

$$h(p_1, p_2) = x_1 x_2 + y_1 y_2$$

or

$$h(p_1, p_2, \dots, p_m) = 0$$

or

$$h(p_1) = x_1 y_1$$

or others. Nodes combination is chosen individually for each curve. Formula (1) represents curve parameterization as $\alpha \in [0;1]$:

$$x(\alpha) = \alpha \cdot x_i + (1-\alpha) \cdot x_{i+1}$$

and

$$y(\alpha) = F(\alpha) \cdot y_i + (1-F(\alpha))y_{i+1} + F(\alpha)(1-F(\alpha)) \cdot h(p_1, p_2, \dots, p_m)$$

$$y(\alpha) = F(\alpha) \cdot (y_i - y_{i+1} + (1-F(\alpha)) \cdot h(p_1, p_2, \dots, p_m)) + y_{i+1}$$

Proposed parameterization gives us the infinite number of possibilities for curve calculations (determined by choice of F and h) as there is the infinite number of human signatures, handwritten letters and symbols. Nodes combination is the individual feature of each modeled curve (for example a handwritten letter or signature). Coefficient $\gamma = F(\alpha)$ and nodes combination h are key factors in PNC curve interpolation and shape modeling.

A. Probability Distributions in PNC Modeling

Points settled between the nodes are computed using PNC method. Each real number $c \in [a;b]$ is calculated by a convex combination $c = \alpha \cdot a + (1-\alpha) \cdot b$ for

$$\alpha = \frac{b-c}{b-a} \in [0;1].$$

Key question is dealing with coefficient γ in (1). The simplest way of PNC calculation means $h = 0$ and $\gamma = \alpha$ (basic probability distribution). Then PNC represents a linear interpolation. MHR method [30] is not a linear interpolation. MHR [31] is the example of PNC modeling. Each interpolation requires specific distribution of parameter α and γ (1) depends on parameter $\alpha \in [0;1]$:

$$\gamma = F(\alpha), F: [0;1] \rightarrow [0;1], F(0) = 0, F(1) = 1$$

and F is strictly monotonic. Coefficient γ is calculated using different functions (polynomials, power functions, sine, cosine, tangent, cotangent, logarithm, exponent, arc sin, arc cos, arc tan or arc cot, also inverse functions) and choice of function is connected with initial requirements and curve specifications. Different values of coefficient γ are connected with applied functions $F(\alpha)$. These functions $\gamma = F(\alpha)$ represent the examples of probability distribution functions for random variable $\alpha \in [0;1]$ and real number $s > 0$:

$$\begin{aligned} &\gamma = \alpha^s, \gamma = \sin(\alpha^s \cdot \pi/2), \gamma = \sin^s(\alpha \cdot \pi/2), \gamma = 1 - \cos(\alpha^s \cdot \pi/2), \\ &\gamma = 1 - \cos^s(\alpha \cdot \pi/2), \gamma = \tan(\alpha^s \cdot \pi/4), \gamma = \tan^s(\alpha \cdot \pi/4), \gamma = \log_2(\alpha^s + 1), \\ &\gamma = \log_2^s(\alpha + 1), \gamma = (2^\alpha - 1)^s, \gamma = 2/\pi \cdot \arcsin(\alpha^s), \gamma = (2/\pi \cdot \arcsin \alpha)^s, \\ &\gamma = 1 - 2/\pi \cdot \arccos(\alpha^s), \gamma = 1 - (2/\pi \cdot \arccos \alpha)^s, \gamma = 4/\pi \cdot \arctan(\alpha^s), \\ &\gamma = (4/\pi \cdot \arctan \alpha)^s, \gamma = \operatorname{ctg}(\pi/2 - \alpha^s \cdot \pi/4), \gamma = \operatorname{ctg}^s(\pi/2 - \alpha \cdot \pi/4), \end{aligned}$$

$$\gamma = 2 - 4/\pi \cdot \arccos(\alpha^s), \gamma = (2 - 4/\pi \cdot \arccos(\alpha))^s.$$

Functions above, used in γ calculations, are strictly monotonic for random variable $\alpha \in [0;1]$ as $\gamma = F(\alpha)$ is probability distribution function. Also inverse functions $F^{-1}(\alpha)$ are appropriate for γ calculations. Choice of function and value s depends on curve specifications and individual requirements. Considering nowadays used probability distribution functions for random variable $\alpha \in [0;1]$ - one distribution is dealing with the range $[0;1]$: beta distribution. Probability density function f for random variable $\alpha \in [0;1]$ is:

$$f(\alpha) = c \cdot \alpha^s \cdot (1-\alpha)^r, \quad s \geq 0, r \geq 0. \quad (3)$$

When $r = 0$ probability density function (3) represents $f(\alpha) = c \cdot \alpha^s$ and then probability distribution function F is like $f(\alpha) = 3\alpha^2$ and $\gamma = \alpha^3$. If s and r are positive integer numbers then γ is the polynomial, for example $f(\alpha) = 6\alpha(1-\alpha)$ and $\gamma = 3\alpha^2 - 2\alpha^3$. Beta distribution gives us coefficient γ in (1) as polynomial because of interdependence between probability density f and distribution F functions:

$$f(\alpha) = F'(\alpha), \quad F(\alpha) = \int_0^\alpha f(t) dt. \quad (4)$$

For example (4): $f(\alpha) = \alpha \cdot e^\alpha$ and $\gamma = F(\alpha) = (\alpha - 1)e^\alpha + 1$.

What is very important in PNC method: two curves (for example a handwritten letter or signature) may have the same set of nodes but different h or γ results in different interpolations (Fig.6-14).

Algorithm of PNC interpolation and modeling (1) looks as follows:

Step 1: Choice of knots p_i at key points.

Step 2: Choice of nodes combination $h(p_1, p_2, \dots, p_m)$.

Step 3: Choice of distribution $\gamma = F(\alpha)$.

Step 4: Determining values of α : $\alpha = 0.1, 0.2 \dots 0.9$ (nine points) or $0.01, 0.02 \dots 0.99$ (99 points) or others.

Step 5: The computations (1).

These five steps can be treated as the algorithm of PNC method of curve modeling and interpolation (1).

Curve interpolation has to implement the coefficients γ . Each strictly monotonic function F between points $(0;0)$ and $(1;1)$ can be used in PNC interpolation.

III. MODELING AND RECOGNITION

PNC method enables signature and handwriting recognition. This process of recognition consists of three parts:

1. Modeling – choice of nodes combination and probabilistic distribution function (1) for known signature or handwritten letters;
2. Unknown writer - choice of characteristic points (nodes) for unknown signature or handwritten word and the coefficients of points between nodes;
3. Decision of recognition - comparing the results of PNC interpolation for known models with coordinates of unknown text.

A. Modeling – the Basis of Patterns

Known letters or symbols ought to be modeled by the choice of nodes, determining specific nodes combination and characteristic probabilistic distribution function. For example a handwritten word or signature “rw” may look different for persons A, B or others. How to model “rw” for some persons via PNC method? Each model has to be described by the set of nodes for letters “r” and “w”, nodes combination h and a function $\gamma = F(\alpha)$ for each letter. Less complicated models can take $h(p_1, p_2, \dots, p_m) = 0$ and then the formula of interpolation (1) looks as follows:

$$y(c) = \gamma \cdot y_i + (1-\gamma)y_{i+1}.$$

It is linear interpolation for basic probability distribution ($\gamma = \alpha$). How first letter “r” is modeled in three versions for nodes combination $h = 0$ and $\alpha = 0.1, 0.2 \dots 0.9$? Of course α is a random variable and $\alpha \in [0;1]$.

Person A

Nodes (1;3), (3;1), (5;3), (7;3) and $\gamma = F(\alpha) = \alpha^2$:

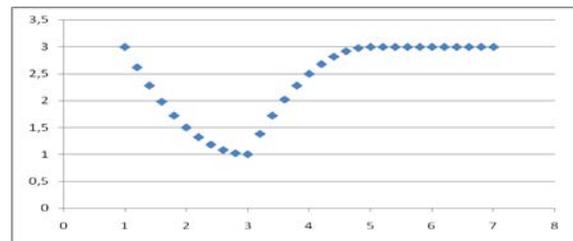


Fig. 1. PNC modeling for nine reconstructed points between nodes.

Person B

Nodes (1;3), (3;1), (5;3), (7;2) and $\gamma = F(\alpha) = \alpha^2$:

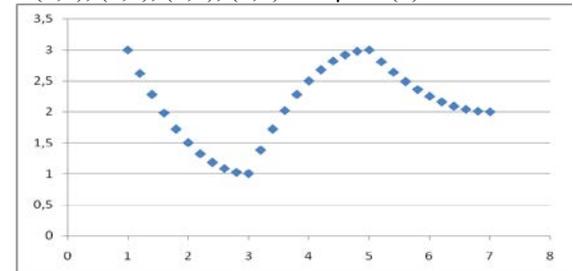


Fig. 2. PNC modeling of letter “r” with four nodes.

Person C

Nodes (1;3), (3;1), (5;3), (7;4) and $\gamma = F(\alpha) = \alpha^3$:

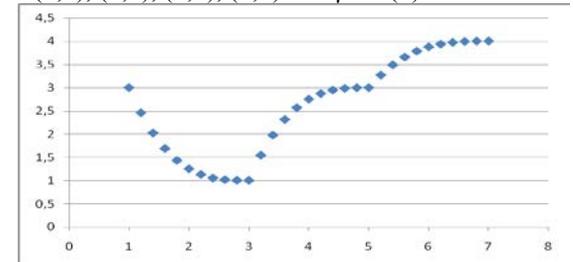


Fig. 3. PNC modeling of handwritten letter “r”.

These three versions of letter “r” (Fig.1-3) with nodes combination $h = 0$ differ at fourth node and probability distribution functions $\gamma = F(\alpha)$. Much more possibilities of modeling are connected with a choice of nodes combination

$h(p_1, p_2, \dots, p_m)$. MHR method [32] uses the combination (2) with good features because of orthogonal rows and columns at Hurwitz-Radon family of matrices:

$$h(p_i, p_{i+1}) = \frac{y_i}{x_i} x_{i+1} + \frac{y_{i+1}}{x_{i+1}} x_i$$

and then (1)

$$y(c) = \gamma \cdot y_i + (1 - \gamma)y_{i+1} + \gamma(1 - \gamma) \cdot h(p_i, p_{i+1}).$$

Here are two examples of PNC modeling with MHR combination (2).

Person D

Nodes (1;3), (3;1), (5;3) and $\gamma = F(\alpha) = \alpha^2$:

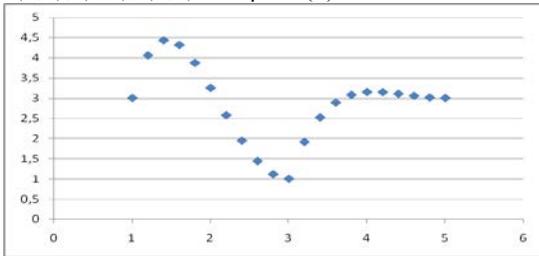


Fig. 4. PNC modeling of letter "r" with three nodes.

Person E

Nodes (1;3), (3;1), (5;3) and $\gamma = F(\alpha) = \alpha^{1.5}$:

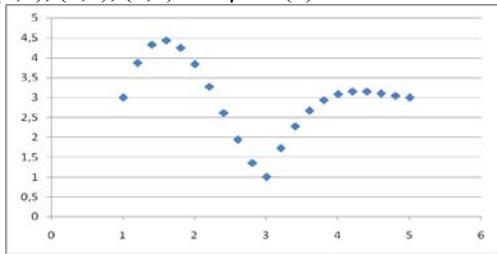


Fig. 5. PNC modeling of handwritten letter "r".

Fig.1-5 show modeling of letter "r". Now let us consider a letter "w" with nodes combination $h = 0$.

Person A

Nodes (2;2), (3;1), (4;2), (5;1), (6;2) and $\gamma = F(\alpha) = (5^\alpha - 1)/4$:

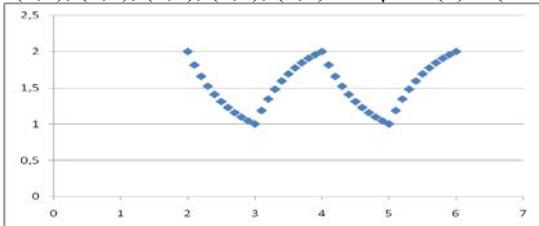


Fig. 6. PNC modeling for nine reconstructed points between nodes.

Person B

Nodes (2;2), (3;1), (4;2), (5;1), (6;2) and $\gamma = F(\alpha) = \sin(\alpha \cdot \pi/2)$:

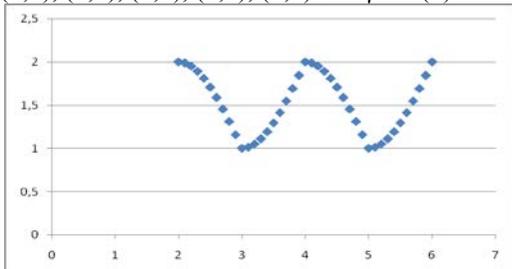


Fig. 7. PNC modeling of letter "w" with five nodes.

Person C

Nodes (2;2), (3;1), (4;2), (5;1), (6;2) and $\gamma = F(\alpha) = \sin^{3.5}(\alpha \cdot \pi/2)$:

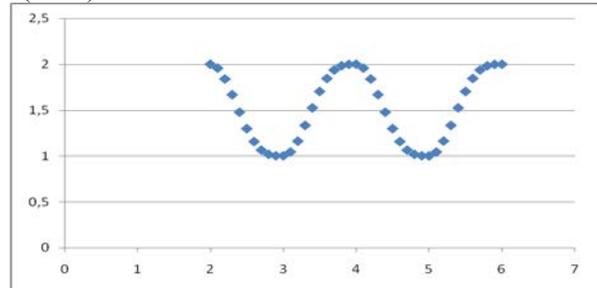


Fig. 8. PNC modeling of handwritten letter "w".

These three versions of letter "w" (Fig.6-8) with nodes combination $h = 0$ and the same nodes differ only at probability distribution functions $\gamma = F(\alpha)$. Fig.9 is the example of nodes combination $h(2)$ from MHR method:

Person D

Nodes (2;2), (3;1), (4;1), (5;1), (6;2) and $\gamma = F(\alpha) = 2^\alpha - 1$:

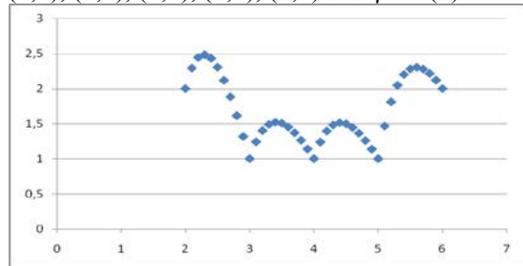


Fig. 9. PNC modeling for nine reconstructed points between nodes.

Examples above have one function $\gamma = F(\alpha)$ and one combination h for all ranges between nodes. But it is possible to create a model with functions $\gamma_i = F_i(\alpha)$ and combinations h_i individually for a range of nodes $(p_i; p_{i+1})$. It enables very precise modeling of handwritten symbol between each successive pair of nodes.

Each person has its own characteristic and individual handwritten letters, numbers or other marks. The range of coefficients x has to be the same for all models because of comparing appropriate coordinates y . Every letter is modeled by PNC via three factors: the set of nodes, probability distribution function $\gamma = F(\alpha)$ and nodes combination h . These three factors are chosen individually for each letter, therefore this information about modeled letters seems to be enough for specific PNC curve interpolation, comparing and handwriting identification. Function γ is selected via the analysis of points between nodes and we may assume $h = 0$ at the beginning. What is very important - PNC modeling is independent of the language or a kind of symbol (letters, numbers or others). One person may have several patterns for one handwritten letter. Summarize: every person has the basis of patterns for each handwritten letter or symbol, described by the set of nodes, probability distribution function $\gamma = F(\alpha)$ and nodes combination h . Whole basis of patterns consists of models S_j for $j = 0, 1, 2, 3, \dots, K$.

B. Unknown Person – Points of the Curve

Choice of characteristic points (nodes) for unknown letter or handwritten symbol is a crucial factor in object recognition. The range of coefficients x has to be the same like the x range in the basis of patterns. Knots of the curve (opened or closed) ought to be settled at key points, for example local minimum or maximum (the highest point of the curve in a particular orientation), convexity changing or curvature maximum and at least one node between two successive key points. When the nodes are fixed, each coordinate of every chosen point on the curve $(x_0^c, y_0^c), (x_1^c, y_1^c), \dots, (x_M^c, y_M^c)$ is accessible to be used for comparing with the models. Then probability distribution function $\gamma = F(\alpha)$ and nodes combination h have to be taken from the basis of modeled letters to calculate appropriate second coordinates $y_i^{(j)}$ of the pattern S_j for first coordinates $x_i^c, i = 0, 1, \dots, M$. After interpolation it is possible to compare given handwritten symbol with a letter in the basis of patterns.

C. Recognition – the Author

Comparing the results of PNC interpolation for required second coordinates of a model in the basis of patterns with points on the curve $(x_0^c, y_0^c), (x_1^c, y_1^c), \dots, (x_M^c, y_M^c)$, we can say if the letter or symbol is written by person A, B or another. The comparison and decision of recognition [33] is done via minimal distance criterion. Curve points of unknown handwritten symbol are: $(x_0^c, y_0^c), (x_1^c, y_1^c), \dots, (x_M^c, y_M^c)$. The criterion of recognition for models $S_j = \{(x_0^c, y_0^{(j)}), (x_1^c, y_1^{(j)}), \dots, (x_M^c, y_M^{(j)})\}, j=0, 1, 2, 3 \dots K$ is given as:

$$\sum_{i=0}^M |y_i^c - y_i^{(j)}| \rightarrow \min.$$

Minimal distance criterion helps us to fix a candidate for unknown writer as a person from the model S_j .

IV. CONCLUSION

The method of Probabilistic Nodes Combination (PNC) enables interpolation and modeling of two-dimensional curves [34] using nodes combinations and different coefficients γ : polynomial, sinusoidal, cosinusoidal, tangent, cotangent, logarithmic, exponential, arc sin, arc cos, arc tan, arc cot or power function, also inverse functions. Function for γ calculations is chosen individually at each curve modeling and it is treated as probability distribution function: γ depends on initial requirements and curve specifications. PNC method leads to curve interpolation as handwriting or signature identification via discrete set of fixed knots. PNC makes possible the combination of two important problems: interpolation and modeling in a matter of writer identification. Main features of PNC method are:

- the smaller distance between knots the better;
- calculations for coordinates close to zero and near by extremum require more attention because of importance of these points;
- PNC interpolation develops a linear interpolation into other functions as probability distribution functions;

- PNC is a generalization of MHR method via different nodes combinations;
- interpolation of L points is connected with the computational cost of rank $O(L)$ as in MHR method;
- nodes combination and coefficient γ are crucial in the process of curve probabilistic parameterization and interpolation: they are computed individually for a single curve.

Future works are going to: application of PNC method in signature and handwriting recognition, choice and features of nodes combinations and coefficient γ , implementation of PNC in computer vision and artificial intelligence: shape geometry, contour modelling, object recognition and curve parameterization.

REFERENCES

- Schlabach, A., Bunke, H.: Off-line writer identification using Gaussian mixture models. In: International Conference on Pattern Recognition, pp. 992–995 (2006).
- Bulacu, M., Schomaker, L.: Text-independent writer identification and verification using textural and allographic features. IEEE Trans. Pattern Anal. Mach. Intell. 29 (4), 701–717 (2007).
- Djeddi, C., Souici-Meslati, L.: A texture based approach for Arabic writer identification and verification. In: International Conference on Machine and Web Intelligence, pp. 115–120 (2010).
- Djeddi, C., Souici-Meslati, L.: Artificial immune recognition system for Arabic writer identification. In: International Symposium on Innovation in Information and Communication Technology, pp. 159–165 (2011).
- Nosary, A., Heutte, L., Paquet, T.: Unsupervised writer adaption applied to handwritten text recognition. Pattern Recogn. Lett. 37 (2), 385–388 (2004).
- Van, E.M., Vuurpijl, L., Franke, K., Schomaker, L.: The WANDA measurement tool for forensic document examination. J. Forensic Doc. Exam. 16, 103–118 (2005).
- Schomaker, L., Franke, K., Bulacu, M.: Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recogn. Lett. 28 (6), 719–727 (2007).
- Siddiqi, I., Cloppet, F., Vincent, N.: Contour based features for the classification of ancient manuscripts. In: Conference of the International Graphonomics Society, pp. 226–229 (2009).
- Garain, U., Paquet, T.: Off-line multi-script writer identification using AR coefficients. In: International Conference on Document Analysis and Recognition, pp. 991–995 (2009).
- Bulacu, M., Schomaker, L., Brink, A.: Text-independent writer identification and verification on off-line Arabic handwriting. In: International Conference on Document Analysis and Recognition, pp. 769–773 (2007).
- Ozaki, M., Adachi, Y., Ishii, N.: Examination of effects of character size on accuracy of writer recognition by new local arc method. In: International Conference on Knowledge-Based Intelligent Information and Engineering Systems, pp. 1170–1175 (2006).
- Chen, J., Lopresti, D., Kavallieratou, E.: The impact of ruling lines on writer identification. In: International Conference on Frontiers in Handwriting Recognition, pp. 439–444 (2010).
- Chen, J., Cheng, W., Lopresti, D.: Using perturbed handwriting to support writer identification in the presence of severe data constraints. In: Document Recognition and Retrieval, pp. 1–10 (2011).
- Galloway, M.M.: Texture analysis using gray level run lengths. Comput. Graphics Image Process. 4 (2), 172–179 (1975).
- Siddiqi, I., Vincent, N.: Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recogn. Lett. 43 (11), 3853–3865 (2010).
- Ghiasi, G., Safabakhsh, R.: Offline text-independent writer identification using codebook and efficient code extraction methods. Image and Vision Computing 31, 379–391 (2013).
- Shahabinejad, F., Rahmati, M.: A new method for writer identification and verification based on Farsi/Arabic handwritten texts, Ninth

- International Conference on Document Analysis and Recognition (ICDAR 2007), pp. 829–833 (2007).
- [18] Schlapbach, A., Bunke, H.: A writer identification and verification system using HMM based recognizers, *Pattern Anal. Appl.* 10, 33–43 (2007).
- [19] Schlapbach, A., Bunke, H.: Using HMM based recognizers for writer identification and verification, 9th Int. Workshop on Frontiers in Handwriting Recognition, pp. 167–172 (2004).
- [20] Marti, U.-V., Bunke, H.: The IAM-database: an English sentence database for offline handwriting recognition, *Int. J. Doc. Anal. Recognit.* 5, 39–46 (2002).
- [21] Collins II, G.W.: *Fundamental Numerical Methods and Data Analysis*. Case Western Reserve University (2003).
- [22] Chapra, S.C.: *Applied Numerical Methods*. McGraw-Hill (2012).
- [23] Ralston, A., Rabinowitz, P.: *A First Course in Numerical Analysis – Second Edition*. Dover Publications, New York (2001).
- [24] Zhang, D., Lu, G.: Review of Shape Representation and Description Techniques. *Pattern Recognition* 1(37), 1-19 (2004).
- [25] Schumaker, L.L.: *Spline Functions: Basic Theory*. Cambridge Mathematical Library (2007).
- [26] Dahlquist, G., Bjoerck, A.: *Numerical Methods*. Prentice Hall, New York (1974).
- [27] Jakóbczak, D.: 2D and 3D Image Modeling Using Hurwitz-Radon Matrices. *Polish Journal of Environmental Studies* 4A(16), 104-107 (2007).
- [28] Jakóbczak, D.: Shape Representation and Shape Coefficients via Method of Hurwitz-Radon Matrices. *Lecture Notes in Computer Science* 6374 (Computer Vision and Graphics: Proc. ICCVG 2010, Part I), Springer-Verlag Berlin Heidelberg, 411-419 (2010).
- [29] Jakóbczak, D.: Curve Interpolation Using Hurwitz-Radon Matrices. *Polish Journal of Environmental Studies* 3B(18), 126-130 (2009).
- [30] Jakóbczak, D.: Application of Hurwitz-Radon Matrices in Shape Representation. In: Banaszak, Z., Świć, A. (eds.) *Applied Computer Science: Modelling of Production Processes* 1(6), pp. 63-74. Lublin University of Technology Press, Lublin (2010).
- [31] Jakóbczak, D.: Object Modeling Using Method of Hurwitz-Radon Matrices of Rank k . In: Wolski, W., Borawski, M. (eds.) *Computer Graphics: Selected Issues*, pp. 79-90. University of Szczecin Press, Szczecin (2010).
- [32] Jakóbczak, D.: Implementation of Hurwitz-Radon Matrices in Shape Representation. In: Choraś, R.S. (ed.) *Advances in Intelligent and Soft Computing* 84, Image Processing and Communications: Challenges 2, pp. 39-50. Springer-Verlag, Berlin Heidelberg (2010).
- [33] Jakóbczak, D.: Object Recognition via Contour Points Reconstruction Using Hurwitz-Radon Matrices. In: Józefczyk, J., Orski, D. (eds.) *Knowledge-Based Intelligent System Advancements: Systemic and Cybernetic Approaches*, pp. 87-107. IGI Global, Hershey PA, USA (2011).
- [34] Jakóbczak, D.: Curve Parameterization and Curvature via Method of Hurwitz-Radon Matrices. *Image Processing & Communications- An International Journal* 1-2(16), 49-56 (2011).

Technical University of Koszalin, Poland and since October 2007 he has been an Assistant Professor in the Chair of Computer Science and Management in this department. His research interests connect mathematics with computer science and include computer vision, artificial intelligence, shape representation, curve interpolation, contour reconstruction and geometric modeling, numerical methods, probabilistic methods, game theory, operational research and discrete mathematics.



Dariusz Jacek Jakóbczak was born in Koszalin, Poland, on December 30, 1965. He graduated in mathematics (numerical methods and programming) from the University of Gdansk, Poland in 1990. He received the Ph.D. degree in 2007 in computer science from the Polish – Japanese Institute of Information Technology, Warsaw, Poland.

From 1991 to 1994 he was a civilian programmer in the High Military School in Koszalin. He was a teacher of mathematics and computer science in the Private Economic School in Koszalin from 1995 to 1999. Since March 1998 he has worked in the Department of Electronics and Computer Science,

Mutations of Adjacent Amino Acid Pairs are not Always Independent

Jyotsna Ramanan and Peter Z. Revesz

Abstract—Evolutionary studies usually assume that the genetic mutations are independent of each other. This paper tests the independence hypothesis for genetic mutations with regard to protein coding regions. According to the new experimental results the independence assumption generally holds, but there are certain exceptions. In particular, the coding regions that represent two adjacent amino acids seem to change in ways that sometimes deviate significantly from the expected theoretical probability under the independence assumption.

Keywords— amino acid, independent probabilities, nucleotide, genetic mutation, protein

I. INTRODUCTION

Biological evolution depends on random mutations accompanied by natural selection for the more fit genes.

That simple statement does not imply that the observed mutations are independent from each other. It is possible that if a nucleotide changes, then it is biologically beneficial to have some of the adjacent or near by nucleotides change as well. For example, if in some protein-coding region within some triplet that encodes a hydrophilic amino acid a nucleotide changes such that the triplet would encode a hydrophobic amino acid, then a mutation of another nucleotide in the same triplet may be advantageous if with that mutation the triplet would again encode a hydrophilic amino acid (or preserve another key property of amino acids). In other words, some mutations within a triplet slightly increase the probability that some accompanying mutation with a readjusting effect would survive in the offspring.

With the greatly increasing number of decoded genes currently available in a number of genome libraries and online databases, it is now possible to have a large-scale computer-based study to test whether the independence assumption holds. One difficulty, however, is to find the coding regions and coding triplets. Hence it seems more convenient to investigate proteins derived from the coding regions.

The mutations in the coding regions of the DNA are usually reflected in the mutations of amino acids. Therefore, instead of the evolution of genes, one may talk about the evolution of proteins within a closely related set of proteins, which is called a *protein family*.

The PFAM library [4] records a growing number of protein families. Each protein in a protein family can be assumed to be genetically related to the other proteins in that family and to have evolved from a single ancestor protein.

For any set of DNA strings and any set of proteins, there are several algorithms that can be used to find a hypothetical evolutionary tree (see the textbooks by Baum and Smith [1], Hall [2], and Lerney et al. [3] for an overview of these algorithms.) Revesz [5] has proposed recently a new phylogenetic tree-building algorithm called the *Common Mutation Similarity Matrixes* (CMSM) algorithm. This algorithm finds a hypothetical evolutionary tree. The first step of the CMSM algorithm is to find a hypothetical common ancestor, which is denoted by μ .

In this paper, we will use the idea of a hypothetical common ancestor. We can compare the hypothetical common ancestor of a family of proteins with each of the proteins in the family to test where the mutations occur. We also can test for each adjacent pair of amino acids how many times that pair changed into another pair of amino acids. The resulting experimental statistics can be compared with the theoretical probability under the independence assumption. If the deviation from the theoretical probability is significant, then the independence assumption fails to provide a satisfying explanation for the experimental results.

Evolutionary studies usually assume that the genetic mutations are independent of each other. This paper tests the independence hypothesis for genetic mutations with regard to protein coding regions. As discussed in Section IV, according to our experimental results the independence assumption generally holds, but there seem to be certain exceptions. We give examples in Section IV of some particular adjacent amino acid pairs that seem to change in ways that deviate significantly from the expected theoretical probability under the independence assumption.

This paper is organized as follows. Section II describes some background concepts about hypothetical common ancestors. Section III describes our method with an extended example. Section IV presents our experimental results. Finally, Section V gives some conclusions and directions for further research.

Jyotsna Ramanan (jramanan@cse.unl.edu) and Peter Z. Revesz (revesz@cse.unl.edu) are with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

II. BACKGROUND CONCEPTS RELATED TO HYPOTHETICAL COMMON ANCESTOR

Consider the seven amino acid sequences, $S_1 \dots S_7$ shown in Figure 1 below. These seven amino acid sequences belong to the protein family DiSB-ORF2_chro (PFAM library identification number PF16506) [1]. The sequences as shown in Figure 1 are already aligned with each other.

S_1	SPYMFDRSCLNVYRTNDYLFGECLTLPNCSEPSVVKLDKTFYQETVVCHS
S_2	TPYVFDRECLSVYRTNDWFFSQCSLPPNCTNPSVVKLERTFFGQETVVCHS
S_3	SPFEFDPEECIEVHRTHSWFFQGCTLPPSCGDVHTKILDSSF-GFKELMCYS
S_4	SPYMFDRSCLNVYRTNDYLFGECLTLPNCSEPSVIKLDKTFYQETVVCHS
S_5	SPYHTDPTCVSVYRTNDWFFAGCELPCHLGKVVSIIEKKWYQETVFCYS
S_6	SPFEFDPEECIEVHRTHSWFFQGCTLPPSCGDVHTKILDSSF-GFKELMCYS
S_7	SPYVFDRECLNVYRTNDYLFGECLTLPNCSEP-----

Fig.1 Seven example proteins from the protein family DiSB-ORF2_chro

For the above set of amino acid sequences, the CMSM algorithm [5] generates the hypothetical evolutionary tree shown in Figure 2.

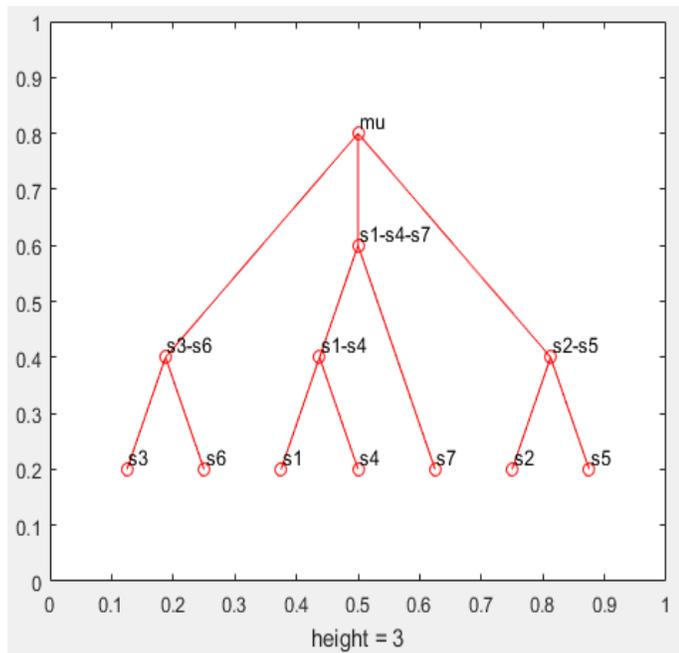


Fig. 2 The hypothetical evolutionary tree for the DiSB-ORF2_chro generated using the CMM algorithm

In Figure 2, the variables $S_1, S_2 \dots S_7$ correspond to the seven amino acid sequences that are listed in Figure 1. The variable μ is the hypothetical common ancestor that was generated using the CMSM algorithm. The value of μ was found to be the following string of amino acids.

$\mu =$ SPYEFDRSCLNVYRTNDWFFGECTLPPNCSEPSVK
ILDKTFYGGQETVVCHS

III. THE INDEPENDENCE TESTING METHOD

In this section, we describe the step-by-step procedure that we used to test whether among the surviving descendants of the hypothetical common ancestor μ the adjacent pairs of amino acids are mutated independently of each other.

As an artificial and simplified example, suppose that there exists an ancestor protein μ that is made up of only the amino acids A, D, N and R as shown in Figure 3. Further assume during evolution each of these four amino acids either remains unchanged or is mutated into only one of the other three amino acids within this group of four amino acids. Suppose that the seven descendants are S_1, \dots, S_7 as shown also in Figure 3.

S_1	RNARDANDRADNRDANRARA
S_2	NRARDANRADADNANARNAD
S_3	RADNRANDANDRANDRDRAN
S_4	DNARDNARDNRDANRANR
S_5	RNDRANRDRDANDNANDRAN
S_6	RNARDANDRADNRDANRARA
S_7	RNARDADDRADNRDANDADA
μ	RNADRANRDRDANDRNADNAN

Fig. 3 A set of seven artificial sequences and their hypothetical common ancestor

Our testing method consists of the following five steps.

1. Construct the hypothetical common ancestor for the proteins in the given set of protein family using the method that is also used by the Common Mutation Similarity Matrix algorithm in the case of amino acid sequences. In the case of amino acid sequences, the hypothetical common ancestor, μ , is constructed by taking an alignment of the amino acid sequences, and in each column of the alignment finding the amino

acid (out of the twenty possible amino acids that are used in almost every protein in all organisms) that is *overall closest* to the all the amino acids in that column. The overall closest amino acid is by definition the one for which the sum of the PAM250 matrix distance values between it and the amino acids in the column considered is minimal. If there are two or more values that are minimal, then we make a random selection.

- Next, we calculate a *mutation probability matrix*. The mutation probability matrix contains the probabilities of any amino acid changing into another amino acid. For the running example with the data shown in Figure 3, the mutation probability matrix is shown in Table 1.

Table 1 The mutation probability matrix for the data in Figure 3.

	A	D	N	R
A	18/35	6/35	7/35	4/35
D	7/35	11/35	4/35	15/35
N	12/35	3/35	18/35	5/35
R	4/35	11/35	3/35	11/35

- Based on the mutation probability matrix values, we estimate the probability of the changes of any adjacent pair of amino acids into another pair of amino acids assuming that the mutations are independent of each other. For example, the probability of AN changing into an NN can be computed as follows:

$$Prob(AD, NR) = Prob(A, N) * Prob(N, N) = \frac{3}{35}$$

- The actual probabilities of changes are calculated for each pair of amino acids. The results for our example are shown in Table 2.
- We compare the theoretical and the actual probabilities and note the most important discrepancies. The *percentage probability difference* in the theoretical and actual probabilities of the mutations of amino acid pairs is the absolute value of the difference between the two types of probabilities divided by the maximum of the two probabilities. Let $T(p1, p2)$ and $E(p1, p2)$ be the theoretical and the experimental probabilities, respectively, that the amino acid pair $p1$ changes into the amino acid pair $p2$. Let also $PD(p1, p2)$ be the percent probability difference defined as follows:

$$PD(p1, p2) = \frac{|T(p1, p2) - E(p1, p2)|}{Max(T(p1, p2), E(p1, p2))}$$

IV. EXPERIMENTAL RESULTS

The experimental results shown in Table 3 are based on the protein family DiSB-ORF chro, which has a PFAM identification number PF16506. Table 3 displays only the top ten highest percentage probability differences that we found.

Table 3 Experimental results using the amino acid sequences in the DiSB-ORF chro protein family

Pair of amino acids		Theoretical probability under independence assumption $T(p1, p2)$	Actual probability $E(p1, p2)$	Percent probability difference $PD(p1, p2)$
From $p1$	To $p2$			
SP	SP	289/19600	6/7	1.0528
SP	PP	289/19600	6/7	1.0528
SP	TP	34/19600	1/7	0.329
SP	PS	119/19600	1/7	0.975
ER	ER	17/19600	2/7	0.997
ER	DV	17/19600	3/7	0.998
ER	GK	12/19600	1/7	0.995
DR	DP	260/19600	3/7	0.969
DR	DR	281/19600	4/7	0.975

V. CONCLUSION AND FUTURE WORK

The experimental results suggest that adjacent pairs of amino acids in the surviving descendants are sometimes mutated in a dependent instead of an independent way. However, the experimental data is based only on one protein family. In the future we plan to use our independence testing method for many other protein families. We also plan to experiment with using other amino acid substitution matrixes beside the PAM250 matrix [6]. We also plan to look at longer sequences, that is, consider adjacent N-mers of amino acids for $N > 2$.

Table 2 The actual probabilities of changes for each pair of amino acids for the artificial example protein family in Figure 3.

	RR	RN	RD	RA	NR	NN	ND	NA	DR	DN	DD	DA	AR	AN	AD	AA	DR
RR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RN	0	$4/7$	0	$1/7$	$1/7$	0	$1/7$	$1/7$	0	$1/7$	0	$2/7$	0	0	0	0	0
RD	$1/7$	0	0	$3/7$	0	0	0	0	$1/7$	0	0	0	0	$1/7$	$1/7$	0	$1/7$
RA	0	0	0	$1/7$	0	0	0	0	0	$1/7$	0	$4/7$	0	$1/7$	0	0	0
NR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ND	0	0	$1/7$	0	$1/7$	0	$3/7$	0	0	0	$1/7$	0	$1/7$	0	0	0	0
NA	0	0	0	$1/7$	0	0	$1/7$	$4/7$	0	0	0	0	0	0	$1/7$	0	0
DR	0	0	$5/7$	$1/7$	$1/7$	0	0	0	0	0	0	0	0	0	0	0	0
DN	0	$1/7$	0	$3/7$	0	0	0	0	$2/7$	0	0	$1/7$	0	0	0	0	$2/7$
DD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DA	0	$1/7$	0	0	0	0	$1/7$	0	0	0	0	$2/7$	0	0	$3/7$	0	0
AR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AN	$2/7$	0	0	$1/7$	$1/7$	0	0	$1/7$	0	0	0	$1/7$	0	$5/7$	$1/7$	0	0
AD	0	0	$1/7$	0	$3/7$	0	$2/7$	0	$1/7$	$1/7$	0	0	$6/7$	0	$1/7$	0	$1/7$
AA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

REFERENCES

- [1] D. Baum and S. Smith, *Tree Thinking: An Introduction to Phylogenetic Biology*, Roberts and Company Publishers. 2012.
- [2] B. G. Hall, *Phylogenetic Trees Made Easy: A How to Manual*, 4th edition, Sinauer Associates, 2011.
- [3] P. Lerney, M. Salemi, and A.-M Vandamme, editors. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edition, Cambridge University Press, 2009.
- [4] The PFAM Protein Library
Available: <http://pfam.xfam.org/family/PF16506>
- [5] P. Z. Revesz, "An algorithm for constructing hypothetical evolutionary trees using common mutations similarity matrices," Proc. 4th ACM International Conference on Bioinformatics and Computational Biology, ACM Press, Bethesda, MD, USA, September 2013, pp. 731-734.
- [6] P. Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, Springer, 2010.

Jyotsna Ramana is currently a graduate student in the Department of Computer Science and Engineering at the University of Nebraska-Lincoln. Her research interests are in bioinformatics, big data and database systems.

Peter Z. Revesz holds a Ph.D. degree in Computer Science from Brown University. He was a postdoctoral fellow at the University of Toronto before joining the University of Nebraska-Lincoln, where he is a professor in the Department of Computer Science and Engineering. Dr. Revesz is an expert in databases, data mining, big data analytics and bioinformatics. He is the author of *Introduction to Databases: From Biological to Spatio-Temporal* (Springer, 2010) and *Introduction to Constraint Databases* (Springer, 2002). Dr. Revesz held visiting appointments at the IBM T. J. Watson Research Center, INRIA, the Max Planck Institute for Computer Science, the University of Athens, the University of Hasselt, the U.S. Air Force Office of Scientific Research and the U.S. Department of State. He is a recipient of an AAAS Science & Technology Policy Fellowship, a J. William Fulbright Scholarship, an Alexander von Humboldt Research Fellowship, a Jefferson Science Fellowship, a National Science Foundation CAREER award, and a "Faculty International Scholar of the Year" award by *Phi Beta Delta*, the Honor Society for International Scholars.

On the Tradeoff between Biomass Exploitation and Exploration within an Interdisciplinary Research Project for Developing a New Product

Dimitrios Batzias, Dimitrios Sidiras, Christina Siontorou, Leonidas Kamarinopoulos, Yiannis Pollalis, Fragiskos Batzias

Abstract— This work deals with the tradeoff between biomass exploitation and exploration within an interdisciplinary research project for developing a new product, which is an innovative adsorbent suitable to remove pollutants from contaminated seawater or freshwater (in rivers, lakes, wetlands). Initially, the normalized ratio exploration/exploitation is defined and used as a basic independent/explanatory variable for determining its optimal value R_{opt} by minimizing an objective cost function C . The constituting partial conflict cost functions C_1 and C_2 , corresponding to expenditure for exploration and exploitation, respectively form the optimization tradeoff that gives R_{opt} at equilibrium, which is shifting by either introducing a novel biomass processing technology or by enhancing the mechanisms of ‘learning by doing’, thus performing sensitivity analysis under simulating conditions. Since the partial costs depend on several parameters, we have designed/developed a methodological framework under the form of an algorithmic procedure as an aid for optimizing R within a changing/dynamic environment. Subsequently, an implementation is presented, including (i) fuzzy multicriteria choice of a relevant R&D project, and (ii) optimal subsidy estimation to support such a project. The exploration-exploitation inter-relation is examined also within a network of biomass processing units (BPUs), where the tradeoff between centralization-decentralization presents another dimension of the same problem from a meta-Cybernetics point of view. A new knowledge acquisition mechanism is presented, as a modification of Nonaka’s SECI model, incorporated within a dynamic project management framework. Finally, R_{opt} is determined by maximizing a total benefit objective function $B = (B_1 + B_2)$, where the partial benefits B_1 and B_2 depend on exploration and exploitation, respectively.

Keywords— fuzzy multicriteria analysis, exploitation, exploration, tradeoff optimization.

This research has been co-financed by the European Union (European Social Fund- ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS. Investing in knowledge society through the European Social Fund. Project: THALIS - University of Piraeus - Development of New Material from Waste Biomass for Hydrocarbons Adsorption in Aquatic Environments (MIS 377356).

F. A. Batzias is with the Dep. Industrial Management & Technology, Univ. Piraeus, 80 Karaoli & Dimitriou, GR18534 Piraeus, Greece (corresponding author phone: +30-210-4142360; fax: +30-210-4142392; e-mail: fbatzias@unipi.gr).

D. K. Sidiras, D. F. Batzias, L. Kamarinopoulos and C. G. Siontorou are with the Dep. Industrial Management & Technology, Univ. Piraeus, 80 Karaoli & Dimitriou, GR18534 Piraeus, Greece (e-mail: sidiras@unipi.gr, dbatzias@unipi.gr, lkamarin@unipi.gr, csiontor@unipi.gr).

Y. Pollalis is with the Dep. Economic Science, Univ. Piraeus, 80 Karaoli & Dimitriou, GR18534 Piraeus, Greece (e-mail: yannis@unipi.gr).

I. INTRODUCTORY ANALYSIS

As quoted in [1], the relation between exploration of new ‘opportunities’ and the exploitation of current ‘strengths’ (following the terminology of SWOT analysis) is very important in order to optimize enterprises’ operation from a techno-economic point of view in the medium/long run. The keywords frequently used to describe/characterize procedures related to exploitation are: process control, product quality, refinement, efficiency increase, carrying out, measurement reliability/precision, production management. On the other hand, keywords frequently used to describe/characterize procedures related to exploration are: research, development, scale up/down, risk, hypothesis setting/testing, experimental design, innovation, discovery, patent.

More specifically, according to [2], exploitation involves the introduction of new combinations that grow out of the old by means of continuous adjustments, in several small steps, and implies a relatively restricted search for alternative solutions to complement an existing technology.

Exploitation thus allows an organization to reduce the likelihood of errors and false starts, and facilitates the development of procedures, making search more reliable/fruitful [3]; in addition, it favors the use of accumulated knowledge, which boosts the enterprises’ ability to introduce new products or new productive processes that in many ways may be not obvious to less experienced competitors [4]. By limiting alternative solutions and increasing efficiency in current operations (i.e., decreasing width or degrees of freedom and increasing depth or information granularity level), the medium/long term performance is significantly improved [5].

On the other hand, exploration refers to distant, system-wide research leading to breakthrough results creating discontinuity, helping to (i) fight organizational inertia, (ii) avoid competency entrapment [6], and extent knowledge/know-how beyond the borders of the topics which are already familiar to the departments of the enterprise [7]-[10]. Moreover, exploration contributes to (i) reducing the probability of equipment and skillfulness to become obsolete [11], and (ii) developing experimental capabilities that may lead to innovative products (and specify/optimize their production conditions) and processes, possibly saving resources [12]. All these factors have a favorable impact to the economic effectiveness [13], taken also into account for determining any

subsidy granted by the Local Authorities or the State.

The strategic significance of balancing exploitation and exploration has been extensively discussed in the scientific and technical literature [1], [14], [15]. Although the definitions of exploitation and exploration vary depending on the authors and their respective topics, there is consensus that the strategic balancing of exploitation and exploration is very interesting in both, the short- and long-term perspective [16]. The most common approach is to search for balance through either ambidexterity or punctuated equilibrium (symbolized herein by AD or PE, respectively). These two approaches are apparently disjunctive, although may coexist in certain time periods, in which case a certain interaction should occur between them (i.e., in Information Theory, the respective logical operator OR connecting these two approaches is rather inclusive than exclusive). The first of them makes use of a tradeoff between exploration-oriented and exploitation-specific units without emphasizing on one of them. On the contrary, according to the PE concept, we put temporarily special emphasis to either exploration or exploitation, and subsequently give priority to the other one, after a certain level of saturation has been reached; therefore, the PE approach is based on a cyclic procedure between exploration and exploitation.

Reich et al. [17] stated/examined three perspectives related to accomplishing the organizational ambidexterity aimed at balancing exploitation and exploration for the sake of sustainable organizational performance: how to obtain ambidexterity, level of pursuing ambidexterity, and temporal perspective (differentiation/ integration, individual/ organizational, static/dynamic, respectively).

An empirical approach used in [18] showed that ambidextrous organizations placing emphasis on technological innovation yield better performance than punctuated organizations. Moreover, two kinds of realizing ambidexterity have been proposed in [19]. The first is architectural ambidexterity, which focuses on the use of organizational structure and strategy to enable proper differentiation of exploitation and exploration; the second is contextual ambidexterity, which involves utilizing more behavioral and social means to integrate exploitation and exploration. In another study [20], the single case of a firm that centered only on exploitation for growth and almost faced bankruptcy, was used to suggest that it is possible for such a firm to renew its growth engine based on exploration, aiming to stay on the balance of exploration and exploitation. Last, the existence of a competency trap, degrading firm performance, was recognized in [21] as the effect of the imbalance (i.e., excessive exploration or exploitation), indicating the significance of organizational ambidexterity.

In general, the tradeoff between exploitation and exploration is similar to the tradeoff between short and long term resources consumption (including effort intensification/duration and knowledge acquisition/processing) for the achievement of a predetermined target, under the assumption of a certain limited budget. Evidently, the future resources consumption should be reduced to present value through forecasting techniques in order to be comparable with short term expenditure. In the

special case that the work to be done is an R&D project for developing a new product, neither the stages to be followed for the achievement of the target are strictly determined *a priori* nor the target itself is known in detail, especially in the case of using a highly heterogeneous raw material, like waste biomass. For this reason, we have adopted/adapted a modified version of dynamic programming [22], where the Bellman's Principle of Optimality dominates either under the form of the 'Bellman equation' in case of solving discrete-time optimization problems or under the form of the 'Hamilton – Jacobi – Bellman equation' in case of solving continuous-time optimization problems.

By denoting with R the normalized ratio exploration/exploitation, i.e. defined in the closed domain $(0,1)$, we can determine the optimal value R_{opt} at minimum cost $C_{min} = (C_1 + C_2)_{min}$, where $C_1(R)$ and $C_2(R)$ are the partial costs corresponding to expenditure for exploration and exploitation, respectively. The former is an increasing function of R with an increasing rate (i.e., $dC_1/dR > 0$, $d^2C_1/dR^2 > 0$), because of moving, with R , from surface to deeper scientific phenomenological level, implying more expensive equipment and human expertise, usually consisted of implicit/tacit knowledge. The latter is a decreasing function of R with an increasing algebraic or a decreasing absolute rate (i.e., $dC_2/dR < 0$, $d^2C_2/dR^2 > 0$ or $d|dC_2/dR| / dR < 0$), since moving towards lower R -values the uncertainty (due to lack of scientific information) as well as the expenditure for exploitation increases. Evidently, R_{opt} is the abscissa of the $C_{min} = (C_1 + C_2)_{min}$ point, at $MC_1 = MC_2$, where $MC_1 = dC_1/dR$ and $MC_2 = |dC_2/dR|$ are the marginal or differential partial costs at the tradeoff equilibrium point, respectively.

In case of introducing a novel biomass processing technology (as a result of an R&D based invention or knowhow), the C_1 -curve will move downwards to C_1' becoming more flat, since the deviation from the original C_1 -curve is expected to be higher in the region of high R -values, where advanced equipment usage and scientific knowledge acquisition contribute more to this partial cost; as a result, R_{opt} is shifting to R'_{opt} , where $R'_{opt} > R_{opt}$, as shown in Fig. 1a. In case of enhancing the mechanisms of 'learning by doing' either for the accumulation of empirical experience or/and for partial transforming of implicit to explicit knowledge, the C_2 -curve will move downwards to C_2' becoming also more flat, since the deviation from the original C_2 -curve is expected to be higher in the region of low R -values, where these mechanisms are dominant; as a result, R_{opt} is shifting to R''_{opt} , where $R''_{opt} < R_{opt}$, as shown in Fig. 1b.

From our point of view, there are four main factors that may increase effectiveness in the medium/long run: (i) determination/selection/acquisition of the means necessary to either reach successfully a preset target or achieve a significant comparative advantage; (ii) recognition of the qualitative/quantitative progress in the to-me course; (iii) selection/application of criteria for changing from AD to PE, and *vice versa*; (iv) construction of an internal Knowledge Base (KB) for either adapting old criteria to new conditions/circumstances or adopting new criteria after consultation provided by interconnecting the internal KB with

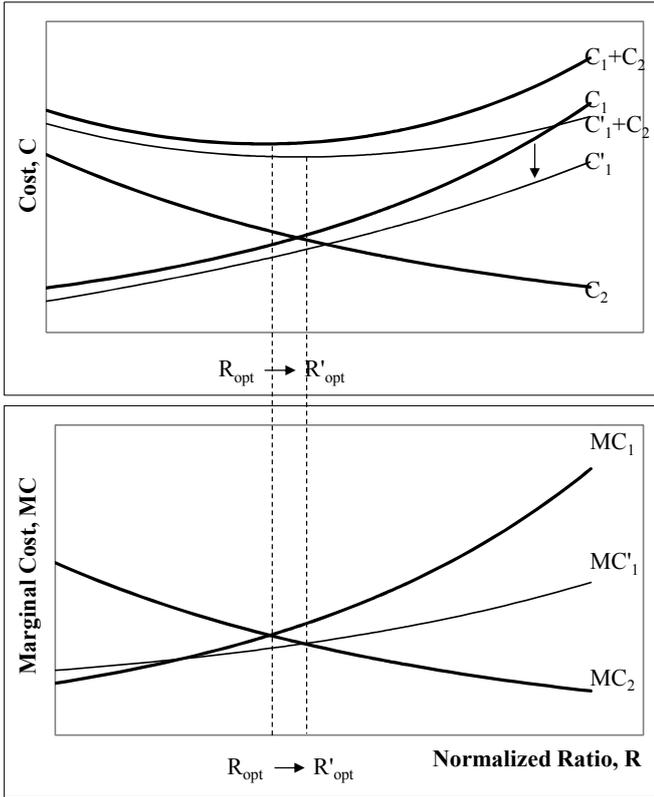


Fig. 1a. Dependence of partial costs C_1 and C_2 , corresponding to expenditure for exploration and exploitation, on the normalized ratio exploration/exploitation, R , and shifting of the optimal value R_{opt} in case of introducing a novel biomass processing technology (as a result of an R&D based invention or knowhow).

an Intelligent Agent (IA) searching in external KBs. Stages (ii) and (iii) include a threefold dialectic scheme of either Fichtean form (thesis – antithesis – synthesis) or Hegelian form (abstract – negative – concrete). The Fichtean model is based on the concept that contradictions or ‘negations’ come from the outside environment in relation with the project under consideration; on the contrary, the Hegelian model is based on the concept that the corresponding procedure is inherent/internal to the project/phenomenon under examination, coming closer to views of the pre-Socratic Greek philosopher Heraclitus. Beyond this disagreement, there is a consensus on the necessity of taking into account the content of a topic in order to apply the proper dialectic formalism upon it. The implication of this consensus is that we have to determine/identify firstly the R&D project under examination (‘biomass based adsorbent’ in our case) with as limited width/depth as possible, in order to maximize effectiveness, under the restriction of a pre-estimated budget, or minimize cost for a preset level of effectiveness, as measured by the corresponding specifications. This can be achieved by (i) decomposing the whole project in quasi-independent sub-projects, and (ii) applying the same criteria vector for ranking these sub-projects, in terms of decreasing preference, to characterize the ranked first as the dominant one. This procedure includes significant difficulties, since several activity stages and decision making nodes should proceed (like

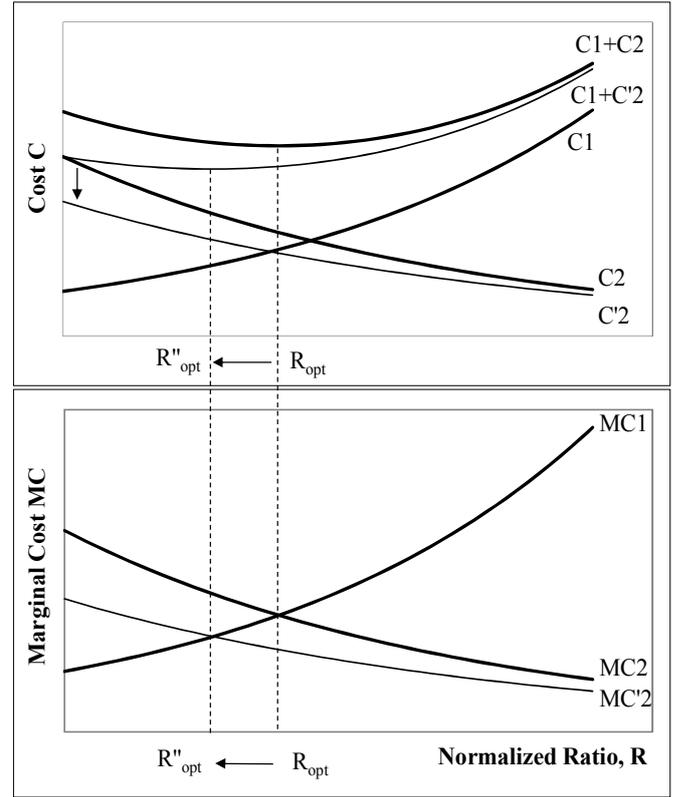


Fig. 1b. Dependence of partial costs C_1 and C_2 , corresponding to expenditure for exploration and exploitation, on the normalized ratio exploration/exploitation, R , and shifting of the optimal value R_{opt} in case of enhancing the mechanisms of ‘learning by doing’ either for the accumulation of empirical experience or/and for partial transforming of implicit to explicit knowledge.

data collection, demand/supply side analysis, Life Cycle assessment, modularization of the production/usage/disposal stages), while other should follow Multicriteria Analysis, abbreviated as MCA (like sensitivity/robustness analysis, ontological integration of information, SWOT analysis, estimation of subsidies) without a predetermined order and a specific interaction pattern.

II. METHODOLOGY

In order to cope with the difficulties quoted above, we have developed a methodological framework, under the form of an algorithmic procedure, including the following 28 activity stages and 9 decision nodes, interconnected as shown in the flow chart of Fig. 2.

1. General description of the biomass-based product intended to be produced and statement of the relevant R_{opt} -estimation problem.
2. Demand side analysis.
3. Required specifications based on marketing research.
4. Determination of the relation between the biomass-based product specifications/properties and the corresponding production conditions.
5. Provision of additional data, mainly through experimentation and simulation.

6. Supply side analysis, including biomass logistics at proper Geographical Information System (GIS) layers.
7. Collection of relevant data through the KB and the IA, mentioned in stage 28.
8. Modularization of the production/usage/disposal stages of the biomass based new product by means of Life Cycle Analysis (LCA), considering this procedure as a partonomy function within a framework of an ontological approach; correspondence of R&D projects to these stages forming the alternatives for the subsequent multicriteria decision making.
9. Determination of criteria to be used for alternatives ranking.
10. Values assignment to the elements of the criteria vector and the preference matrix, in a fuzzy version, in order to count for uncertainty.
11. Performance of multicriteria analysis (MCA).
12. Defuzzification and sensitivity analysis of the ranked first alternative.
13. Robustness analysis of the ranked first alternative.
14. SWOT analysis to connect exploitation with exploration.
15. Estimation of subsidies supporting R&D programs, renewable resources management, energy saving, materials substitution, environmental protection at local level, jobs creation for unemployment decrease, sustainable regional development.
16. Collection/processing of the required techno-economic data at micro-scale and estimation of the respective additional impact at medium/macro-scale.
17. Selection of the most necessary scientific topics (at the closest deeper phenomenological level) to supplement the required information.
18. Design/development/enrichment of the corresponding hierarchical Ontology.
19. Evaluation and thinning of this ontology.
20. Extension of the acquired/used information at the same phenomenological level through observation/experimentation/simulation.
21. Establishment of such a preliminary network.
22. Experimental design and assignment of measurements to (i) the proper members of the existing network and (ii) other laboratories/agents becoming (even temporarily) new members of this network.
23. Measurements performance
24. Statistical processing of the interlaboratory examination and Ontology enrichment/updating (and also restructuring if necessary).
25. Optimization of independent/explanatory variables and model parameter values.
26. Sensitivity/robustness analysis of optimal solutions and application to confirm the results.
27. Revision (if necessary) of the subsequent R&D stages/steps by following either the 'Bellman

equation' or the 'Hamilton-Jacobi-Bellman equation' in case that the R_{opt} -estimation problem we are dealing with is a discrete-time or a continuous optimization problem, respectively.

28. Enrichment of the internal KB by (i) accumulating/evaluating experience and (ii) updating its content through the respective IA.

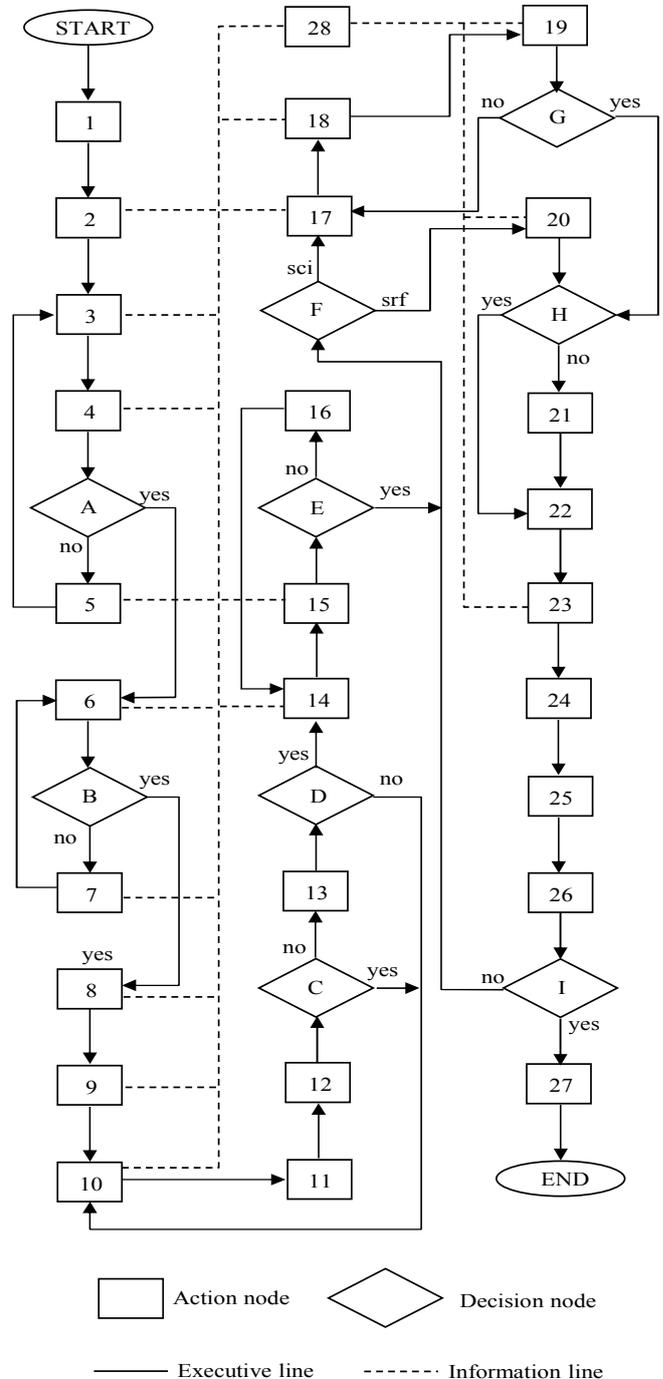


Fig. 2. The flowchart of the algorithmic procedure we have designed /developed in order to (i) overcome the difficulties mentioned in the Introductory Analysis Section, and (ii) upgrade knowledge management towards a holistic approach.

- (A) Are the available data adequate to establish such a relation?
- (B) Are the available supply side data sufficient to support this analysis at the required information granularity level?
- (C) Is this solution sensitive?
- (D) Is this solution robust?
- (E) Are the relevant data qualitatively reliable and quantitatively sufficient for documenting adequately the subsidies claims?
- (F) Is the already incorporate information preferred to be (i) extended at the same (possibly empirical/surface or directly referred to the usage/functionality of the product) phenomenological level or (ii) supplemented at the deep/scientific level (the answer denoted by *surf* or *sci*, respectively, in the flowchart of Fig. 2)?
- (G) Is this Ontology adequate to support the present algorithmic procedure?
- (H) Is there an already established interlaboratory network capable to carry out the required experimental work?
- (I) Are they confirmed?

The SWOT analysis (denoting Strengths, Weaknesses, Opportunities, Threads), quoted in stage 14, is closely related to the subsequent stages, since it contributes to both present situation assessment and forecasting/backcasting.

III. IMPLEMENTATION

According to stages 8-13 of the flowchart presented in Fig.2 and described in the Methodology Section, we have decomposed the whole R&D project (for developing a biomass-based novel adsorbent) in the following modules/topics to form priority alternatives, suitable for MCA. Defuzzification is performed in input and the results are shown in Table I. The procedure of weighting the grades of the alternatives and summing up to give the corresponding total criterion SWA-values is shown in Table II. Where the ranking is BTP>AAI>SAU>PTS>RCT, in order of descending preference. Mono-parametric sensitivity analysis for two cases is presented in Figs. 3a and 3b, where the point of order change between the first two alternatives is also shown.

The alternative R&D topics are: Raw material Collection/Transportation (RCT), Biomass Treatment/Processing (BTP), Product Transportation/Storing (PTS), Adsorbents Application In Citu (AAI), Spent Adsorbents Usage (SAU). The criteria used are: Partial Economic Benefit (PEB), Total Economic Benefit (TEB), Favorable Environmental Impact (FEI), Development of Promising Technology (DPT).

Table I. Defuzzified crisp values assigned to 4x5 preference matrix and the weights vector W_i .

	W_i	RCT	BTP	PTS	AAI	SAU
PEB	0.141	4.07	4.98	4.65	6.01	5.94
TEB	0.258	4.52	6.34	4.27	5.86	5.17
FEI	0.406	5.76	7.09	7.12	6.97	6.62
DPT	0.195	3.61	7.52	3.98	7.25	7.18

Table II. Ranking of Alternatives, according to the general criterion WSA (Weighted Sum for each Alternative) giving the following order in terms of descending preference; BTP > AAI > SAU > PTS > RCT.

	RCT	BTP	PTS	AAI	SAU
PEB	0.57387	0.70218	0.65565	0.84741	0.83754
TEB	1.16616	1.63572	1.0166	1.51188	1.33386
FEI	2.33856	2.87854	2.89072	2.82982	2.68772
DPT	0.70395	1.4664	0.7761	1.41375	1.4001
WSA	4.78254	6.68284	5.42413	6.60286	6.25922

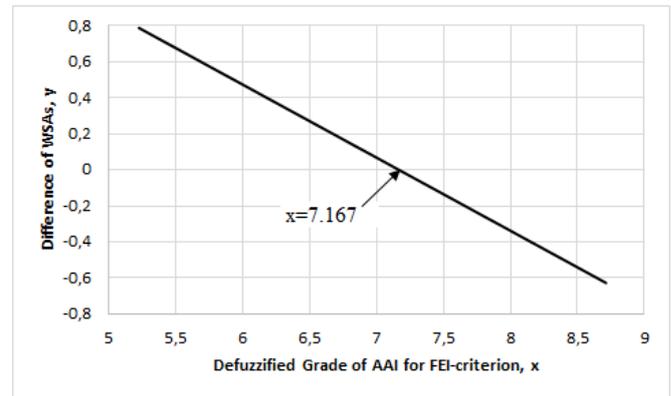


Fig. 3a. Mono-parametric sensitivity analysis of the ranked first alternative Biomass Treatment/Processing (BTP) by changing the corresponding to Favorable Environmental Impact (FEI) grade for the ranked second alternative Adsorbents Application In Citu (AAI) within the range $\pm 25\%$ of the defuzzified value 6.97, expressed as crisp number.

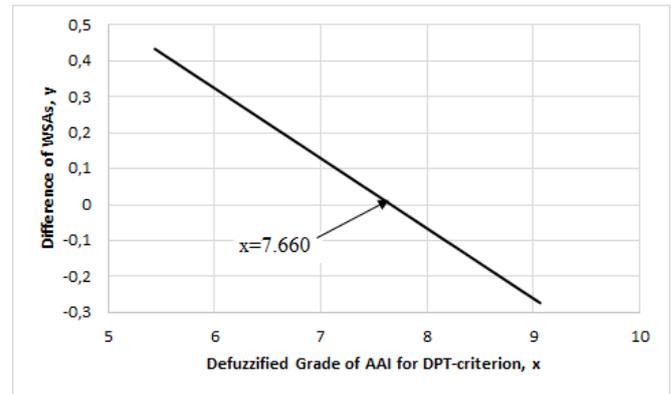


Fig. 3b. Mono-parametric sensitivity analysis of the ranked first alternative Biomass Treatment/Processing (BTP) by changing the corresponding to Development of Promising Technology (DPT) grade for the ranked second alternative Adsorbents Application In Citu (AAI) within the range $\pm 25\%$ of the defuzzified value 7.25, expressed as crisp number.

According to stage 15, we have estimated a subsidy to be granted from the State to the BPUs system as follows. Let K the fraction of annual energy/materials saving and environmental protection benefits (all expressed in monetary

units), which is deducted by the State from each welfare budget, in order to support such activities, that correspond to best exploitation practices. By denoting with a the portion of K agreed to be given for exploration activities, the annual ‘gains’ are presented by the following amounts V_j ($j=1, 2, \dots, t$), expressed in present values (see also [23]-[25]):

$$\begin{aligned} V_1 &= K \cdot F \cdot (1+i)^{-1} \\ V_2 &= K \cdot F \cdot (1+i)^{-2}(1+f)(1+g) \\ &\dots \\ V_t &= K \cdot F \cdot (1+i)^{-t}(1+f)^{t-1}(1+g)^{t-1} \\ V &= \sum_{j=1}^t V_j = K \cdot F \cdot (1+i)^{-1} \cdot (1+z+z^2+\dots+z^{t-1}) \end{aligned} \quad (1)$$

where $z = (1+f)(1+g)/(1+i)$, t is the number of time periods (herein years), corresponding to the useful life of the investment to be subsidized, i is the common interest rate (used for money equivalents in the time course), F is the first year energy/materials savings (including environmental benefits), increased each time period by fractions f and g for energy/environment and materials savings/benefits.

The expression giving V is a geometric series, the sum of which is easily calculated:

$$V = K \cdot F \cdot (1+i)^{-1} (z^t - 1) / (z - 1) \quad (2)$$

Thus the total value of K deducted by the State for determining the corresponding subsidy is given by

$$K = \frac{F(1-a)/V \left[\frac{(1+f)(1+g)-1}{1+i} \right]}{\left[\frac{(1+f)(1+g)}{1+i} \right]^t - 1} \quad (3)$$

The diagrams in Figs. 4 and 5 give the dependence of K (representing subsidy for both exploitation and exploration) on i and a (representing the subsidy fraction granted for exploration).

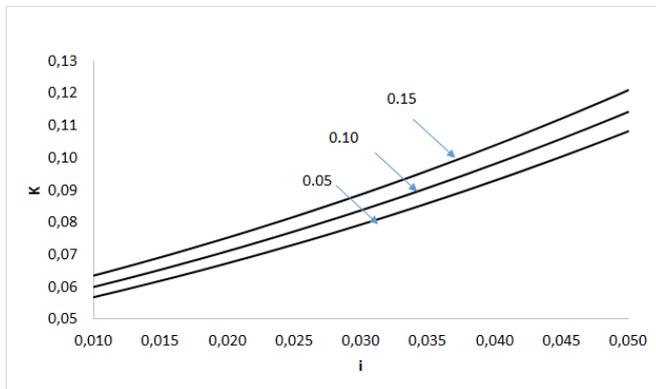


Fig. 4. Dependence of K on interest rate, ranging from 0.01 to 0.05 (corresponding to deflation and inflation, respectively) for different values of a .

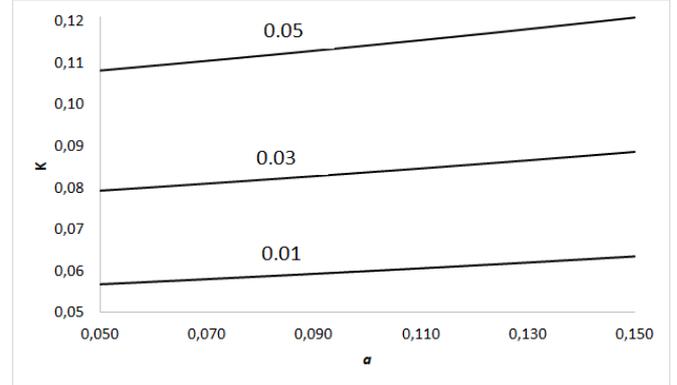


Fig. 5. Dependence of K on fraction a , ranging from 0.05 to 0.15 (corresponding to minimum and maximum values, set a priori, respectively) for different values of interest rate i .

IV. DISCUSSION

Instead of determining the optimal R_{opt} at C_{min} we can use a more general optimization technique by considering total benefit $B(R) = B_1(R) + B_2(R)$ as the objective function, where $B_1(R)$ and $B_2(R)$ are the partial benefits, due to exploration and exploitation, respectively, opposing to each other in the corresponding trade off, like in the case described in Fig. 1. Evidently, the benefits constituting B_1 grow slowly in the beginning (initial stages after the start up, representing an incubation period, where $dB_1/dR > 0$ and $d^2B_1/dR^2 > 0$), rapidly later, because of synthesizing / incorporating the acquired knowledge / knowhow into upgraded structures, and slowly again, as approaching asymptotically a state of saturation. The main part of the resulting sigmoid curve is after the point of inflection (at which $dB_1/dR = 0$ and $d^2B_1/dR^2 = 0$, since this part (where $dB_1/dR > 0$ and $d^2B_1/dR^2 < 0$) is the more fruitful one from the R&D point of view.

On the other hand, the opposing benefits represented by B_2 decrease (i) gradually/slowly in the region of lower R -values, where the resources spent for exploitation are adequate, (ii) rapidly in the region of medium R -values, where the resources are not available in satisfactory quality/quantity, and (iii) very rapidly, since at least one of the input variables is sub-optimal, according to the LDR, or insufficient; in mathematical terms, $dB_2/dR < 0$, $d^2B_2/dR^2 < 0$ or $d|dB_2/dR|/dR > 0$. Consequently the R_{opt} -value is the abscissa of $B_{max} = (B_1 + B_2)_{max}$ at $MB_1 = MB_2$, where $MB_1 = dB_1/dR$ and $MB_2 = |dB_2/dR|$ are the marginal values of B_1 and B_2 , respectively, at the tradeoff equilibrium.

At this point, is important to introduce the knowledge management approach, we have adopted for signifying exploration pathways by using a modification of Nonaka's model, including four stages in cyclic arrangement (i.e., without start/end points): Socialization (S), as the process of converting shared experience into new implicit knowledge; Externalization (E), as the process of converting implicit to explicit knowledge; Combination (C), as the process of converting rather simple to more complex explicit knowledge; Internalization (I), as the process of converting explicit into implicit knowledge, by embodying the former into the experts'

cognitive background. This illustrative mechanism indicates a movement rather on a spiral than on a cycle circumference, since each recursive procedure is changing as a function of time.

By decomposing the R&D project in quasi-independent sub-projects and running several of them in parallel *ab initio* (as suggested in the Introductory analysis section), we may have accelerated progress in exploration (mainly achieved in Nonaka's E-stage) especially in the region of low R -values, where the corresponding resources are limited and the R&D results still poor and / or out of focus. As a result, the B_1 -curve is expected to move upwards to B_1' , becoming more flat, since the benefits increase will be more expressed in the region of higher R -values, where the resources for exploitation are limited; as a result, R_{opt} is shifting to R'_{opt} , where $R'_{opt} < R_{opt}$, as shown in Fig. 6a. If such a progress is not possible, the available resources and the corresponding progress is expected to be re-oriented towards the Nonaka's C stage, favoring exploitation, possibly contributing to a change from AD to PE or vice versa (see Introductory analysis). In this case, the B_2 -curve is expected to move upwards to B_2' , becoming also more flat, since the benefits increase will be more expressed in the region of lower R -values, where the resources for exploration are limited; as a result, R_{opt} is shifting to R''_{opt} , where $R''_{opt} > R_{opt}$, as shown in Fig. 6b.

It is worthwhile noting that in case of the composition, the quasi-independent sub-project with the low rate knowledge conversion enforces the total rate if and only if (denoted by 'IF' in symbolic logic) the independence of sub-projects remains after transition, as shown in Fig. 7a; otherwise, a combined rate conversion results as shown in Fig. 7b.

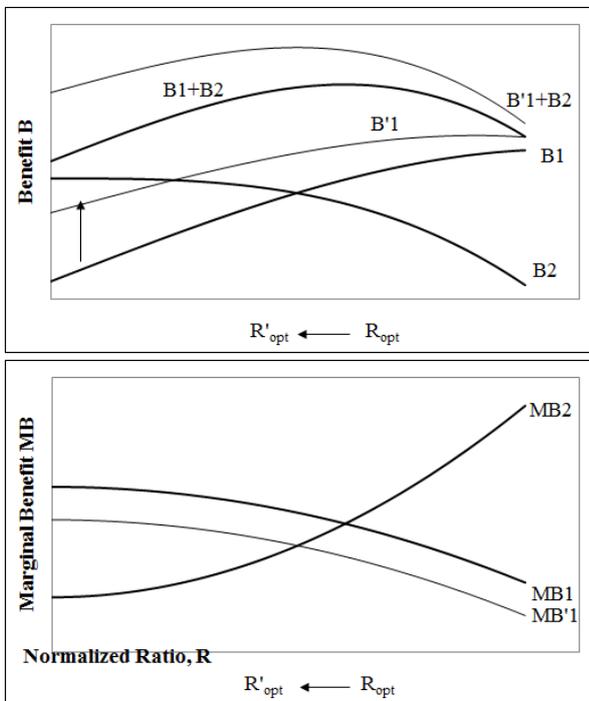


Fig. 6a. Dependence of partial benefits B_1 and B_2 on normalized ratio R , and shifting of the optimal value R_{opt} in case of decomposing the R&D project in quasi-independent sub-projects and running several of them in parallel, possibly resulting to accelerated progress in exploration.

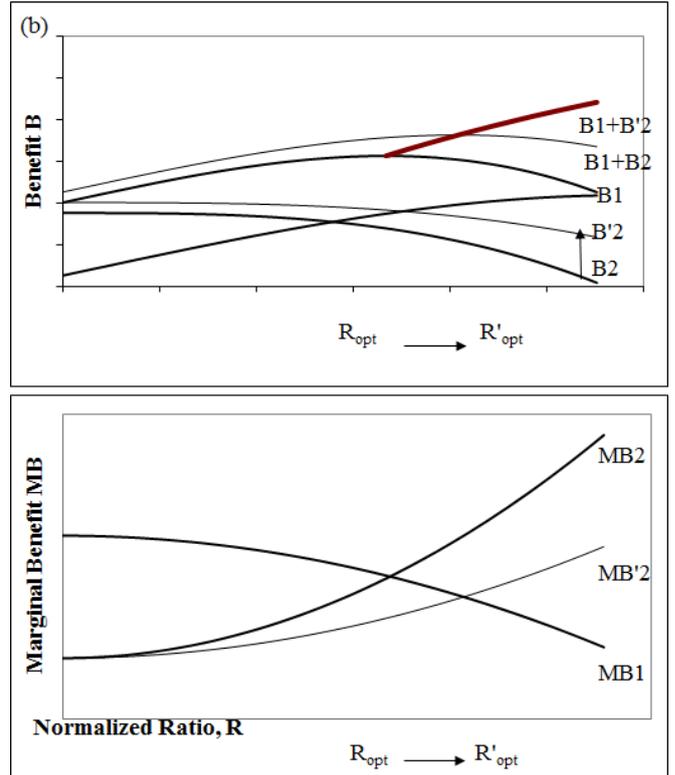


Fig. 6b. Dependence of partial benefits B_1 and B_2 on normalized ratio R , and shifting of the optimal value R_{opt} in case that the available resources and the corresponding progress is expected to be re-oriented towards the Nonaka's C stage, favoring exploitation; the thick curve in the upper diagram is the locus of the B_{max} -points.

V. CONCLUSIONS

We have proved that certain difficulties appear when applying tradeoff optimization for balancing exploitation against exploration under conditions of a limited budget. A methodological framework, we have designed under the form of an algorithmic procedure including 28 activity stages and 9 decision nodes has been successfully implemented, as an add to overcome these difficulties, concerning the following characteristic topics: (i) decomposition of the whole project of R&D referring to a novel biomass-based adsorbent; (ii) multicriteria ranking (including sensitivity/robustness analysis) of sub-projects for giving priority/emphasis in research; (iii) determination of corresponding subsidies to be granted for exploration and exploitation; (iv) statement of a new normalized exploration/exploitation ratio R and estimation of its optimal value by using minimum cost and maximum benefit as optimization criteria, (v) modification of Nonaka's knowledge management model by discriminating sub-models running in parallel for accelerating progress.

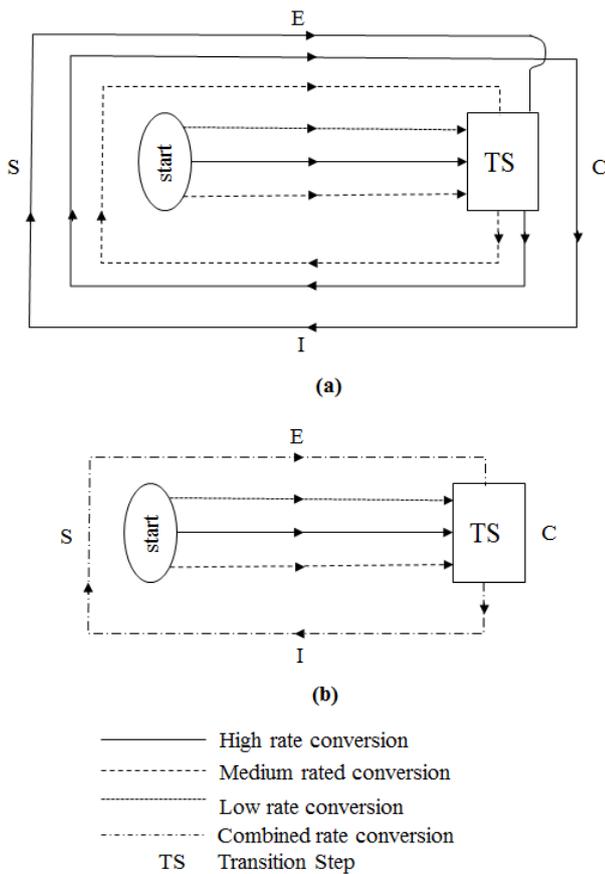


Fig. 7. A modified Nonaka's knowledge conversion model, taking into consideration the decomposition of the R&D project into three quasi-independent sub-projects and running them in parallel (see corresponding text for relevant analysis).

ACKNOWLEDGMENT

The authors would like to thank Prof. A. Kanas (Dep. Economic Science, Univ. Piraeus, Greece). This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES - Investing in knowledge society through the European Social Fund. Project: THALIS – University Of Piraeus – Development of New Material from Waste Biomass for Hydrocarbons Adsorption in Aquatic Environments. MIS 377356.



REFERENCES

- [1] J. G. March, "Exploration and exploitation in organizational learning", *Organization Science*, Vol. 2, 1991, pp. 71–87.
- [2] C. Gatti, L. Volpe, G. Vagnani, "Interdependence among productive activities: Implications for exploration and exploitation", *Journal of Business Research*, Vol. 68, Issue 3, March 2015, pp. 711–722.
- [3] D. Levinthal, J.G. March, "A model of adaptive organizational search", *Journal of Economic Behavior and Organization*, Vol. 2, (1981), pp. 307–333.
- [4] R. Katila, G. Ahuja, "Something old, something new: A longitudinal study of search behavior and new product introduction", *Academy of Management Journal*, Vol. 45, 2002, pp. 1183–1194.
- [5] D. Lavie, U. Stettner, M.L. Tushman, "Exploration and exploitation within and across organizations", *Academy of Management Annals*, Vol. 4, 2010, pp. 109–155.
- [6] B. Levitt, J.G. March, "Organizational learning", *Annual Review of Sociology*, Vol. 14, 1988, pp. 319–340.
- [7] J.G. March, "Rationality, foolishness, and adaptive intelligence", *Strategic Entrepreneurship Journal*, Vol. 27, 2006, pp. 201–214.
- [8] L. Rosenkopf, A. Nerkar, "Beyond local search: Boundary-spanning, exploration, and impact in the optical disc industry", *Strategic Management Journal*, Vol. 22, 2001, pp. 287–306.
- [9] L. Fleming, "Recombinant uncertainty in technological search", *Management Science*, Vol. 47, 2001, pp. 117–132.
- [10] W. J. Abernathy, K. B. Clark, "Mapping the winds of creative destruction", *Research Policy*, Vol. 14, 1985, pp. 3–22.
- [11] D. Leonard-Barton, "Core capabilities and core rigidities: A paradox in managing new product development", *Strategic Management Journal*, Vol. 13, 1992, pp. 111–126.
- [12] L. Fleming, O. Sorenson, "Technology as a complex adaptive system: Evidence from patent data", *Research Policy*, Vol. 30, 2001, pp. 1019–1039.
- [13] A.Y. Lewin, C.P. Long, T.N. Caroll, "The coevolution of new organizational forms", *Organization Science*, Vol. 10, 1999, pp. 535–550.
- [14] D. Y. Choi, K. C. Lee, "Dynamic resource allocation for exploitation and exploration with ambidexterity: Logical mechanism and simulations", *Computers in Human Behavior*, Vol. 42, January 2015, pp. 120–126.
- [15] D. J. Teece, G. Pisano, A. Shuen, "Dynamic capabilities and strategic management", *Strategic Management Journal*, Vol. 18, 1997, pp. 509–533.
- [16] A. K. Gupta, K. G. Smith, C. E. Shalley, "The interplay between exploration and exploitation", *Academy of Management Journal*, Vol. 49, 2006, pp. 693–706.
- [17] S. Raisch, J. Birkinshaw, G. Probst, M. L. Tushman, "Organizational ambidexterity: Balancing exploitation and exploration for sustained performance", *Organization Science*, Vol. 20, 2009, pp. 685–695.
- [18] Z. He, P. Wong, "Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis", *Organization Science*, Vol. 15, 2004, pp. 481–494.
- [19] C. Andriopoulos, M. W. Lewis, "Exploitation–exploration tensions and organizational ambidexterity: Managing paradoxes of innovation", *Organization Science*, Vol. 20, 2009, pp. 696–717.
- [20] P. McNamara, C. Baden-Fuller, "Lessons from the Celltech case: Balancing knowledge exploration and exploitation in organizational renewal", *British Journal of Management*, Vol. 10, 1999, pp. 291–307.
- [21] W. Liu, "Knowledge exploitation, knowledge exploration, and competency trap", *Knowledge and Process Management*, Vol. 13, 2006, pp. 144–161.
- [22] D. Sidiras, D. Batzias, C. Siontorou, L. Kamarinopoulos, Y. Pollalis, F. Batzias, "Introducing Dynamic Project Management in a Novel Adsorbent Production/Application Scheme for Oil Spill Response in Aquatic Environments" in *WSEAS Proc. Intern. Conf. on Energy, Environment, development and Economics (EED 2015)*, by A. Bulucea (Ed.), 2015, pp. 246-253.
- [23] A.P. Oliveira Francisco, F.A. Batzias, H.A. Mato. "Computer aided determination of optimal subsidy for installing energy saving

- equipments within an industrial complex". *Computer Aided Chemical Engineering*, Vol. 18, 2004, pp. 409-414.
- [24] F. Batzias, A. Bountri, "Determination of maximum allowable subsidy for natural resources optimal exploitation and recycle", in proc. WSEAS Intern. Conf. on Energy, Environment, Devices, Systems, Communications and Computers (EEDSCC '11), Eds. S. Chen, N. Mastorakis, F. Rivas-Echeverria, V. Mladenov, Venice, Italy, 2011, pp 19-24.
- [25] F. Batzias, I. Salapa, C. Siontorou "On the tradeoff between reliability and uncertainty when combining bioreactors for wastewater treatment", in Proc. of 10th WSEAS Intern. Conf. on Environment, Ecosystems and Development, Eds. S. Oprisan et al., Montreux, Switzerland, Dec. 2012, pp 181-186.

A-Maze-D: Advanced Maze Development Kit Using Constraint Databases

Shruti Daggumati, Peter Z. Revesz, and Corey Svehla

Department of Computer Science and Engineering

University of Nebraska-Lincoln

Lincoln, Nebraska 68588-0115

Email: sdagguma@cse.unl.edu, revesz@cse.unl.edu, csvehla@unl.edu

http://cse.unl.edu/revesz Telephone: (1+) 402 472-3488

Abstract—In this paper, we describe the A-Maze-D system which shows that constraint databases can be applied conveniently and efficiently to the design of maze games. A-Maze-D provides a versatile set of features by a combination of a MATLAB library and the MLPQ constraint database system. A-Maze-D is the first system that uses constraint databases to build maze games and opens new ideas in video game development. *Keywords*—animation, constraint database, maze, MLPQ, moving objects, video gamification, constraint database, maze, MLPQ, moving objects, video game

I. INTRODUCTION

The rapidly growing video game industry has a revenue of approximately twelve billion U.S. dollars per year in the United States alone (Statista [8]). The efficient development of new video game products is needed to supply the growing demands for video games that have become ubiquitous on our computers, consoles, and phones.

A large set of video games require the representation of a map, usually some kind of maze, and other spatial objects. In addition, video games also routinely require the representation of moving objects. Hence video games have a strong connection with geographic, spatial and moving object (also called spatio-temporal) databases. Since these types of databases can be viewed as special cases of constraint databases (Kanellakis et al. [2], Revesz [3]), we propose *A-Maze-D*, an *Advanced Maze Development* kit with a novel design based on the MLPQ system (Revesz et al. [4]). The MLPQ system, developed at the University of Nebraska-Lincoln, is one of the systems that implements and visualizes constraint databases and has been used already in many spatial and moving object database applications.

Our proposed A-Maze-D system provides a large set of useful features that enable game development where the main objective is to find the way out of a maze with limited viewing distance from an overhead view. We describe in detail the features that are the most important in developing maze games. We envision that with a growing set of features, the development can be extended from maze games to a plethora of different types of video games.

This paper is organized as follows. Section 2 describes some related work on mazes and constraint databases. Section 3 presents the A-Maze-D system. Section 4 provides some conclusions and future work.

II. RELATED WORK

A. Mazes and Maze Games

A maze is a complex passage with multiple branches where the user needs to find the best route. In most mazes, the walls are fixed and do not change as the user progresses through the maze. Maze solving computer algorithms that try to find the best path to an exit or the fastest way to attain a prize have been implemented many times before. However, the typical solutions do not take into account a user's limited vision and inability to mark the walls as hindrances for solving the mazes. In Section 3, we specifically allow the option of representing the limited vision of a user.

In games like Super Mario World, there are levels where the user needs to solve a maze and has limitations as to what they can see. We see this type of maze in many different games including some Legend of Zelda games. In addition, in video games such as Super Mario World if the user has failed a level numerous times then the system offers the ability for the system to show the proper way to finish the level where obstacles of all kinds are taken into account and the shortest path is used.

In well-known games such as Pac-Man, the objective is to collect all the pellets and to live as long as the ghosts do not eat the user. In other games, the goal is to rescue a princess or to find the treasure and the end of the journey. We choose to favor the idea of there being some form of treasure at the end of the levels.

Mazes form also an important element in biology in the study of learning. Rodents were first used in mazes by Willard Small from Clark University (Small [5]). Using rodent burrows as a design a maze was created to test the cognitive abilities of rodents. Soon after these early experiments, different animals were used for testing purposes ranging from monkeys to birds (Watson [7]). James Watson also sent rodents through a maze with some sensory deprivations. Fleming Perrin of the University of Chicago tested humans where he blindfolded each person and let them solve a dodecagonal maze (Perrin [6]). However, rats remain in biology the primary test subject for mazes. Many of these types of experiments can be modeled in our maze choice game.

B. Constraint Databases

Constraint databases (Kanellakis et al. [2], Revesz [3]) provide an extension of relational databases where the input data consists of constraint relations. Each constraint relation is a set of constraint tuples. In constraint tuples, the attributes are referred to by variables and the possible values of the attribute variables are restricted by constraints. In particular, the MLPQ constraint database system uses linear constraints on the attribute variables (Kanellakis et al. [2], Revesz [3]).

III. THE A-MAZE-D (ADVANCED MAZE DEVELOPER) KIT

The A-Maze-D, short for Advanced Maze Developer, kit is a versatile system that enables efficient development of complex maze games. The A-Maze-D system consists of a MATLAB library that allows the easy translation of input data into MLPQ system input files. Once all stationary and moving objects are translated into MLPQ input files, the input files can be opened and animated in the MLPQ system. The animation lasts until some choice point is reached. The choice point requires that the user enter some input parameters, such as whether to turn left or right at the current location in the maze, to fire some bullets, to start a conversation or some other action. Once the input parameters are entered the corresponding animation can continue until the next choice point.

We organize the description of the A-Maze-D system into the following subsections. Section III-A shows how A-Maze-D can specify stationary objects such as mazes with either straight or curved walls. Section III-B describes the specification of moving objects such as persons, exploding objects and shields.

A. Stationary Objects

The main stationary object in a maze game is the maze itself. Figure 1 shows two mazes with walls that are composed of straight lines, while Figure 4 illustrates a maze with curved lines. The implementation of mazes, especially with curved walls, can be a tedious software engineering task. However, we show below that using the A-Maze-D system mazes can be developed efficiently whether the mazes have straight or curved walls.

Mazes with straight walls: Suppose that we would like to implement the maze shown on the left side of Figure 1. At first, we record the corner points for each wall as shown with highlights in Figure 2.

Each wall needs to have a different id. In the case of the maze in Figure 2 we need six walls with the (x, y) corner points shown in Table III-A.

A-Maze-D has a library of MATLAB scripts that contains a function called *buildWalls* that takes as input the set of points and turns them into an MLPQ input file that represents the walls. The basic idea behind the *buildWalls* function is illustrated in Figure 3, which shows on the left a simple table with two points and on the right their implied meaning. The *buildWalls* function also needs a parameter that specifies the width for the wall. In this case for simplicity we chose a width

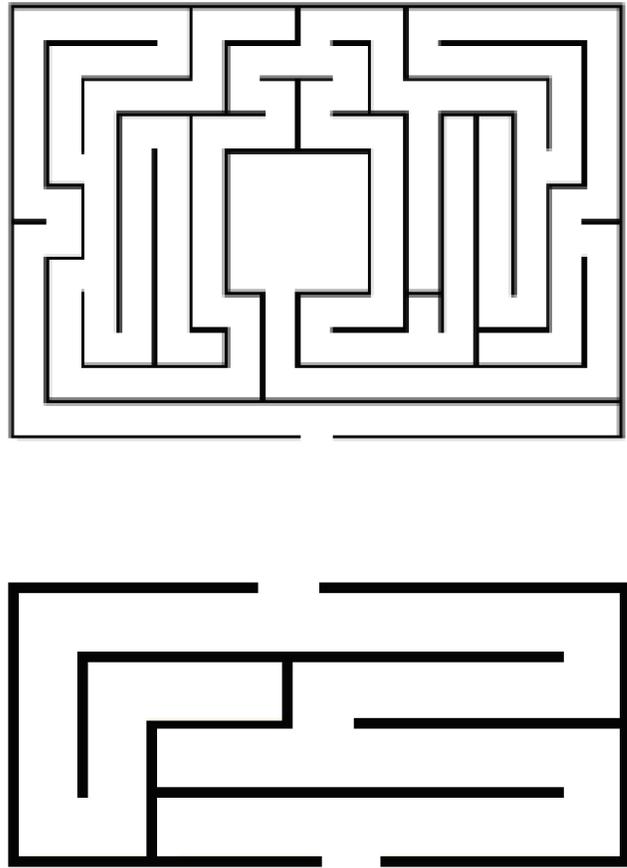


Fig. 1. Examples of mazes with straight walls.

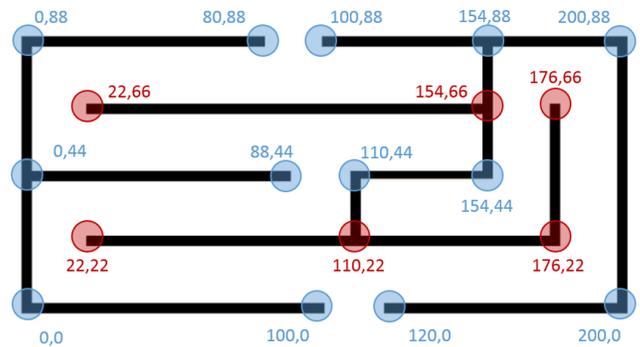


Fig. 2. Examples points on a maze.

of one. As can be seen, the (a, c) and (b, d) points define a parallelogram with width one and whose lower boundary line is the line segment from (a, c) to (b, d) .

The *buildWalls* function transforms the points described in Table III-A into the following MLPQ input file:

TABLE I
INPUT: CORNER POINTS ON THE MAZE.

id	X	Y
1	100	0
1	0	0
1	0	88
1	80	88
2	120	0
2	200	0
2	200	88
2	100	88
3	0	44
3	88	44
4	22	22
4	176	22
4	176	66
5	110	22
5	110	44
5	154	44
5	154	88
6	22	66
6	154	66

X	Y
a	c
b	d

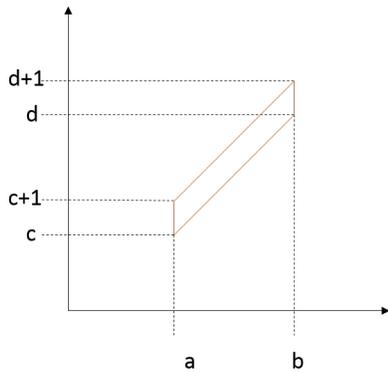


Fig. 3. The points from the table are represented visually on the right.

```
begin %MLPQ%
R(id,x,y) :- id=1, x>=0, x<=100, y=0.
R(id,x,y) :- id=1, x=0, y>=0, y<=88.
R(id,x,y) :- id=1, x>=0, x<=80, y=88.
R(id,x,y) :- id=2, x>=120, x<=200, y=0.
R(id,x,y) :- id=2, x=200, y>=0, y<=88.
R(id,x,y) :- id=2, x>=100, x<=200, y=88.
R(id,x,y) :- id=3, x>=0, x<=88, y=44.
R(id,x,y) :- id=4, x>=22, x<=176, y=22.
R(id,x,y) :- id=4, x=176, y>=22, y<=66.
R(id,x,y) :- id=5, x=110, y>=22, y<=44.
R(id,x,y) :- id=5, x>=110, x<=154, y=44.
R(id,x,y) :- id=5, x=154, y>=44, y<=88.
R(id,x,y) :- id=6, x>=22, x<=154, y=66.
end %MLPQ%
```

Mazes with curved walls: The A-Maze-D system uses multiple MATLAB scripts in order to attain the smoothest-looking curved walls. For example, Figure 4 shows a maze where all walls are curved except for six straight walls that are all horizontal and are used to block further passage in maze at dead ends.

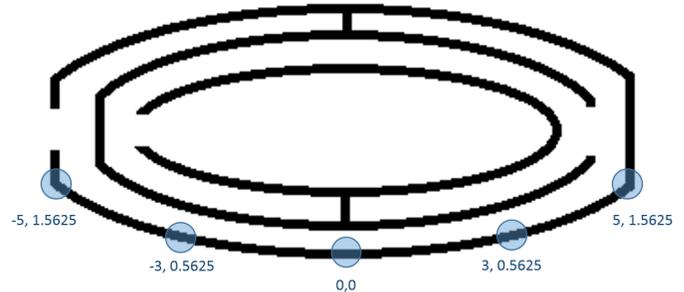


Fig. 4. Example of a curved maze.

Let us illustrate the A-Maze-D functions using the bottom curved wall, which is a segment of the parabola $y = (x/4)^2$. An approximation of the parabola can be specified by five points as shown in Figure 4. Our MATLAB script takes the x coordinates and the parabolic function $y = (x/4)^2$ to generate Table III-A. The script can take any other polynomial function. In case the boundary of the wall cannot be described by the user as a polynomial function, the A-Maze-D system also provides an alternative MATLAB script that makes a cubic spline interpolation for the given sample points of the wall.

TABLE II
EXAMPLE CORNER POINTS ON A MAZE.

id	X	Y
1	-5	1.5625
1	-3	0.5625
1	0	0
1	3	0.5625
1	5	1.5625

The more points we choose for the approximation, the smoother-looking parabolic curved wall we obtain. However, the smoother representation generates a larger MLPQ input file, which means generally a slower visualization and animations of the maze game. Given the above input data, the A-Mazed-D system already generates a large MLPQ output file, which begins as follows:

```
begin %MLPQ%
R(id,x,y) :- id=1, x>=-4.0000, x<=-3.9200,
-0.640043x - y = 1.560172.
R(id,x,y) :- id=1, x>=-3.9200, x<=-3.8400,
-0.588340x - y = 1.357495.
R(id,x,y) :- id=1, x>=-3.8400, x<=-3.7600,
-0.541344x - y = 1.177033.
...
end %MLPQ%
```

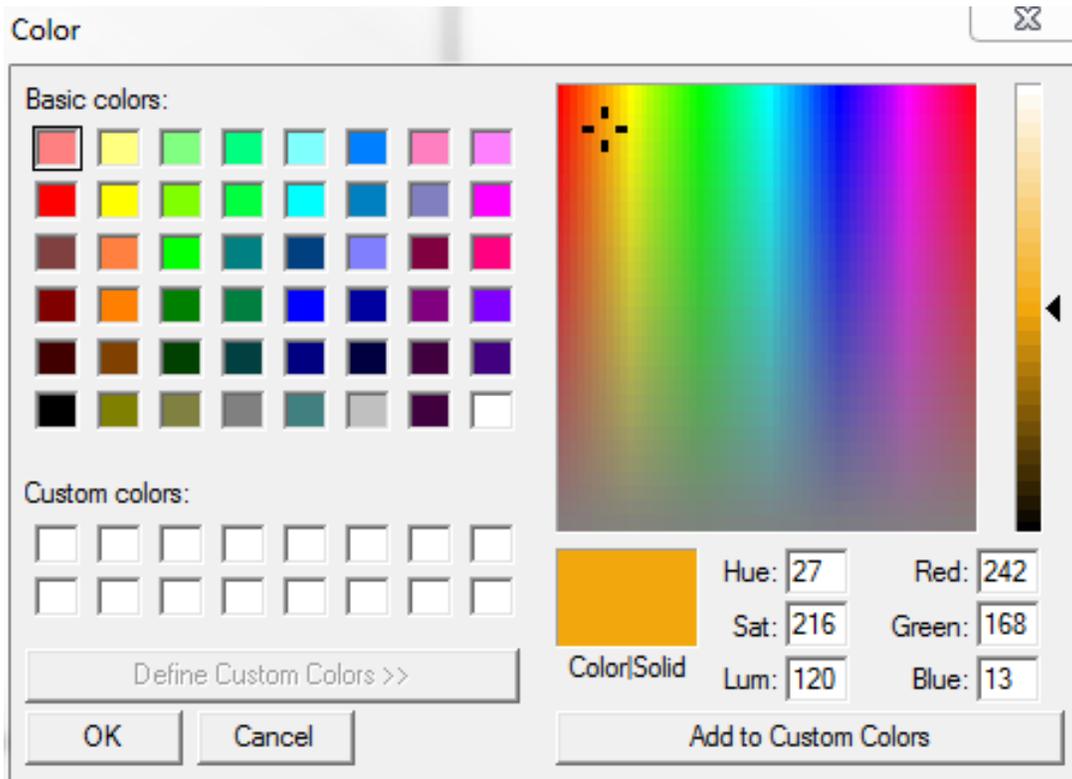


Fig. 6. Color Scheme Window

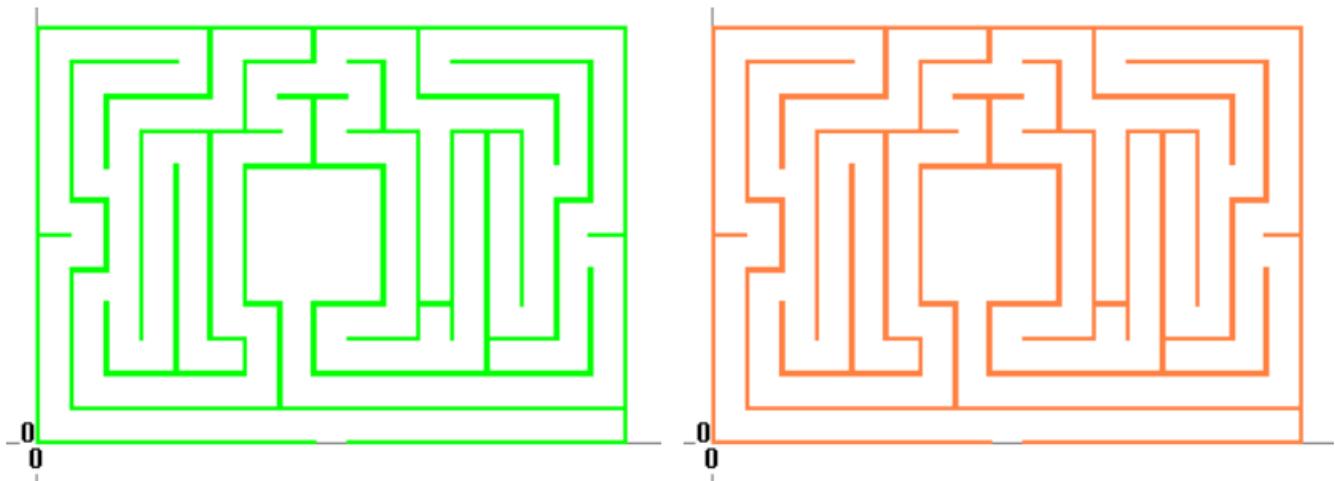


Fig. 7. Maze color changes from green (on the left) to orange (on the right)



Fig. 5. Color Example

Color Change: We call *color animation* when stationary or moving objects change color over time. Color animation can provide important visual cues to the users. For example, in our maze we could make the walls change color that would show some type of time limit where the walls will start to change to a specific color when the user is running out of time. MLPQ

handles color animation by allowing the user to choose two different colors for each relation.

When an MLPQ input file is loaded into the MLPQ system, then next to each spatial or moving object there are two colored boxes as shown in Figure 5. The first box represents the starting color, and the second box the ending color for the displayed relation during the animation. The MLPQ system provides a random pair of colors for each relation after a new file is uploaded. To change the colors to the desired values, we can double click on one of the boxes. Then a color scheme window will pop up as shown in Figure 6. This window allows the user to create the specific color of their

choice for the selected box. For example, Figure 7 shows two snapshots of the color animation of a maze, which changes from green to orange.

B. Moving Objects

Player Movement: Movement of the player is guided by choices that the player can select in MLPQ. When the player chooses a movement option in the maze, the A-Maze-D system animates the player moving through the maze until the player reaches the next choice point where another decision is required. At any moment in time, each player is assumed to occupy a 4 by 4 square area. The movement of the square is represented by the attributes x , y and t , which denoted time. For a moving object, x and y are functions of time represented by linear constraints in the MLPQ system input files. The A-Maze-D system uses another MATLAB script to generate the MLPQ file.

The MATLAB script will automatically compile this and return the movements for your player to take. Once the constraints are created from the script you then have to put them in your MLPQ maze file with the proper relations. The code below is one example that shows how the constraints are stored inside the MLPQ file.

The A-Maze-D system generates the following MLPQ file for the input shown in Table III-B.

```
begin %MLPQ%

player(id,x,y,t) :- id=1,
                    x + t >= 119,
                    x + t <= 121,
                    y > 10, y <= 11,
                    t >= 10, t <= 109.

player(id,x,y,t) :- id=1,
                    x >= 10, x <= 12,
                    y - t > -99,
                    y - t <= -98,
                    t >= 109, t <= 131.

player(id,x,y,t) :- id=1,
                    x - t >= -121,
                    x - t <= -119,
                    y > 32, y <= 33,
                    t >= 131, t <= 219.

player(id,x,y,t) :- id=1,
                    x >= 98, x <= 100,
                    y - t > -187,
                    y - t <= -186,
                    t >= 219, t <= 241.

end %MLPQ%
```

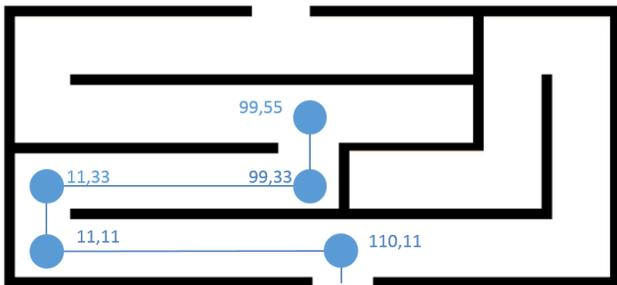


Fig. 8. A moving point within a maze.

For example, Figure 8 shows the movement of the player from location (110, 11) to location (99, 55). Besides these beginning and ending points, we also record some of the other points that are on the way with the restriction that all turning points need to be included in the list. In this case, the selected points can be represented in the following table:

TABLE III
INPUT: CENTER POINTS FOR PLAYER MOVEMENT ON THE MAZE.

id	X	Y
1	110	11
1	11	11
1	11	33
1	99	33
1	99	55

The A-Maze-D system generates an animation using MLPQ as a basis.

Searchlight: The A-Maze-D system also allows a limited field of view for the game player by adding a larger square relation around the moving player. The complement of that larger square will be displayed in black over the actual maze. For example, Figure 9 shows the limited field of view provided by a search light. The searchlight follows the player as he or she moves in the maze.

Explosions: The A-Maze-D system can represent explosions and other expanding objects. For fireworks can be represented as an object that expands until it ceases to exist. The A-Maze-D system provides a function that gives as input the beginning and the ending shapes of the exploding object and gives as output a parametric rectangles representation of the expanding object that is an accepted MLPQ input file.

Color Change: Color animation can be applied to moving objects too similarly as it is applied to stationary objects. For example, the firework could be black at the beginning and gradually lighted up and become completely red in the end.

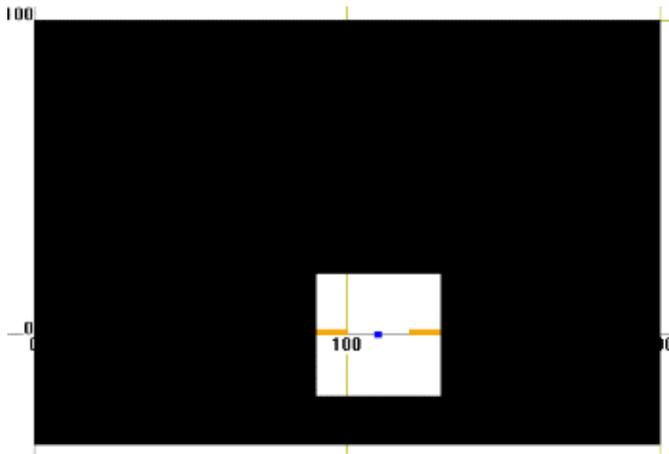


Fig. 9. Searchlight Example

- [7] John B. Watson, Kinesthetic and organic sensations: Their role in the reactions of the white rat to the maze. *The Psychological Review: Monograph Supplements*, 8.2, 1907.
- [8] "Monthly U.S. Video Game Industry Revenue 2015 — Statistic." Statista. Web. 13 Apr. 2015. <http://www.statista.com/statistics/201093/revenue-of-the-us-video-game-industry/>

IV. CONCLUSION AND FUTURE WORK

Constraint databases and video games have been around for a long time, but they have never been combined together before. The A-Maze-D system shows that constraint databases can be applied conveniently and efficiently to the design of maze games. A-Maze-D provides a versatile set of features by a combination of a MATLAB library and the MLPQ constraint database system. This is the first time that maze games have been created using constraint databases.

In the future we would like to try implementing different game types in constraint databases and see how well they can be converted over from the normal game development into constraint databases. We believe the best way to do this is testing out different types of games and see how well they perform. A few examples to name would be to test games that are more like Tetris where one needs a faster reaction or one's actions are limited by the time allotted.

Examining new possible features that constraint databases can provide for game development is another area we would like to research more and test out. Constraint databases have been used before in the animation of human faces. Hence an intriguing possibility is to allow players to speak to each other. Whenever a player wants to say something a pop-up window would open and show an animation of the player's face as he or she speaks.

REFERENCES

- [1] M. Lewis and J. Jacobson, Game engines, *Communications of the ACM*, 45.1, 27, 2002.
- [2] P. C. Kanellakis, G. M. Kuper and P. Z. Revesz, Constraint query languages, *Journal of Computer and System Sciences*, 51 (1), pp. 26-52, 1995.
- [3] P. Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, New York, USA: Springer, 2010.
- [4] P. Z. Revesz, R. Chen, P. Kanjamala, Y. Li, Y. Liu and Y. Wang, The MLPQ/GIS constraint database system, *ACM SIGMOD International Conference on Management of Data*, ACM Press, 2000.
- [5] W. S. Small, Experimental study of the mental processes of the rat. II *The American Journal of Psychology*, pp. 206-239, 1901.
- [6] F. A. C. Perrin, An experimental and introspective study of the human learning process in the maze. *Psychological Monographs: General and Applied*, 16.4, i-97, 1914.

International Trade of Environmental Goods: Is Trade Liberalization Fostering the Mexican Environmental Industry?

Petr Sauer, René Fernando Lara Cervantes

Abstract— the promotion of trade liberalization encompasses a wide range of products including environmental goods (EGs). However, the efforts to encourage a further liberalization of EGs have been criticized for being considered more favourable for developed countries. This paper studies the performance of the Mexican environmental industry in the NAFTA region where it competes with the United States and Canada. Our findings show that Mexico remains as a net importer of high-tech EGs which is the most profitable segment in the region. The recurrent economic crises limited the financing for the modernization of the industry, thus, is difficult to seize the benefits from the trade liberalization of environmental goods.

Keywords—environmental goods, international trade negotiations, international trade policy, NAFTA.

I. INTRODUCTION

AS globalization progresses economic blocs become stronger. Some examples are the European Union (EU) that encompasses 27 countries which is the main trading power of the world, followed up by the North America Free Trade Agreement (NAFTA) that is formed by United States (US), Canada and México (Romero and Vera-Colina, 2012). Within the NAFTA region the transition of Mexico from an inward economic policy featured by a strong state intervention to an open and liberalized economy requires a deeper analysis.

Mexico has shown a historical preference for an import substitution industrial policy. For decades the industry was supported by governmental subsidies and protected with import tariffs until the 1980s where the trade policy changed drastically (Sánchez, 2006). According to Morales (1997), the economic restructuring was featured by trade liberalization, thus, from 1986 to 1994 the government reduced import tariffs, joined the General Agreement on Tariffs and Trade (GATT) and ultimately the NAFTA. The implementation of NAFTA was the most crucial step in the transition from the

import substitution model to an open economy with trade liberalization in order to seize the economic gains of the exchange with other countries, especially the US (Hanson, 2010). Also, Puyana (2003) stressed that in the year 2000, 92% of the total Mexican exports had as main destination industrialized countries where the US showed the largest share with 85% of the Mexican exports. Although the Mexican industry has some large companies, it is composed mainly of small and medium size firms. After the assumption of NAFTA in the period of 1993 to 1999, the number of export-oriented participants grew 67% to reach over 30,000 firms; however, most of them were either small or medium size business that represented 89% of the exporting sector (Muñoz, 2005). The composition of the sector had a significant impact in the distribution of the benefits of international trade. In 1994, the 80% of the non-oil exports were concentrated in 573 firms, whereas 20% was distributed among the 20,846 business remaining in the sector (Morales, 1997). This situation is not exclusive of a particular industry but encompasses all branches involved in international trade.

In this regard, the environmental industry and the liberalization of the trade of environmental goods (EGs) is becoming an important topic in the agenda of several countries. In the Fourth World Trade Organization (WTO) Ministerial Conference in Doha in November 2001, it was agreed to negotiate the reduction or elimination of tariffs for EGs. The purpose is to develop a win-win strategy to expand the flows of trade and simultaneously foster the path towards sustainable development (SD), by broadening the access to environmental technology and eco-friendly goods at lower costs (Mathew and Fernández, 2009). Nevertheless, the strategy has been subjected to discussion for being considered more favourable for developed countries in terms of exports EGs. According to Sang and Jisun (2011), Mexico is listed as a key player in the international trade of EGs, however, its share in the world trade is small if compared to the European Union and the United States. Therefore, given the strategic value of the NAFTA region for the Mexican exporting sector, it is important to assess the size of the benefits for the environmental industry.

The aim of this paper is to determine how the liberalization of international trade of EGs is benefiting the Mexican environmental industry within the NAFTA region. In the first

This work was supported by project IGA No. F2/26/2014 „Impact of economic liberalization on environmental policy“

Petr Sauer is director of the Institute for Sustainable Business at the University of Economics in Prague. Phone: +420 224 095 298; e-mail: sauer@vse.cz

René Fernando Lara Cervantes is with the Institute for Sustainable Business at University of Economics in Prague. He is also a journalist of the Mexican newspaper El Diario NTR. e-mail: lararene83@yahoo.com.mx

section we analyze the discussion regarding the liberalization of environmental goods and their main challenges. In the second section, we compare the trade balance of EGs of Mexico with the other members of NAFTA in order to assess the participation of the country in the region. Also, we discuss the factors that determine the performance of the industry. The last part shows conclusions and policy recommendations.

II. BENEFITS OF THE TRADE OF ENVIRONMENTAL GOODS AND CONCEPTUALIZATION CHALLENGES OF THE LISTS AND DEFINITIONS

A. *The discussion about the benefits of the liberalization of environmental goods*

The beginning of the discussion started in the late 1990s, but it was until 2001 when the topic was addressed in a separate agenda. Since the Fourth WTO ministerial Conference in Doha in 2001, the WTO members have been seeking agreements to liberalize the trade of environmental goods and services and the related import tariffs (Znamenackova et.al., 2014). However, the benefits of the reduction of tariffs are being subjected to discussion. According to Sang and Jisun (2011), most of the promoters of the liberalization are key players in the international trade of EGs; their main argument is the growth in the demand for clean technologies and products in both developed and developing countries. Thus, lowering the current tariffs will result in price reductions to become environmental technology more affordable for developing countries and simultaneously developed countries will increase their trade flows (Vikhlyaev, 2004). In this regard the strategy has been criticized for being disadvantageous for developing countries in terms of international trade balance, mainly because the competitiveness the environmental industries of developed countries are more competitive (Mathew and Fernández, 2009).As stressed by Balineau and De Melo (2013), in fact EGs are less protected than other goods, in the period 1996-2010, tariffs were reduced approximately 50% from the initial levels but remain high in low-income countries. Also, the size of the contribution to the improvement in the environmental performance is vague.

Thus, the economic and development gap between developed and developing countries raises doubts about who are the true winners of the trade liberalization of EGs.

B. *Problems in the agreement of lists and definitions for environmental goods*

The definitions of environmental goods and lists agreement are a key element for the progress of the liberalization negotiations. There are two reasons for the slow progress in the negotiations, one are the perceptions of the benefits of EGs liberalization which was described above and the other to the political economy of trade policy formulation (Balineau and De Melo, 2013). Therefore, the future of negotiations will depend in the ability of the negotiators to show data and favourable arguments of EGs liberalization (Znamenackova

et.al., 2014).

The formulation of trade policy depends largely on the agreement of a definition of the environmental goods to be sorted in a list. However, there are two crucial problems for the elaboration of EGs definitions, one associated to the production process and the other linked to the dual-use dilemma. The first problem refers to whether a good is ecologically manufactured or if the process generates pollution (Znamenackova et.al., 2014). The second matter is linked to the end use of goods; since the same a product could be used for environmental purposes but also in polluting activities (Sang and Jisun, 2011). Because of both problems it has not been possible to create a globally accepted definition of EGs. Although several international organizations have proposed lists and definitions, none of this has been universally accepted. An early definition of EGs was developed in the 1990s by experts of the Organization for Economic Cooperation and Development (OECD) and Eurostat: "the environmental goods and services industry consists of activities which produce goods and services to measure, prevent, limit, minimize or correct environmental damage to water, air and soil, as well as problems related to waste, noise and eco-systems." (OECD, 1999, p. 10). The OECD list includes goods spanning 6-digit Harmonized Commodity Description and Coding System (HS codes). The items include goods for water treatment, waste management, renewable energy and sound technologies. Among all the conceptualization efforts, the definitions and lists of EGs of the Asia-Pacific Economic Cooperation (APEC) and the OECD have been the baseline for the current APEC and WTO negotiations on environmental goods (Znamenackova et. al., 2014). The APEC and WTO definitions and categories of EGs are shown in Table 1.

Table 1. APEC and WTO definitions and categories of environmental goods.

Organization	Definition	Number and categories of listed items
Asia-Pacific Economic Cooperation (APEC)	Environmental goods and services (EGS) is an industry sector devoted to solving, limiting or preventing environmental problems. The industry is involved in manufacturing and/or services related to water or air pollution, waste management, recycling,	The EGs list consists of 54 official environmental goods. The items include renewable and clean energy technologies, wastewater treatment technologies, environmental monitoring and assessment equipment. Biofuels are not included.

	renewable energy, monitoring, analysis and assessment.	
World Trade Organization (WTO)	Environmental goods are those activities that produce materials, equipment and technologies used to address particular environmental problems; or products considered preferable than less ecological similar goods.	Several lists were proposed by the members of the WTO, the last cover 26 items of core EGs. In this regard the members consider the list to be helpful but not definitive. The list includes goods for waste management, and water treatment, renewable energy, environmental technologies, carbon capture and storage and air pollution control

Source: APEC, 2012, Annex C, p. 1; Hamwey et al., 2003.

An important feature of the lists and definitions of EGs developed by international organizations is that they are based on the interests of their members. Thus, the best positioned countries in the international trade of these products could have a greater influence in the outcome of the categorization of EGs. As stressed by Balineau and De Melo (2013), until 2010 three broad approaches were proposed in order to foster trade liberalization of EGs. The authors showed that until 2010, some efforts like the List approach, the Request and Offer approach; the Integrated-Project approach and the hybrid approach were proposed by different countries in order to reduce import tariffs within the WTO. The List approach was proposed by some key players in the trade of EGs such as the EU, Japan and the US and it was broadly accepted, however, the listed items in which large developed countries had a comparative advantage. Most of the EGs sorted in the core list of the WTO are high-tech and capital goods, therefore, given the gap between developed and developing countries, the latter might remain only as net importers (Vykhlyaev, 2004). Thus, this conflict of interests helps to explain why developing countries have been more reluctant to establish negotiation for tariffs reduction.

III. PERFORMANCE OF THE MEXICAN ENVIRONMENTAL INDUSTRY IN THE TRADE OF ENVIRONMENTAL GOODS

A. The importance of NAFTA and the Mexican market

The international trade of environmental goods of Mexico will be analysed in the framework of NAFTA. In the year 2000, the 92% of the Mexican exports had as destination industrialized countries, being the US the largest receiver with a share of 85% of the exported goods (Puyana, 2003). This places Mexico close to a single market context which has been disadvantageous in terms of international trade and is also the reason why the trade of EGs with Canada is not deeply discussed. According to Muñoz (2005), based on the definition of environmental goods and services of the OCDE-Eurostat, the environmental Mexican environmental industry can be sorted in three categories: goods and services for pollution control, intermediate goods and services linked to cleaner technologies and goods and services devoted to the rational use of natural resources. However the number and size of the participants in the environmental industry is small. Although the environmental industry has been growing in Mexico it has only 8,300 businesses in the sector, whereas in the United States the number reaches 117,000 firms (Ferrier, 2010). However, both countries are key players in the international trade of EGs. This is shown in Table 2.:

Table 2. World key players of trade in Environmental Goods according to the WTO Core List (millions USD, 2009)

Country	Export	World Share	Import	World Share
Europe an Union	34,248.0	19.31	28,802.2	16.3
Japan	22,842.5	12.92	5,170.8	2.97
United States	17,651.9	10.04	21,228.9	12.03
China	21,813.2	12.33	19,552.7	11.1
Korea	6,786.0	3.85	11,068.1	6.35
Mexico	4,811.7	2.76	3,327.9	1.9
World Total	177,187.1		176,877.5	

Source: Sang and Jisun, 2011, P.12

The gap among the trade balances of Mexico and the US suggest a big difference in terms of competitiveness which affects the performance of Mexico in the NAFTA region. According to Zhang (2010), since the implementation of the NAFTA in the period of 1995-2005 the market of EGs within the region of Mexico, US and Canada doubled its size. Additionally, the Mexican market became important for being the largest single market for environmental goods in the Americas. In this regard, Avery and Boadu, (2004) estimated

a regression model for the exports' demand of the EGs industry of the US, showing that the highest increases in this demand would occur in the developing areas of the world, including Mexico. Since Mexico joined the NAFTA, the exports of EGs have increased but the trade balance is in deficit (Muñoz, 2005). This is shown in Table 3.:

Table 3. Mexico's trade balance of EGS in 2001 based on the Eurostat-OECD definition with US and Canada (million USD).

Type of goods	Exports to the US	Imports from the US	Exports to Canada	Imports from Canada
<i>Equipment:</i>				
Equipment and chemicals for water	1.3	300.5	0.09	4.44
Equipment for control of air pollution	1.3	98.4	0.13	4.47
Instruments and information systems	0.00	35.0	0.00	0.16
Equipment for waste management	10.5	62.1	1.62	4.26
Technology for processes and prevention	0.2	3.2	0.02	0.02
<i>Services:</i>				
Solid waste management	3.3	40.8	0.00	0.59
Hazardous waste management.	0.4	31.8	0.00	0.13
Consultancy and engineering	0.5	58.7	0.05	0.59
Sanitation and industrial services	0.7	23.8	0.34	0.90
Analytical treatment	0.0	3.7	0.00	0.19
Water treatment	5.4	46.9	0.00	0.08
<i>Resources:</i>				
Public companies for water management	0.3	26.1	0.00	0.07
Recovery of resources	0.3	142.2	0.07	1.15
Systems and services of	1.1	41.1	0.01	0.29

clean energy				
Sustainable agriculture	32.1	4.3	4.45	1.26
Sustainable forestry	5.6	0.0	0.70	0.00
Ecotourism	53.6	0.1	5.36	0.33
<i>Total</i>	<i>116.6</i>	<i>918.7</i>	<i>12.84</i>	<i>18.93</i>

Source: *Análisis del comercio de bienes y servicios ambientales en la región del TLCAN*, pg.10,12. Comisión para la Cooperación Ambiental (2004)

The data show a generalized trade deficit of Mexico in the NAFTA region; this could also be the reason of why the country maintains higher average import tariffs. According to data of the World Trade Organization (2011), Mexico has an average tariff of 6.75% for the goods of the core lists of the WTO which is the highest among the members of NAFTA.

The trade balance also reveals the sectors where Mexico is more competitive. The activities where Mexico has an advantage are ecotourism, sustainable agriculture and sustainable forestry. However, it is important to determine the share of the total value correspondent to each category in order to find the core segments of the trade of EGs. This is shown in Table 4.

Table 4. Share of the total trade by sector in the NAFTA region in 2001

Type of goods	Total trade in USD millions	Share (%)
<i>Equipment:</i>		
Equipment and chemicals for water	1,469.8	
Equipment for control of air pollution	590.7	
Instruments and information systems	123.2	
Equipment for waste management	535.8	
Technology for processes and prevention	11.7	
<i>Services:</i>		
Solid waste management	130.6	
Hazardous waste management.	48.7	
Consultancy and engineering	372.1	
Sanitation and industrial services	63.6	
Analytical treatment	15.4	
Water treatment	69.5	
<i>Resources:</i>		
Public companies for water management	42.5	
Recovery of resources	235.0	
Systems and services of	82.2	

clean energy		
Sustainable agriculture	197.6	
Sustainable forestry	19.5	
Ecotourism	78.5	
<i>Total</i>	<i>4,086.6</i>	<i>100%</i>

Source: *Análisis del comercio de bienes y servicios ambientales en la región del TLCAN*, pg.13. *Comisión para la Cooperación Ambiental (2004)*

The data shows that Equipment is the main segment of the trade of EGs which is also the most profitable. As shown before, the participation of Mexico in this category is not very significant and it exposes the lack of specialization in the environmental industry. The causes of the underdevelopment of the Mexican environmental industry will be discussed in the next subsection.

B. Why is the Mexican environmental industry underdeveloped?

Along with the progress of trade liberalization a transition towards specialization started as well. According to Morales (1997), since Mexico joined the GATT the country shifted from exporting raw materials to high-tech manufactured goods, however, only a few firms had sufficient resources to perform this change. This affected also the modernization of the environmental industry. The size of firms represents an obstacle for the competitiveness and the development of the sector mainly because of financial restrictions (Muñoz, 2005). The economic crises of the 1980s and the 1990s reduced considerably the public budget for the improvement of environmental infrastructure and also led the banks to bankruptcy (Feagans, 1997; Fairchild and Sosin, 1987). Although the production of EGs and the number of firms grew during the crisis of 1994, the growth responded to reduction of costs due to a policy of low wages but not to a modernization strategy (Muñoz, 2005). Also, Dutrénit et.al. (2003) which performed a study about the technological transition for three Mexican firms, the economic shocks substantially decreased the investment on research and development. Thus, if the lack of resources to foster the competitiveness of the environmental industry is not tackled Mexico will not be able to seize a larger share of the benefits of the international trade of EGs.

IV. CONCLUSION

There is no doubt that international trade has fostered the exports of the Mexican environmental industry. According to the WTO, México is a key player in the trade of environmental goods, however, in the NAFTA region the country remains as a net importer of capital environmental goods which are the most profitable segment of the trade of EGs. Additionally, the lack of specialization of the Mexican environmental industry limits the economic and environmental benefits that can be obtained from the liberalization of the trade of EGs.

Given the importance of the American market the trade

policy must be revised. New approaches to develop definitions and categories are needed, as well as the development of new industrial policy to foster the growth and competitiveness of the firms participating in the environmental industry. Also, greater specialization will increase the exports destination of Mexican EGs and at the domestic level it will reduce the costs of environmental protection.

Future research should be focused on comprehensive studies about possible strategies to develop the environmental industry and simultaneously reduce the dependency of imported environmental goods and technology.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank” Instead, write “F. A. Author thanks” **Sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page.**

REFERENCES

- [1] APEC List of Environmental Goods, 20th APEC Economic Leaders' Declaration, Annex C, 2012. Available: http://www.mofa.go.jp/policy/economy/apec/2012/pdfs/aelm_declaration_en.pdf
- [2] B. Avery O.F, Boadu, “Global Demand for U.S. Environmental Goods and Services”, *Journal of Agricultural and Applied Economics*, Vol. 36, pp. 49-62, 2004
- [3] G. Balineau, J. De Melo, “Removing barriers to trade on environmental goods: an appraisal”, *World Trade Review*, Vol.12, pp. 693-718, 2013.
- [4] G. Dutrénit, O.A. Vera-Cruz, A.A. Navarro, “Diferencias en el perfil de acumulación de capacidades tecnológicas de tres empresas mexicanas”, *El Trimestre Económico*, vol.70, pp 109-162, 2011.
- [5] Environmental Business International. *Análisis del comercio de bienes y servicios ambientales en la región del TLCAN*. Comisión para la Cooperación Ambiental, Montreal, diciembre de 2004.
- [6] L.G. Fairchild, K. Sosin, “Manufacturing Firms in Mexico's Financial Crisis: Determinants of Severity and Response”, *Mexican Studies*, vol. 3, No. 1, pp. 127-150, 1987
- [7] B. Feagans, “Green Lighs”, *Business Mexico*, vol. 7, pp.38-42, 1997
- [8] G. Ferrier, “The evolution of the environmental industry in the post-NAFTA era in Mexico”, *International Environmental Agreements*, vol. 10, pp. 147-164, 2010.
- [9] R. Hamwney, U. Hoffman, R. Vossenar, “Liberalization of International Trade in Environmental Goods and Services”, UNCTAD, Division of International Trade in Goods and Services, 2003
- [10] E. Lendo, “Defining Environmental Goods and Services: A Case Study of Mexico”, ICTSD Trade and Environment Series Issue Paper No. 1, CEC and ICTSD, Geneva, Switzerland, 2005.
- [11] A.J. Mathew, S. Fernández de Córdorba, “The Green Dilemma about Liberalization of Trade in Environmental Goods”, *Journal of World Trade*, London, vol. 43, pp.379-416, 2009.
- [12] I. Morales, “The Mexican Crisis and the Weakness of the NAFTA Consensus”, Sage Publications Inc., Vol. 550, pp. 130-152,1997.
- [13] V.C. Muñoz, “Bienes y servicios ambientales en México: caracterización preliminar y sinergias entre protección ambiental, desarrollo el mercado y estrategia comercial”. Cepal: Serie Medio Ambiente y Desarrollo, Santiago de Chile, pp.31-86, 2005.
- [14] The Environmental Goods and Services Industry: Manual for Data Collection and Analysis, Eurostat, OECD and the Statistical Office of the European Communities, 1999. Available: http://unstats.un.org/unsd/envaccounting/ceea/archive/EPEA/EnvIndustry_Manual_for_data_collection.PDF

- [15] A. Puyana, "El camino mexicano hacia el regionalismo abierto: los acuerdos de libre comercio de México con América del Norte y Europa". Cepal: Serie Comercio Internacional, Santiago de Chile, pp.5-17, 2003.
- [16] A. Romero, M.A., Vera-Colina, "La globalización posible: límites y alternativas", Cuadernos de Economía, Vol. 31, pp. 49-53, 2012
- [17] L. Sánchez Mier," Grupos de interés y reforma comercial en México", El Trimestre Económico, Vol. 73 (2), pp.337-361, 2006.
- [18] A. Vhyklyayev, "Environmental Goods and Services: Defining Negotiations or Negotiating Definitions?", Journal of World Trade, London, Vol. 38 No.1, pp. 93-122, 2004
- [19] Y. H. Sang, K. Jisun "Trade Liberalization in Environmental Goods: Major Issues and Impacts", Korea and the World Economy, Vol.12 No.3, pp580-594, 2011
- [20] Z. Zhang, "Services the environment and the NAFTA, International Environmental Agreements", Vol. 10, pp. 89-91, 2010
- [21] J. Znamenacková, P. Sauer, and R.F. Lara Cervantes, "Environmental Goods Market Liberalization: Case study of Czech Republic and possible Way to follow for Ukraine", Actual Problems of Economics, vol. 151, pp. 312-322, 2014.

A Computational Study of the Evolution of Cretan and Related Scripts

Peter Z. Revesz

Abstract— Crete was the birthplace of several ancient writings, including the Cretan Hieroglyphs, the Linear A and the Linear B scripts. Out of these three only Linear B is deciphered. The sound values of the Cretan Hieroglyph and the Linear A symbols are unknown and attempts to reconstruct them based on Linear B have not been fruitful. In this paper, we compare the ancient Cretan scripts with four other Mediterranean and Black Sea scripts, namely Phoenician, South Arabic, Greek and Old Hungarian. We provide a computational study of the evolution of the three Cretan and four other scripts. This study encompasses a novel translation of the scripts to a DNA encoding, which enables the use of hypothetical evolutionary tree reconstruction algorithms from the area of bioinformatics.

Keywords—Cretan Hieroglyph, Linear A, Linear B, Evolution, Neighbor Joining, Phylogenetic tree.

I. INTRODUCTION

CRETE was the birthplace of several ancient writings that were first categorized by Arthur Evans, the explorer of Knossos Palace, as the Cretan Hieroglyph, the Linear A and the Linear B scripts [5]. Linear A, which dates back to about 2500 BC, was the main script used in the Minoan palaces of ancient Crete. The Cretan Hieroglyph script, which may predate Linear A, was used for centuries simultaneously with Linear A. Linear A was replaced around 1450 BC by Linear B, which was used in Mycenaean Greece and is the oldest known Greek writing [10]. In 1952 Michael Ventris gave a decipherment of Linear B as described in Chadwick [2]. However, the Cretan Hieroglyph and the Linear A scripts are still not deciphered.

In order to understand better these three ancient Cretan scripts, in this paper we study their relationship with four other scripts. The other scripts are the Phoenician, the South Arabic, the Greek and the Old Hungarian alphabets.

The Phoenician alphabet [25] was a major influence on the development of many other alphabets due to the Phoenicians' widespread commercial influence in the Mediterranean area. Both the Phoenician and the South Arabic alphabets derive from the Proto-Sinaitic alphabet, which is assumed to have originated in the Sinai Peninsula sometime between the mid-19th and mid-16th century BC [26]. Phoenician represents a northern branch while South Arabic represents a southern branch of Proto-Sinaitic.

The classical Greek alphabet from about 800 BC had a

major influence for many other European alphabets. The classical Greek alphabet derives from the Phoenician alphabet except for the letters Φ , X , Ψ and Ω [24].

The Old Hungarian alphabet is the alphabet used by Hungarians before the adoption of the Latin alphabet. Parallel with the Latin, it was used sporadically until the 20th century in some Hungarian ethnic minority areas of Rumania. The origin of Old Hungarian is still debated. Hosszú [11] presents a detailed view of the development from Phoenician via Aramaic and Turkish and Proto-Rovas scripts. On the other hand, Forrai [8] and Varga [23] claim that the Old Hungarian script already existed in the Bronze Age and cite putative translations of engraved artifacts going back to 1000 BC.

In computational biology, the question of evolutionary relationships is greatly facilitated by the wide availability of genomic data and the development of a growing number of phylogenetic tree construction algorithms. Some of the best-known phylogenetic tree algorithms are Saitou and Nei's *neighbor-joining* method [19] and Sokal and Michener's UPGMA method [21]. The books by Baum and Smith [1], Hall [9] and Lerney et al. [12] review the *maximum likelihood* and several other methods. Recently, Revesz [15] also proposed the *Common Mutations Similarity Matrix* or CMSM method for phylogenetic tree construction. The CMSM method derives from a series of previous evolutionary biology studies, including [14], [16]-[18], [20] and [22].

Some of the efficient phylogenetic tree algorithms are able to reconstruct hypothetical evolutionary trees in a few minutes of computational time. Moreover, they are based on statistical techniques that are free of human bias, which sometimes prevent the objective evaluation of linguistic artifacts. Human translation attempts are inherently prone to error. For example, the Phaistos Disk, which contains some form of Cretan Hieroglyph writing, was translated in numerous contradictory ways by a large number of professional and amateur linguists. Fauconau [6] and Fisher [7] are example decipherment attempts, and Duhoux [4] is a critique of previous decipherment attempts. In this paper, we strongly advocate computerized approaches to the study of linguistic questions in order to eliminate human bias.

This paper is organized as follows. Section II presents a comparative table of the script symbols. Section III describes the DNA encoding of scripts. Section IV presents a computational reconstruction of the evolutionary tree of the scripts and a discussion of the results. Finally Section V gives some conclusions and directions for future work.

Peter Z. Revesz is with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA (revesz@cse.unl.edu).

Table 1 A comparison of the script symbols

Hieroglyph and Phaistos	Linear A	Linear B	Value	Phoenician	Value	South Arabic	Value	Greek	Old Hungarian	Value
					?		?	A		A
			SE		B		B	B		P
					G		G	Γ		G
			DA		D		D	Δ		D, T
					H		H	E		E
					W		W	Υ		US
					Z		Z	Z		U
					H		d ^h	H		GY
			KA		T ^s		T ^s	Θ		TY
					Y		Y	I		J
			WE		K		K	K		G ^H , K
			PU		L		L	Λ		L
			TWE		M		M	M		M
					N		N	N		NT
			TE		S		S	Ξ		H
			QE		ς		ς	O		J, L
					P			Π		
			ZO		ς		ς	Μ		C
			QA		Q		Q	ϙ		K
					R		R	P		*R
			TI		š		š	Σ		š
			RO		T		T	T		D
							F	Φ		F
								X		H
			RE				H	Ψ		ZS
			TA					Ω		O

II. A COMPARATIVE TABLE OF SCRIPT SYMBOLS

As a first step, we built a comparative table of script symbols as shown above in Table 1. In Table 1, the Phoenician alphabet and the South Arabic alphabet columns are taken from [25] with minor modifications. The Greek alphabet column is taken from [24]. The Old Hungarian alphabet column is our addition. The sound values of the Old Hungarian alphabet are from [8], [11] and [23]. The symbols marked with a star * are Proto-Rovas symbols that were used in the early phases of Old Hungarian according to Hosszú [11]. Our reconstruction assumed that the * symbols represent the more archaic form of Old Hungarian. It is possible that these archaic forms were changed to the latter forms due to Turkish or other influences.

Our reconstruction of Old Hungarian was guided by a combination of visual and sound value correspondences. The visual and the sound value correspondences almost always support each other. There are a few exceptions. For example, the Old Hungarian “US” sound value is different from the Phoenician and South Arabic semivowel “W” sound value. However, in languages where the “W” was not used, it was commonly translated as the vowel “U,” including in ancient Greek, where the symbol was named “UPSILON.” The Old Hungarian “US” may be a similar adaptation of “W” to “U.”

Linear B and its values are from Chadwick [2] and Hooker [10]. The Cretan Hieroglyph and Linear A correspondences to Linear B are our reconstructions but are based in part on previous observations by Evans [5], Fisher [7] and Young [28]. Since the sound values of the Cretan Hieroglyph and Linear A symbols are unknown, their correspondences are based only on visual observations. Finally, the correspondence between the Linear B symbols and the four alphabets is also based primarily on visual observations. For example, the Linear B

wheel symbol  has an obvious parallel in Phoenician  even though they denote different sound values. The major difficulty here is not simply that Linear B is a syllabary while Phoenician is an alphabet. A syllabary with consonant-vowel syllable combinations can have a natural evolution into an alphabet when either the consonant or the vowel is dropped. However, in this case the Linear B sound value is “RA” which cannot be reconciled with the Phoenician sound value or the corresponding Greek Θ or “THETA.” When there was a conflict between the visual and the sound value correspondences, we always took the visual correspondence as having the primary significance.

III. THE DNA ENCODING OF SCRIPT SYMBOLS

After the alignment of the script symbols as shown in Table 1, we took a careful look at each row. In each row, we divided the set of symbols into groups such that in each group the symbols were closer together than they were to members of other groups. For example, in the first row both the Hieroglyph symbol  and the Linear A symbol  clearly denote persons. Hence they are grouped together. The

Phoenician and the Greek letters are only rotations of each other. Hence they also are grouped together. The South Arabic and the Old Hungarian are much more different than the others. Hence we placed the South Arabic into the third group, and the Old Hungarian into the fourth group.

We call the first group the A group, the second group the C group, the third group the G group, and the fourth group the T group. These groups are named after the four DNA nucleotides. After the grouping of the symbols in a row of Table 1, we wrote down the group labels in column where the rows corresponded to the seven scripts. The final result is shown in Figure 1.

CLUSTAL 2.1 multiple sequence alignment

```

Linear_A      AAAACA-CC-AAAA-A-ACATA--AG
Linear_B      -A-A---CC-AAAACA-AC-TA--AG
Hieroglyph    AAAAAAAAAAAAAAAAAA-AAATAG-AG
O_Hungarian  TGAACAACCAGAAAAA-ATATAGTAG
S_Arabic      GGCACAAGAAGCAAAA-ACATAC-A-
Phoenician    CCCC GCCC CCCC CACA ACGCCA ----
Greek         CCCC GCGT CGCC CACA ACGCCAGTAG
    
```

Fig. 1 The DNA encoding of the seven alphabets

IV. COMPUTATIONAL RECONSTRUCTION OF AN EVOLUTIONARY TREE USING PHYLOGENETICS

A. Computational Reconstruction Using ClustalW2

We used ClustalW2’s phylogenetic algorithms because they are currently among the most frequently used in bioinformatics and are available free to all users from the website <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

For the DNA encoding in Figure 1, ClustalW2 computed a hypothetical phylogenetic tree as shown in Figure 2. This type of phylogenetic trees is also called a *cladogram*. ClustalW2 provides two phylogenetic tree generation algorithms the Neighbor Joining (NJ) method, which is the default, and the UPGMA method. We chose the NJ method.

B. Discussion of the Results

The results shown in Figure 2 suggest that the seven scripts have a common ancestor from where three branches descend.

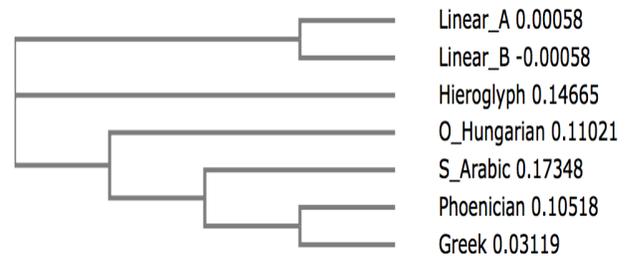


Fig. 2 The ancient scripts evolutionary tree reconstructed by the Neighbor Joining algorithm in ClustalW2

These branches are as follows:

1. Hieroglyph
2. Linear A and Linear B
3. Greek, Old Hungarian, Phoenician, South Arabic

Since the Hieroglyph script is the oldest attested writing, the common ancestor was likely very close to the Hieroglyph writing. Linear A and Linear B were developed in Crete although they spread with Minoan and Mycenaean Greek cultures to other areas. Many researchers already noted the close visual relationship between the Linear A and the Linear B script symbols.

In the third branch some relationships are completely as expected. For example, the close relationship between Phoenician and Greek is due to widely recognized ancient Greek adoption of the Phoenician script. Phoenician and South Arabic are also recognized to have a common ancestor called Proto-Sinaitic [26]. Therefore the Proto-Sinaitic could be the common root of the Greek, Phoenician and South Arabic scripts. However, it seems new information in the third branch that Old Hungarian and Proto-Sinaitic are also related as sister scripts. This result seems to contradict the view presented in Hosszú [11] that Old Hungarian is a derivative of Phoenician because then Old Hungarian would be placed closer to Phoenician. The results split the third branch into an Old Hungarian sub-branch and a Proto-Sinaitic sub-branch. Therefore, the results are consistent with the view of Forrai [8] and Varga [23] in Old Hungarian being an ancient script whose archaic form existed concurrently with Proto-Sinaitic.

Previously, Naddeo [13] suggested a relationship between the Ugaritic and the Old Hungarian scripts. Colless [3] claims the Ugaritic script is derived from Proto-Sinaitic. If Colless' and Naddeo's claims are true, then they provide another link between Old Hungarian and Proto-Sinaitic. Further, Ugaritic may be a root of the third branch of the evolutionary tree in Figure 2. However, the Ugaritic script is a cuneiform abjad (consonants only alphabet), and the translations provided by Naddeo and by Colless from Ugaritic cuneiform to Old Hungarian and Proto-Sinaitic, respectively, are not very convincing. Hence more research needs to be done to decide whether the Ugaritic script also belongs to the third branch of the evolutionary tree.

V. CONCLUSIONS AND FUTURE WORK

In this study, we focused on the shape of the script symbols and the objects that they may depict because the sound values of the Cretan Hieroglyph and Linear A symbols are unknown. When the sound values are known, the correspondence is close for Phoenician, South Arabic, Greek and Old Hungarian. Only the Linear B sound values are markedly different. It has been attempted to read Cretan Hieroglyph and Linear A scripts using Linear B sound values without any fruitful result. Instead of the Linear B sound values, now it is possible to read Cretan Hieroglyph and Linear A scripts using some possible common ancestor sound values of the corresponding

Phoenician, Greek, South Arabic, and Old Hungarian alphabet symbols. The common ancestor sound values may correspond well to the initial sounds of the syllables that the Cretan Hieroglyph and the Linear A symbols stand for, in case they also denote syllables. We believe that the third branch of the evolutionary tree of Figure 2 may have preserved the original sound values better than Linear B did. We hope that this realization will open a new phase in the understanding of the ancient Cretan scripts.

REFERENCES

- [1] D. Baum and S. Smith, *Tree Thinking: An Introduction to Phylogenetic Biology*, Roberts and Company Publishers, 2012.
- [2] J. Chadwick, *The Decipherment of Linear B*, Cambridge University Press, 1958.
- [3] B. Colless, "Cuneiform alphabet and picto-Proto-alphabet," <https://sites.google.com/site/collesseum/cuneiformalphabet>, downloaded July 5, 2015.
- [4] Y. Duhoux, "How not to decipher the Phaistos Disc," *American Journal of Archaeology*, Vol. 104, No. 3 (2000), pp. 597–600.
- [5] A. J. Evans, *Scripta Minoa: The Written Documents of Minoan Crete with Special Reference to the Archives of Knossos*, Volume II, Classic Books, 1909.
- [6] J. Fauconnau, *Le Décifrement du Disque de Phaistos: Preuves et conséquences*. L'Harmattan, Paris/Montreal 1999.
- [7] S. R. Fisher, *Glyph-Breaker*, Springer, 1997.
- [8] S. Forrai, The Old Hungarian Writing from Ancient Times to the Present, (in Hungarian), Antológia Kiadó, 1994.
- [9] B. G. Hall, *Phylogenetic Trees Made Easy: A How to Manual*, 4th edition, Sinauer Associates, 2011.
- [10] J. T. Hooker, *Linear B: An Introduction*, Bristol Classical Press, 1980.
- [11] G. Hosszú, *Heritage of Scribes: The Relation of Rovas Scripts to Eurasian Writing Systems*, Rovas Foundation Hungary, 2013.
- [12] P. Lerney, M. Salemi, and A.-M. Vandamme, editors. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edition, Cambridge University Press, 2009.
- [13] M. Naddeo, *The Ugarit Abjad ... A Rovás Alphabet*, self-published book, 2007.
- [14] P. Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, Springer, New York, 2010.
- [15] P. Z. Revesz, "An algorithm for constructing hypothetical evolutionary trees using common mutations similarity matrices," *Proc. 4th ACM International Conference on Bioinformatics and Computational Biology*, ACM Press, Bethesda, MD, USA, September 2013, pp. 731-734.
- [16] P. Z. Revesz and C. J.-L. Assi, "Data mining the functional characterizations of proteins to predict their cancer relatedness," *International Journal of Biology and Biomedical Engineering*, 7 (1), 2013, pp. 7-14.
- [17] P. Z. Revesz and T. Triplet, "Classification integration and reclassification using constraint databases," *Artificial Intelligence in Medicine*, 49 (2), 2010, pp. 79-91.
- [18] P. Z. Revesz and T. Triplet, "Temporal data classification using linear classifiers," *Information Systems*, 36 (1), 2011, pp. 30-41.
- [19] N. Saitou and M. Nei, "The neighbor-joining method: A new method for reconstructing phylogenetic trees," *Molecular Biological Evolution*, 4, 1987, pp. 406-425.
- [20] M. Shortridge, T. Triplet, P. Z. Revesz, M. Griep, and R. Powers, "Bacterial protein structures reveal phylum dependent divergence," *Computational Biology and Chemistry*, 35 (1), 2011, pp. 24-33.
- [21] R. R. Sokal, and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Science Bulletin*, 38, 1958, pp. 1409-1438.

- [22] T. Triplet, M. Shortridge, M. Griep, J. Stark, R. Powers, and P. Z. Revesz, "PROFESS: A protein function, evolution, structure and sequence database," *Database -- The Journal of Biological Databases and Curation*, 2010, Available: <http://database.oxfordjournals.org/content/2010/baq011.full.pdf+html>
- [23] G. Varga, *Bronzkori Magyar Irásbeliség*, Irástörténeti Kutató Intézet publication, 1993.
- [24] Wikipedia, "History of the Greek alphabet," downloaded July 6, 2015. Available: https://en.wikipedia.org/wiki/History_of_the_Greek_alphabet
- [25] Wikipedia, "Phoenician alphabet," downloaded July 6, 2015. Available: https://en.wikipedia.org/wiki/Phoenician_alphabet
- [26] Wikipedia, "Proto-Sinaitic script," downloaded July 6, 2015. Available: https://en.wikipedia.org/wiki/Proto-Sinaitic_script
- [27] Wikipedia, "South Arabian alphabet", downloaded July 5 2015. Available: https://en.wikipedia.org/wiki/South_Arabian_alphabet
- [28] J. G. Young, "The Cretan Hieroglyphic script: A review article," *Minos* 31-32 (1996-1997[1999]) 379-400.

Peter Z. Revesz holds a Ph.D. degree in Computer Science from Brown University. He was a postdoctoral fellow at the University of Toronto before joining the University of Nebraska-Lincoln, where he is a professor in the Department of Computer Science and Engineering. Dr. Revesz is an expert in databases, data mining, big data analytics and bioinformatics. He is the author of *Introduction to Databases: From Biological to Spatio-Temporal* (Springer, 2010) and *Introduction to Constraint Databases* (Springer, 2002). Dr. Revesz held visiting appointments at the IBM T. J. Watson Research Center, INRIA, the Max Planck Institute for Computer Science, the University of Athens, the University of Hasselt, the U.S. Air Force Office of Scientific Research and the U.S. Department of State. He is a recipient of an AAAS Science & Technology Policy Fellowship, a J. William Fulbright Scholarship, an Alexander von Humboldt Research Fellowship, a Jefferson Science Fellowship, a National Science Foundation CAREER award, and a "Faculty International Scholar of the Year" award by *Phi Beta Delta*, the Honor Society for International Scholars.

Computer Controlled Low Cost System for Professional 360° Photography

Krzysztof Szklanny and Alicja Wieczorkowska

Abstract—Commercial products can be three-dimensionally visualized using 360° photography, and then displayed as a web site. There are professional systems available for 360° photography, but they are expensive. The goal of this work was to create a professional 360° photography system that can be controlled by a computer application, at the cost of no more than 100 euro. The schemata available at a dedicated web site [1], together with the computer application, will help the interested users to construct a system as do-it-yourself.

Keywords—360° photography, Arduino, stepper motor.

I. INTRODUCTION

THIS paper presents a low-cost system for 360° photography, i.e. combining photography and interactive video. Each picture can be considered a movie frame, and the user selects (by means of dedicated tools) which frame she or he wants to see. This way of presenting products is not only intuitive, but also advantageous for the user, as the product can be seen from an arbitrary perspective, and the view can be zoomed in. Therefore, the users can see more than a single or a few pictures provided by the seller, and make informed decision about buying the presented product or not.

A. Existing 360° Systems

There exist many 360° photography systems, e.g. mode360° [2], at the price of about 4000 euro. In Poland, the most popular among the buyers of 360° photography systems is the „Orbitvu” manufacturer [3], selling 3 models: MINI, MIDI, and MAXI. MINI device is intended to be used with objects of weight up to 9 kg; the price is about 2500 euro. MIDI version can be used with objects weighing up to 45 kg, and its price is about 4800 euro. The most expensive version (MAXI) can be used with objects of weight up to 250 kg, and it costs more than 5000 euro.

B. Our Proposed System

The system described in this paper is intended to be used with small objects, so it cannot replace Orbitvu versions for

big and heavy objects. Still, it can be an inexpensive alternative for smaller and yet expensive systems.

The implemented 360° photography system is based on Arduino Uno board, bipolar stepper motor 42BYGH118-01 [5] and Adafruit Motor Shield [6], [7]. The shield contains a motor driver (2 L293D motor drivers – 2 stepper motors can be connected) for communication between the microcontroller and the motor, and also connectors to hook up wires (10-22AWG) and power. We also used a digital single-lens reflex camera DSLR Nikon d700, but other professional Nikon camera can be used as well. Another camera brand could be used, but then a different connecting cable is needed. Currently, the system is set to work in the configuration as described above.

All that the user has to do is to connect the system to a camera and to a computer. Afterwards, the user can place the object of interest on the rotary platform (part of the system), select basic settings in the application, i.e. the total number of pictures to be taken and pause time after each step. After clicking START button, Arduino sends signals through the Motor Shield driver to the stepper motor, and the stepper motor rotates the platform by the required angle. After rotation the platform stops for the time specified by the user in the application. Pictures are taken during the time the platform is stopped.

For example, if the user demands 40 pictures, rotation 9° is performed each step. After each rotation, the signal is sent to the camera, and the picture is taken and saved, according to the exposition parameters defined earlier. This is repeated as many times as needed, in order to obtain the number of pictures demanded by the user.

After finishing taking pictures, the dedicated application is automatically launched. The application saves all the pictures taken as object presentation, which can be displayed on the Internet. The entire system is in plastic housing. All parts can be easily accessed and replaced if needed.

II. ARDUINO

Arduino is an electronic prototyping platform designed to create interactive electronic objects [8], [9]. Technically speaking, Arduino is an integrated circuit containing a microcontroller ATMEGA AVR, a power jack, USB interface (to program the device, for communication with a computer, and as power provider) and input/output (I/O) pins. External circuits can be connected to Arduino using the pins provided.

This work was partially supported by the Research Center of Polish-Japanese Academy of Information Technology, supported by the Ministry of Science and Higher Education in Poland.

K. Szklanny is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (corresponding author, phone: +48 22-58-44-500; fax: +48 22-58-44-501; e-mail: kszklanny@pjwstk.edu.pl).

A. Wieczorkowska, is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (e-mail: alicja@poljap.edu.pl).

In our systems, a photo camera is connected to Arduino through I/O pins, and also the stepper motor driver is connected to Arduino. This is a convenient solution, as the system can be extended without interfering with the original setting.

III. THE 360 DEGREE PHOTOGRAPHY SYSTEM

A. Stepper Motor

Stepper motors are powered with power pulses. They convert a train of input pulses into a precisely defined increment in the shaft position. Each pulse moves the shaft through a fixed angle. As a result, such a motor divides a full rotation into a number of equal steps, rather than rotate continuously. The rotation angle and speed depends on the number and frequency of the pulses. Usually, a single step rotates the shaft by a few up to a few dozen degrees. Maximum speed can reach several hundred rotations per minute. The stepper motor used in our system is 42BYGH118-01, with a single axis (5mm diameter and 20mm length). The dimensions of the motor body are: 42mm x 42mm x 38mm. It is a small engine, so it is easily portable. It weighs only 0.24kg. In the front part, the motor has 4 holes with M3 thread (pitch 31mm,



Fig. 1 The stepper motor used in our system

depth 4.5mm). Minimal step is 1.8 degree. The recommended power supply is 0.5A/12V power. This motor has high torque, 4.4kg/cm, which is quite impressive for such a small stepper motor. The motor is shown in Figure 1.

B. Rotary Platform

The rotary platform consists of three elements: a metal structure, a laser-cut poly (PMMA, Polymethyl methacrylate) base of 25 cm diameter, and a laser-cut poly base 40 cm diameter. The smaller base can be used for taking pictures of small objects, like jewelry or cosmetics. Figure 2 shows the metal structure and the smaller poly base.

The metal structure is placed on the stepper motor pivot. Next, it is screwed, to assure the proper hold. The smaller base is glued to the structure, whereas the bigger one is mounted using small magnets. Such a construction allows choosing



Fig. 2 The rotary platform: the metal structure and a small base

which base is to be used, depending on the object to be photographed. This ensemble can be disassembled, and the user has easy and intuitive access to all parts.

C. Motor Shield

The stepper motor is driven by a driver, which is integral part of the entire system. The Adafruit Motor Shield is used for this purpose, as it contains the L293D driver. This driver is a dual H-Bridge motor driver, which can run our bi-polar stepper motor with up to 600mA per channel. The shield is a ready-made system which allows increasing the stepper power through additional power supply. The driver can be used with direct power stepper motors, up to 36V. It can drive four bi-directional DC motors, two stepper motors, unipolar or bipolar. The shield size is 70mm x 55mm. AF_Motor Arduino library is provided by the producer, and it must be installed before using the motor shield. This library was very useful in our work and made it much faster. The motor shield is shown in Figure 3.



Fig. 3 The motor shield

D. Shutter Release

In order to connect a photo camera to Arduino, we decided to use one of the least expensive remote controls for shutter

release, bought at an internet auction. To assure the appropriate work of the entire system, the cable connecting the remote with the plug was cut, and the 3 wires inside were soldered onto the motor shield. They cannot be directly connected to Arduino pins, as they are covered by the shield. The shutter release works in 2 steps. Two connected wires allow setting focus, and the third one triggers taking picture. After the wires were soldered, the signals sent through these wires can be interpreted and the picture is taken after each step of the motor.

E. Housing

In order to protect all parts of our system, housing was created and the system parts were placed inside housing. It is a simple housing of size 22 cm (length) x 17 cm (width) x 8 cm (height). In order to mount the structure supporting the stepper motor, holes were drilled in the housing. This also allows dismounting the entire system, in the case of failure of any of the system parts. This way our system is aesthetic and ergonomic. Logo and icons were placed on the housing, to guide connecting cables for the camera, power, USB. Labels were printed on eco solvent printer in 1440x1440 dpi



Fig. 4 The 360 degree photography system

resolution on reflective foil (of color matching the housing), thick and raised, so it is more durable than regular transparent or monomer stickers

The entire system in the housing is shown in Figure 4.

F. Problems

While constructing the entire system we tried various solutions. For instance, servomotor was tested, but in this case we only achieved 180 degree rotation. We also had problems with overheating of the stepper motor. It was caused by low level signals sent from Arduino; these signals had to be released. Another problem was to place centrally the photographed object. To facilitate this, we determined precisely the center of the rotary platform and clearly marked it.

IV. APPLICATION

The application is an integral part of the system. The user

interface is in Polish, as it was customized for Polish users. The application is launched by the user after connecting the camera to Arduino, and the driver (the shield) to the computer.

In our application, the user can select how many pictures are to be taken in one cycle; one full rotation of the rotary platform and taking the specified number of pictures is considered to be one cycle. The more pictures, the more smooth the animation will be, but then the final file size increases considerably.

The second parameter chosen by the user is the stop time duration between pictures. It is especially important in the case of studio photography. This time should be adjusted to the flash lamps we have, their charging time, and to the camera (depending on the camera buffer).

The third parameter is the COM port, to which Arduino is connected. Our application automatically detects all available ports.

After completing basic settings, the user clicks „start” button. Next, the rotary platform rotates by the required number of steps. At each stop (of the duration specified by the user), a picture is taken. By default, 50 pictures are taken (rotation by 7.2 degree). Afterwards, the pictures are rescaled and saved (menu: Edycja - Wczytaj zdjęcia). The application automatically finds *.jpg files in the selected folder. The saved pictures are presented in the main window of the application. The user can remove bad pictures (e.g. too dark) and change the order of pictures. Also, an internet presentation of the photographed object can be created as the internet gallery, after clicking „Generuj galerię”. The user is requested to enter the gallery location, and then the gallery is automatically generated. The gallery presents the product rotating by 360 degrees.

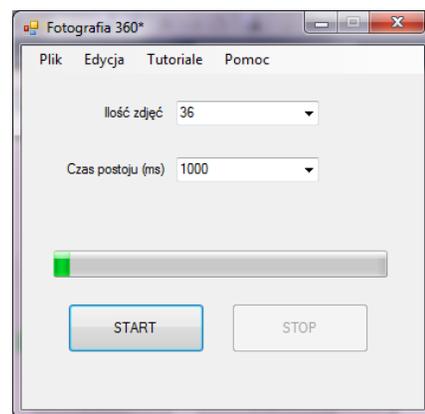


Fig. 5 The user interface of the 360 degree photography system

The application was written in C# programming language, using Microsoft Visual Studio 2010.

V. SUMMARY

Since internet became a very popular selling medium, there exist numerous virtual shops, without having counterparts in

real life. Therefore, sometimes there is no possibility to physically see the object of interest that the potential buyer would like to buy, and he or she has to rely on photographs. Many users just cannot imagine an internet auction without photographs. Thousands of articles are offered for sale every day. Some of them have not chance to be sold because of not being accompanied by photographs, or because of too low technical quality of the illustrating photographs. Although many users have high quality cameras, they often make errors when taking pictures. Usually these errors are caused by insufficient knowledge of the camera used, and lighting techniques. The photographer uses just the built-in flash lamp, whereas the picture taken using natural or artificial light will look better. The sellers take pictures in such a way that the potential buyer does not know what actually is in the picture. Low quality pictures are result of the photographer having no idea how to make these pictures. Using a 360 degree photography system, like the one presented in this paper (or any other one, see [10] for example), can be very helpful in this case. Professional systems are expensive, so the described system is an affordable alternative for ever user of internet auctions.

ACKNOWLEDGMENT

The authors would like to thank the students who worked on this project: Łukasz Romaniuk, Krzysztof Marcinkowski, and Jakub Nalewajko.

REFERENCES

- [1] Fotografia 360 [Online]. Available: <http://fotografia360.pjwstk.edu.pl/>
- [2] Mode360° [Online]. Available: <http://mode360.eu/en>
- [3] Orbitvu [Online]. Available: <http://orbitvu.com/>
- [4] Arduino Uno [Online]. Available: <https://www.arduino.cc/en/Main/arduinoBoardUno>
- [5] Motors [Online]. Available: <http://www.silniki.pl/index.php?site=products&type=250&details=1076>
- [6] Motor Shield [Online]. Available: <http://playground.arduino.cc/Main/AdafruitMotorShield>
- [7] Adafruit Motor Shield [Online]. Available: <https://learn.adafruit.com/adafruit-motor-shield>
- [8] M. McRoberts, "Beginning Arduino", New York, Springer Science+Business Media 2013
- [9] B. Evans, "Beginning Arduino Programming", New York, Springer Science+Business Media 2011
- [10] T. Tseng, "Spin: a photography turntable system for creating animated documentation". Proceedings of the 14th International Conference on Interaction Design and Children IDC '15, ACM New York, 2015, pp. 422-425

Alicja A. Wiczorkowska obtained her PhD. in 1999 and DSc. in 2012. She is the head of the Multimedia Department at the Polish-Japanese Academy of Information Technology. She published about 90 papers including conference papers, journal papers, and book chapters. She is also the author of the handbook on multimedia (in Polish).

Krzysztof Szklanny obtained his PhD. in 2010. He is the head of the photography laboratory at the Multimedia Department at the Polish-Japanese Academy of Information Technology. He published about 20 papers including conference papers, journal papers, and book chapters. He is also an author of high-speed photographs, including commercial photos, exhibited in the Polish-Japanese Academy of IT.

Prediction of Surface Roughness in CNC Milling of Al7075 alloy: A case study of using 8mm slot mill cutter

J. Kechagias, P. Kyratsis, K. Kitsakis and N. Mastorakis

Abstract—The current study investigates the surface roughness of slots produced using an 8mm slot mill cutter during milling of Al7075 alloy. Twenty seven slots were cut using all the different cutting conditions by a KC633M 8mm drill-slot end mill cutter. The three independent variables considered were the depth of cut (a_p , mm), cutting speed (V_c , m/min), and feed rate (f , mm/rev); each one having three different levels. Process performance is estimated using the statistical surface texture parameters R_a , and R_z ; both measured in microns. To predict the surface roughness within the limits of the parameters involved, an artificial feed forward back propagation neural network model was designed for the data obtained.

Keywords—Face Milling, Neural Networks, Modelling, Surface Texture Parameters

I. INTRODUCTION

Aluminum alloys are extensively used as a main engineering material in various industries automotive and aerospace industries, the mould and die components manufacturers and every case in which weight is the most important factor [1]. Surface roughness has an important role in the performance of finished components. It refers to the third up to the sixth order deviation from the nominal surface and all different order deviations are superimposed and form the surface roughness profile [2]. Surface properties dominate the quality of the finished component, since they influence features like dimensional accuracy; tribological characteristics such as the friction coefficient and wear; post processing requirements; appearance and cost. Besides the obvious problems related to correct dimensions, one of the more significant problems is achieving the appropriate finish or surface smoothness on the workpiece. Surface quality is important for a number of reasons i.e. aesthetic (a smooth and free from scratches surface is more likely to give a favorable impression to the customer), surfaces affect safety and they

interact with the environment due to their influence on mechanical properties [3]. Surface roughness or texture constitutes a measure for achieving finer surface irregularities in the finished product, while three components i.e., roughness, waviness, and form are required for its determination [4].

A number of methodologies investigating the relations of the cutting parameters with the produced surface quality are reported in literature. Response surface methodology (RSM) is one of the mostly used in order to build mathematical models based on the Taguchi theory [5]. Other researchers are combining the application of fuzzy logic with the Taguchi method and optimise the surface roughness achieved [6, 7].

The present paper deals with the effects of different process parameters: depth of cut (a_p), cutting speed (V_c), and feed rate (f) on the surface quality, when slot milling Al7075 alloy workpieces. A set of experiments using design of experiments and Neural Networks modelling were used and the surface texture parameters measured during this study were the following: R_a (the arithmetic mean surface roughness) and R_z (the mean of 5 maximum peak-to-valley roughness depths in 5 successive sampling lengths), all measured in μm . Experimental results are used in order to train a feed forward back propagation neural network (FFBP-NN) and predict the surface texture parameters in finish slot milling of Al7075 alloy parts. The use of the FFBP-NN together with the performed experiments resulted in a successful way to model the process and predict the surface texture parameters when different cutting parameters apply.

II. EXPERIMENTAL SETUP

The material used for performing the experiments was Al 7075 (90% Al, 5.6% Zn, 2.5% Mg, 1.6% Cu, and 0.23% Cr). A two flute end mill cutter (KC633M) made by Kennametal was used to perform 27 slots upon three plates. The cutter diameter was 8mm, the length 63mm and the helix angle 30° (Fig. 1).

Three Al7075 plates with a thickness of 12mm (150mm in length and 50mm in width) were used to cut the slots (Fig. 2). A four axis HAAS VF1 CNC machining center with continuous speed and feed control within their boundaries was used for twenty seven slotting operations (Fig. 3). During cutting operations an appropriate coolant fluid was used.

J. Kechagias is with Department of Mechanical Engineering, Technological Educational Institute of Thessaly, Larisa, Greece (e-mail: jkechag@teithessaly.gr).

P. Kyratsis is with Department of Mechanical Engineering and Industrial Design, Technological Education Institution of Western Macedonia, Kila Kozani, Greece (e-mail: pkyratsis@teiw.m.gr).

K. Kitsakis is with Department of Mechanical Engineering, Technological Educational Institute of Thessaly, Larisa, Greece (e-mail: kitsakis@teithessaly.gr).

N. Mastorakis is with Department of Industrial Engineering, Technical University of Sofia, Sofia, Bulgaria (e-mail: mastor@tu-sofia.bg).

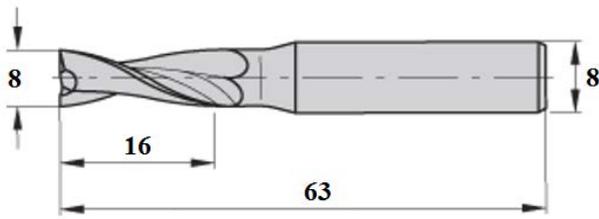


Fig. 1: Cutting tool main dimensions



Fig. 2: Machined specimens of Al7075 alloy



Fig. 3: HAAS VF1CNC machine centre (7500 rpm, 15 KW)

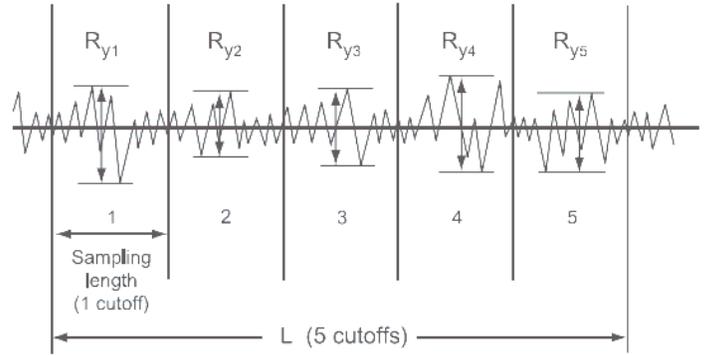
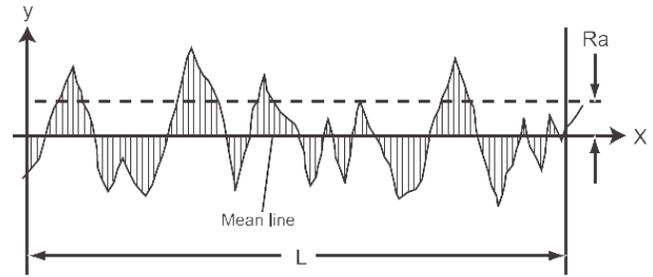


Fig. 4: Surface texture parameters



Fig. 5: Machined specimen and surface tester

Surface roughness is a widely used index characterising a product’s quality, and is measured off-line, when the component is already machined. The surface texture parameters measured during this study were the following (Fig.4):

- $R_a(\mu\text{m})$: the arithmetic average height of roughness irregularities measured from a mean line within the evaluation length (L) [$R_a=(y_1+y_2+\dots+y_n)/n$]
- $R_z(\mu\text{m})$: the mean of 5 maximum peak-to-valley roughness depths in 5 successive sampling lengths [$R_z=(R_{y1}+R_{y2}+R_{y3}+R_{y4}+R_{y5})/5$]

Surface roughness measurements were taken using a RUGOserf tester. For the purposes of the current research, a full factorial experiment plan was used [8, 9, 12, 13].

The main cutting parameters (depth of cut - a_p , in mm; cutting speed - V_c , in m/min; and feed rate - f , mm/rev) were assigned on a standard orthogonal array in order to explore the entire parametric space with a limited number of experiments. Three levels were specified for each of the three cutting parameters (Table 1, 2).

Table 1: Parameter design.

No	Process Parameters	Levels		
		1	2	3
1	Depth of cut (a_p , mm)	0.5	1	1.5
2	Cutting Speed (V_c , m/min)	50	100	150
3	Feed Rate (f , mm/rev)	0.05	0.08	0.11

Table 2: Matrix of experiments

Ex. No.	a_p mm	V_c m/min	f mm/rev	R_a μm	R_z μm
1	0.5	50	0.05	0.247	2.167
2	0.5	50	0.08	0.717	3.067
3	0.5	50	0.11	1.080	4.933
4	0.5	100	0.05	0.335	2.033

5	0.5	100	0.08	0.427	3.500
6	0.5	100	0.11	0.813	5.400
7	0.5	150	0.05	0.250	1.933
8	0.5	150	0.08	0.377	3.467
9	0.5	150	0.11	0.423	3.933
10	1	50	0.05	0.307	2.100
11	1	50	0.08	0.557	3.567
12	1	50	0.11	0.867	4.267
13	1	100	0.05	0.387	2.500
14	1	100	0.08	0.693	3.567
15	1	100	0.11	1.017	4.800
16	1	150	0.05	0.447	3.133
17	1	150	0.08	0.617	3.700
18	1	150	0.11	0.907	5.000
19	1.5	50	0.05	0.307	2.133
20	1.5	50	0.08	0.497	3.433
21	1.5	50	0.11	0.773	4.467
22	1.5	100	0.05	0.343	2.600
23	1.5	100	0.08	0.620	3.867
24	1.5	100	0.11	0.853	5.000
25	1.5	150	0.05	0.657	4.400
26	1.5	150	0.08	0.567	3.667
27	1.5	150	0.11	0.990	4.467

III. NEURAL NETWORK ARCHITECTURE

Aiming in the prediction of the produced surface roughness parameters (R_a , and R_z) during slot milling of an AL7075 alloy, a NN model has been developed. The three factors studied were used as input parameters of the NN model, together with the constant value one (1). Previous studies indicate that by using DoE methods, a structured method of NN parameter-setting can be implemented [14]. It identifies NN and training parameter settings, resulting in enhanced NN performance. Training samples are presented to the NN during training, and the network is adjusted according to its error. The twenty seven (27) experimental data samples (Table 2), were separated into three groups, namely the training, the validation and the testing samples (70%, 15%, and 15% respectively). Training samples were presented to the network during training and the network was adjusted according to its error. Validation samples were used to measure network generalization and to halt training, when generalization stopped improving. Testing samples have no effect on training and so they provide an independent measure of network performance during and after training (confirmation runs).

In general, a standard procedure for calculating the proper number of hidden layers and neurons does not exist. For complicated systems the theorem of Kolmogorov or the Widrow rule can be used for calculating the number of hidden neurons [15]. In this work, the feed-forward with back-propagation learning (FFBP) architecture has been selected to analyze the surface texture parameters. These types of

networks have an input layer of X inputs, one or more hidden layers with several neurons and an output layer of Y outputs. In the selected ANN, the transfer function of the hidden layer is hyperbolic tangent sigmoid, while for the output layer a linear transfer function was used. The input vector consists of the three process parameters of Table 1 and the constant value one (1). The output layer consists of the performance measures, namely the R_a and R_z surface texture parameters. According to ANN theory FFBP-NNs one hidden layer is appropriate to model each mapping between process parameters and performance measures in engineering problems [16].

In the present work, five trials using FFBP-NNs with one hidden layer were tested having 5, 6, 7, 8, and 9 neurons each (Fig. 6). The one with 6 neurons on the hidden layer gave the best performance, as indicated from the results tabulated in Table 3.

The one-hidden-layer 6-neurons FFBB-NN was trained using the Levenberg-Marquardt algorithm (TRAINLM) and the mean square error (MSE) was used as the objective function. The data used were randomly divided into three subsets, namely the training, the validation and the testing samples.

Back-propagation ANNs are prone to the overtraining problem that could limit their generalization capability [15]. Overtraining usually occurs in ANNs with a lot of degrees of freedom [16, 17] and after a number of learning loops, in which the performance of the training data set increases, while the performance of the validation data set decreases. Mean squared error (MSE) is the average squared difference between network output values and target values. Lower values are better. Zero means no error. The best validation performance is equal to 0.306 at epoch 2 (Fig. 7).

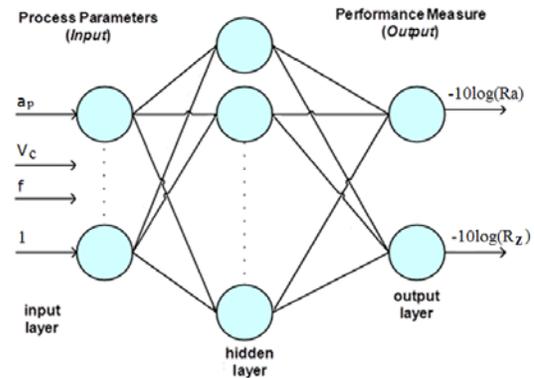


Fig. 6: The selected ANN architecture (feed-forward with back-propagation learning).

Table 3. Best performance of ANN architecture.

	ANN Architecture				
	4x5x2	4x6x2	4x7x2	4x8x2	4x9x2
Training	0.997	0.993	0.998	0.996	0.996
Validation	0.794	0.821	0.580	0.767	0.705
Test	0.897	0.791	0.784	0.705	0.691
All	0.896	0.919	0.883	0.881	0.827
Best val.	0.612	0.306	1.584	2.0541	1.774

epoch	3	2	6	4	3
-------	---	---	---	---	---

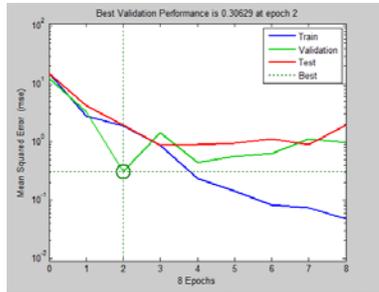


Fig. 7: The selected ANN architecture (feed-forward with back-propagation learning).

Another performance measure for the network efficiency is the regression (R) (Figs 8-11). Regression values measure the correlation between output values and targets. The acquired results show a good correlation between output values and targets during training (R=0.993), validation (R=0.821), and testing procedure (R=0.791).

The trained ANN model can be used for the optimization of the cutting parameters during slot milling of Al7075 alloy.

This can be done by testing the behavior of the response variable (Ra and Rz) under different variations in the values of depth of cut (ap), cutting speed (Vc), and feed rate (f) (Fig. 12-13).

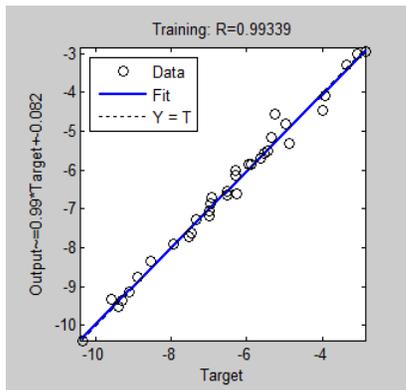


Fig. 8: Regression plots-Training

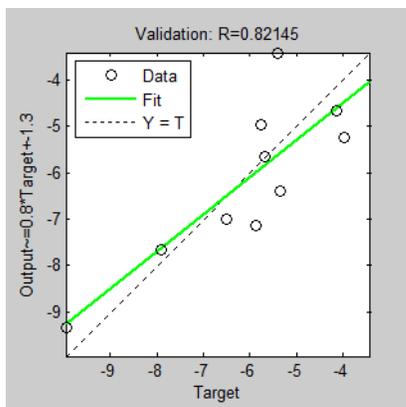


Fig. 9: Regression plots-Validation

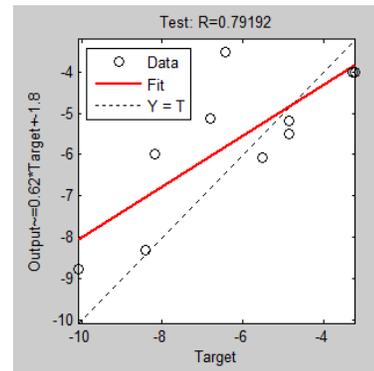


Fig. 10: Regression plots-Test

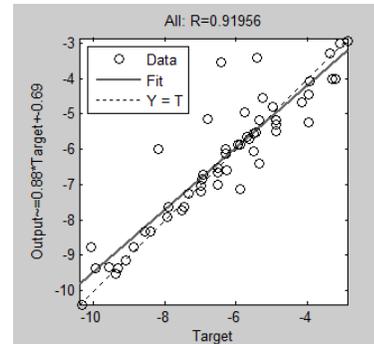


Fig. 11: Regression plots-Validation

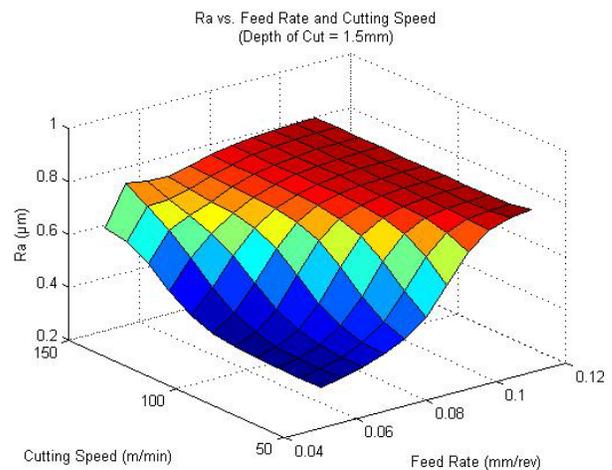


Fig. 12: Ra vs. feed rate and cutting speed (ap=1.5mm)

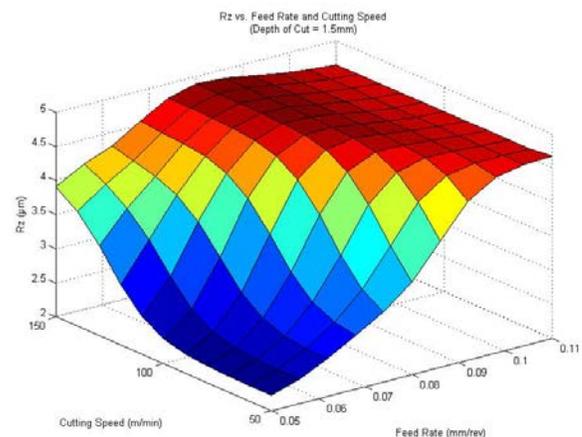


Fig. 13: Rz vs. feed rate and cutting speed (ap=1.5mm)

IV. CONCLUSIONS

The surface texture parameters (Ra and Rz) of Al7075 parts during slot milling were measured according to an orthogonal matrix of experiments. The results were used to train a feed forward back propagation neural network with a topology of 4X6X3 neurons. The proposed NN can be used to predict the surface texture parameters as well as to optimize the process according to each one of the surface texture parameters. As a future work, authors plan to improve the performance of FFBP-NN incorporating more experiments as well as investigate the performance of alternatives training algorithms. In addition, a comparison among other approaches such as regression and additive modeling will be performed. Using the extracted NN, the surface response of Ra and Rz can be drawn and the effects of process parameters can be estimated inside the experimental region in which the designed experiments were conducted. This methodology could be easily applied to different materials and initial conditions for optimization of other material removal processes.

ACKNOWLEDGMENT

The authors express special thanks to Mr Konstantinos Sirgkanis for their support during the developed laboratory activities.

REFERENCES

- [1] S. Rawangwong, J. Chatthong, R. Burapa, W. Boonchouytan, An investigation of optimum cutting conditions in face milling semi-solid AA7075 using carbide tool, *International Journal of Innovation, Management and Technology*, 3(6) (2012) 692-696.
- [2] D. Vakondios, P. Kyratsis, S. Yaldiz, A. Antoniadis, Influence of milling strategy on the surface roughness in ball end milling of the aluminum alloy Al7075-T6, *Measurement*, 45(6) (2012) 1480-1488.
- [3] P. Munoz-Escalona, P. Maropoulos, A geometrical model for surface roughness prediction when face milling Al 7075-T7351 with square insert tools, *Journal of Manufacturing Systems*, (2014) doi:10.1016/j.jmsy.2014.06.011.
- [4] S. Karagiannis, P. Stavropoulos, C. Ziogas, J. Kechagias, Prediction of surface roughness magnitude in computer numerical controlled end milling processes using neural networks, by considering a set of influence parameters: An aluminium alloy 5083 case study, *Proc IMechE Part B: J Engineering Manufacture*, 228(2) (2014) 233–244.
- [5] M.Y.Wang, H.Y. Chang, Experimental study of surface roughness in slot end milling AL2014-T6, *International Journal of machine Tools & Manufacture*, 44 (2004) 51-57.
- [6] T.P. Mahesh, R. Rajesh, Optimal selection of process parameters in CNC end milling of Al7075-T6 aluminum alloy using a Taguchi-Fuzzy approach, *Procedia Materials Science*, 5 (2014) 2493-2502.
- [7] S. Karagiannis, V. Iakovakis, J. Kechagias, N. Fountas, N. Vaxevanidis, Prediction of Surface Texture Characteristics in Turning of FRPs using ANN, *Communications in Computer and Information Science* 383 (2013) 144-153.
- [8] M.S. Phadke, *Quality Engineering using Robust Design*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [9] N. Vaxevanidis, J. Kechagias, N. Fountas, D.E. Manolagos, Three component cutting force system modeling and optimization in Turning of AISI D6 tool steel using design of experiments and Neural Networks, In *Proceedings of the World Congress on Engineering 2013 London, U.K.*, Vol. I, July 2-6, (2013).
- [10] J. Kechagias, V. Iakovakis, A neural network solution for LOM process performance, *The International Journal of Advanced Manufacturing Technology*, 43(11) (2009) 1214-22.
- [11] M. Pappas, J. Kechagias, V. Iakovakis, S. Maropoulos, Surface roughness modelling and optimization in CNC end milling using Taguchi design and Neural Networks, *ICAART 2011 – In Proceedings of the 3rd International Conference on Agents and Artificial Intelligence 1* (2011) 595-598.
- [12] D. Montgomery, *Design and Analysis of Experiments*, 7th Edition, John Wiley & Sons (2008).
- [13] C.C. Tsao, Grey-Taguchi method to optimize the milling parameters of aluminum alloy, *The International Journal of Advanced Manufacturing Technology*, Vol. 40 (2009) 41–48.
- [14] W. Sukthomaya, J. Tannock, The optimisation of neural network parameters using Taguchi's design of experiments approach: an application in manufacturing process modelling, *Neural Computing and Applications*, 14(4) (2005) 337-344.
- [15] S.G. Tzafestas, P.J. Dalianis, G. Anthopoulos, On the overtraining phenomenon of backpropagation NNs., *Mathematics and Computers in Simulation*, 40 (1996) 507-521.
- [16] C.T. Lin, G.C.S. Lee, , *Neural fuzzy systems- A neuro-fuzzy synergism to intelligent systems*. Prentice Hall (1996).
- [17] L. Prechelt, Automatic early stopping using cross validation: quantifying the criteria, *Neural Networks*, 11(4) (1998) 761–767.

Why an Economy needs more than one Currency: the Scientific Evidence

Bernard Lietaer

Abstract - The paradigm of a single currency per country which has been prevailing in the entire field of economics for centuries needs to be challenged if we want to achieve a more resilient economic system. This finding is based on the recent discovery of a general law that determines the structural conditions of stability of any complex flow network. This law applies equally to all natural ecosystems because they are complex flow networks in which biomass circulates; and to economies because they are complex flow networks in which money circulates. Graphics provide illustrations to this argument.

Keywords - resilience vs efficiency, monetary ecosystems, complementary currencies

I. INTRODUCTION

A complementary currency is any standardized medium of exchange used in an economy other than the official legal tender. Examples include commercial loyalty currencies such as airline miles; local and regional currencies of which several thousands small-scale systems have emerged in many countries around the world such as Time Banks or Local exchange trading systems (LETS); and Business-to-Business systems such as the WIR in Switzerland. Conventional economists tend to regard all such complementary currencies as an anomaly, to be dismissed as an irrelevant or romantic distraction. Regulators tolerate them, as long as they remain marginal. If ever any were to grow to a relevant size, they believe such monetary innovations should be suppressed because of concerns that they might disturb monetary policy or reduce economic efficiency (Rösl, 2006).

A recent scientific breakthrough provides the evidence that a systemic cause for the prevailing monetary and financial instability happens to be the monopoly of one single type of currency. In fact, it will be shown that, far from being a disturbance, a minimum variety of monetary media is a *necessary condition* for sustainable economic and financial stability in a society. Furthermore, this isn't a conclusion based on some personal preferences or opinions, but is backed by a law that is as universal as gravity!

Indeed, there exists a structural flaw in our Modern monetary system, a flaw that has been with us for centuries. Actually, it

was already “doing its thing” when the Dutch tulip bubble burst in 1637, and it played an unacknowledged role in every crash since that time, including the one whose fallout we are still experiencing now! The proof for this claim comes from a fundamental law that governs all complex flow systems, including all natural ecosystems, as well as economic and financial systems. It is based on a recent theoretical breakthrough which makes it possible to measure quantitatively with a single metric the sustainability of any complex flow system as an emergent property of its structural diversity and interconnectivity. Furthermore, it is revealed that whenever diversity in a complex flow network is being sacrificed because of too much emphasis on efficiency, structural collapses are a totally predictable consequence.

From this perspective, it is obvious that we have been living worldwide with a monoculture of the same type of media of exchange, in the form of a single national currency monopoly in each country, created everywhere through bank-debt. Both the Marxist and the various capitalist schools of economic thought have been blind to this issue: they all presuppose as self-evident a monopoly of a single national currency. The main difference in this respect between communism of the Marxist-Leninist variety on the one side, and capitalism on the other, was that in the former governments were the owners of the banks all the time, while in the capitalist system this is the case only for banks that have dramatically failed. But the money system itself is in fact the same: a single national currency created through bank-debt... The structural solution to economic and financial sustainability becomes also clear: we need to diversify the types of currencies available in a society and the types of agents that are creating them, specifically through complementary currencies.

II. SUSTAINABILITY OF COMPLEX FLOW SYSTEMS

For those desiring a fully documented step by step mathematical evidence of what is being claimed here, please refer to the seminal paper (Ulanowicz, Goerner, Lietaer and Gomez, 2009).¹ Only the most relevant points can be summarized here.

Information is any “difference that makes the difference” (Gregory Bateson) and, as the binary logic of the digital age has popularized, such difference almost always involves the absence of something. In coming to terms with the working of whole systems, information theory (IT) is a means for apprehending and quantifying what is missing. The key point is that if one is to address the issue of sustainability, then the

inchoate, undetermined “potentiality” of a system also becomes an indispensable focus of inquiry, because it is the source of the resilience that allows the system to persist (Conrad, 1983).

What IT tells us is that a system’s capacity to undergo change has two components: order and the absence of order. The first component, called “mutual constraint”, quantifies all that is regular, orderly, coherent and efficient. It encompasses basically all the concerns of conventional science. This first component is an analogue of Newton’s Third Law of motion, or of the *Yang* construct familiar in Chinese philosophy.ⁱⁱ By contrast, the second component represents the lack of those same attributes, or the irregular, disorderly, incoherent and potentially inefficient behaviours that have escaped the scrutiny of science mainly because they cannot easily be described, and even less readily repeated or measured, or all of the above. It corresponds to the Chinese *Yin*.

In the jargon of IT, this second, usually overlooked component of system change is called “conditional entropy”; it can also be thought of as uncommitted potential. Critically what this says is that the very absence of order (even if its potential is never activated, and therefore unnoticed and unmeasured) plays the key role for a system to persist over the long run, to adapt to a changing environment, or survive unexpected challenges. It will be shown next why this absence of order happens to be even more significant than the first variable, order, if we are to understand sustainability.

A living system adapts in homeostatic fashion to buffer performance by expending what Odum called “reserves” (Odum, 1953). The reserve in this case is not some palpable storage, like a cache of some material or energy resource. Rather, it is a characteristic of the system structure that reflects its flexibility both to survive change and to adapt to new circumstances – and it usually requires some loss of efficient performance (Ulanowicz, 2009, 2010). Systems that endure – that is, are sustainable – lie in dynamic balance somewhere between these two poles of order and disorder, efficient performance and adaptive resilience.

Let us now define more precisely our terminology:

- *Efficiency* is defined as the capacity of a complex flow network to process volume of whatever flows through it, per unit of time (e.g. grams of biomass per square meter per year for a natural ecosystem; GNP per capita in an economy; billions of dollars per day in an electronic payment system; or KWh in an electrical distribution network).
- *Resilience* is the capacity of a complex flow network to survive an attack, a disease, or adapt to a change in the environment.

So crucially, efficiency is definitely not a sufficient metric for sustainability. This confirms the often repeated point that GNP/capita is not sufficient to assess economic sustainability. Indeed, this metric cannot distinguish between a healthy

sustainable growth and a short-term bubble doomed to collapse. For a complex flow system to be sustainable, it is also necessary that it possesses enough resilience, an undefined and contingent responsiveness to the unpredictable challenges thrown up by its own workings or its environment. It is thanks to this feature that a resilient flow network of any kind can withstand shocks and adapt itself when necessary.

In summary, natural ecosystems exist because they have *both* sufficient self-directed identity *and* flexibility to change. The polarities necessitate each other in an appropriate balance in harmonious complementarity. Over time, nature must have solved many of the structural problems in ecosystems. Otherwise, these ecosystems simply wouldn’t exist today. They are our best living examples of large scale and long-term sustainability in action.

Furthermore, it has been shown that the driving force behind this evolution happens to be entropy, which is generally accepted as the most universal law in our universe. It is even more fundamental than gravity, given that modern thermodynamics have proven that gravity is one of its consequences! (Dewar, 2003) (Chaison, 2003) (Roddier, 2012)

III. EMPIRICAL ECOLOGICAL EVIDENCE

The key observation is that nature does not select for maximum efficiency, but for a balance between these two opposing poles of efficiency and resilience. Because both are indispensable for long-term sustainability and health, the healthiest flow systems are those that are closest to an optimal balance between these two opposing pulls. Conversely, an excess of either attribute leads to structural instability. Too much efficiency (excess Yang) leads to brittleness and too much resilience (excess Yin) leads to stagnation: the former is caused by too little diversity and connectivity and the latter by too much diversity and connectivity.

Sustainability of a complex flow system can therefore be defined as the optimal balance between efficiency and resilience of its network. With these distinctions we were now able to define and precisely quantify a complex system’s sustainability in a single metric. However, the object of our interest involves in reality four dimensions, and is therefore hard to visualize mentally or graphically.

The following 3 dimensional graphic analysis provides an insight into the shape of this 4 dimensional object.ⁱⁱⁱ

The vertical axis represents the sustainability *F* of a complex flow network. The two horizontal axes are respectively the interconnectivity and the diversity of the same complex network. In natural ecosystems the weight on efficiency has empirically been determined from a regression analysis using actual data from a wide variety of ecosystems to have a value of $\beta = 1.288$. Fig. 1 illustrates the conditions where natural

ecosystems seen as complex flow networks are the most sustainable (color yellow).

What the graph shows is that a natural ecosystem will tend to evolve towards sustainability preferably by increasing diversity, and secondarily through increasing interconnectivity. For instance, a forest will tend to become more sustainable by multiplying the diversity of fauna and flora that can survive in the environmental conditions of the area. To a lesser extent will it encourage animals to feed on other plants and prey than what they had evolved to feed from in their specific ecological niche (e.g. by having rabbits become carnivores, or pandas learn to eat something else than their usual specific type of bamboo).

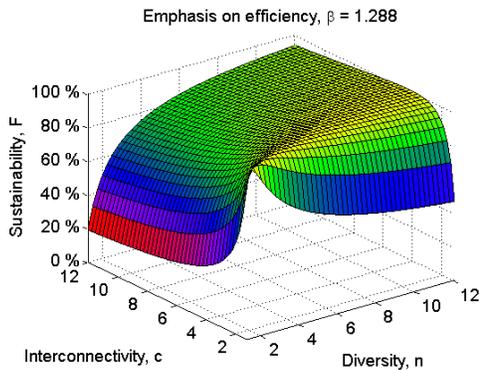


Fig. 1 : Sustainability as a function of diversity and interconnectivity in natural ecosystems (with $\beta = 1.288$ determined empirically). The shape of the highest sustainability zones (colored in yellow) shows that a natural ecosystem will tend to evolve towards sustainability primarily by increasing diversity, and secondarily through increasing interconnectivity.

In many of our human designed systems, there is a tendency to over-emphasize growth of volume, at the expense of resilience. In short, to emphasize efficiency more than resilience. As we saw earlier, the key technical argument from conventional economists against multiple currencies is that it would decrease efficiency.

One consequence is the unleashing of autocatalytic forces that the theory warns can be problematic for the sustainability of the whole network. In an economy, that takes the form for instance of the larger economic actors that succeed through lobbying to obtain laws and rules that favor their growth at the expense of the smaller ones. In the financial domain, it takes the form of banks that become “too big to fail”, and thereby obtain implicit guarantees from governments that they will be bailed out in case of trouble.

Fig. 2 shows what tends to happen when such a bias towards efficiency is introduced. The sustainability is dramatically reduced. Only a narrow zone, requiring a lot of diversity remains sustainable. However, in the monetary domain, as

long as we remain stuck with the idea of a single national currency, this diversity is systematically eliminated. Such a system will predictably collapse, and if it is restored to “normal” will collapse again and again.

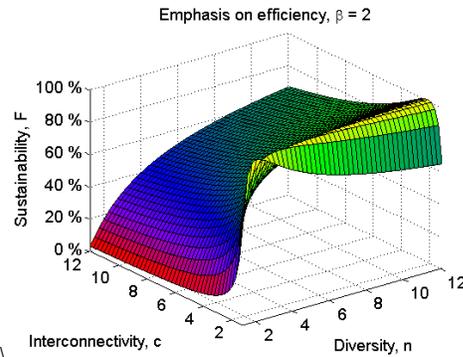


Fig. 2 Reduction of sustainability when efficiency is overemphasized ($\beta = 2$). Sustainability is reduced to a narrow band requiring a lot of diversity.

When this bias towards favoring efficiency is pushed still further, the collapse of interconnectivity and sustainability gets still more accentuated, as illustrated in Fig. 3. In other words, the more the bias towards efficiency prevails, the quicker the system will collapse.

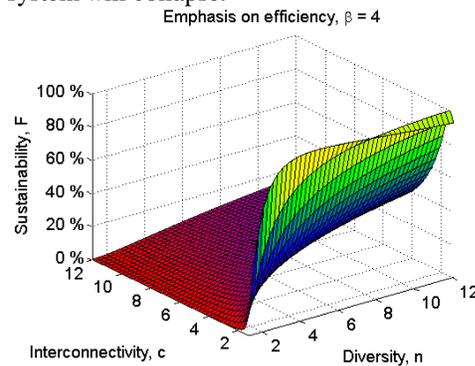


Fig. 3 When overemphasis on efficiency is pushed still further ($\beta = 4$), a complex flow network becomes a totally unsustainable system.

Fig. 4 is a two dimensional graph of Sustainability as a function of Efficiency and Resilience in complex flow networks.

Observe that there is an asymmetry: in natural ecosystems optimality requires more resilience than efficiency! (The optimal point lies closer to resilience than efficiency on the horizontal axis).

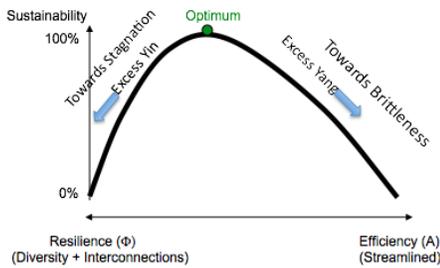


Fig. 4: Sustainability curve mapped between the two polarities of Efficiency and Resilience. Nature selects not for a maximum of Efficiency, but for an optimal balance between both requirements. Too much efficiency will lead to collapses, too much resilience leads to stagnation.

Moving beyond information theory, ecologists have empirically measured the transfer of biomass and energy (“trophic exchanges”) within ecosystems. They have also found ways to derive values for an ecosystem’s throughput efficiency and resilience by estimating network size and network connectedness in terms of the two structural variables: diversity and interconnectivity. It turns out that there is a specific zone of optimal robustness, into which all observed natural ecosystems fall. This zone has been named the “window of viability” (in ecological literature the “window of vitality”). (See Fig. 5).

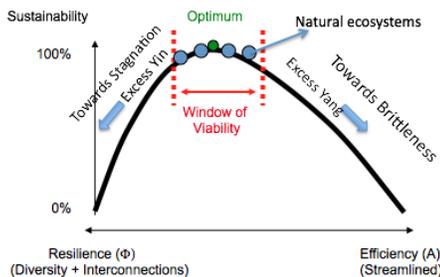


Fig. 5: The “Window of Viability” in which all sustainable natural ecosystems operate. Complex natural ecosystems invariably operate within a specific range on each side of the Optimum point.

IV. APPLICATION TO OTHER COMPLEX FLOW SYSTEMS

The question will undoubtedly be raised whether what we learn from ecosystems still is relevant when applied to other complex flow systems, such as economic or financial systems. This question is typically the result of the interpretation that natural ecosystems are used here as a metaphor for economies.

We can’t emphasize enough that the relevant link is *not* a metaphor! Instead, what applies is a specific universal law that happens to apply to the structure of *all* complex flow network systems, independently of the nature of what flows in the network. Just as Newton’s discovery was that the same law of gravity applies to both the moon and an apple falling in his garden! In our particular case, this is even literally true, because the driving force behind our theoretical framework is entropy, which happens to be at the origin of gravity itself. In other words, we were conservative at the beginning of this paper when stating that the law backing our claims is as fundamental as gravity. It is in fact even more fundamental than gravity!

It is indeed critical to understand that the findings arising from natural ecosystems involve only the two *structural* variables of variety and interconnectivity. They remain therefore valid for any complex flow network with the same structure, regardless of what is flowing through the network. It can be biomass in an ecosystem, information in a biological system, electrons in an electrical power network, or money in an economic system.^{iv} This is precisely one of the strong points of using a web-like network approach instead of a machine-like metaphor.

The fields of engineering, business and economics have all been focusing almost exclusively on efficiency, and therefore constitute a wide-open field to explore the validity of the proposed metrics to improve sustainability. For example, electrical power grids had been systematically optimized for decades towards ever greater technical and economic efficiency. It has come as a surprise to many engineers that, as they have approached higher efficiencies, suddenly large-scale blackouts have been breaking out with a vengeance “out of nowhere”. For instance, a few decades ago several blackouts hit large areas of the United States and Northern Germany. The data should be available to model these systems as flow networks, because that is what they literally are. One could then quantify their efficiency and resilience, and their Window of Viability. The solution on how to rebalance such a system to make it less brittle, and to determine its optimal sustainability, would be an obvious “hard science” test application of the concepts and metrics described here.

The point being made here is truly profound and has wide-reaching implications for all complex systems, natural or human-made. Placing too much emphasis on efficiency tends to automatically maximize flows, size and consolidation at the expense of choice, connectivity and resilience until the entire system becomes unstable and collapses. In contrast, conventional engineering, economics and finance assumes invariably assumes that more efficiency is always better!

Until this finding, total throughput and efficiency have been the only means for us to identify the relative success of a system, whether in nature or in economics. For example, in ecosystems, as in economies, size is generally measured as the total volume of system throughput/activity. Total System

Throughput (TST) measures size in this way for ecosystems, exactly as Gross Domestic Product (GDP) does in economies. Many economists urge endless growth in size (GDP) because they assume that growth in size is a sufficient measure of health. TST and GDP, however, are both poor measures of sustainability because *they ignore network structure*. They cannot, for example, distinguish between a healthily thriving resilient economy; and a bubble that is doomed to burst. Or between healthy “development,” as Herman Daly (1997) describes it, and explosive growth in monetary exchanges simply due to runaway speculation. Now, however, we can distinguish whether a particular increase in throughput and efficiency is a sign of healthy growth or just a relatively short-term bubble that is doomed to collapse.

V. APPLICATION TO FINANCIAL AND MONETARY SYSTEMS

Applying the above complex flow framework specifically to financial and monetary systems, we can predict that excessive focus on efficiency will tend to create exactly the kind of bubble economy which we have been able to observe repeatedly in every boom and bust cycle in history, including the biggest bust of them all, the one triggered in 2007-8 from which we are still experiencing the consequences today.

If we view economies as flow systems, this ties directly into money’s primary function as medium of exchange. In this view, money is playing in the real economy the identical role of biomass in an ecosystem: it is an essential vehicle for catalyzing processes, allocating resources, and generally allowing the exchange system to work as a synergetic whole. The connection to structure is immediately apparent. In economies, as in ecosystems and living organisms, the health of the whole depends heavily on the structure by which the catalyzing medium, in this case, money, circulates among businesses and individuals. Money must continue to circulate in sufficiency to all corners of the whole because poor circulation will strangle either the supply side or the demand side of the economy, or both.

Our global monetary system is itself an obvious flow network structure, in which monopolistic national currencies flow within each country (or group of countries in the case of the Euro), and interconnect on a global level. The technical justification for enforcing a monopoly of a single currency within each country is to optimize the efficiency of price formation and exchanges in national markets. Tight regulations are in place in every country to maintain these monopolies. Banking regulations further ensure that banks tend to be institutional carbon copies of each other both in terms of their structure and behaviour. This was demonstrated among the world’s bigger banks, most recently and with a vengeance, with the simultaneous crash of 2008.

Furthermore, in a seminal 1953 paper, Milton Friedman proposed that letting markets determine the value of each national currency would further improve the overall efficiency

of the global monetary system (Friedman, 1953). This idea was actually implemented by President Nixon in August 1971, to avoid a run on the dollar at that time. Since then, an extraordinarily efficient and sophisticated global communications infrastructure has been built to link and trade these national currencies. The trading volume in the foreign exchange markets reached an impressive \$5.3 trillion *per day* in 2013, to which trillions of currency derivatives should be added (Bank of International Settlements, 2013). Over 95% of that trading volume is speculative, and less than 5% is in fact used for actual international trade of goods and services.

Speculation can play a positive role in any market: theory and practice show that it can improve market efficiency by increasing liquidity and depth^v in the market. But current speculative levels are clearly out of balance. Although over half a century old, John Maynard Keynes’ opinion has never been as appropriate as it is today. “Speculators may do no harm as bubbles on a steady stream of enterprise. But the position is serious when enterprise becomes the bubble on a whirlpool of speculation. When the capital development of a country becomes a by-product of the activities of a casino, the job is likely to be ill-done.” (Keynes, 1936)

Nobody questions the efficiency of these huge markets; but their lack of resilience has also been amply demonstrated, for instance during the Asian crisis of the late 1990s, and dozens of other monetary crashes. In short, our global network of monopolistic national moneys has evolved into an overly efficient and therefore dangerously brittle system. This system’s lack of resilience shows up not in the technical field of the computer networks (which all have backups which too narrow minded accountants may see as “redundant” or “inefficient”), but in the financial realm. Such a crisis, particularly a combined monetary and banking crash, is - other than war - the worst thing that can happen to a country.

Even more ironically, whenever a banking crisis unfolds, governments invariably help the larger banks to absorb smaller ones, believing that the efficiency of the system is thereby further increased. This make banks that are “too big to fail” into still bigger ones, until they become “too big to bail”. This situation of the current financial system can therefore be graphically illustrated as in Fig. 6.

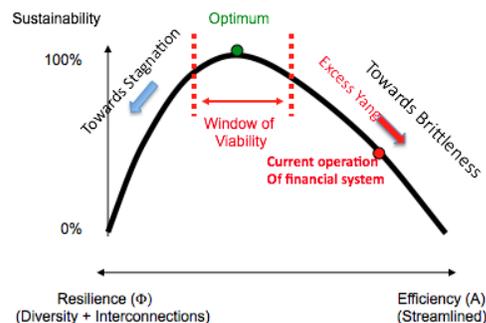


Fig. 6: Today’s global monetary ecosystem is significantly overshooting the Window of Viability because of its exclusive emphasis on efficiency. It is careening toward brittleness and

collapse because a general belief prevails that all improvements need to go further in that the direction of increasing growth and efficiency (red downward arrow on the right). For instance, the global monoculture of bank-debt money as legal tender is technically justified on the basis of efficiency of price formation and exchanges within each country. Internationally, floating exchanges were also justified because they are “more efficient”. This is part and parcel of how we are building up an “Excess Yang” unbalance.

Similarly, the substance that circulates in our global economic network – money – is also maintained as a monopoly of a single *type* of currency: bank-debt money, created with interest.

Imagine a planetary ecosystem where only one single type of plant or animal is tolerated and artificially maintained, and where any manifestation of successful diversity is eradicated as an inappropriate “competitor” because it would reduce the efficiency of the whole.

Such an overly efficient system - as described in Fig. 6 - is “an accident waiting to happen”. It is condemned to a sudden crash and collapse whatever heroic efforts will be made even by competent people who try to manage it. What happens after a collapse? Let us take as example the most extreme cases of total systemic meltdown. For instance: an atomic bomb burns a forest down to ashes; or a complete meltdown of a financial system.

As examples in the monetary domain, total collapses happened in Germany in the 1920s, in the United States during the “bank holidays” of the Great Depression, in Russia in the 1990s and in Argentina in 1999-2002. All these cases have in common to be simultaneous monetary and banking crisis. A dollar or a Euro crisis could again make this phenomenon part of our near future.

The process of a collapse shows up graphically with a drop of sustainability to close to 0%. (see Fig. 7) and an extreme fragmentation, without any efficiency at all. In a forest, this takes the form of seedlings of any type trying to sprout randomly. This is gradually followed by the successful multiplication of the flora and fauna that happen to be best adapted to the prevailing environment gradually giving structure to the newly emergent natural ecosystem until it slowly climbs back into the window of viability (see green arrow in Fig. 7).

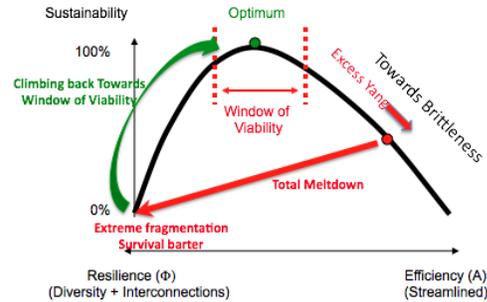


Fig. 7: Dynamics of a total collapse and recovery of a complex flow-network in natural ecosystems. First, extreme fragmentation takes place, without any efficiency. All kinds of species try to emerge. In the case of an economy, this takes the form of survival barter where everything is potentially used as currency. In natural ecosystems, the recovery sees the less inefficient species starting to thrive, and together they generate an ecosystem operating within the Window of Viability.

The same Fig. 7 can be used to explain what could happen in a total collapse of the monetary and banking system. After a period of total fragmentation in the form of survival barter, some standards and some informal agreements would start to emerge. In Argentina this took the form of the multiplication of local exchange mechanisms, under the names of “*ruedes de trueque*” in which locally issued *creditos* currencies were used as medium of exchange in weekly neighbourhood markets. Assuming that the designs of these emergency systems had been sound (which unfortunately wasn’t the case in Argentina!), then the better systems would have had a chance to emerge as models for others, and gradually more efficient exchange systems could have evolved. Over time, a more diversified and more interconnected economy would rebuild, which would return the system back into the Window of Viability.

In Modern monetary practice, however, what has invariably happened until now is that as soon as possible, under the ideology of monetary orthodoxy and as a result from bank lobbying, a monopoly of bank debt money as the only medium of exchange is re-established back as before. This took place for instance in Germany in the 1920s and in the US in the 1930s, when all the “emergency currencies” were outlawed; or in Argentina through a massive falsification of *credito* paper currencies.

However, we now know that such a monoculture is not a sustainable structure in the long run, so that such a return to “normalcy” is in fact missing the window of sustainability. As a consequence, we are getting back on the next cycle of pushing for more efficiency within a monoculture environment, which will lead to the next crash a few decades later.

This process is illustrated in Fig. 8.

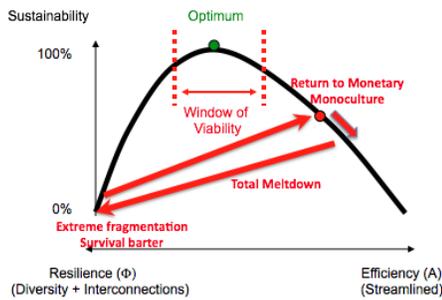


Fig. 8: Under the pressure of the monetary orthodoxy and powerful lobbies, the monopoly of bank-debt money as medium of exchange is re-established after a meltdown. This brings the system back to the situation described in Fig. 6. The autocatalytic pressures in such a monoculture will tend to gradually push the system towards more excessive efficiency and therefore its next crisis. As soon as memories of the previous one has started to fade...

Of course, in natural ecosystems, such a systematic, repetitive and artificial human intervention isn't taking place, and natural systems follow normally the dynamic of the green arrow in Fig. 7.

Fortunately, most crises are less extreme than total financial and monetary meltdowns. However, the exercise of exploring "pure" extremes gives some ideas of the power and nature of the dynamics that are involved. Less extreme crises simply manifest only some of the features of the process. Just like a partial forest fire, one that doesn't reduce everything to ashes, manifests only some of the attributes of a total burnout.

As stated earlier, nature has over billions of years selected the conditions under which complex ecosystems are sustainable, otherwise they wouldn't exist today. In contrast, humanity still struggles with the issue of how to create sustainable economies. We know that the same theoretical framework applies to both natural and man-made complex systems...

We have been going through this loop many times by now. To be precise, according to IMF data, between 1970 and 2010, there have been no less than 145 countries experiencing a banking crises, 208 monetary crashes and 72 sovereign debt crises: a staggering total of 425 systemic crises, an average of more than ten per year! (Capri & Klingebiel, 1996; Laevan & Valencia, 2010). The Greek crisis increases the number of sovereign debt crisis from 72 to 73...

How many more crises do we want to undergo? Lots of countries in Europe are now candidates for being considered "overendebted"...

These crises have hit more than three-quarters of the 180 countries that are members of the IMF, many of them several times. How many more crises do we need before humanity is willing to learn that this is a structural issue, and that only structural changes will avoid repeating the same patterns?

It may be useful to remember during the next monetary or financial crisis that Albert Einstein defined insanity as doing

the same thing over and over again and expecting different result...

VI. A STRUCTURAL MONETARY SOLUTION

Conventional economic thinking assumes the *de facto* monopolies of national moneys as an unquestionable given. In contrast, the clear lesson from nature is that systemic monetary sustainability requires a diversity of currency systems, so that multiple and more diverse agents and channels can emerge for a richer range of exchanges to emerge.

It is important to realize that another way than waiting for a total meltdown is available now to get back towards the window of vitality. That other way is to let complementary currency system grow, or even encourage the soundest of them to blossom, and gradually and gently push back the excesses of the monoculture, as seen in Fig. 9. There is an complementary currency movement slowly gathering steam which should be encouraged as it is a gradual way to bring our economy into more stability.

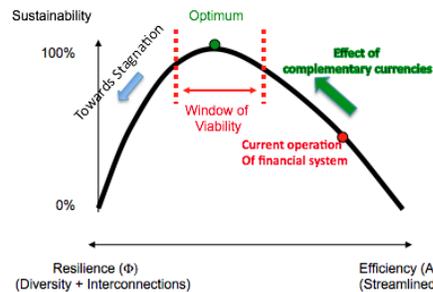


Fig. 9: The Effect of Complementary Currencies

The operation of complementary currencies of diverse types enables the economy to flow back towards greater sustainability (thick upward green arrow). While it is true that this reduces efficiency, that is the price to pay for increased resilience of the whole. Complementary currencies facilitate transactions that otherwise wouldn't occur, linking otherwise unused resources to unmet needs, and encouraging diversity and interconnections that otherwise wouldn't exist.

Conventional economists are therefore correct when they claim that a diversity of media of exchange is less *efficient* than a monopoly. However, it has now been proven that such a drop in efficiency is the necessary price for increasing the *resilience* of the economic system.

At the other extreme, some complementary currency enthusiasts claim that we should encourage very large numbers of complementary currency systems. A warning can and should now be issued: More is not always better... At a caricature extreme, if every citizen issued his or her currency, the result would predictably be total stagnation!

Furthermore, we now have also empirical proof from 75 years of data from the WIR system in Switzerland that business-to-

business complementary currencies actually help central banks in their task of stabilizing the national economy in terms of employment and in smoothing the swings in the business cycle (Stodder, 1998, 2008, 2010). In a period when unemployment, poverty and economic exclusion are all increasing in the developed world, it would be important that central banks revisit this issue with a more open mind than has been the case so far...

VII. CONCLUSION

A monetary monoculture has been legally imposed for centuries in the name of market efficiency. Furthermore, governments everywhere enforce this monopoly most importantly by requiring that all taxes be paid exclusively with conventional bank-debt money.

It is important to understand that the only real power and leverage that governments have in the monetary domain is the capacity to specify the kind of currency or currencies it accepts in payment of fees and taxes. This is critically important because the choice of that currency also dictates the kind of efforts that governments require from businesses and citizens. If conventional money created by bank debt is the only currency acceptable in payment of taxes, de facto the efforts that governments require is to be “commercially successful” regardless of the environmental consequences.

Humanity has become the steward of this planet’s biosphere. This has become so, involuntarily, unconsciously and reluctantly. Ultimately, humans have no choice but to learn how to learn to live more sustainably on this planet, or they will cease to exist. A new balance between efficiency and resilience is needed to attain that. Nature has shown us how this can be achieved. Why not listen?

REFERENCES

- Bank of International Settlements (BIS) (2010) Stephen G. Cecchetti, Madhusudan S. Mohanty and Fabrizio Zampolli, *The Future of Public Debt: Prospects and Implications*
- Bank of International Settlements (BIS). (2013) *Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity 2013*
- Caprio, Gerard Jr, and Daniela Klingebiel. (1996). *Bank Insolvencies: Cross Country Experience* Policy Research Working Papers No.1620. Washington, DC, World Bank, Policy and Research Department.
- Chaisson, Eric “Non-equilibrium Thermodynamics in an Energy-Rich Universe”, in A. Kleidon and R.D. Lorenz (eds), *Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond* (Berlin/ New York: Springer, 2005), pp. 21-33.
- Conrad, Michael. (1983). *Adaptability: The Significance of Variability from Molecule to Ecosystem*. New York, Plenum Press.
- Daly, Herman. E. (1997). *Beyond Growth: The Economics of Sustainable Development*. Boston, Beacon
- Dewar, Roderick “Information Theory Explanation of the Fluctuation Theorem, Maximum Entropy Production and Self-Organized Criticality in Non-Equilibrium Stationary States”, *Journal of Physics A: Math. Gen.* 36 #3 (2003), pp. 631-641.
- Freire Vasconcellos, Marusa (2009) “Social Economy and Central Banks: Legal and Regulatory Issues on Social Currencies (social money) as a Public Policy consistent with Monetary Policy” *International Journal of Community Currency Research* Vol 13 (2009) pp.76 – 94
- Friedman, Milton (1953). "The Case for Flexible Exchange Rates". In *Essays in Positive Economics* (pp.157-203). Chicago: University of Chicago Press.
- Goerner, Sally J., Bernard Lietaer, and Robert E. Ulanowicz. (2009). Quantifying Economic Sustainability: Implications for free enterprise theory, policy and practice. *Ecological Economics*, 69(1), 76-81.
- IMF (2012) Jaromir Benes and Michael Kumhof *The Chicago Plan Revisited* (IMF Working paper 12/202 (August 2012).
- Keynes, John Maynard. (1936). *The General Theory of Employment, Interest and Money* London: Macmillan. P. 159.
- Laevan, Luc and Valencia, Fabian, 2010, Resolution of Banking Crises: The Good, the Bad, and the Ugly, IMF Working Paper 10/146 (Washington: International Monetary Fund).
- Lietaer, Bernard, Robert E.Ulanowicz, and Sally J.Goerner. (2009). Options for Managing a Systemic Bank Crisis. *Sapiens*, 2 (1). Available online at <http://sapiens.revues.org/index747.html>
- Odum, Eugene. P. (1953). *Fundamentals of Ecology*. Philadelphia: Saunders.
- Roddiar, Francois *Thermodynamique de l'Evolution: un Essai de Thermo-bio-sociologie* (Editions Parole, 2012)
- Rösl, Gerhard. (2006) *Regional Currencies in Germany: Local Competition for the Euro?*. Discussion Paper, Series 1: Economic Studies, No 43/2006, Deutsche Bundesbank Eurosystem. Available for download at http://www.bundesbank.de/download/volkswirtschaft/dkp/2006/200643dkp_en.pdf
- Stodder, James. (1998). Corporate Barter and Economic Stabilization. *International Journal of Community Currency Research*, 2.
- Stodder, James. (2000). “Reciprocal Exchange Networks: Implications for Macroeconomic Stability”. *Conference Proceedings, International Electronic and Electrical Engineering (IEEE)*, Engineering Management Society (EMS), Albuquerque, New Mexico. An updated version (2005) is available at http://www.rh.edu/~stodder/Stodder_WIR3.htm
- Stodder, James. (2009). Complementary Credit Networks and Macro-Economic Stability: Switzerland’s *Wirtschaftsring*. *Journal of Economic Behavior and Organization*, 72, 79–95. Available for download at http://www.rh.edu/~stodder/BE/WIR_Update.pdf

- Ulanowicz, Robert E. (2009). *A Third Window: Natural Life beyond Newton and Darwin*. West Conshohocken, PA: Templeton Foundation Press.
- Ulanowicz, Robert.E., Sally J. Goerner, Bernard Lietaer, and Rocio Gomez. (2009). Quantifying sustainability: Resilience, efficiency and the return of information theory. *Ecological Complexity* 6(1):27-36.
- Werner, Richard A. (2012) *How to end the European crisis – at no further cost and without the need for political changes*. Southampton, GB, University of Southampton, 12pp. (Centre for Banking, Finance and Sustainable Development Policy Discussion Paper, 2-12).
- Werner, Richard A. (2014) Enhanced Debt Management: solving the eurozone crisis by linking debt management with fiscal and monetary policy. *Journal of International Money and Finance*, 1-27. (doi:10.1016/j.jimonfin.2014.06.007).

NOTES

- ⁱⁱ Available for download on https://www.researchgate.net/publication/222401950_Quantifying_sustainability_Resilience_efficiency_and_the_return_of_information_theory
- ⁱⁱ What we have defined here as “Efficiency” is a quantitative measure of Yang; and “Resilience” is a quantitative measure of Yin. As far as we know, this is the first time that Western science is proving the validity of the Chinese insight about the need for a balance between Yin-Yang polarities. The origin of this insight has been traced back to prehistoric Siberian shamanism, and was fully developed by Lao Tse in the 5th century BC. If we are using this Yin-Yang vocabulary, at the risk of appearing exotic, it is simply because we don’t have any equivalent words in any of our Western languages.
- ⁱⁱⁱ The 4 dimensions are identified in the graphs as respectively: sustainability (F) , diversity (n), interconnectivity (c) and relative emphasis on efficiency (beta). These 3D graphs were graciously provided by Maria-Lovisa Amundadotir
- ^{iv} The type of dynamics are indeed universal for any complex flow network. However, it is possible that the parameters of the window of viability may be different depending on what flows in the system. If we could have access to real-life data, these parameters would be able to be determined empirically. The necessary data exist in both electrical distribution systems and in the banking system. Because of their competitive relevance, we have not been able to obtain such data. We would be happy to collaborate with anybody who has access to such information...
- ^v “Liquidity” and “Depth” of a financial market refers to the possibility of moving large volumes of money without significantly affecting prices. In a deep market, a lot of people are buying and selling. By contrast, in a thin market, because fewer people are trading, even one single sufficiently large transaction could significantly affect prices.

Aphastory for Google Glass

Krzysztof Szklanny, Marcin Wichrowski, and Alicja Wiczorkowska

Abstract— Aphasia is a partial or total loss of the ability to articulate ideas or comprehend spoken language, resulting from brain damage due to injury or disease, in the person whose language skills were normal. The applications dedicated for aphasiac persons and designed for wearable devices, especially Google Glass, are still underexplored topic [1]. Our previous experience in application design for aphasiac persons and in application prototyping for Google Glass [2] inspired us to implement a project combining these two domains, and prepare a storytelling application for Google Glass. The main goal of this work was to test to what extent wearable devices like Google Glass can support aphasiac persons in the future in their rehabilitation and in improving the quality of their lives. Conclusions from the experiments performed are presented in the concluding section of this paper.

Keywords—aphasia, wearable devices, Google Glass.

I. INTRODUCTION

APHASIA is most often caused by stroke, but a mechanical injury to the brain or the head can also cause it. Aphasiac persons suffer from a sudden inability to naturally use language.

Interpersonal communication is one of the most important factors of social life, and it usually takes a verbal form. Meetings and phone conversations are good examples of verbal communication, but written or graphical forms are also used. As a result of an impairment of language, aphasiac persons have great difficulty communicating verbally. This influences negatively their everyday lives, especially when interaction involving speaking and speech understanding is needed. This significantly degrades the quality of their lives [3]. Communication ability impairment causes isolation; the person is withdrawn and suffers from increasing frustration and depression because of losing social bonds [4]. Aphasiac persons cannot easily reconcile themselves to the fact that their communication skills are far below the level before aphasia.

There exist devices dedicated to aphasiac persons, namely Augmentative and Alternative Communication (AAC) devices,

This work was partially supported by the Research Center of Polish-Japanese Academy of Information Technology, supported by the Ministry of Science and Higher Education in Poland.

K.. Szklanny is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (corresponding author, phone: +48 22-58-44-500; fax: +48 22-58-44-501; e-mail: kszklanny@pjwstk.edu.pl).

M.. Wichrowski is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (corresponding author, phone: +48 22-58-44-500; fax: +48 22-58-44-501; e-mail: mati@pjwstk.edu.pl).

A. Wiczorkowska., is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (e-mail: alicja@poljap.edu.pl).

helping in basic communication, and then later in rehabilitation. However, these devices have numerous limitations, for instance symbols the aphasiac persons have to learn, in order to use the device in a very simple communication, regarding their basic needs [5]. This is why most of such devices are useful only in the beginning phase of aphasia. Aphasiacs are aware of these limitations and they lack of support in storytelling and sharing events of their everyday lives.

II. RELATED WORKS

Storytelling applications are a possible solution to the problems aphasiacs suffer from. These applications consist in creating simple stories, made of photographs or movies taken by the user. The pictures or movies can be marked with additional symbols or icons, and shared with other users [5]. Such a communication, based on stories created by the users, plays many roles, but first of all it stimulates the aphasiac person and encourages her or him to be more creative in keeping her or his social bonds. The use of the user's photographs allows remembering past events, sets them in time line, supports narration, and rebuilds social relationships. This is why well-designed storytelling application can directly improve the aphasiac's mood, reduce the feeling of social isolation and the passive attitude towards her or aphasia [4, 6]. There exist storytelling applications that can be used for this purpose.

The Storehouse application [7] allows creation of everyday life stories using text, photographs, and video. This application uses very small space for icons controlling the use of the application. The icons are shown in a bar in the lower part of the interface. When a story is selected, the icons become hidden. Choosing a picture or a movie launches a pleasant and aesthetic transition to displaying the selected material in full screen. In order to go back to the previous display, the user simply has to touch the screen. The interaction within this interface is intuitive, and the use/adjustment of settings is not necessary.

Glossaic [8] is a social network service integrated with Google Glass. After registering and launching the service, the user can share pictures or movies directly from Google Glass. The pictures published can also have descriptions of their content. The users can rate materials created by the others (through messages, or they can like them) and the materials of highest rates can be placed on the main site of the service. The virtue of this service is its full integration with the Google Glass device, as no installation of the application is needed,

and the use of this service is intuitive. The drawback of this service is the lack of the possibility to edit the pictures, the required registration of the device, and the necessity to log on.

Memoirs [9] is an application integrated with Google Glass. After activating, voice comments can be created and placed on the official service web site for download. The application also allows saving pictures and placing them in this web site. The goal of this application is to document everyday life (a diary) using Google Glass. The virtue of this application is its integration with Google Glass (no installation needed), the possibility to create and save comments and pictures, and downloading them from the web site. The lack of comment editing is the drawback of this application.

III. GOOGLE GLASS

Google Glass [10] is a development platform based on the Android 4.4.2 mobile operating system. The device itself is constructed similarly to other mobile devices (phones, tablets, etc.), but taking its specific requirements into account. The hardware architecture is based on a SoC solution OMAP 4430, containing a 2-core Cortex A9 1-1.2 GHz CPU, a PowerVR SGX540 GPU and 2 GB of RAM. The device uses a 16GB, non-removable Flash memory for non-volatile storage. It has a camera capable of recording 720P video and taking 5MP images, an audio interface that uses a bone-conduction speaker built into the frame, and also WiFi and Bluetooth connectivity. The screen is displayed on a small prismatic crystal using a 640x320 resolution, with a viewing angle corresponding to a 25'' screen viewed from a distance of 2.4 meters. Apart from the voice interface, the main interaction device is a touchpad built into the side of the frame, with an estimated resolution of 1366x187 pixels and multi-touch capability detecting as much as 5 simultaneous touches. Just like many other mobile devices, it contains a 3-axis gyroscope, accelerometer, magnetometer, light sensor and a proximity sensor that probably also doubles as a blinking detector (which is not documented anywhere). The device does not have a GSM or a GPS module, but it uses a Bluetooth paired mobile device to seamlessly access these features when necessary.

The software development for the Google Glass platform is very similar to other Android devices, but given the UI (user interface) limitations of the device, the programming library was severely modified and exists as a separate SDK, called Glass Development Kit (or GDK). The major changes include application presentation layer (using cards in a timeline instead of regular activities), changes in layout (due to limitations of the screen), user interaction (lack of keyboard) and voice interaction (this was rather new for Android devices at the time).

IV. APPLICATION CONCEPT AND ITS IMPLEMENTATION

The AphaStory application concept is based on observations and experience gained by one of the authors, Marcin Wichrowski, while he was designing and implementing his

storytelling application Aphasias Create. This application allows:

- Free drawing and taking notes,
- Taking photos with front and back camera,
- Accessing camera roll for loading and saving images,
- Reading QR codes for opening websites in a mobile browser,
- Accessing websites in a mobile browser,
- Adding text with full OS system features (spell checking, speech synthesis etc.),
- Adding current time and date,
- Accessing maps with Geolocation,
- Saving created images to camera roll and printing,
- Publishing canvas to Facebook with easy to use dedicated interface and sending emails.



Fig. 1 Exemplary story from Aphasias Create

An example of a story created using Aphasias Create is shown in Figure 1.

We decided to use some of these functions with Google Glass and address the limitations of this device. Because of great communications problems of aphasiacs, we decided to acquire the knowledge needed from experts working with aphasiacs, and from the literature [5, 6]. As a result we created the AphaStory application for Google Glass. It is a prototype, and the following functions have been implemented, among others:

- Creation of a story, built of several consequent pictures or movies,
- Adding pictures to the history using a built-in camera or an external device,
- Adding video to the story, using the built-in camera, or from a gallery in this device,
- Adding text descriptions to pictures or movies (simple sentences),
- Adding predefined emoticons to pictures or movies,
- Adding geolocation to pictures or movies,
- Sharing a story via email.

Interaction with the application is performed through tap

and swap gestures on the device’s touchpad. Voice interaction was not implemented, for obvious reasons.

The stories are created in the form of a show of consequent pictures/movies, which can be accompanied with predefined phrases and icons, helping in narration. Geotagging is also available for each story element. Such a package can be sent

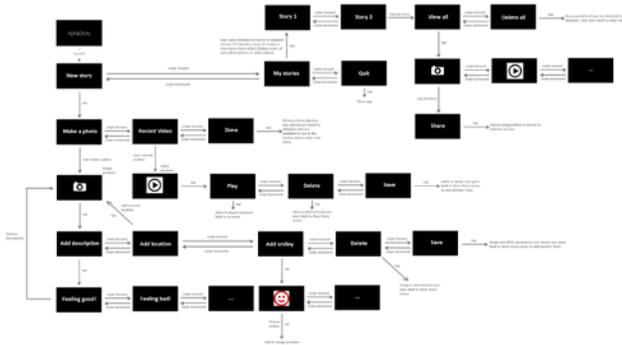


Fig. 2 Workflow of Aphastory

via email. The detailed workflow of this application is shown in Figure 2.

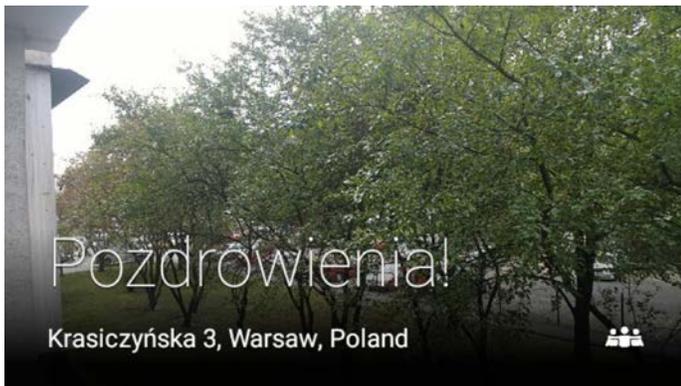


Fig. 3 Workflow of Aphastory

Figure 3 presents a story fragment, consisting of a picture, description, icon, and geolocation. The application is prepared for Polish users, so the interface and text elements are all in Polish.

The main problem encountered while working on this application was the interface of the device. The interface does not allow freely designing the screens, and standard position and size of icons in the screen had to be used.

V. EVALUATION BY PHYSICIANS

The Aphastory application for Google Glass was presented as a prototype to a physician, namely a neuropsychologist specializing in aphasia, in order to hear his medical opinion.

After introductory presentation of Google Glass, the physician used the application for 15 minutes, to check its usability and test how it works in practice. We keep cooperating also with aphasiacs, and we are looking forward for their opinions about the application.

VI. SUMMARY

The Aphastory application offers many functions, which are easy to use, and it allows sharing stories. The main difference between Aphastory and other applications is that the user is not required to record voice comments, he/she has predefined choices (to reduce confusion), and simplification of the application. Aphastory also offers a personal gallery with documented materials, available without using external devices. Other applications additionally required registering the Google Glass device.

Additionally, we are planning a simple usability test, with participation of aphasiac persons, in order to have the application tested in practice by the target users. However, the Google Glass device is not popular, and its interface is new for the users, so it can be a problem for the first time users. Early experiments with Aphasia Create for tablet computers proved that the main problem is that the users were not familiar with portable devices. This is why young users more easily used this application, as they used portable devices in their everyday lives. For this reason, persons familiar with new technologies and interfaces will be selected for further usability tests. The test will be designed as a quality test, in order to find most difficult parts of the interface, resulting in its potentially low usability, and also to get the users’ opinions about the application for Google Glass. All proposals of adding new functions to the application will be welcomed. The observations from the initial test will be taken into account in the next versions of the application. Especially, we would like to improve the way of creating stories, in order to prepare effective and intuitive method for reconstructing everyday events. We would also like to observe the use of our application for longer time, to see how the aphasiac gets used to the application and the interface.

We hope that our observations can also help developing applications for other special needs users, including persons suffering from dementia or Alzheimer disease.

ACKNOWLEDGMENT

The authors would like to thank Michał Jaworski for his help in implementing the Aphastory application.

REFERENCES

- [1] A. Roper, J. Marshall, S. M. Wilson, “Assessing technology use in aphasia”, Proceeding ASSETS '14 Proceedings of the 16th international ACM SIGACCESS Conference on Computers & Accessibility, ACM New York, 2014 pp. 239-240
- [2] M. Wichrowski, D. Korżinek, K.Szklanny, “Google Glass Development in Practice: UX Design Sprint Workshops”, MIDI Conference, Warsaw, Poland, 2015

- [3] R. Chapey, "Language intervention strategies in aphasia and related neurogenic communication disorders", Lippincott Williams & Wilkins, Philadelphia, 2001
- [4] J. Boyd-Graber, S. Nikolova, K. Moffatt, K. Kin, J. Lee, L. Mackey, M. Tremaine, M. Klawe, "Participatory Design with Proxies: Developing a Desktop-PDA System to Support People with Aphasia", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, 2006
- [5] A. Al Mahmud, J. Martens, "'I Have Something to Say': Supporting Aphasics for Organizing and Sharing Personal Experiences by Photos". Proceedings of the 12th IFIP TC 13 international Conference on Human-Computer interaction: Part I (Uppsala, Sweden, August 24 – 28, 2009) Lecture Notes in Computer Science, vol. 5726. Springer-Verlag, Berlin, Heidelberg, pp. 81-84
- [6] K. Williams, K. Moffatt, D. McCall, L. Findlater, "Designing Conversation Cues on a Head-Worn Display to Support Persons with Aphasia", Proceedings of CHI 2015
- [7] Storehouse [Online]. Available: <https://www.storehouse.co/>
- [8] Glossaic [Online]. Available: <http://www.glossaic.com/home/index.php>
- [9] Memoirs [Online]. Available: <http://www.hensonism.com/memoirs-for-google-glass/>
- [10] Google Glass [Online]. Available: <https://www.google.com/glass/start/>

Krzysztof Szklanny obtained his PhD. in 2010. He is the head of the photography laboratory at the Multimedia Department at the Polish-Japanese Academy of Information Technology. He published about 20 papers including conference papers, journal papers, and book chapters. He is also an author of high-speed photographs, including commercial photos, exhibited in the Polish-Japanese Academy of IT.

Marcin Wichrowski is a PhD. candidate at the Multimedia Department at the Polish-Japanese Academy of Information Technology. He published many papers, especially about usability engineering.

Alicja A. Wieczorkowska obtained her PhD. in 1999 and DSc. in 2012. She is the head of the Multimedia Department at the Polish-Japanese Academy of Information Technology. She published about 90 papers including conference papers, journal papers, and book chapters. She is also the author of the handbook on multimedia (in Polish).

Design of an intelligent system for disasters management

Theodora Dumitrescu, Razvan Popescu, Radu Dobrescu

Abstract— The paper proposes an architecture for an Intelligent System for Disaster Management. It is envisioned as a Multi-Agent System. By including a Model Integration component to form a hybrid system, it aims to offer support for an as wide as possible range of decisions.

Keywords— Decision Support System, Disasters Management, Intelligent System, Multi Agent System.

I. INTRODUCTION

DISASTERS management is a combined term encompassing all aspects of preparation for and responding to disasters, including prevention, mitigation, preparedness, response, and recovery [1]. The coordination of these emergency operations must be performed by a command center which needs to efficiently allocate the available resources, communicate information and take decisions regarding the planning and execution of the operations. There are different Decision Support Systems (DSS) developed for various categories of disasters and these systems are based on specific models. Due to different decision support needs that arise in disaster management area, one single model is not sufficient to cope with all of them. The first objective of this paper is to present a framework for a hybrid DSS model, which integrates different DSS models and propose the adaptation to a given scenario. Based on this model, intelligent techniques will be used to improve disaster management processes such as monitoring, controlling, and decision making. Therefore the second objective is to propose the architecture of an Intelligent System for Disasters Management (ISDM) able to offer support for various decisions.

The work has been funded by the Sectoral Operational Programme Human Resources development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU 187/1.5/S/155536.

Theodora Dumitrescu is PhD student at the Doctoral School of the Faculty of Control and Computers, Politehnica University of Bucharest, in the field of Systems Engineering (corresponding author: phone: +40214029105; e-mail: theodora.dumitrescu@gmail.com).

Razvan Popescu is PhD student at the Doctoral School of the Faculty of Control and Computers, Politehnica University of Bucharest, in the field of Systems Engineering (e-mail: codecrz@gmail.com).

Radu Dobrescu, PhD, is professor in the Department of Automatic Control and Industrial Informatics of the Faculty of Control and Computers, Politehnica University of Bucharest, (e-mail: rd_dobrescu@yahoo.com).

II. MODEL INTEGRATION

Model integration can be described as a way of developing decision models from existing models, by adapting a specific paradigm according to a given disaster situation. It produces a composite model, developed by merging or combining two or more models. To address this problem, we propose a dynamic integrated model which is based on a group of subroutines selected by an intelligent technique. Such a group, based on a specific disaster scenario, can be considered as a dynamic integrated model for disaster management decision support systems.

In order to provide effective decision making, modularity has been suggested as one of the possible solutions to the problems in developing decision support systems for disaster management [2]. The design of this system offers a variety of technical and theoretical aspects such as modularity and model reusability approaches to model decomposition. In order to improve this model, we suggest considering the modular routines utilized in the integrated DSS model as agents in a multi-agent system (MAS). In this new framework, the integrated model is realized in three steps:

1. Selection of the intelligent technique proper for events representation
2. Events correlation
3. Implementation of a knowledge base with dynamic relationships between the subroutines for a particular disaster scenario and the subsequent development of the domain base.

The integrated model will represent from here an Intelligent System for Disasters Management (ISDM).

We consider event correlation to be one of the key technologies in recognizing complex multi-source events. The task of event correlation can be defined as a conceptual interpretation procedure in the sense that a new meaning is assigned to a set of events that happen within a predefined time interval. The conceptual interpretation procedure could stretch from a trivial task of event filtering to perception of complex situational patterns. The act of recognition of a new situation by the correlation procedure could be formally handled as a synthetic event, and as such, it is a subject for further correlation. The process of building correlations from correlations allows the formation of complex inter-connected processes. In Fig. 1 are shown several basic connections between different correlation processes, proposed in [3], which can be mixed to create a flexible and scalable environment for complex situation modeling.

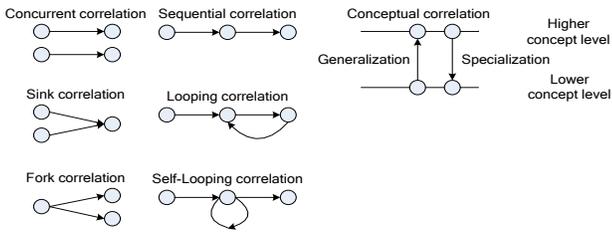


Fig. 1. Interconnections between correlation processes (after [3])

The conventional ISDM practice involves various coupled databases, event monitoring and public emergency announcement systems. During the last years significant progress was made in the development and deployment of integrated disaster information monitoring systems, and new emerging solutions of cognitive information processing, situation management, distributed computing and agent technologies has opened opportunities for new architectures for ISDMs.

III. USING A MULTI-AGENT SYSTEM FOR DISASTERS MANAGEMENT

A. Basic Principles of the Approach

The Multi-Agent System (MAS) has been widely recognized as an effective solution in modeling large number of dynamic interacting entities due to (a) the distributed organization of MAS, (b) the use of perceptual and reasoning models of mobile intelligent agents, and (c) the natural fit to model collaboration between the teams of agents. Such characteristics of MAS directly fit the requirements of ISDM. Several different architectures of MAS have been proposed, including the Belief-Desire-Intention (BDI) agent architecture [4]. Since its introduction, the BDI model has experienced several functional advancements; however especially for large-scale distributed dynamic systems this model presents some weakness, namely the lack of an adequate capability to cope with complex operational situations [5].

In the case of the proposed ISDM we focus on its cognitive aspects that require a cognitive-level MAS that is organized in a reactive situation-driven architecture, supports varying populations of agents, and scales to many interacting agent systems, where each system might have many agents. The difference between this new approach and the previous cited [3-5] is the use of a specific model of reasoning called case-oriented reasoning (COR), where each case is a template for a generic situation. The library of standard case templates that represent typical generic situations allows the construction of specific ISDM models by selecting the appropriate case templates and modifying (adapting) the selected cases with actual parameter values deduced from previous experience. In fig. 2 is represented the relation between two main processes involved in decision making, one for Situation Recognition (SR) enabled by Event Correlation (EC) which operate with the Correlation Memory and the other for Plan Reasoning (PR) driven by Case-Oriented Reasoning (COR) which operates with the Case Memory. Both processes work in a

main loop, where the primary situations recognized by EC might be refined and combined by the COR and EC might get context-sensitive meta-situations in order to proceed with the event correlation process. In case of incomplete information, EC might pass queries to event collection procedures for additional information. A secondary loop appears in the PR process, where sections of a plan can trigger an iterative deliberation process.

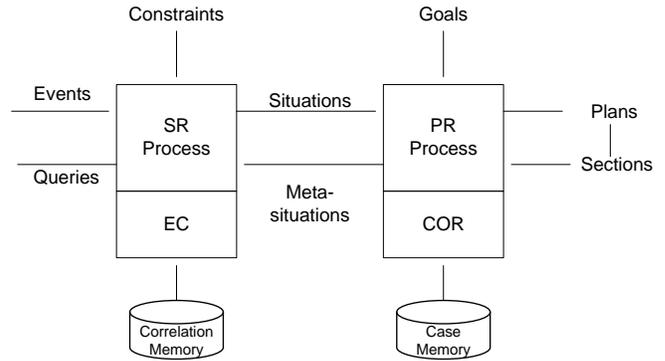


Fig. 2. Reactive relations in the decision making process

One can consider the structure in fig.2 as a situation-aware agent for Disaster Situation Management (DSM). One of the important aspects of using MAS for DSM is that the concept of an agent takes two embodiments: the physical embodiment of material resources and the virtual embodiment of software agents. Consequently, the DSM environment allows mapping the physical agents (vehicles, robots, human teams, etc.) into the abstract framework of MAS. This task involves several engineering considerations, including energy consumption, relative autonomy of physical agents, information sharing, security, etc. In particular, this architecture allows the application of the Event-Situation-Plan (ESP) paradigm, which drives invocation of a plan in a Belief-Desire-Intension (BDI) model [6].

B. BDI agent architecture

The Belief-Desire-Intension (BDI) model was conceived as a relatively simple rational model of human cognition. It operates with three main mental attitudes: beliefs, desires and intentions, assuming that human cognitive behavior is motivated by achieving desires (goals) via intentions providing the truthfulness of the beliefs [7].

Beliefs are the knowledge about the managed operational space (the World) that the agent possesses and believes to be true. Beliefs could be specifications of the World entities, their attributes, relations between entities, and states of the entities, relations. In many cases, the agent's beliefs include the knowledge about other agents as well as models of itself.

Desires are agent's motivations for actions. Two kinds of activities are associated with the desires: (a) to achieve a desire, or (b) prove a desire. In the first case, by applying a sequence of actions the agent wants to reach a state of the World, where the corresponding desire formula becomes true,

while in the second case, the agent wants to prove that the World is or isn't in a particular state by proving that the corresponding belief formula is true or not. Often desires are called goals or tasks.

Plans are operational specifications for an agent to act. An agent's plan is invoked by a trigger event (acquisition of a new belief, removal of a belief, receipt of a message, and acquisition of a new goal). When invoking a plan, an agent tests whether the plan invocation preconditions are met, and tests run-time conditions during the plan execution. Actions could be external ones, essentially procedure calls or method invocations, or internal ones of adding and removing of beliefs. Abstract plans are stored in the agent's plan library.

Intentions are sequences of instantiated plans that an agent is committed to execute. Always while responding to a triggering external event, an agent is invoking a plan from the plan library, instantiating it and pushing into a newly created stack of intentions. Contrary to that, when an agent responds to an internal triggering event, i.e., an event created by an internal action of some previous plan instance, then the new plan instance is pushed onto the stack of the previous plan that caused the invocation of the new plan instance.

The architecture of a BDI agent is presented in Figure 3.

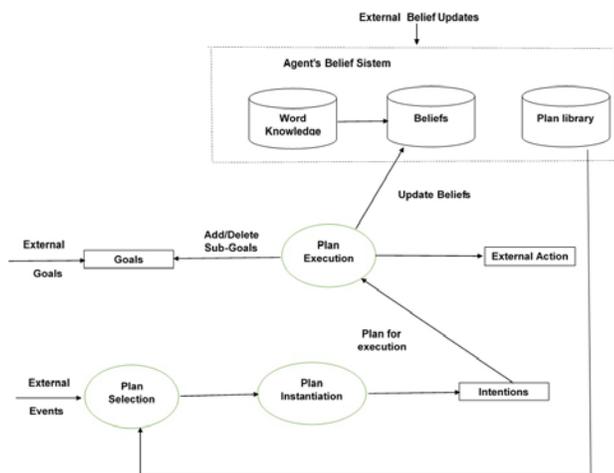


Fig. 3. Architecture of a Situation – Aware BDI Agent

IV. ISDM CONCEPT AND DESIGN

A. Main Tasks of ISDM

The major activities with decision-making needs (shown inside brackets) in disaster management are as follows:

- Hazard assessment (vulnerability analysis, frequency of hazard occurrences)
- Risk management (analysis of disaster risks, evaluating risks and treating risks)
- Mitigation (developing mitigation plan, analysis of measures)
 - Preparedness (planning and resource management)
 - Response (emergency response plans, analysis and evaluation)
 - Recovery (assessments, re-settlement issues)

In order to offer the possibility to combine and adapt different strategies, we decided to build our Intelligent System for Disaster Management (ISDM) using a Decision Support System (DSS) able to help decision-makers by cooperative work of several intelligent agents, included in a multi-agent structure. Multi-agent systems are ideally suited to representing problems that have multiple problem solving methods and multiple perspectives. Intelligent agents take initiative where appropriate, and socially interact, where appropriate, with other artificial agents and humans in order to complete their own problem solving and to help others with their activities.

The most important responsibilities of the agents involved in Decision Support for disaster management usually are:

- Monitoring: observe the environment and detect problematic behaviors;
- Alarm generation: raise alarms if there is a critical situation;
- Warning: warning respecting undesired consequences of “bad” actions and potentially suggesting better ones.

The specific tasks for the implementation of the above mentioned activities must be carefully designed. In methodologies that go back to the knowledge engineering field, a task is usually conceived as an abstract description of how the world (or an agent's “mental model” of it) needs to be transformed in order to achieve a desired behavior or functionality. To generate answers for the different classes of actions in our management framework, we have identified four essential tasks:

- Problem identification: From the analysis of the information received from a communication infrastructure or directly from the operator, the classifier chooses the state of the monitored system;
- Diagnosis: The presence of unacceptable events or situations requires an explanation in terms of causal features of the situation.
- Action planning: Once a problem has been identified, a possible sequence of actions applicable on the causes may be established.
- Prediction: The consequences of events and operator actions are simulated.

Let us consider a set of system components S , a set of external events E and a set of operator actions A .

By combining the above tasks in different manners, several questions that a decision maker typically faces can be answered. For instance:

“What is happening in S ?” represents: *problem identification + diagnosis*.

A diagnosis D for some potential malfunction is produced: “What to do on D in S ?” represents: *action planning + prediction*

Decision options are shaped and their potential effects evaluated:

“What may happen if E in S ?” represents: *prediction + problem identification + diagnosis*

Potential future problems in evolution of the system are identified:

“What to do if E in S ?” represents: *prediction + problem identification + diagnosis + planning*.

A common way of dealing with these issues is to conceive the ISDM itself as a multi-agent system, where each distributed entity is controlled by an agent. Hence, any of the aforementioned tasks of problem identification, diagnosis, action planning and prediction can be performed locally by each agent within the multi-agent system. These local tasks will be of less complexity, but they are also interdependent. The co-ordination task that such a multi-agent system faces, refers to the management of these dependencies between local tasks.

B. Specific Methods for ISDM

Most knowledge-oriented methodologies make use of the concept of problem-solving methods in order to cope with tasks. In particular, such methods indicate how a task is achieved, by describing the different steps by which its inputs are transformed into its outputs. The problem-solving process associated to a task is structured as follows: each of its steps may set up several subtasks, which again are to be solved by simpler methods and so on, until some elementary tasks can be achieved directly.

Problem identification methods

A classification method with two options may be applied:

- Identification of a reference situation and classification of the differences between the reference and the current situation.
- Direct classification of the current situation based on a predefined taxonomy where problems of different types are described.

The first approach requires: (1) to infer from the current situation the evolution of parameters consistent with the functional and structural constraints which optimizes a collection of predefined criteria and (2) to classify the differences between the observed situation and the resulting class of situations according to a hierarchy similar to the one previously commented.

For the first subtask, the method is applied in two steps. The first step derives a possible new state from the current situation that may be supported by an ad hoc procedure, adapted to the characteristics of the domain model. For the second subtask a primary representation based on rules and/or frames may be applied in a hierarchical establish & refine model. The second approach is similar to the first one, but in this case a complete description of the situation is required, not only the differences with the reference situation.

Diagnosis methods

This task infers a collection of causes explaining the problems identified by the previous one. Several methods may be directly applied:

(1) The classification method, which extends problem type frames by additional cause attributes in such a way that once a problem pattern has been selected, the cause features assumed for this problem type are assumed.

(2) A version of the cover & differentiate method [8] where a hierarchical approach to an explanatory set of causes is generated through the following reasoning steps: (i) from the attributes of the type of problem detected a collection of possible causes may be inferred covering these values; (ii)

since this first set of causes may be too large, a deeper analysis to differentiate subsets explanatory enough is necessary.

Action planning methods

After the problem identification and diagnosis tasks, some scenarios of causes of problems have been deduced together with its impacts. The action planning task must generate a consistent set of actions oriented toward the reduction or elimination of causes and/or toward the reduction of impact damages where no possible cause reduction may be produced. Specifying this task in a general way requires defining the elementary actions that will be the basis for definition of acceptable decision plans together with their models [9].

Behavior prediction methods

This task has as main goal to propose scenarios of short-term future behavior of the different components of the model. There may be specific simulation methods performing this type of task. A library could be considered to support a class of applications including a collection of typical physical components. The model of reasoning may take the current state from the information system and the assumptions about the external actions and match it with some node in the graph [10]. As a result, for every matched situation the predictable short-term changes are described by the downstream connected states.

Coordination methods

Coordination is best conceived of as the management of dependencies between activities. Methods that perform this type of management usually comprise three steps:

- Dependency detection: using domain knowledge about the different dependencies that may occur (producer-consumer relationships, resource limitations etc.) positive and negative relationships between the different local tasks of the agents are detected.
- Option generation: for every dependency, the set of possible management actions is generated.
- Management decision: finally, a decision must be taken respecting the dependency management action to be applied.

V. SYSTEM ARCHITECTURE

In a well-known publication [11], Meystel and Albus stipulate that any intelligent system consists of two parts:

1. Internal, or computational, which can be decomposed into four internal subsystems of intelligence as follows:

a) Sensor processing - inputs are provided to an intelligent system via sensors and are processed to create a consistent state of the world. Sensors are used to monitor the state of the external world and the intelligent system itself.

b) World modeling - is the estimation of the state of the world; it includes knowledge databases about the world and contains a simulation module that provides information about future states of the world.

c) Behavior generation – is the decision making module that selects goals and plans, and executes tasks.

d) Value judgment – it evaluates both the observed state and predicted state; it provides the basis for decision making.

2. External, or interfacing; input and output from the internal part of the intelligent systems are generalized via sensors and actuators that can be considered external parts.

We adapted the system architecture referenced in [11] which is based on the real control system techniques. Figure 4 shows the basic components of the ISDM.

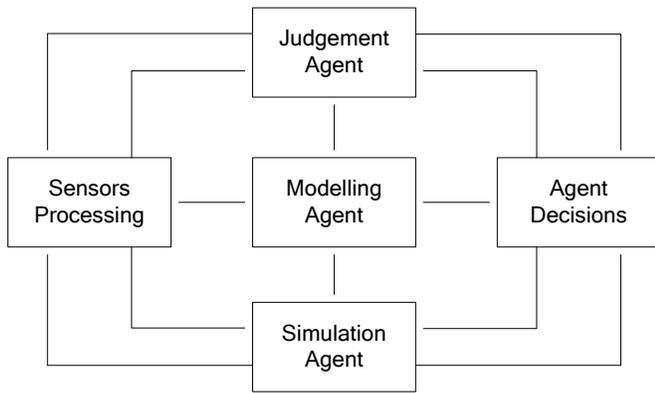


Figure 4. Block scheme of ISDM

The general diagram in Fig. 5 is constructed in concordance with the principles of a multi-agent system organization. Software agents are computational units that are repeated many times within an intelligent system at many different levels as the units of information in all of the subsystems are aggregated into entities, events, situations, and goals are decomposed into sub-goal tasks and generate actions or commands. Within each loop, data processing maintains a knowledge database with a characteristic range and resolution. At each level, plans are made and updated with different planning horizons. At each level, short term memory traces sensory data over different historical intervals. At each level, feedback control loops have a characteristic. This model of a multi-resolution hierarchy of computational loops ushers deep insights into the phenomena of behavior, perception, cognition, problem solving and learning.

The architecture of an intelligent system is a specific framework of agents and each agent has its own architecture. In the core of any intelligent system, there is also the concept of a generalized agent. Agents with similar functions can be gradually lumped in a group type agent, which basically is a generalized agent. The group agent gives a new world representation in terms of granularity or resolution. Furthermore, group agents can be aggregated into an even more generalized agent, in a hierarchical structure.

The proposed architecture includes elements of intelligence to create functional relationships and information flows between different subsystems. The elements of intelligence are based on components using one or more AI techniques: natural language processing, artificial neural networks, fuzzy logic, cellular automata (in particular Boolean Networks for solving Boolean Delay Equations - BDE).

Figure 5 depicts the proposed architecture of the intelligent assistance system, based on the integration of traditional statistical methods and various AI techniques to support a general system that operates automatically, adaptively, and proactively.

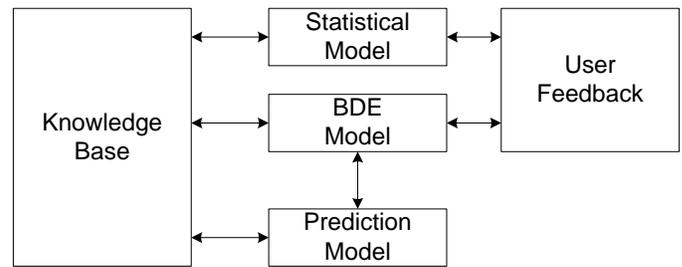


Figure 5. Intelligent model components

The system's architecture is based on a hybrid approach that yields both the robustness and depth of understanding of decision making using intelligent models. This system aims to improve monitoring and decision making processes with an effect size that is higher than a human expert. In addition, this system provides mechanisms to enhance the active construction of knowledge about threats, policies, procedures, and risks. The model is adaptive and supports processing and classification of events and data that leads to the prediction of anomalies or even extreme events. One major component in design is the development of an intelligent model for the analysis and correlation of events and data in real-time to increase the detection and prevention capabilities. The hybrid system is an integration of different models for the disaster and extreme events management to include AI techniques and other methods based on statistical and traditional procedural approach. The basic idea of the multiple models is to independently perform different functions with different measures, and to complement the weaknesses of one model with the strengths of another model.

Furthermore, the ISDM can interpret the data for human. Because the outputs of the models are uncertain and imprecise in some situation, and because human experts may have some intuition or additional knowledge on the characteristics of the presented information, ISDM could interpret the outcomes of other models in a form that humans relate better to. The system should include functions for the automation of tasks such as data collection, data reduction, filtering, and event correlation based on multi-agent technologies. The system may generate commands to end processes or move the processing to another device when signs of suspicious behavior or failures are detected.

In addition, our system is an intelligent assistant to provide feedback to the user such as help on making decisions and taking actions. The system includes a user interface based on multimedia for supporting network administrator's operations, and a knowledge base for maintaining trustworthiness as systems change and adapt. This knowledge base must be adaptive and shared via the network. The validation of the computer generated decisions can be performed by comparing with the decisions of experts. The user feedback module provides different feedback to a network administrator. The type of feedback available is important. Direct feedback entails specific information about the results and impact of each possible feedback. Indirect feedback is situated on a higher level, with no specific information about individual

change or predictions but rather proposing new strategies and systemic changes.

An important issue is to maintain the functionality of the system in situation of risk and hazards, so to provide the management of uncertainties to ensure improved design, robust operation, accountable performance and responsive risk control. Therefore, ISDM can provide the ability to adaptively assign emergency-role and permissions to specific subjects and inform subjects without explicit access requests to handle emergency situations in a proactive manner. In this aim the concepts of emergency-group and emergency dependency [12] were introduced. Emergencies are processed in sequence within the group and in parallel among groups.

VI. CONCLUSION

One can conclude that because of the complexity of information management tasks, the proposed system is based on the integration of different types of intelligent agents, i.e. a hybrid architecture under real-time constraints. That means that the ISDM can be considered a cyber-physical system which ensures a Cognition-Adaptive Human-Computer Interface for any type of mission-critical systems, in particular for disasters management. Our research based on cognitive load measurement (the BDI approach) has yielded promising outcome and validated the feasibility of ISDM. At the moment, we are working on the improvement of the software architecture and the development of other key modules.

REFERENCES

- [1] R.J.Craddock, "Crisis Management Models and Timelines", *Thales Research and Technology*, United Kingdom (2006)
- [2] S. Asghar, D. Alahakoon, L. Churilov, "A Comprehensive Conceptual Model for Disaster Management". In *Journal of Humanitarian Assistance* (2006)
- [3] G. Jakobson, N. Parameswaran, J. Buford, L. Lewis, and P. Ray, "Situation-Aware Multi-Agent System for Disaster Relief Operations Management". In *Proceedings of the 3rd International ISCRAM Conference*, pp.313-324 (2006)
- [4] J. F. Buford, G. Jakobson, L. Lewis, "Multi-Agent Situation Management For Supporting Large-Scale Disaster Relief Operations". In *International Journal of Intelligent Control and Systems*, Vol.11, No. 4, pp. 284-295 (2006)
- [5] G. Jakobson, J. Buford, L. Lewis, "A Framework of Cognitive Situation Modeling and Recognition." In *Workshop on Situation Management (SIMA 2006)*. MILCOM (2006)
- [6] M. Bratman, "Intension, Plans, and Practical Reason". Harvard University Press, 1987.
- [7] M. d'Inverno, M. Luck, M. Georgeff, D. Kinny, and M. Wooldridge, "The dMARS Architecture: A Specification of the Distributed Multi-Agent Reasoning System", in *Journal of Autonomous Agents and Multi-Agent Systems*, 9(1-2):5-53, 2004
- [8] S. Ossowski, J. Z. Hernández, C. A. Iglesias, A. Fernández, "Engineering Agent Systems for Decision Support". In *Proceedings of the 3rd Int. Conf. on Engineering societies in the agents world ESAW'02* (2002)
- [9] G. P. Alexandris and N. G. Bardis, "Comparing accuracy and solution times in coverage models", *Latest Trends on Systems - Volume II, Proceedings of the 18th International Conference on Systems*, July 2014, pp.408-413.
- [10] A. Desmet and E. Gelenbe, "Graph and Analytical Models for Emergency Evacuation", *Future Internet*, 5(1), 2013, pp. 46-55
- [11] A. M. Meystel, J. M. Albus, "Intelligent systems architecture, design, and control". John Wiley & Sons (2002)
- [12] G. Wu, D. Lu, F. Xia and L. Yao, "A Fault Tolerant Emergency-Aware Access Control Scheme for Cyber-Physical Systems", *Information Technology and Control*, vol.40, iss. 1, Dec. 2011, pp. 29-44.

Using Chroma Subsampling in Lossy Compression

P. Pokorny

Abstract—Nowadays, compression of raster image is frequently used. The reasons for this are the quite high compression ratios and also, the usually high compression/decompression speed (the success of these two major parameters of the compression algorithms depends on the image information content). In addition, the images can often allow one to use lossy compression algorithms which cause a substantially greater reduction of the image data size. The JPEG graphic format is the most widely used. Here, the loss rate is determined in two ways by the subsampling color components and the quantization of the coefficients that are obtained by the Discrete Cosine Transform calculation. In this paper, attention is focused on the subsampling of the color components. A comparison of the effects of the most frequently used sampling techniques to the quality and size of the processed image is then performed.

Keywords—Image Compression, Lossy Compression, Image Quality, Sampling, Colors.

I. INTRODUCTION

Programs using simple or complex graphics are appearing in virtually every area of computing applications: e.g. games, education, desktop publishing or graphical design, just to mention a few. These programs all have one factor in common. The images they use consume prodigious amounts of memory or disk storage space. The raster representation method is very often used for images (a display mode for the vast majority of hardware). The amount of image data depends on the resolution and color depth in this case. For example, a single image with a resolution 800 x 600 pixel and 224 color depth consumes over 1.4 MB of memory. It isn't hard to imagine applications that would require literally hundreds of these images to be accessed. [6]

Compression offers the solution to this problem. In images, its objective is to reduce the irrelevance and redundancy of the image data in order to be able to store or transmit data in an efficient form. There are two types of compression algorithms - lossless and lossy. Lossless compression is mostly used or preferred for archival purposes or technical drawings, clip art, or comics. Lossy methods are especially suitable for natural images such as photographs in applications where minor (sometimes imperceptible) loss of fidelity is acceptable in order to achieve a substantial reduction in bit rate.

A range of different lossy compression methods are used currently. One of them is to reduce color space to the most

common colors in the image (i.e. the selected colors are specified in the color palette that is included in the image file). An other method is based on transform coding. This is the most commonly used method and forms the core of the jpeg images. [8]

The third lossy compression method is chroma subsampling, which takes advantage of the fact that the human eye perceives spatial changes of brightness more sharply than those of color. [1][5] The attention of this paper will be focused on this compression method.

II. CHROMA SUBSAMPLING

A. Color Transformation

As mentioned above, the subsampling algorithms concern only the color components of the picture. The images themselves are mostly represented by the RGB or CMYK color models, or by the internal color space of CCD cameras. For this reason, the first step is to separate the brightness (sometimes called luma) and color components. This conversion process usually means a transformation to one of the models used in color analog television techniques e.g. YIQ (the PAL TV standard), YUV (the NTSC TV standard) or YCbCr (the SECAM TV standard).

This transformation between two color spaces is lossless, i.e. there is no possibility of losing any image information. So, if one converts an image into a second one with any color space used in TV techniques, it is possible to convert it back to the RGB or CMYK color model at any time and we always get a result identical to the original image. [7]

The main concern of this article is the last-mentioned color model (YCbCr), because it is used very often, for example, in the JPEG graphic format. Its transformation equation is:

$$\begin{aligned} Y &= 0,299 \times R + 0,587 \times G + 0,114 \times B \\ Cb &= -0,1687 \times R - 0,3313 \times G + 0,5 \times B + 128 \\ Cr &= 0,5 \times R - 0,4187 \times G - 0,0813 \times B + 128 \end{aligned} \quad (1)$$

The reverse conversion is performed according to the following formula:

$$\begin{aligned}
 R &= Y + 1.402 \times (Cr - 128) \\
 G &= Y - 0.34414 \times (Cb - 128) - 0.71414 \times (Cr - 128) \\
 B &= Y + 1.772 \times (Cb - 128)
 \end{aligned}
 \tag{2}$$

Figure 1 shows the original image and the same image with the separated components Y, Cb and Cr. [2] It is possible to clearly see that the most significant component is precisely Y in this picture. In contrast, the Cb and Cr components reported low intensity. For this reason, the subsampling algorithms refer only to these components.



Fig. 1 Top left: original image; Top right: the Y component; Bottom left: the Cb component; Bottom right: the Cr component [9]

B. Subsampling

There are several ways how to make the chroma subsampling process. It is possible to find them in all analog color television standards (SECAM, PAL, NTSC) in a certain form. Usually, two subsampling methods are performed - these methods are often used in the JPEG graphic format. The first method calculates the ratio of the color of the neighboring pixels that are placed on the screen in a row. The second algorithm calculates the average color of the four neighboring pixels (these pixels are placed in the two adjacent rows and columns on the screen). Both methods are lossy; when the algorithm calculates the final color component, in the first case, the size of the raster data is reduced by about 33%, in the second case by about 50%. [9]

In practice, subsampling is designated by a string of 3 (or sometimes 4) integers separated by colons, e.g. 4:2:2:4. The relationship among the integers denotes the degree of vertical and horizontal subsampling. At the outset of digital video, subsampling notation was logical; unfortunately, technology

outgrew the notation. [4] Today's notation means:

- The first number represents the horizontal sampling reference (the width of the area in pixels, which are subsampled). Usually, this value is equal to 4.
- The second number is the horizontal factor of the Cb and Cr components (i.e. number of chrominance samples in the rows of the processed area).
- The third number describes the vertical factor of the Cb and Cr components (i.e. number of chrominance samples in the columns of the processed area).
- The fourth number is like the first number. The difference is, it represents the horizontal sampling reference for the alpha channel (i.e. transparency) in the case where the image supports it. This number is not used without the transparency.

The most commonly used leading digit of 4 is a historical reference to a sample rate roughly four times the NTSC or PAL color subcarrier frequency; the notation originated when subcarrier-locked sampling was under discussion for component video. [4] The most common chroma subsamplings are:

- 4:4:4 – without subsampling. Each of the three YCbCr components has the same sample rate, like the input resolution. This scheme is commonly used in cinematic post-production and a different compression method is usually used in the following steps.
- 4:2:2 - both chroma components are each subsampled by a factor of 2 horizontally; their effective positions are coincident (co-sited) with alternate brightness. Many digital video formats and interfaces use this scheme.
- 4:1:1 - in this scheme, the Cb and Cr components are each subsampled by a factor of 4 horizontally and are coincident with every fourth brightness sample. Currently, this subsampling is used for low-end and consumer applications.
- 4:2:0 – the Cb and Cr components are each subsampled by a factor of 2 horizontally and a factor of 2 vertically. This scheme is used in JPEG/JFIF still-frames in computing, in H.261 (for video-conferencing), in MPEG-1, in consumer 576i25 DV25, and in most variants of MPEG-2. [4]

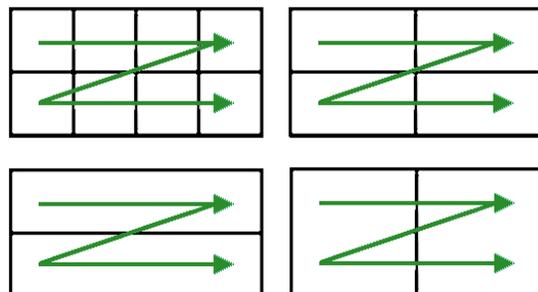


Fig. 2 Top left: subsampling 4:4:4; Top right: subsampling 4:2:2; Bottom left: subsampling 4:1:1; Bottom right: subsampling 4:2:0

The above described differences among these chroma subsamplings are shown in Figure 2.

C. A Testing Application

In order to evaluate the results of the subsampled color components, an application based on HTML5 and the Javascript programming language was created. [10] The advantage of these techniques is their simplicity and flexibility because this application can be run in any web browser.

The user interface of this application is shown in Figure 3. Its largest part is a pair of two windows that are defined by the HTML5 canvas element. In the left window, the original image is shown, while the right window represents the output image, which is generated according to the set parameters. The lower part of the user interface contains the whole control commands and parameters. The first button allows one to load any image from a file (bmp, gif, png and jpg are supported image formats). A pop-up menu (drop-down list) is to the right of this button, where the user can choose the subsampling method. The third button includes text “Calculate” and clicking the mouse on it starts a subroutine in the Javascript language, which ensures the subsampling calculation according to the selected method and then displays its result in the right window. The last button, called “Reset Output”, erases subsampling results and clears the right window.



Fig. 3 User interface of application to test the image subsampling

III. RESULTS AND DISCUSSION

To test the created application, several different images were used. The commonly used jpg graphic format was avoided because it already includes lossy compressed data and the results would be inaccurate. That is the reason why images were used, which were rendered in the Blender software [12]. The rendered images were exported to the PNG graphic format because this format uses the lossless compression algorithm. Figure 4 shows a zoomed selected part of one image - a roof with a chimney in 64 x 64 pixel resolution. The original image is in the top left corner, the

other images represent outputs created using the 4: 2: 2, 4: 1: 1 and 4: 2: 0 subsampling algorithms. The differences are not very noticeable in the full resolution of these images. So, this zoomed part of the image was created in order to see more visible differences in the various types of the subsampling algorithms. The most significant changes are logically evident after using the 4: 1: 1 and 4: 2: 0 subsampling. [11]

The subsampling process causes merging (usually by using the averaging calculation) of the color components of the neighboring pixels. In essence, this leads to the blurring effect that is more pronounced with the increasing number of merged pixels. This implies that this method is particularly suitable for photographs or rendered images which don't contain sharp edges or color gradients. In these cases, the subsampling of the color components is the most visible.

Table 1 shows the data saving for each type of subsampling. This is the total savings of the original data, i.e. the percentage numbers include all three components.

Tab. 1 Data saving using different subsampling methods

Subsampling method	4:4:4	4:2:2	4:1:1	4:2:0
Save	0%	33%	50%	50%

A comparison of Figure 4 and Table 1 evokes the idea of seeking a compromise between image quality and the quantity of image data. If the highest image quality is preferred, avoiding the losing compression algorithms including the subsampling processes is the best solution. The second thing is, if the images are intended for common user purposes, some loss rate can usually be tolerated, because the image quality can be decreased very slightly. And, Table 1 shows that the user can obtain very interesting memory storage savings.

There is a problem to define the term “acceptable image quality”. How much image information can be lost so one ever could say that image still provides sufficiently high quality image information? Classical methods that identify an image quality (expressed errors) are not applicable because their solution offers strange results. For example, a comparison of the differences between the original image and the image with the subsampled color components can often give the result that none of the pixels have the same value. So, the typical method for error measurement (i.e. the sum of the squares of the differences of color pixel values) can return high values, even where the differences between these images are almost unnoticeable. [8]

IV. CONCLUSION

Subsampling algorithms offer a very interesting option to reduce the amount of image data. Due to the fact that the loss concerning image information is the least noticeable to the human eye, this process is preceded by the conversion process, where separated brightness and color components are

separated. Then, subsampling is only used for these color components.

Currently, several different types of subsampling algorithms are used in the image processing process and - depending on its type, this leads to the corresponding compression results. Practical experience shows that the subsampling process is suitable for images with high color depths and larger resolutions; and lost colors are more visible in images that contain sharp edges or color gradients.

In order to test the subsampling algorithms, the application in HTML5 in combination with the JavaScript programming language was created. This application allows one to subsample components by using 4:4:4, 4:2:2, 4:1:1 and 4:0:0 methods. Currently, these standards are often used in some graphics images and many video formats. This application that was created could be further extended in the future. This extension can contain further subsampling algorithms including transparency or the support of different color representations (e.g. YUV, YIQ). In addition, this application could support other user-friendly features such as zoom in/out and navigation of the images, or save the output into a file.

The increased compression ratios can be further reached by using it in combination with other compression algorithms - just as the JPEG standards do. Future work will also proceed in this direction.

REFERENCES

- [1] H. Chen, S. Mingzhe and Eckehard Steinbach, "Compression of Bayer-pattern video sequences using adjusted chroma subsampling." in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, 2009, pp. 1891-1896.
- [2] Ch. Glenn, "Toward Better Chroma Subsampling Recipient of the 2007 SMPTE Student Paper Award." in *SMPTE Motion Imaging Journal*, vol. 117, 2008, pp. 39-45.
- [3] C. Hass (2008). JPEG Chroma Subsampling [Online]. Available: <http://www.impulseadventure.com/photo/chroma-subsampling.html>
- [4] Ch. Poynton (2008). Chroma Subsampling notation [Online]. Available: http://w.poynton.com/PDFs/Chroma_subsampling_notation.pdf
- [5] Prepressure contributors (2015). The JPEG file format [Online]. Available: <http://www.prepressure.com/library/compression-algorithm/jpeg>
- [6] M. Nelson, "Lossy Graphics Compression" in *The Data Compression Book*, 2 edition, Wiley, Cambridge, UK, 1995, pp 216-255.
- [7] Y. Wiseman (2013). The still image lossy compression standard – JPEG [Online]. Available: <http://u.cs.biu.ac.il/~wiseman/jpeg2.pdf>
- [8] P. Tisnovsky (2006). Lossy compression with JPEG [Online]. Available: <http://www.root.cz/clanky/ztratova-komprese-obrazovych-dat-pomoci-jpeg/#ic=serial-box&icc=text-title>
- [9] P. Tisnovsky (2006). Programming JPEG – color transformation and subsampling [Online]. Available: <http://www.root.cz/clanky/programujeme-jpeg-transformace-a-podvzorkovani-barev/#ic=serial-box&icc=text-title>
- [10] W3C contributors (2014). HTML5 - a vocabulary and associated APIs for HTML and XHTML [Online]. Available: <http://www.w3.org/TR/html5/>
- [11] Red contributors (2015). Chroma subsampling techniques [Online]. Available: <http://www.red.com/learn/red-101/video-chroma-subsampling>
- [12] Blender Foundation (2015). Blender – Free and Open 3D Creation Software [Online]. Available: <http://www.blender.org/>

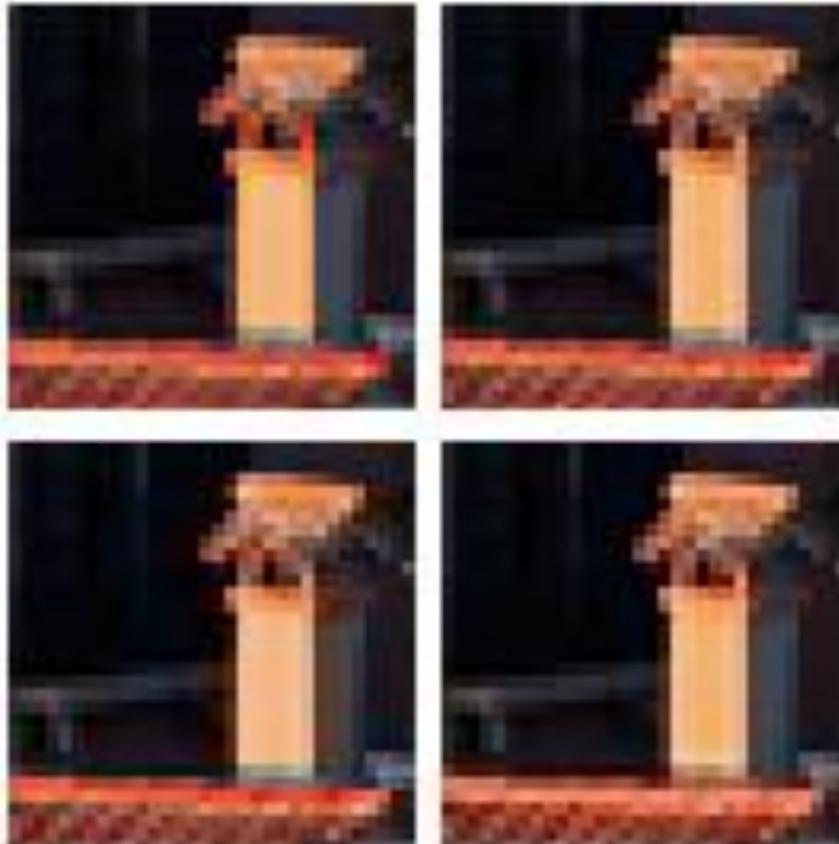


Fig. 4 Top left: original image; Top right: the 4:2:2 image subsampling; Bottom left: the 4:1:1 image subsampling; Bottom right: the 4:0:0 image subsampling

Application of Ostwald ripening to a prediction of grain mean radius in grain coarsening and coalescence

Aliki D. Muradova

School of Mineral Resources Engineering
 Institute of Computational Mechanics and Optimization
 School of Production Engineering and Management
 Technical University of Crete
 Chania, GR-73100, Greece
 Email: aliki@mred.tuc.gr

Abstract—The Ostwald ripening formula is applied to a prediction of grain mean radius in grain coalescence and coarsening. A grain mass evolution is described by a nonlinear kinetic model. The model is a system of nonlinear ordinary differential equations that determines a transport of mass between grains. Initial conditions are set up. The initial value problem is solved, numerically by means of the fourth order Runge-Kutta method. The algorithm involving mass coalescence and mean radius computation is presented. The values of mean radius on each time step of the numerical simulation are used in the least square method (nonlinear regression) for finding parameters in the Ostwald formula for best fitting. The techniques are tested for lognormal initial grain distribution, different values of activation energy and coalescence ratio.

I. INTRODUCTION

In this paper the classic Ostwald-ripening theory [1], [2] for a prediction of grain radius during coarsening and coalescence process is applied. The Ostwald ripening is an observed phenomenon in solid solutions or liquid sols that describes the change of an inhomogeneous structure over time. Namely, small crystals or sol particles dissolve, and redeposit onto larger crystals or sol particles.

Modeling of the sintering kinetics is a topic of continuing research. Recent computational approaches involve the Direct Multiscale Modeling [3], the Discrete Element Modelling [4], [5] and generalized Monte Carlo simulations [6]. Recently, a self-consistent, mean-field kinetic theory has been proposed to describe atomic diffusion in non uniform alloys [7]. This theory uses thermally activated transition rates between species and corrects the Cahn-Hilliard model [8] in the presence of non-uniformities.

Here the physical process of grain coalescence is modeled by a system of nonlinear ordinary differential equations that determines the transport of mass between grains [9], [10]. The rate of mass transfer is controlled by an Arrhenius factor leading to a nonlinear model of mass transfer and grain coarsening. The resulting dynamical system of coupled nonlinear differential equations with random initial conditions

(initial grain mass configuration) is solved, numerically by means of the fourth Runge-Kutta method (RK4).

The most important parameters in the sintering process are the mass and the arrangement of the grains. A mass can be calculated from a grain radius, whose distribution can be obtained experimentally via laser diffraction analysis, or theoretically. The exact spatial properties of the sintering process, however, are impossible to characterize.

In order to compare the outcome specific realizations of grain size distributions for all the experiments are given. Different realizations give almost similar results.

The paper is organized as follows. In Section II we present the nonlinear kinetic model with initial conditions. A grain coalescence algorithm involving mean radius computation is described in Section III. The Ostwald ripening formula is presented in Section IV. Section V focuses on comparison of the mean values of grain radius, obtained after numerical treatment of the model, with the mean values of grain, computed by the Ostwald ripening. The nonlinear regression (least square method) is used for finding the fitting parameters in the Ostwald formula. Main results are reported in Section VI.

II. INITIAL VALUE PROBLEM

A. Kinetic model

A collection of spherical grains such that the surface of each grain (particle) is in contact with at least one other particle is considered. A one-dimensional geometry in which N grains are assumed to be located along a line (the x -axis), and each grain has two neighbors, one on each side, except for the two boundary grains (Figure 1a). The grains are indexed by i ($i = 1, 2, \dots, N$). The first grain ($i = 1$) does not have a left neighbor and the last grain ($i = N$) does not have a right neighbor. The model reads as the following system of ODEs

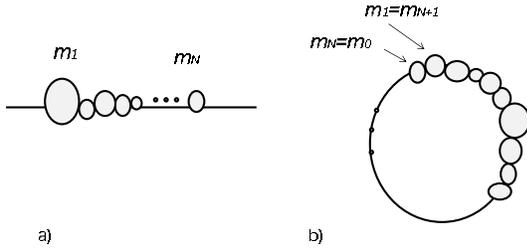


Fig. 1. Illustration of the geometry of the one-dimensional N -grain system: a) grains arranged along a straight line; b) grains arranged along a ring with cyclical boundary conditions.

([9], [10]):

$$\dot{m}_1(t) = e^{-um_2(t)}m_2(t) - e^{-um_1(t)}m_1(t), \quad t \in (0, T], \quad (1)$$

$$\begin{aligned} \dot{m}_i(t) = & e^{-um_{i-1}(t)}m_{i-1}(t) + e^{-um_{i+1}(t)}m_{i+1}(t) \\ & - 2e^{-um_i(t)}m_i(t), \quad i = 2, 3, \dots, N-1, \end{aligned} \quad (2)$$

$$\dot{m}_N(t) = e^{-um_{N-1}(t)}m_{N-1}(t) - e^{-um_N(t)}m_N(t). \quad (3)$$

In the above, $m_i(t)$ is the mass of i -th grain, N is the number of grains, T is the final time and u represents the normalized grain activation energy. The terms $\pm e^{-um_i(t)}m_i(t)$ in (1)-(3) denote transition rates, i.e. rates of getting and losing some mass, respectively. Equations (1) and (3) correspond to the mass transfer process for the boundary grains ($i = 1, N$). The first grain is in contact with the right neighbor and the last one is in contact with the left neighbor. Equations (2) describe the mass transfer process for the grains in the bulk. In the bulk each grain has two neighbors.

For the mass of i -th grain we suppose

$$m_i(t) = l_d r_i^d(t) \rho, \quad (4)$$

where $r_i(t)$ is the radius of i -th grain and ρ is the mass density of the grains (assumed to be constant). For spherical grains $d = 3$ and $l_d = 4\pi/3$.

If we assume that the grains are arranged along a ring with cyclical boundary conditions (Figure 1b) so that $m_0(t) := m_N(t)$, $m_{N+1}(t) := m_1(t)$, then the original model (1)-(3) is written in the following simpler form, whereas the final result of the simulation is not significantly affected by the change,

$$\begin{aligned} \dot{m}_i(t) = & e^{-um_{i-1}(t)}m_{i-1}(t) + e^{-um_{i+1}(t)}m_{i+1}(t) \\ & - 2e^{-um_i(t)}m_i(t), \quad i = 1, 2, \dots, N. \end{aligned} \quad (5)$$

B. Initial conditions

We assume that the initial grain masses, i.e., the initial conditions for (1)-(3) are randomly distributed numbers,

$$m_i(0) = m_{i0}, \quad i = 1, 2, \dots, N. \quad (6)$$

In order to have initial values for masses we must consider initial grain radii. We get initial radii by using normal, lognormal or some empirically determined probability distributions. The choice of the specific boundary conditions depend on the input particle size distribution of the considered real system related with sintering, crystallization or other processes. For every chosen initial distribution we set up an optimal activation parameter u which controls the simulation.

III. PROCEDURE OF A GRAIN COALESCENCE AND MEAN RADIUS COMPUTATION

The kinetic model (1)-(3) with the initial conditions (6) is discretized with using the RK4. An analogous scheme is obtained for the system of equations (5). For the numerical implementation of the RK4 scheme the Matlab ordinary differential equation toolbox, function **ode45 solver**, is used. The statistical analysis of the resulting grain distributions is based on an ensemble of simulated states.

Three regimes in behaviour of the solution of the problem can be observed. The first regime is diffusive, when the masses of grains tend to equal values. The second one is mass growing (decay), the largest grain captures the smallest one. The third regime is trapping, when the masses are almost not being changed from their initial values. A detail analysis about the thresholds (boundaries) between the regimes is given in [9].

The grain coalescence happen when the mass of two neighboring grains differ significantly, the larger grain incorporates the smaller one while the total number of grains is reduced by one (Figure 2). Our goal is to investigate a dynamics of

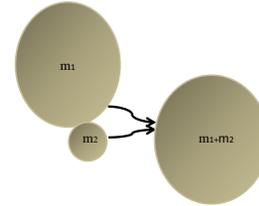


Fig. 2. Illustration of grain coalescence.

changing of mean value of grain radius. The mean values at discrete time steps are used in construction of the Ostwald ripening formula for mean grain radius prediction. The simulation consists of the following stages.

- 1) Choose an initial configuration of grain radii $r_i(t_0)$ ($t_0 = 0$) from a specified probability distribution (e.g., normal, lognormal) and set up the activation energy parameter u .
- 2) Compute grain masses $m_i(t_0)$ by the formula (4).
- 3) Put $t = t_j = \tau j$, $j = 1, 2, \dots, K$, $\tau = K/T$, $K_1 = 1$, $K_2 = K_d$, $1 \leq K_d \leq K$ and $q = 1$.
- 4) Use the RK4 (or other finite difference) formula for computing $m_i(t_j)$ for the time steps $j = K_1, K_1+1, \dots, qK_2$,

- 5) Check the neighboring mass ratios and implement grain coalescence if needed. Namely, if $m_i(t_j)/m_{i+1}(t_j) < \delta$, where $0 < \delta < 1$ is the coalescence parameter, then the $(i+1)$ -th grain becomes $m_{i+1}(t_j) + m_i(t_j)$ whereas the i -th grain is removed. If $m_i(t_j)/m_{i+1}(t_j) > \delta$ the i -th grain accumulates the mass of both grains whereas the $(i+1)$ -th grain is removed. (The check on coalescence can be done on each time step, $K_d = 1$ or in some time steps, $K_d > 1$).
- 6) Reduce the number of the equations in the system (1)-(3), or equivalently (5), by the number of grains removed according to the rules in Step 5.
- 7) Check if the updated number of grains is greater than 1: If it is, go to the next step; otherwise, the simulation terminates.
- 8) Check the number of time iterations for solving (1)-(3) or (5). If it does not exceed the specified limit K go to the next step. Otherwise, go to Step 10.
- 9) Put $K_1 = qK_2 + 1$, increase q by 1 and go to Step 4.
- 10) Compute the radii by using (4), i.e.

$$r_i(t_j) = \left(\frac{m_i(t_j)}{l_d \rho} \right)^{1/d}, \quad j = 1, 2, \dots, K. \quad (7)$$

After obtaining the radii of grains at discrete time we can compute the mean value of them

$$\bar{R}(t_j) = \frac{1}{N} \sum_{i=1}^N r_i(t_j), \quad (8)$$

where $r_i(t_j)$ is defined by (7).

IV. OSTWALD RIPENING FOR A MEAN RADIUS EVOLUTION

According to the model (1)-(3) the grain mass is changing with time and the rate of mass transfer is controlled by an Arrhenius factor, activation energy u , and also depends on the coalescence factor δ . The activation parameter controls the transition rate, although by changing initial conditions, taking them as $um_i(0)$ similar behaviour can be observed, since u can be absorbed in the initial conditions.

If $\bar{R}(t_0)$ is the arithmetic mean value of grain radius at the initial time $t = t_0$ then the Ostwald ripening formula for the mean radius evolution in grain coalescence is written as

$$R_{ost}(t)^n - \bar{R}(t_0)^n = kt, \quad (9)$$

where the parameters n and k must be found. After numerical implementation of the numerical techniques for the model (1)-(3) or (5) with the initial conditions (6) we obtain discrete values of the mean values $\bar{R}(t_j)$ by the formula (8). These values are used in the nonlinear least square method (nonlinear regression) for finding the values of the parameters $k = k(u, \delta)$ and $n = n(u, \delta)$ which provide best fitting of the Ostwald formula (9) for the mean radius evaluation. For implementation of the method we use **nlinfit** function of MatLab optimization tools, which estimates the coefficients (k, n) of the nonlinear regression function

$$R_{ost}(t) = (\bar{R}(t_0)^n + kt)^{1/n}, \quad (10)$$

or

$$R_{ost}(t) = e^{\frac{\ln(\bar{R}(t_0)^n + kt)}{n}}. \quad (11)$$

We take the discrete values of mean radius $\bar{R}(t_j)$ at the discrete time t_j , $j = 1, 2, \dots, K$ (K is the number of time steps), computed by the above described numerical procedure and use them as input data for **nlinfit**. After finding the optimal coefficients (k_{ost}, n_{ost}) and substituting them in (10) or (11) we construct the Ostwald formula with the help of which we can predict the mean radius $R_{ost}(t)$. The values $R_{ost}(t_j)$ can be compared with $\bar{R}(t_j)$, obtained as result of the numerical simulation.

For large activation energy u , coalescence ratio δ and initial masses $m_i(0)$ the trapping regime, explained in Section III, can come rapidly and an error of approximation by the Ostwald formula may be large on the whole time segment. In these cases we can consider an application of the Ostwald fitting in phases. We divide the segment $[0, T]$ into q parts $[t^0, t^1], [t^1, t^2], \dots, [t^{q-1}, t^q]$, where $t^0 = 0$, $t^q = T$ and solve the system (1)-(3) or (5) in each time segment $[t^i, t^{i+1}]$, $i = 0, 1, \dots, q-1$, where t^i is the starting (initial) time and t^{i+1} is the final time. First, we implement simulation for (1)-(3) or (5), (6) on the segment $[t^0, t^1]$ and construct the Ostwald formula for mean radius evolution, $R_{ost}^1(t)$ by using $\bar{R}(t_j)$, $t_j \in [t^0, t^1]$. The obtained values of masses at the end of the first phase of simulation, $m_i(t^1)$ will be initial conditions for (1)-(3) or (5) on the next phase of simulation. We proceed to the second phase of simulation, in $[t^1, t^2]$ in order to compute $m_i(t^2)$. The values $\bar{R}(t_j)$, $t_j \in [t^1, t^2]$ are applied for constructing $R_{ost}^2(t)$ with new k and n . We continue so on for the next time segments.

V. NUMERICAL EXPERIMENTS

In this section we give some numerical examples with fitting Ostwald ripening formula. We experiment with random, lognormal distribution.

Figure 3 shows initial lognormal distribution. The values of k and n are presented in Table I for different activation energy u and coalescence factor δ . The time step $\tau = 0.1$ and $K = 500$. The coalescence is acted on each two time steps. The parameter z in the table is the maximum of relative error at the discrete time t_j , $j = 1, 2, \dots, K$ of the Ostwald fitting formula, i.e.

$$z = \max_{1 \leq j \leq K} \frac{|R_{ost}(t_j) - \bar{R}(t_j)|}{\bar{R}(t_j)}.$$

The parameters used for lognormal distribution, i.e. grain radius distribution whose natural logarithm is normally distributed, are $\mu = 0.3$ and $\sigma = 0.7$. Figures 4-6 show the results of the simulation and Ostwald fitting.

From the figures we can see that the best fitting for the lognormal distribution is reached with $u = 0.3$, $\delta = 0.001$. A good fitting is reached for small activation energy in virtue of the chosen initial distributions. Increasing u demands decreasing δ . In general, large coalescence ratio δ brings rapid increasing mean radius in the first time steps and almost

TABLE I
THE PARAMETERS k AND n FOR DIFFERENT VALUES OF u AND δ . N SHOWS THE NUMBER OF GRAINS AT THE END OF SIMULATION.

u	δ	k	n	N	z
0.02	0.05	191.0	4.96	742	0.393
0.02	0.02	0.77	2.00	826	0.131
0.1	0.002	0.002	0.15	2941	0.052
0.2	0.001	0.025	1.03	4032	0.044
0.2	0.002	0.16	2.10	3673	0.042
0.25	0.002	0.321	2.68	3966	0.034
0.3	0.002	0.82	3.59	3566	0.031
0.3	0.005	12.36	5.79	4072	0.041
0.6	0.002	35.04	7.80	5426	0.031
1.2	0.0001	9.94	8.01	6982	0.034

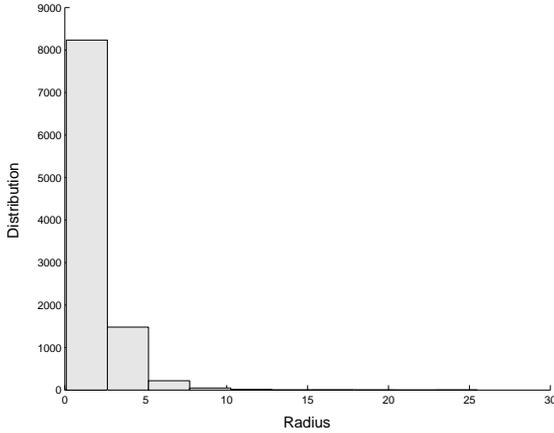


Fig. 3. The initial lognormal distribution of masses.

stationary process on the rest of time (the first rows in Table I). In these cases we use the Ostwald ripening in phases.

VI. CONCLUSIONS

A nonlinear dynamic model of mass transfer of grains has been considered. The numerical techniques involving grain coalescence with radius computation have been presented. The Ostwald ripening formula has been applied for mean radius

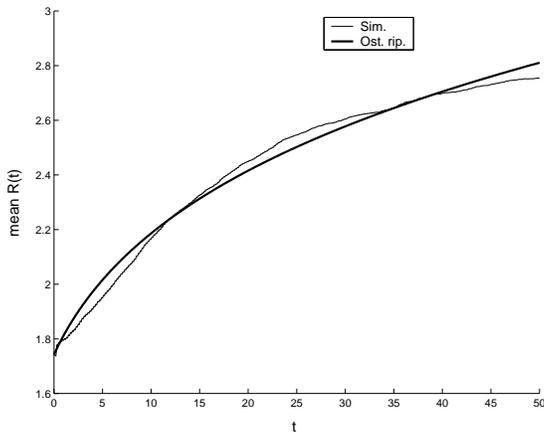


Fig. 4. Fitting with the Ostwald ripening: $u = 0.4$, $\delta = 0.002$, $\tau = 0.1$, $N = 4738$, $k = 4.46$, $n = 5.31$.

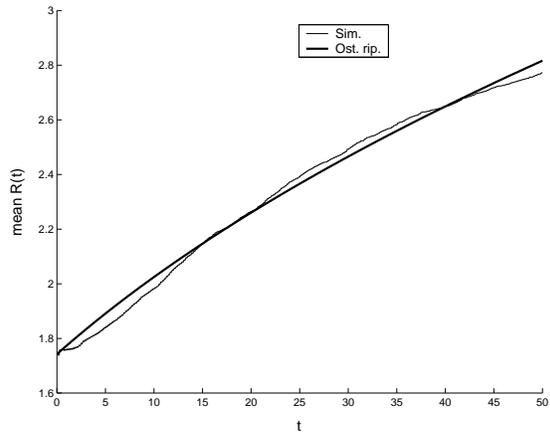


Fig. 5. Fitting with the Ostwald ripening: $u = 0.3$, $\delta = 0.001$, $\tau = 0.1$, $N = 4511$, $k = 0.17$, $n = 2.41$.

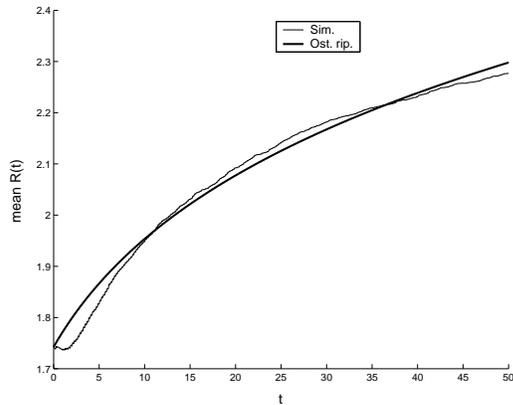


Fig. 6. Fitting with the Ostwald ripening: a) $u = 1$, $\delta = 0.0002$, $\tau = 0.1$, $N = 6598$, $k = 7.36$, $n = 7.27$.

prediction. The efficiency of the Ostwald fitting depends on the choice of the initial distribution, activation energy and coalescence factor. The best fitting is achieved for a short time consideration if the coalescence ratio is large (close to 1) and for small coalescence ratio larger time range can be taken. In general, for long time consideration and in case of non-uniform behavior of the solution the simulation is implemented in phases in order to get more accurate approximations. The procedure of computations of the Ostwald formula on phases has been described. The numerical examples have been illustrated.

ACKNOWLEDGMENT

This work has been funded by the project *NAMCO: Development of High Performance Alumina Matrix Nanostructured Composites*. NAMCO is implemented under the “THALIS” Action of the operational programme “Education and Lifelong Learning” and is co-funded by the European Social Fund and National Resources.

The author would like to thank to Prof. Dionisios Hristopulos for his valuable comments and suggestions, and Prof.

Athena Tsetsekou (National Technical University of Athens) for helpful discussions. The contributions of former students Spyros Blanas and Ioannis Kardaras in earlier numerical investigations are also acknowledged.

REFERENCES

- [1] H. G. J. H. Yao, K. R. Elder, and M. Grant, "Theory and simulation of Ostwald ripening", *Phys. Rev.*, B 47, 14110, 1993.
- [2] A. Pototsky, U. Thiele and A. J. Archer, "Coarsening modes of clusters of aggregating particles", *Phys. Rev.*, E 89, 032144, 2014.
- [3] A. Maximenko, A. Kuzmov, E. Grigoryev, and E. Olevsky, "Direct Multi-Scale Modeling of sintering", *J. Amer. Cer. Soc.*, 95, 2012, pp. 2383-2388.
- [4] C. L. Martin, L. C. R. Schneider, L. Olmos, and D. Bouvard, "Discrete element modeling of metallic powder sintering", *Scripta Mater.*, 55, 2006, pp. 425-428.
- [5] L. Olmos, C. L. Martin, D. Bouvard, D. Bellet, and M. D. Michielz, "Investigation of the sintering of heterogeneous powder systems by synchrotron microtomography and Discrete Element Simulation", *J. Amer. Cer. Soc.*, 92, 2009, pp. 1492-1499.
- [6] H. Cetinel, O. Kayacan, and D. Ozaydin, "Investigation of nucleation and grain growth in 2-dimensional systems by using generalized Monte Carlo simulations", *Physica A*, 392, 2013, pp. 4121-4126.
- [7] M. Nastar, "Atomic diffusion theory challenging the Cahn-Hilliard method", *Phys. Rev.*, B 90, 144101, 2014.
- [8] J. W. Cahn, and J. E. Hilliard, "Free energy of a nonuniform system. I. interfacial free energy", *J. Chem. Phys.*, 28, 1958, 258.
- [9] D. T. Hristopulos and A. D. Muradova, "Kinetic Model of Mass Exchange with Dynamic Arrhenius Transition Rates, *Phys. A* (submitted for publication).
- [10] A. D. Muradova and D. T. Hristopulos, "Numerical Investigation of Grain Coarsening and Coalescence Model", *J. Phys.: Conf. Series*, 574, 012160, 2015.

An intelligent diagnosis system based on extreme learning machine for diabetes diseases

R. Coteli

Abstract— The many peoples have been suffered from diabetes. The diagnosis of diabetes is very important task for physician. In this study, an expert system based on Extreme Learning Machine (ELM) is suggested to diagnose diabetes disease. The parameters of ELM such as type of activation function and the number of hidden neurons is tuned to obtain the best possible performance. For this aim, the number of hidden neurons and activation functions are determined by trial and error. Training and testing dataset used in this study are obtained from UCI. The optimum performance of ELM are achieved in hidden neurons with 50 and tangent-sigmoid activation function. The obtained results for different types of activation function and different number of hidden neurons are also presented. To validate efficiency of the proposed method, a comparison is also given with previous studies. Performance of ELM classifier is calculated by using sensitivity and specify analysis, classification performance. The classification accuracy of this intelligent diagnosis system is found as 84.61 %.

Keywords— extreme learning machine; diabetes diseases dataset; intelligent diagnosis system.

I. INTRODUCTION

THE insulin which is a hormone types converts food into energy. If this hormone in body isn't properly produced or used, diabetes disease has been occurred. There are two main types of diabetes as type-1 and type-2. Also, other type of diabetes develops only during pregnancy. Both types of the diabetes could be seen in some people [1]. The obesity is one of the most important reasons of diabetes. Although there is no exact treatment, it is possible to delay and control by means of diet, exercise and eating habits. There are many factors to make difficult diagnosis of diabetes.

In the diagnosis of diabetes, physicians commonly decides by means of blood test of patient or compares patient's symptoms in other one with same condition. Therefore, diagnose of diabetes is more difficult task for physicians [1].

There have been performed many studies on diagnosis of diabetes in literature. Various methods such as Levenberg-Marquardt, radial basis function, general regression neural network, gradient descent, generalized discriminant analysis based least square support vector machine, SVM, Semi-Naive Bayes, Naive Bayes have been proposed for diagnosis of diabetes [1-4].

The past studies in the intelligent diabetes diagnosis system areas have shown that slow learning speed is one of major

disadvantage of feed-forward neural networks. In addition, all parameters of the neural networks are tuned iteratively by means of such learning algorithms [5-9]. To solve these disadvantages of feed-forward neural networks, a new learning algorithm called Extreme Learning Machine (ELM) is suggested for single-hidden layer feedforward neural networks (SLFNs) [5-9]. In this learning algorithm, hidden nodes are chosen randomly and output weights of SLFNs are determined analytically. This algorithm has not only good generalization performance but also fast learning speed [5].

This study presents an intelligent diagnosis system using Extreme Learning Machine (ELM) to facilitate diagnose of diabetes. The proposed intelligent diagnosis system is an auxiliary tool to help physician for diagnose of disease. Extreme Learning Machine (ELM) for single-hidden layer feed-forward neural networks (SLFNs) is used to efficiently diagnose diabetes diseases. The diabetes dataset is obtained from UCI and it consists of two classes as healthy and patient. A total instances of 768 are utilized for training and testing. The compared results with previous methods on diagnosis of diabetes disease are given. The classification performance of the ELM based intelligent system is also estimated by using sensitivity, specify analysis, classification accuracy and confusion matrix.

The rest of the paper is organized as follows. Section II presents basic principle of the pattern diagnosis. Section III explains single hidden layer feed forward networks and ELM. The application of ELM for diabetes diagnosis is given in Section IV. Conclusion are presented in Section V.

II. PATTERN DIAGNOSIS CONCEPT

Fig. 1 shows basic pattern diagnosis concept. It include feature extraction and classification stages. The feature extraction stage is one of the most important part for pattern diagnosis [10, 11]. The suitable features are extracted by a feature extractor. If suitable features are not chosen, the classification performance will be poor even though using the best classifier. In addition, feature extractor should make reduction in the dimension of pattern vector. These reduced vector should also cover useful information of the original vector. The last stage covers classification. In this step, reduced feature vector is fed to inputs of classifier [12, 13].

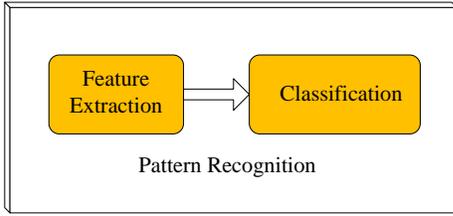


Fig.1 the block diagram of the pattern diagnosis concept.

III. SINGLE HIDDEN LAYER FEED FORWARD NETWORKS AND EXTREME LEARNING MACHINE

SLFN is the simplest kind of neural network and its architecture is given in Fig. 2. It is well-known that a SLFN having enough neurons in the hidden layers can overcome any finite input-output mapping problem [14]. However, increase in the number of neurons in the hidden layers causes some drawbacks.

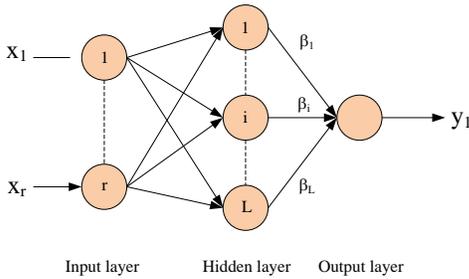


Fig. 2 structure of SLFN

The learning speed of feed forward neural networks is very slowly [5-9]. This property of feed-forward neural networks limits their application areas. In [5-9], it is explained that this behavior takes place two main reasons. One of them is slow learning speed of gradient based algorithms used in training of neural network. Other reason is the iterative tuning of the parameters of the networks in these learning algorithms. A novel learning algorithm called ELM is proposed to overcome these limitations in Ref. [5]. In this learning algorithm, the input weights are randomly selected and output weights of SLFN are analytically determined. It has both much better generalization performance and much faster learning speed when compared with conventional learning algorithms.

The output of hidden nodes for SLFN could be calculated by:

$$y = a_i \cdot x + b_i \tag{1}$$

a is weight matrix connecting hidden and input nodes, b is biases matrix, f is activation function and x is inputs.

The output 'o' of the SLFN with N-hidden neurons is calculated as follows:

$$o_n = \sum_{i=1}^N K_i \cdot (f(a_i \cdot x + b_i)) \tag{2}$$

K is weight matrix connecting hidden and output nodes. Eq.(2) can be re-written as follows:

$$H \cdot K = T \tag{3}$$

Where, T is target. In theory of ELM, input weights and biases are randomly selected and K values are calculated analytically to make zero error [15]. If H matrix is known, the K matrix can be obtained as follows:

$$K = H^+ \cdot T \tag{4}$$

Where, H^+ is Moore-Penrose generalized inverse of H matrix and can be calculated using several methods [16].

Some significant features of ELM can be summarized as follows:

- The ELM is extremely fast when compared with conventional learning algorithm.
- The generalization performance of ELM is better.
- The ELM overcomes problems encountered in conventional learning algorithm.

ELM algorithm could be applied for training of the SLFN even in case of non-differentiable activation functions. ELM randomly chooses weights between input and hidden neurons. Then, it determines analytically the weights between hidden neurons and output neurons of the SLFN.

Realization of ELM method for SLFNs can be briefly explained as follows:

1. Assign input weights and biases randomly.
2. Calculate the hidden layer output matrix H and its Moore-Penrose generalized inverse.
3. Obtain the output weights.

IV. APPLICATION OF ELM FOR DIABETES DIAGNOSIS

The block diagram of the ELM based intelligent auxiliary system is presented in Fig. 3. Diabetes dataset includes a total cases of 768. These dataset are used for training and testing of the ELM classifier. Firstly, ELM classifier are trained with dataset of 576. Then, remaining dataset are used for testing of ELM. The outputs of classifier represent disease and healthy cases.

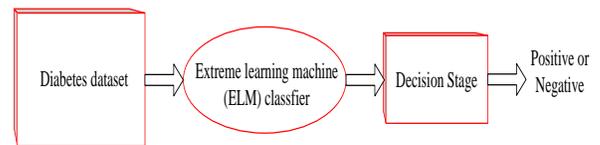


Fig. 2 the block diagram of ELM based intelligent diagnosis system

In this studies, the diabetes dataset obtained from UCI is used for training and testing of ELM. This dataset has instances of 768 and attributes of 8. The attributes also include number of the times pregnant, plasma glucose concentration a 2 hours in an oral glucose tolerance test, diastolic blood pressure, triceps skin fold thickness, hour serum insulin, body mass index, diabetes pedigree function and age. Table I shows statistical analysis of this dataset.

Class variable: 0 or 1

Class Distribution: Class value 1 is interpreted as "tested positive for diabetes"

The studies are carried out by means of MATLAB 7.7.0 software [17]. The performance of the ELM based intelligent system is evaluated with overall test accuracy obtained for diabetes diagnosis. In addition, obtained results are compared with previous studies reported in literature. In these studies, 75% of dataset is used for training of ELM based intelligent diagnosis system for diabetes whereas remaining dataset is utilized for testing of ELM classifier. As previously stated, Five hundred instances belongs to healthy class. Four hundred instances of healthy class are used for training and remaining healthy samples are utilized for testing. Similarly, 176 instances for disease class are utilized for training and 92 instances for testing.

Table I brief statistical analysis of database

Attribute number	Mean	Standard Deviation
1	3.8	3.4
2	120.9	32.0
3	69.1	19.4
4	20.5	16.0
5	79.8	115.2
6	32.0	7.9
7	0.5	0.3
8	33.2	11.8

In order to conduct a valid comparison, ELM is trained with the optimized parameters, on the same fixed training set (75 %) and evaluated on the same fixed test set (25 %). Different processes are performed to find the optimum parameters of ELM such as type of the activation function and the number of hidden layer neurons. For this aim, different types of activation functions are used to obtain the best classification performance. In addition, the number of hidden layer neurons (NHLN) are tuned increasing by step with 10 from 1 to 100 gradually.

Type of activation function and the number of hidden layer neurons are determined based on classification accuracy. Table I shows performance of ELM for different combinations of parameters in diabetes diagnosis. Further, inputs to ELM classifier have been normalized into [-1; +1]. It is shown from this table that the optimum performance is achieved in tangent sigmoid activation function and the NHLN with 50. Further, the training time of ELM is closely related to the number of hidden neurons.

As shown in Table II, training time of ELM has been increased when the number of hidden neurons is increased. In Table II:

LS: Logarithmic sigmoid activation function

TS: Tangent sigmoid activation function

S: Sinusoid activation function

CA: Classification accuracy

AF: Type of the activation function

Table II performance of ELM for different parameters

AF	NHLN	Time (s)		CA (%)	
		Training	Testing	Training	Testing
LS	10	0.0468	0.0149	79.69	76.02
	20	0.0584	0.0162	78.65	75.00
	30	0.0927	0.0168	77.60	76.50
	40	0.0936	0.0176	77.08	76.04
	50	0.0962	0.0192	77.08	76.04
	60	0.1102	0.0214	80.73	75.52
	70	0.1286	0.0232	77.60	75.52
	80	0.1302	0.0284	80.73	79.17
	90	0.1384	0.0324	80.73	72.40
	100	0.1404	0.0452	81.25	74.48
S	10	0.0468	0.0158	78.99	75.52
	20	0.0468	0.0122	80.21	77.60
	30	0.0628	0.0134	81.60	75.52
	40	0.0762	0.0232	80.73	75.00
	50	0.0842	0.0342	82.47	76.04
	60	0.0936	0.0443	84.03	76.56
	70	0.1092	0.0452	82.29	76.04
	80	0.1404	0.0462	83.33	76.56
	90	0.0936	0.0468	83.16	78.65
	100	0.0936	0.0468	83.85	75.00
TS	10	0.0468	0.0132	78.47	78.13
	20	0.0468	0.0141	81.03	80.21
	30	0.0559	0.0154	82.05	81.25
	40	0.0936	0.0168	80.73	77.08
	50	0.0974	0.0172	85.02	84.61
	60	0.0984	0.0181	80.73	80.21
	70	0.0996	0.0188	82.64	76.56
	80	0.1248	0.0254	84.03	76.56
	90	0.1248	0.0344	84.03	75.52
	100	0.1262	0.0376	84.55	79.69

The testing of the ELM classifier are realized after training. The obtained results are also compared with previous studies [2, 4, 17-19] and given in Table III. It is shown that maximum accuracy is 80.21% in the previous studies. This study achieves a high classification accuracy of 84.61%.

Table III comparison results with previous methods

Method	Accuracy %	Reference
Linear Discr. Anal.	77.5-77.2	Statlog; Ster & Dobnikar
Fisher discr. analysis	76.5	Ster & Dobnikar
MLP+BP	76.4	Ster & Dobnikar
MLP+BP	75.8	Zarndt

LVQ	75.8	Ster & Dobnikar
RBF	75.7	Statlog
BP	75.2	Statlog
Bayes	72.2	Zarndt
Kohonen	72.7	Statlog
kNN	71.9	Ster & Dobnikar
BFGS quasi Newton	77.08	(Yildirim et al. 2003)
Levenberg-Marquardt	77.08	(Yildirim et al. 2003)
GDA-LS-SVM	79.16	(K. Polat et al., 2008)
Gradient descent	77.60	(Yildirim et al. 2003)
SVM (5xCV)	77.6	Bennet and Blue
Semi-Naive Bayes	76.0	Friedman, 5xCV
Naive Bayes	74.5	Friedman, 5xCV
ELM	84.61	(used in this study)

In addition, several analysis are performed to show the efficiency of the proposed diagnosis system. The obtained results for classification accuracy, sensitivity and specificity analysis by using ELM based intelligent diagnosis system for diabetes are given in Table IV. More detail information on sensitivity and specificity analysis could be founded in Ref. [19-23].

Table IV. The obtained sensitivity and specificity results

	CA (%)	Sensitivity (%)	Specificity (%)
LS-SVM [1]	78.82	73.91	80
GDALS-SVM [1]	79.16	83.33	82.05
This study	84.61	84.67	85.33

V. CONCLUSION

This paper proposes an auxiliary intelligent system based on ELM to diagnose diabetes. The proposed intelligent diagnosis system has advantages such as fast training and testing time and generalization capability over conventional neural networks with back propagation. The experimental results show that ELM based intelligent diagnosis system for diabetes achieves a considerable classification accuracy for possible diabetes patients. It is concluded that this satisfactory results can be very helpful for final decision of physicians on patients. Thus, the physicians can decide more accurately for diabetes diagnosis by means of efficient tool. In addition, structure of diagnosis system is quite simple.

REFERENCES

[1] K. Polat, S. Gunes and A. Arslan, (2008). A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine, *Expert Systems with Applications* 34, p.p.482-487.
 [2] K. Kayaer and T. Yildirim, (2003). Medical diagnosis on Pima indian diabetes using general regression neural networks, artificial neural

networks and neural information processing (ICANN/ICONIP) (pp. 181-184), Istanbul, Turkey, June 26-29.
 [3] K. P. Bennett, J. Blue, A Support Vector Machine Approach to Decision Trees, R.P.I Math Report No. 97-100, Rensselaer Polytechnic Institute, Troy, NY, 1997.
 [4] N. Friedman, D. Geiger, M. Goldszmit (1997). Bayesian networks classifiers. *Machine Learning* 29: p.p.131-163.
 [5] G. B. Huang, Q. Y. Zhu, C. K. Siew, *Neurocomputing*, Volume 70, Issues1-3, Pages 489-501, December 2006.
 [6] G. B. Huang, Q. Y. Zhu, C. K. Siew, Universal approximation using incremental networks with random hidden computation nodes, *IEEE Trans. Neural Networks* 17 (4) 2006.
 [7] G. B. Huang, C. K. Siew, Extreme learning machine with randomly assigned RBF kernels, *Int. J. Inf. Technol.* 11 (1) (2005).
 [8] G. B. Huang, Q. Y. Zhu, C. K. Siew, Real-time learning capability of neural networks, Technical Report ICIS/45/2003, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, April 2003.
 [9] M. Leshno, V.Y. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Networks* 6 (1993) 861-867.
 [10] E. Avci and Z. H. Akpolat, (2006). Speech Recognition Using A Wavelet Packet Adaptive Network Based Fuzzy Inference System, *Elsevier Expert Systems with Applications*, vol. 31(3), p.p. 495-503.
 [11] E. Avci, R. Coteli, A new automatic target recognition system based on wavelet extreme learning machine, *Expert Systems with Applications*, Volume 39 (16), 2012.
 [12] H. Dickhous and H. Heinrich, (1996). Classifying Biosignals with Wavelet Networks. *IEEE Engineering in Medicine and Biology*, 103-111.
 [13] C.M. Bishop, (1996). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.
 [14] N. Y. Liang, G. B. Huang, P. Saratchandran, N. Sundararajan, A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks, *IEEE Transactions on Neural Networks*, Vol. 17, No. 6, November 2006.
 [15] A. Balbay, E. Avci, O. Sahin, R. Coteli, Modeling of Drying Process of Bittim Nuts (pistacia terebinthus) in a Fixed Bed Dryer System by Using Extreme Learning Machine, *International Journal of Food*, Volume: 8: Issue: 4, DOI: 10.1515/1556-3758. 2737.
 [16] Rampal Singh, S. Balasundaram, "Application of Extreme Learning Machine Method for Time Series Analysis, *International Journal of Intelligent Technol*G. O. Young, "Synthetic structure of industrial
 [17] *Matlab 7.7.0*, MATLAB Company, 2011.
 [18] K. Polat and S. Gunes, (2007). An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease, *Digital Signal Processing*, Volume 17, Issue 4, Pages 702-710.
 [19] J. Beyer, J. Schrezenmeir, G. Schulz, T. Strack, E. Küstner, G. Schulz, The influence of different generations of computer algorithms on diabetes control, *Computer Methods and Programs in Biomedicine*, Vol. 32, Issues 3-4, July-August 1990, Pages 225-232.
 [20] <http://www.phys.uni.torun.pl/kmk/projects/datasets.html> ((last accessed: April 18, 2011).
 [21] K. Polat and S. Gunes, (2007). An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease, *Digital Signal Processing*, Volume 17, Issue 4, Pages 702-710.
 [22] K. Polat and S. Gunes, (2008), Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer, *Expert Systems with Applications*, Volume 34, Issue 1, Pages 214-221.
 [23] A. Watkins. (2001). AIRS: A resource limited artificial immune classifier, Master thesis, Mississippi State University.

Simulation studies for coalescence of carbon nanotubes from graphene using controlled methods

D Fülep, I Zsoldos and I László

Abstract — In the present work we study the topological conditions for the growing of nanotubes from graphene nanoribbons using graphite substrate. For this purpose Density Functional based Tight-Binding (DFTB) Molecular Dynamics (MD) simulations were performed for producing carbon nanotubes.

Keywords — carbon nanotube, graphene, graphite, Density Functional Tight Binding method, Lennard-Jones potential, molecular dynamics simulation, substrate, van der Waals force

I. INTRODUCTION

Although the outstanding electric properties of carbon nanotubes has already been proved in several publications [1], until now only very few electric devices were presented or realized [2-4]. This fact can be explained by the lack of well controlled reliable technology for construction of either a standalone nanotube or a complex nanotube network. Nanotube construction from nanoribbons is a promising possibility [5]. Nanoribbons can be produced with the help of nanolithography [6] and various chemical compounds [7].

For the time being the accuracy of nanoribbon cutting from graphene is about few nm, and only one order of magnitude is missing to the atomic accuracy. There are studies for the instabilities at nanoribbon edges and nanotubes are obtained in molecular dynamics simulations from two nanoribbons [8]. It was demonstrated in molecular dynamics simulations that graphene patterns with atomic accuracy can develop in a self organizing way to the predetermined fullerenes or nanotubes [5, 9-11].

The importance of nanotube production from two nanoribbons comes from the fact, that in this way open ended carbon nanotubes are developed but the one pattern nanotubes are always closed at one end [9]. Coalescence of nanotube

from nanoribbons is not a simple process although the idea has already been published [8, 12]. In our previous work we obtained the topological and energetical conditions [13] for growing straight nanotube self-assembling from two superposed parallel nanoribbons [5]: For armchair nanotubes the critical ribbon width is 9.23 Å corresponding to the critical curvature energy of 0.18eV. For zigzag nanotubes we obtained the critical ribbon width of 15.99 Å (14,0) and the corresponding critical curvature energy of 0.1eV. These values concern models without any substrate, as discussed later.

In the present work we study how the presence of a substrate changes these experiences. We also made Density Functional based Tight-Binding (DFTB) Molecular Dynamics (MD) simulations. The constant temperature simulations were controlled with the help of Nosé-Hoover thermostat. Interaction between the graphene model itself and its substrate were determined by van der Waals forces.

The two nanoribbons are placed between two blocks of graphite substrates. This new type of theoretical experimental setup is closer to practical realization than previous models. Thus, the new experiences gained with the improved model are more valuable. In this work we summarize our first experiences with the new model, and determine a lower diameter nanotube which can be constructed.

II. THE METHOD

The method we used is based on our previous models but has some improvements which ensure more accuracy and helps our model approaching reality.

The interatomic interaction was calculated with the help of Density Functional Tight Binding method [14]. The nanoribbons were cut out from a graphene sheet of interatomic distance $r=1.42$ Å. The two nanoribbons are placed facing each other in parallel position. During the molecular dynamics calculation constant environmental temperature was provided [15-16]. Verlet algorithm [17] was used to calculate velocity. The initial atomic displacements during the time step of $\Delta t = 0.7$ fs were sorted randomly and they gave the initial velocities by appropriate scaling. In this scaling we supposed an initial kinetic temperature T_{init} . This initial temperature was chosen from the range of $T_{init} = 1000\text{K}$ and 1100K . We have found that the final structure was depending more strongly on the direction of the initial velocities than the actual value of T_{init} . That is by scaling of the initial temperature in the above

Dávid Fülep: Faculty of Technology Sciences, Széchenyi István University, H-9126 Győr, Hungary (fulep@sze.hu)

Ibolya Zsoldos: Faculty of Technology Sciences, Széchenyi István University, H-9126 Győr, Hungary (zsoldos@sze.hu)

István László: Department of Theoretical Physics, Institute of Physics, Budapest University of Technology and Economics, H-1521 Budapest, Hungary (laszlo@eik.bme.hu)

E-mail: fulep@sze.hu, zsoldos@sze.hu, laszlo@eik.bme.hu

mentioned range the final structure was not strongly changing. As the formation of new bonds decreased the potential energy and increased the kinetic energy we had to keep the temperature constant. In a constant energy calculation the kinetic energy obtained by forming new bonds destroyed other bonds of the structure. We used Nosé-Hoover thermostat [15-16, 18-19] for the constant temperature simulation. It is evident that in the Nosé-Hoover thermostat there is an oscillation of the temperature but it cannot destroy the structure formation. In the following the temperature of the calculation will mean the temperature of the thermostat. If the constant temperature were realized with the help of random scaling of the kinetic energy we could not distinguish the temperature of the environment and the structure. This is why we can speak about the T_{init} temperature and the temperature of the Nosé-Hoover thermostat (the environment temperature).

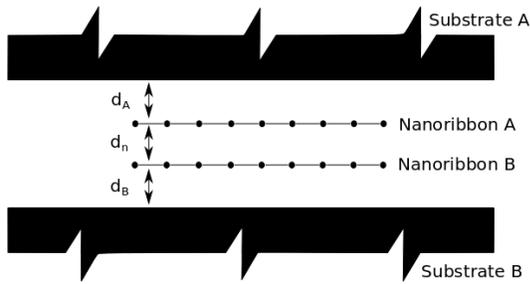


Figure 1. Initial model between two blocks of substrates.

The graphene nanoribbons are placed between two blocks of graphite substrates as we can see on Figure 1. This improvement was inspired by possible production technology and aims to achieve better control of growing perfect nanotubes. Our models contain two parallel nanoribbons of d_n distance. The distance of the model is d_A from Substrate A and d_B from Substrate B. Usually $d_A=d_B=d_n$. The long-range van der Waals interaction between the substrates and the ribbons is characterized by Lennard-Jones (LJ) term [20]. The LJ potential is as follows:

$$U_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

The parameters were calculated as $\sigma = 2.9845 \text{ \AA}$ and $\epsilon = 0.002 \text{ eV}$. Parameter r is the distance of the given model atom and the substrate atom.

On Figure 2 the Lennard-Jones potential can be seen as a function of the atomic distance.

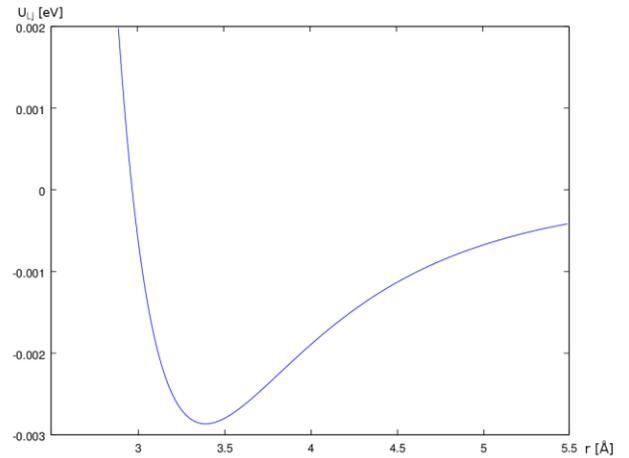


Figure 2. Van der Waals Lennard-Jones Potential (U_{LJ}) versus atomic distance (r). When the slope of the function is negative, it means repellent effect. In case of positive slope attractive force will present.

III. RESULTS

We were studying armchair and zigzag nanotubes. We examined all the cases of different orientations and nanotube diameters to know, how the existence of the substrates influences the self-organized growing of nanotubes. The initial structure contained two parallel (coincident or similar size) graphene nanoribbons $d_n=3.35 \text{ \AA}$ from each other. We also set the initial distance of the nanoribbons and the substrates the same value: $d_A=d_B=3.35 \text{ \AA}$. We calculated the interatomic forces between the carbon atoms considering repulsive and attractive van der Waals forces of the substrates.

It was expected that larger models would show similar behavior so we started creating nanotubes due to diameter of critical size and above. Figure 3 shows such a large model. Figure 3.a shows initial model and Figure 3.b shows the developed structure after 2 ps. It can be seen easily that the structure fits compressed is the two substrates. Then we started to pull apart the two substrates so the structure reached its final shape as seen on Figure 3.c. For better understanding Figure 3.d shows initial and the flattened model between the substrates. Note that on the figure only one graphite layer can be seen instead of each block of graphite substrates.

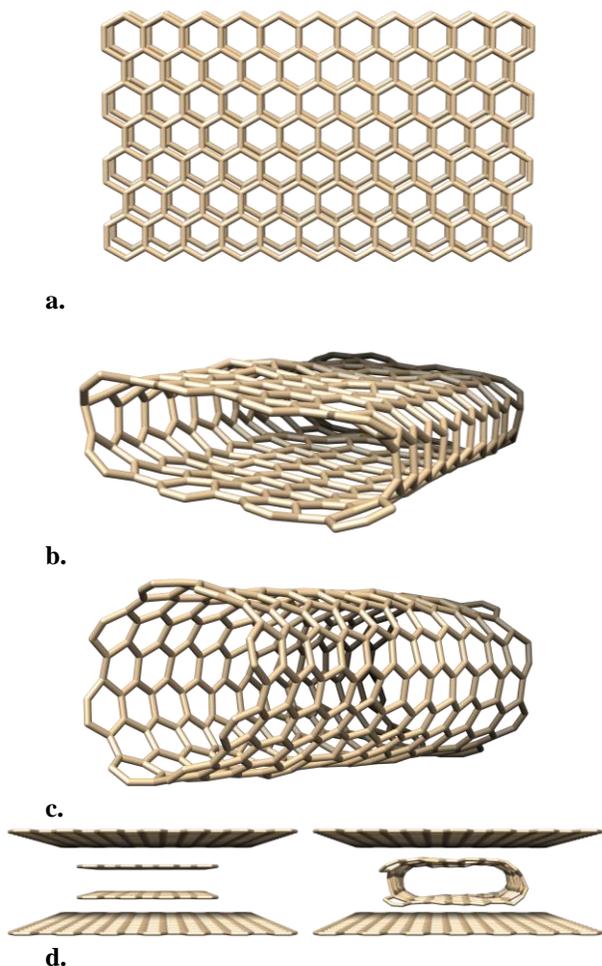


Figure 3. Simulation of armchair nanotube.
a. The initial model
b. The flattened shape model at 2 ps. (Flattening caused by the two substrates)
c. The final shape of the model after the substrates were removed
d. The initial and the flattened model between the substrates (note only one graphite layer of each substrate block is shown)

On Figure 4 two narrow nanoribbons can be seen. These width of 7.1 Å would be below critical width if we did not use substrates. Using the two graphite substrates in the described setup, perfect nanotubes can be grown as seen on the figure.

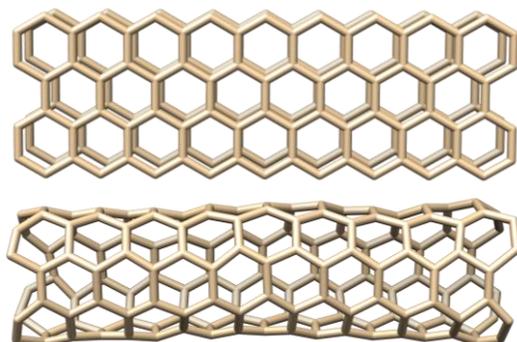


Figure 4. Simulation of armchair nanotube. The initial (upper) and the final (lower) structures. The simulation parameters are the followings: 1000K simulation temperature, 22.14 Å of length and 7.10 Å of width for the parallel nanoribbons.

We made several experiments with even smaller nanoribbons. We can discover the tendency to form a graphene sheet. Even with using substrates self organized growing of perfect nanotubes of such a small diameter could not be done.

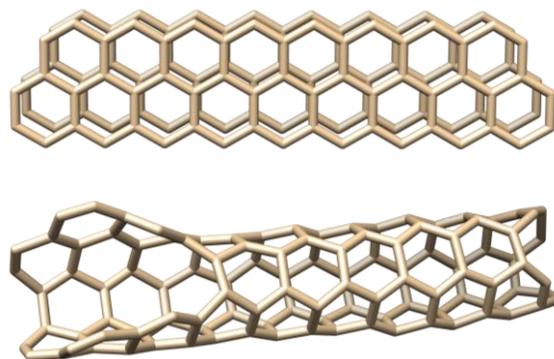


Figure 5. Simulation of armchair nanotube. Initial (upper) and final (lower) structures. Initial nanoribbons are 4.97 Å wide.

In the case of zigzag nanotubes the critical curvature energy is less, the critical ribbon width is greater than the same value at the armchair nanotubes. These simulations were performed again with using substrates. In Figure 6 a wide model is shown. The initial structure would be above critical size if substrates were not used but with the substrates perfect nanotube structure has been developed.

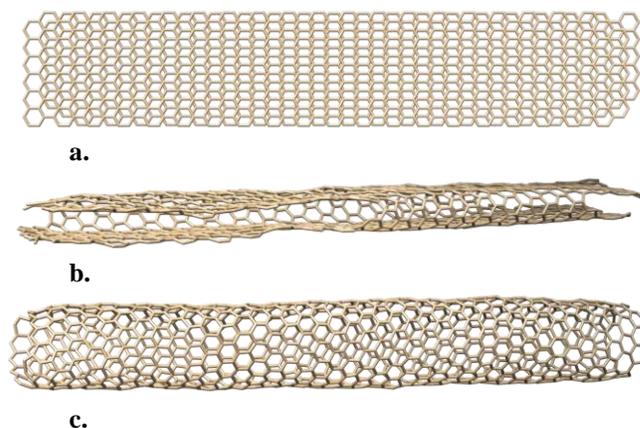


Figure 6. Simulation of zigzag nanotube.
a. The initial model
b. The flattened structure after a simulation time of 0.8 ps (Not all the bonds are formed yet)
c. The final shape of the model at 3.6 ps. After all the bonds are formed, the substrates were removed.
 The simulation parameters are the followings: 1000 K simulation temperature 85.91 Å of length and 15.99 Å of width for the parallel nanoribbons.

We experienced that behavior of models wider than or equal to the model shown in Figure 6 show the same behavior. We wanted to know how the structures with smaller width behave when substrates are in use. In Figure 7 a narrower model (initial width: 13.53 Å) is shown. Although this model cannot

REFERENCES

coalesce to nanotube without substrates, it develops well with proper setup, as can be seen in the middle of the figure. When the self-organized growing of the nanotube has finished after 0.4 ps, we gradually pulled off the two substrates to allow the nanotube forming a perfect round shape. This can be seen on the bottom of the figure.

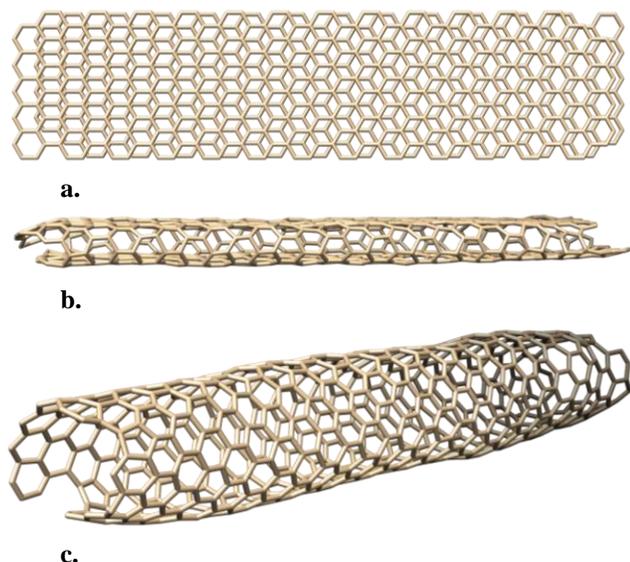


Figure 7. Simulation of zigzag nanotube.

a. The initial model

b. The flattened shape model at 0.4 ps

c. The final round shape model at 0.67 ps. After 0.4 ps the substrates were pulled off gradually to give enough space for perfect round nanotube.

The simulation parameters are the followings: 1050 K simulation temperature, 56.10 Å of length and 13.53 Å of width for the parallel nanoribbons.

Building from parallel graphene nanoribbons can give chances for controlled reliable technology in the case of more complicated carbon nanostructures, according to molecular dynamics simulations.

IV. CONCLUSION

From our molecular dynamics simulations we obtained the following conditions for straight nanotube formation from two parallel nanoribbons placed between two graphite substrate blocks:

- For armchair nanotubes the critical ribbon width is 7.10 Å.
- For zigzag nanotubes we obtained the critical ribbon width of 13.53 Å.

Both critical widths are significantly narrower than critical widths without using substrates. Further conditions will be researched to find even lower values as critical width for self-organized building nanotubes.

- [1] Avouris P 2002 “Molecular electronics with carbon nanotubes” *Accounts Chem. Res.* 35 1026-1034
- [2] Tans S J, Verschueren A R M, Dekker C 1998 “Room-temperature transistor based on a single carbon nanotube” *Nature* 393 49-52
- [3] Yao Z, Postma H W C, Balents L, Dekker C 1999 “Carbon nanotube intramolecular junctions” *Nature* 402 273-276
- [4] Keren K, Berman R S, Buchstab E, Sivan U, Braun E 2003 “DNA-templated carbon nanotube field-effect transistor” *Science* 302 1380-1382
- [5] Fülep D, Zsoldos I, László I “Molecular dynamics simulations for lithographic production of carbon nanotube structures from graphene” 2015 *Mathematics in Computer Science and Engineering Series 42*, pp 253-256
- [6] Tapasztó L, Dobrik G, Lambin P, Biro L P 2008 “Tailoring the atomic structure of graphene nanoribbons by scanning tunnelling microscope lithography” *Nature Nanotechnology* 3 397-401
- [7] Nemes-Incze P, Magda G, Kamarás K, Biró L P 2010 “Crystallographically selective nanopatterning of graphene on SiO₂” *Nano Research* 3 110-116
- [8] Han S S, Lee K S, Lee H M 2004 “Nucleation mechanism of carbon nanotube” *Chemical Physics Letters* 383 321-325
- [9] László I, Zsoldos I 2012 “Graphene-based molecular dynamics nanolithography of fullerenes, nanotubes and other carbon structures” *Europhysics Letters* 99 63001
- [10] László I, Zsoldos I 2012 “Molecular dynamics simulation of carbon nanostructures: The C₆₀ buckminsterfullerene” *Phys. Status Solidi B* 249 2616-2619
- [11] László I, Zsoldos I 2014 “Molecular dynamics simulation of carbon nanostructures: The D_{5h} C₇₀ fullerene” *Physica E* 56 427-430
- [12] He L, Lu J Q, Jiang H 2009 “Controlled Carbon-Nanotube Junctions Self-Assembled from Graphene Nanoribbons” *Small* 5 2802-2806
- [13] D Fülep, I Zsoldos, I László: 2015 “Topological and energetic conditions for lithographic production of carbon nanotubes from graphene” *Journal of Nanomaterials*, *accepted, in press*
- [14] Allen M P, Tildesley D J 1996 “Computer Simulation of Liquids” Clarendon Press, Oxford
- [15] Frenkel D, Smit B 1996 “Understanding Molecular Simulation – From Algorithms to Applications” Academic Press, San Diego
- [16] Porezag D, Frauenheim T, Köhler T, Seifert G and Kaschner R 1995 “Construction of tight-binding-like potentials on the basis of density-functional theory: Application to Carbon” *Phys. Rev. B* 51 12947-12957
- [17] Verlet L, 1967 “Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules” *Phys. Rev.* 159 98-103
- [18] Nosé S, 1984 “A molecular dynamics method for simulation in the canonical ensemble” *Mol. Phys.* 52 255-268
- [19] Hoover W G 1985 “Canonical dynamics: Equilibrium phase-space distributions” *Phys. Rev. A* 31 1695-1697
- [20] Xian-Hong Meng, Ming Li, Zhan Kang, Jian-Liang Xiao 2014 “Folding of multi-layer graphene sheets induced by van der Waals interaction” *Acta Mechanica Sinica* 30(3): 410-417

An investigation of dimensional accuracy of Multi-Jet Modeling parts

K. Kitsakis, Z. Moza, V. Iakovakis, N. Mastorakis, and J. Kechagias

Abstract—Additive Manufacturing (AM), also called 3D Printing, is a process where a three dimensional physical part is created by printing or adding very thin layers of material on top of each other, until the part is complete. The parts are created with a computer aided drawing software which produces the appropriate 3D print file. When the 3D printer reads the file, it processes or slices the 3D computer generated part into hundreds or thousands of layers. When the first layer is done printing, the printer’s head moves up, or the print bed moves down allowing the next layer to be printed. 3D printing processes can be used in Rapid Prototyping, Rapid Manufacturing, Direct Manufacturing, Health, Dental and Jewelry Industry, Architecture, Quick Fixtures or by hobbyists. In this study, by using the Multi-jet method, we produced 3D parts to investigate the dimensional accuracy for linear sizes of parallel opposite surfaces and cylinder, and to categorize them, according to the International Tolerance Grade. The results indicated essential size errors connected with each measured dimension, confirming that further research is needed to optimize the specific 3D process.

Keywords—3D Printing, Additive Manufacturing, Dimensional Accuracy, International Tolerance Grade, Multi-jet Modeling, Rapid prototyping.

I. INTRODUCTION

According to the American Society for Testing Materials (ASTM), Additive Manufacturing is “A process of joining materials to make objects from 3D model data, usually layer upon layer, as opposed to subtractive manufacturing technologies” [1]. It is a key technology for the future, with strong impact on the addition of business value for manufacturers, over the next decade. Its history begins in 1983 when Chuck Hull invented a process called Stereolithography, enabling a 3D object to be printed from CAD data and took almost 20 years before the first consumer devices began to appear. 3D printing is an Additive Manufacturing process

K. Kitsakis, is with the Department of Mechanical Engineering, Technological Educational Institute of Thessaly, Larissa, 41110, Greece (e-mail: kitsakis@teithessaly.gr).

Z. Moza, is with the Department of Mechanical Engineering, Technological Educational Institute of Thessaly, Larissa, 41110, Greece (e-mail: zoigr79@yahoo.com).

V. Iakovakis is with the Department of Mechanical Engineering, Technological Educational Institute of Thessaly, Larissa, 41110, Greece (e-mail: iakovakis@teithessaly.gr).

N. Mastorakis, is with the Department of Industrial Engineering, Technical University of Sofia, Sofia, Bulgaria (e-mail: mastor@tu-sofia.bg).

J. Kechagias is with the Department of Mechanical Engineering, Technological Educational Institute of Thessaly, Larissa, 41110, Greece (e-mail: jkechag@teithessaly.gr).

invented and patented by the Massachusetts Institute of Technology in 1993 [2]. In 2009 the technology patent for the material extrusion process called Fuse Deposition Modeling (FDM) expired, so the technology became much cheaper and available to a much broader cross section of producers.

The 3D printing process begins with a representation of an object as a 3D model in computer aided design (CAD) based software. That model can be created directly in the software, or it can be input into the software, through the use of a laser scanning device, that will take a physical object and bring it into the system. Once that design is created, the Standard Tessellation Language (STL) file is generated by the software. STL is the most popular file format for Additive Manufacturing. To tessellate something, as in Standard Tessellation language, means to break it down into a series of polygons -and to be more specific in triangles- to represent not just its external structure, but its internal structure, too. Once that file is created, the system slices it into many different layers and passes that information to the Additive Manufacturing device.

The Additive Manufacturing system itself creates the object, by definition layer by layer, until we have a finished object. Very often, post production is required. That might be the removal of dust or other material, it might require some machining or a process called sintering where we’re closing voids, or some sort of infiltration process where we’re filling voids within the object itself with other materials. The goal is to create a physical object from a 3D model [3].

Although traditional-subtractive manufacturing has its advantages and is preferred when producing very high volumes of similar objects, when we need a diversity of different materials to choose from and when we want to create large parts, 3D printing is preferred when there is high design complexity, the need for rapid production and the necessity for restriction of the material waste [4].

The Additive Manufacturing processes for 3D printing which are used today are [3]:

- Vat Photo polymerization
 - Stereolithography (SLA)
 - Digital Light Processing (DLP)
- Material Jetting
 - Multi-jet Modeling (MJM)
- Material extrusion
 - Fused Deposition Modeling (FDM)
 - Fused Filament Fabrication (FFF)
- Powder Bed Fusion
 - Electron beam melting (EBM)

- Selective Laser Sintering (SLS)
- Selective Heat Sintering (SHS)
- Direct Metal Laser Sintering (DMLS)
- Binder Jetting
 - Powder Bed and Inkjet Head 3D printing (PBIH)
 - Plater-Based 3D printing (PP)
- Sheet Lamination
 - Ultrasonic Consolidation (UC)
 - Laminated Object Manufacturing (LOM)
- Direct Energy Deposition
 - Laser Metal Deposition (LMD)

In Multi-Jet Modeling (MJM) technology a print head selectively deposits material on the platform. These droplets are most often comprised of photopolymers with second materials, such as wax, used to support structures during the building process. At the next step, an ultraviolet light solidifies the photopolymer material to shape the cured parts. At the final pass, the planerizer levels the material to create a nice flat surface (Fig. 1). When the parts have been printed, a post-building process is required to remove the support material. The advantages of this 3D printing technology are the high accuracy and surface finish, the capability to use multiple materials and colors on the same part and the hands-free removal of support material. The disadvantages are the limited range of materials and the relatively slow build process [3], [1].

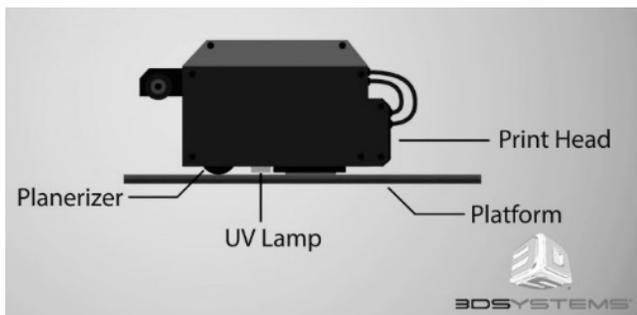


Fig. 1: Multi-Jet Modeling printer parts

Various implementations for this technology can be found in the industries of industrial and consumer products and, more specific, industries of automotive, medical, aerospace and defense. Applications of this technology are Rapid Prototyping (RP), Rapid Tooling (RT) and Direct Digital Manufacturing (DDM) [5]. The parts produced for these sectors need to be firm and fit the prospective functionality, but 3D printing for RP, RT and DDM is a relatively new technology and in the literature there are only a few studies related to dimensional accuracy and categorization [6], [7], [8]. In this paper we studied the coherence and repeatability of linear and circular dimensions of parts produced by a 3D printer using the Multi-jet Modeling (MJM) process. We classified the above dimensions and categorized their quality according to the International Tolerance Grade.

II. SCOPE

The most important applications for MJM 3D printing technology are any situation in which there is a need for a high fidelity mock up, with complex objects, moving parts, constant service finish, fine resolution, and accuracy. The most crucial aspect for ensuring the dimensional repeatability of the printed parts is the dimensional accuracy which represents the degree of agreement between designed prescription and the manufactured product dimensions. The current dimensioning and tolerance standards evaluate the dimensional accuracy of a component part through its size and shape [9].

The variation in size of the manufactured products is very crucial for the applications that this technology is applied for. We examined different linear dimensions in the X, Y and Z axis and circular dimensions at the top and the bottom of a vertical cylinder to categorize the produced samples according to the ISO 286 International Tolerance Grades table reference.

III. EXPERIMENTAL WORK

For our experiment we designed an L like 3D shape combined with a cylinder near one of its corner, but not tangential to the edges, as shown in Fig. 2. This shape gave us the opportunity to examine external linear dimensions (length, width and height of the L shape) and outer circular dimensions (diameter of the cylinder). The dimensions have been chosen so that they belong to different Basic Size ranges of the International Tolerance Grades Table.

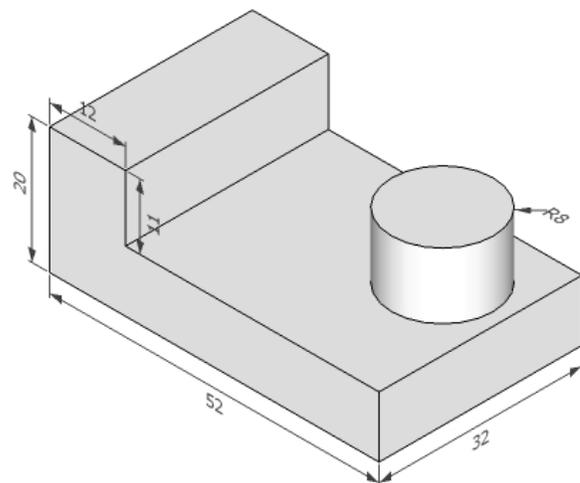


Fig. 2: Designed part dimensions in mm

Eight independent parts (Fig. 3) were produced from the same STL file using an Objet Eden250 3D printer. We selected High speed printing mode (30-micron) and glossy surface at the printing process to have more sleek surfaces and only at the bottom of the parts the FullCure705 support was printed to ground them. For all parts Veroblack Opaque material was used.

The Eden250 printer (Fig. 4) is an ultra-thin-layer, high-resolution 3D printer for rapid prototyping and rapid manufacturing. Some technical specifications of the printer

are: Net build size 250mm X 250mm X 200mm, Build resolution at X-axis: 600dpi, at Y-axis: 300dpi, and Z-axis: 1600dpi, Horizontal build layers down to 16 micron at High Quality and down to 30 micron at High speed printing mode, 0.1-0.2mm typical accuracy varying according to geometry, part orientation and print size, Materials supported: FullCure 720 Model Transparent, VeroWhite Opaque, VeroBlue Opaque, VeroBlack Opaque and Support type FullCure705 Support [10].

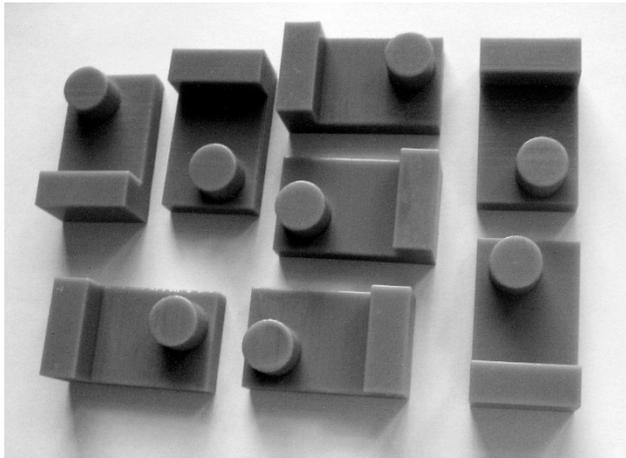


Fig. 3: The eight printed parts



Fig. 4: OBJET EDEN250 MJM 3D printer

The printed parts were measured by three micrometers, with measuring ranges of 0-25mm, 25mm-50mm and 50mm-75mm each. We measured the height (Z-Print Axis) at ten measurement points on the upper top surfaces (rectangle and circle surface) and at nine measurement points on the lower top surface (points HA1-HA10, HB1-HB9 -Fig. 5), the length (Y-Print Axis) at six measurement points on the upper rectangle and at six measurement points on the lower rectangle (LA1-LA6, LB1-LB6 - Fig. 6), the width (X-Print Axis) at ten

measurement points (W1-W10 - Fig. 7) and the diameter at four measurement points (D1-D4 - Fig. 8).

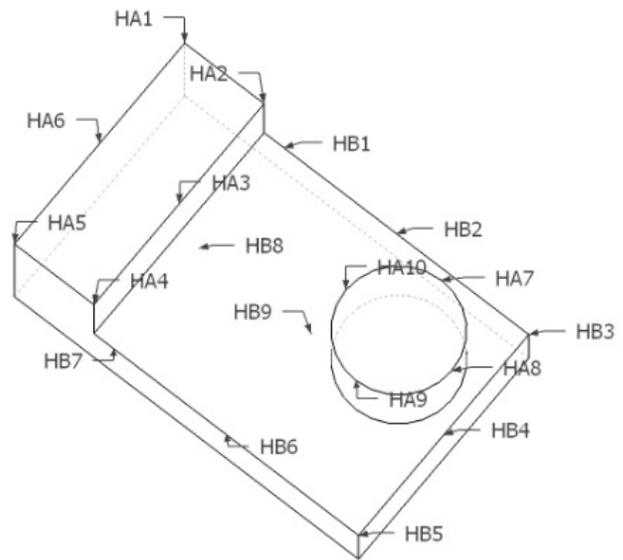


Fig. 5: Measured Height points

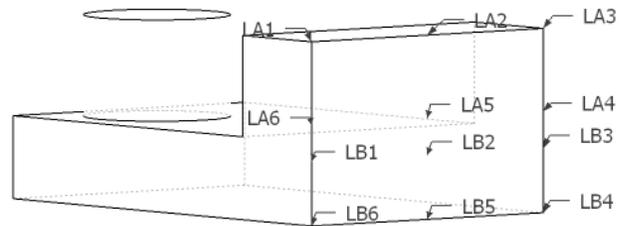


Fig. 6: Measured Length points

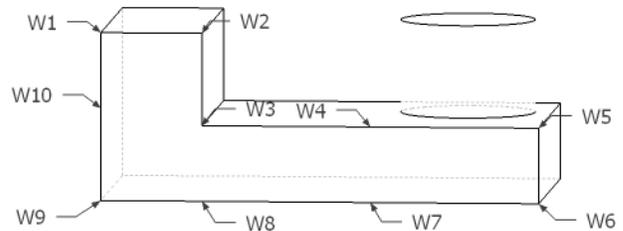


Fig. 7: Measured Width points

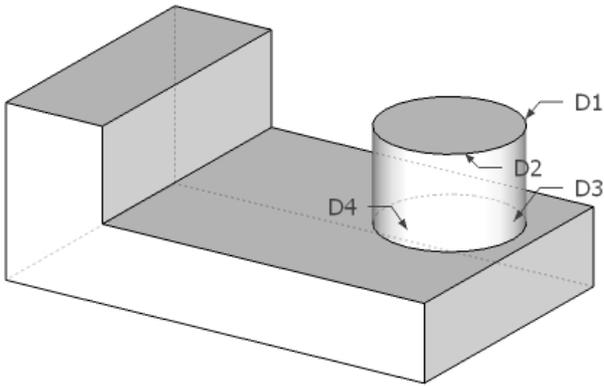


Fig. 8: Measured Diameter points

IV. RESULTS AND ANALYSIS

In Chart 1 and Chart 2 the values of height or Z-Print Axis is presented and although the median is very close to the designed dimension, the variation of errors ($\pm 3\sigma$) sometimes extends the printer’s nominal accuracy of 0.2mm given by the constructor. The measurement points 7-10 of HA, which correspond to the top face of the cylinder have even better median values, but larger variation values too. We used the High Speed print mode and probably at High Quality print mode this phenomenon could be eliminated, so further investigation has to be done with another experiment.

By examining the Y-Print Axis with dimensions LA and LB, values of which are represented in Chart 3 and Chart 4, we note that length is oversized and the variation of errors is very small, excluding measurement points 4, 5 and 6 of LB which are at the lower layer.

This observation occurs also in Chart 5, for measurement points W6 – W9 of width dimension, which represents the X-Print Axis. The oversizing of the base layer dimensions is thought to be caused by the support material layer which is printed prior to the main printing process of the part. The Support material is less compact and causes partial diffusion into the lower levels of the printed parts.

The medians and the variations in diameter of a cylinder at four different measurement points, two on the top and two on the bottom are represented in Chart 6. From the results we ascertain that the bottom of the cylinder is undersized, while the top face is oversized and with bigger variation. This finding confirms that geometry is a very basic aspect considering the variation of accuracy and more investigation about the geometry error is required.

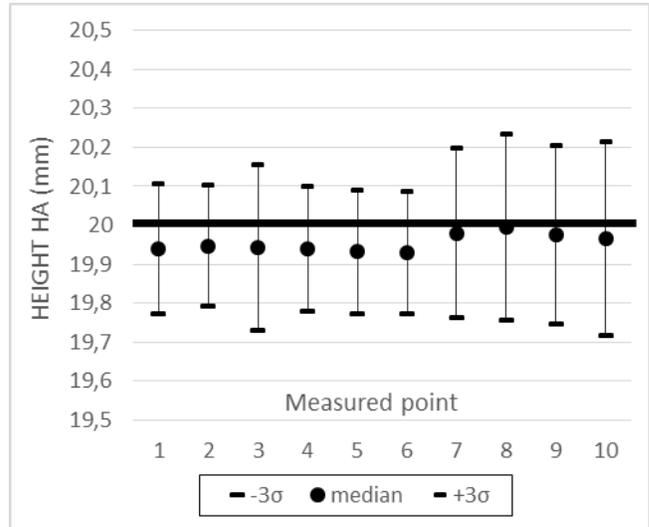


Chart 1: Variations of Height HA

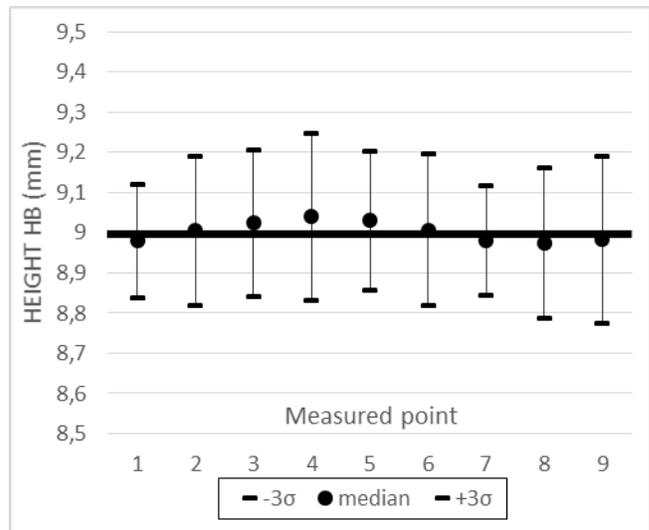


Chart 2: Variations of Height HB

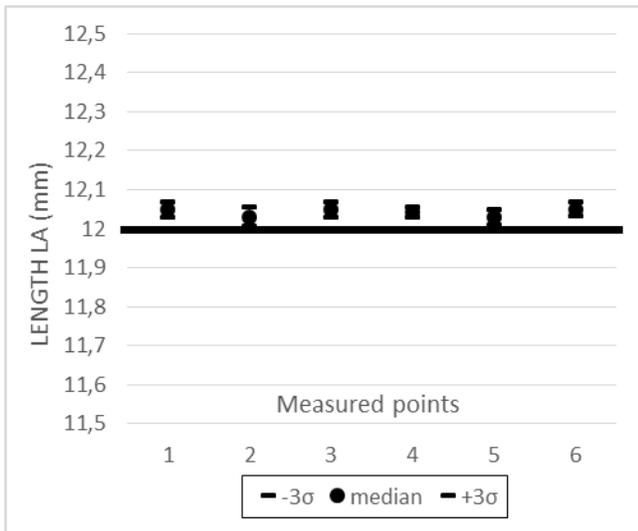


Chart 3: Variations of Length LA

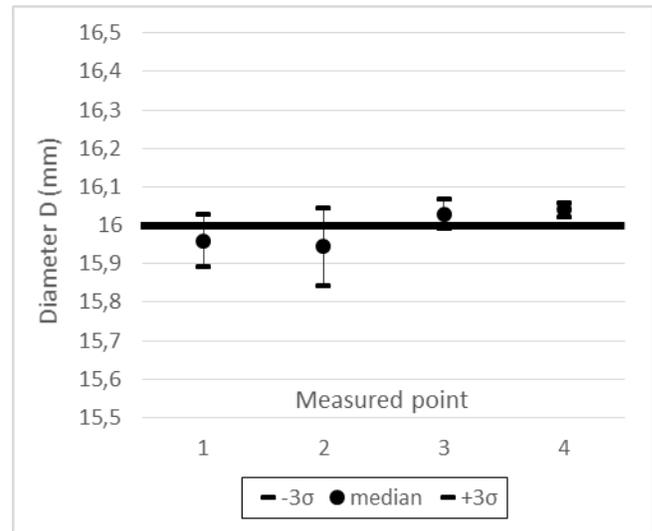


Chart 6: Variations of Diameter D

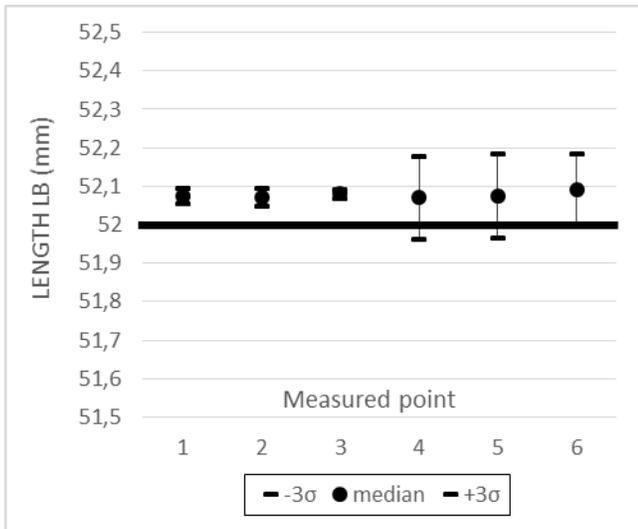


Chart 4: Variations of Length LB

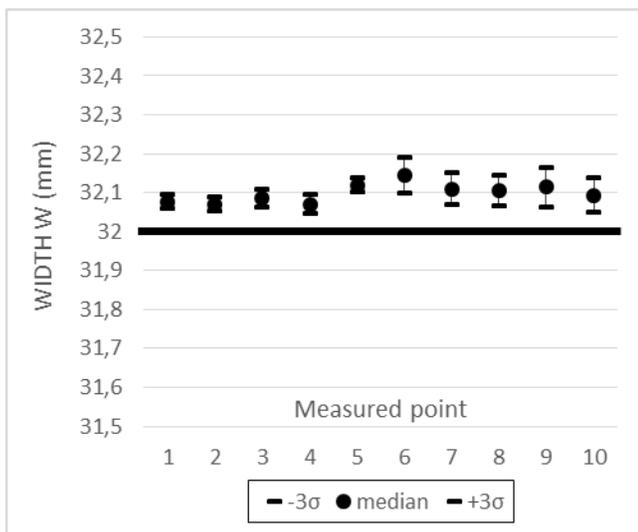


Chart 5: Variations of Width W

The International Tolerance Grade (ITG) specifies tolerances with associated manufacturing processes for a given dimension. It indicates how precise an industrial process is and according to ISO 286-1:2010 for sizes up to 3150mm there are twenty values, IT01, IT0, IT1..IT18. The lower the IT grade number is, the higher is the precision of a machining process. For measuring tools IT01-IT7 is required. IT7-IT11 are grades for Fits and IT12-IT18 for Large Manufacturing Tolerances [11], [12].

The process Tolerance T can be calculated by the following function:

$$T = 10^{\frac{ITG-1}{5}} (0.45\sqrt[3]{D} + 0.001D)$$

where ITG is the IT Grade category value (for IT5-IT18), and D is the Geometrical mean dimension in mm:

$$D = \sqrt{D_{min} D_{max}}$$

where Dmin and Dmax are the limits of the dimension range. There are twenty one ranges for sizes up to 3150mm and the values of standard tolerance grades according to IT grades and Nominal size are shown in Table 4 [13].

To specify the IT grade for the size of two parallel opposite surfaces we examined the shaft through all LA values for one dimension (Y-Axis) and W values 1,2,3,10 for the second dimension (X-Axis). The size of cylinder type was examined with all the values of D. Table 1 - Table 3 show the minimum and maximum values for the examined measurement points and the difference of these in μm. From Table 4:

- For LA size, Y-Print Axis, with the nominal value of 12 mm and maximum difference of 20 μm we assign the IT8 grade.
- For W size, X-Print Axis, with the nominal value of 32 mm and maximum difference of 40 μm we assign the IT9 grade
- For D size, cylinder type, with the nominal value of 16 mm and maximum difference of 90 μm we assign the IT11 grade.

All three sizes are categorized to IT11 grade at maximum which means that they belong to the Fits area of IT grades.

Table 1: Tolerance of the Y-Axis

LA	Min mm	Max mm	(Max-Min) X 1000 $\mu\text{m}/\text{mm}$
1	12,04	12,06	20 μm
2	12,02	12,04	20 μm
3	12,04	12,055	15 μm
4	12,04	12,05	10 μm
5	12,02	12,04	20 μm
6	12,04	12,06	20 μm

Table 2: Tolerance of the Y-Axis

W	Min mm	Max mm	(Max-Min) X 1000 $\mu\text{m}/\text{mm}$
1	32,07	32,085	15 μm
2	32,06	32,08	20 μm
3	32,07	32,09	20 μm
10	32,07	32,11	40 μm

Table 3: Tolerance of the cylinder

D	Min mm	Max mm	(Max-Min) X 1000 $\mu\text{m}/\text{mm}$
1	15,93	16,00	70 μm
2	15,91	16,00	90 μm
3	16,01	16,05	40 μm
4	16,03	16,05	20 μm

V. CONCLUSIONS

From the experimental values and the analysis that followed we made some conclusions, which are summarized below:

- Dimensions in the X and Y axis are always oversized.
- The accuracy of the base of the produced parts is poorer in both X and Y axis.
- The accuracy in Z axis is smaller than the other two axis and it depends on both the geometrical shape, rectangle or cylinder, and the height of layer.
- The cylinder diameter is undersized at the top, oversized at the base and with less accuracy at the top level.
- The X, Y and cylinder size belong to the Fits area of IT grades.

This is a preliminary study on the IT grade specification for the MJM 3D-print process. Further study examining the

tolerance of holes and channels in 3D printed objects by this technology is required.

REFERENCES

- [1] M. Cotteleer and J. Joyce, "3D Opportunity Additive manufacturing paths to performance, innovation, and growth," *Deloitte Review*, no. 14, 2014.
- [2] D. L. Chandler, "MIT research continues to push the boundaries of the burgeoning technology of 3-D printing," 14 September 2011. [Online]. Available: <http://web.mit.edu/newsoffice/2011/3d-printing-0914.html>. [Accessed 25 Mai 2015].
- [3] M. Cotteleer, J. Holdowsky and M. Mahto, *The 3D opportunity primer. The basics of additive manufacturing*, Deloitte University Press, 2013.
- [4] X. Yan and P. Gu, *A review of rapid prototyping technologies and systems*, vol. 28, Computer-Aided Design, 1996.
- [5] J. Kruth, M. Leu and T. Nakagawa, *Progress in additive manufacturing and rapid prototyping*, vol. 47, CIRP Annals-Manufacturing Technology, 1998.
- [6] P. F. Jacobs, *Rapid Prototyping and Manufacturing: Fundamentals of stereolithography*, Dearborn: Society of Manufacturing Engineers, 1992.
- [7] D. Dimitrov, M. van Wijck and K. a. d. B. Schreve, "Investigating the achievable accuracy of three dimensional printing," *Rapid Prototyping Journal*, vol. 1, no. 12, pp. 42-52, 2006.
- [8] D. Dimitrov, M. and K. a. d. B. Schreve, "Advances in three dimensional printing-state of the art and future perspectives," *Rapid Prototyping Journal*, vol. 3, no. 12, pp. 136-147, 2006.
- [9] J. Kechagias, P. Stavropoulos, A. Koutsomichalis, I. Ntintakis and N. Vaxevanidis, "Dimensional Accuracy Optimization of Prototypes produced by PolyJet Direct 3D Printing Technology," in *Proceedings of the 18th International Conference on Circuits, Systems, Communications and Computers*.
- [10] "User Guide," *Objet30 3-D Printer System*, [Online]. Available: <http://www.ece.ubc.ca/~leos/pdf/tools/objet/manual.pdf>.
- [11] H. Conway, *Engineering Tolerances*, 3rd ed., London: Sir Isaac Pitman & Sons Ltd., 1966.
- [12] L. E. Farmer, *Dimensioning and Tolerancing for Function and Economic Manufacture*, Sydney: Blue Publ., 1999.
- [13] European Committee for Standardization, *EN ISO 286-1*, Brussels: CEN, 2010.

Table 4: Values of standard tolerance for nominal sizes up to 120 mm

Nominal size mm		Standard tolerance grades																			
		IT01	IT0	IT1	IT2	IT3	IT4	IT5	IT6	IT7	IT8	IT9	IT10	IT11	IT12	IT13	IT14	IT15	IT16	IT17	IT18
Above	Up to and including	Standard tolerance values																			
		μm														mm					
—	3	0,3	0,5	0,8	1,2	2	3	4	6	10	14	25	40	60	0,1	0,14	0,25	0,4	0,6	1	1,4
3	6	0,4	0,6	1	1,5	2,5	4	5	8	12	18	30	48	75	0,12	0,18	0,3	0,48	0,75	1,2	1,8
6	10	0,4	0,6	1	1,5	2,5	4	6	9	15	22	36	58	90	0,15	0,22	0,36	0,58	0,9	1,5	2,2
10	18	0,5	0,8	1,2	2	3	5	8	11	18	27	43	70	110	0,18	0,27	0,43	0,7	1,1	1,8	2,7
18	30	0,6	1	1,5	2,5	4	6	9	13	21	33	52	84	130	0,21	0,33	0,52	0,84	1,3	2,1	3,3
30	50	0,6	1	1,5	2,5	4	7	11	16	25	39	62	100	160	0,25	0,39	0,62	1	1,6	2,5	3,9
50	80	0,8	1,2	2	3	5	8	13	19	30	46	74	120	190	0,3	0,46	0,74	1,2	1,9	3	4,6
80	120	1	1,5	2,5	4	6	10	15	22	35	54	87	140	220	0,35	0,54	0,87	1,4	2,2	3,5	5,4

Human-centered architecture of a medical cyber-physical system

Razvan Popescu, Theodora Dumitrescu, Radu Dobrescu

Abstract— This paper presents a dedicated architectural design framework for Medical Cyber Physical Systems (MCPS). This framework is a dynamic event based sensing and monitoring system designed by integrating the full functionality of cyber physical system. The system adapts the concept of local and global awareness for intelligent decision making in physical and cyber world. Moreover, the new introduced concept of human-centered architecture allows improving the efficiency of the decisional process by fusion of data perceived by human senses and environment data measured by sensor networks.

Keywords— cyber-physical system, event driven system, human perception, human-centered architecture.

I. INTRODUCTION

IN a previous work [1] an analysis of the current trends in the development and use of medical cyber-physical systems (MCPS) has underlined that medical device industry is undergoing a rapid transformation, embracing the potential of embedded software and network connectivity in a distinct class of distributed systems that simultaneously control multiple aspects of the patient's physiology using a combination of embedded software controlling the devices, networking capabilities, and complicated physical dynamics that patient bodies exhibit. The MCPS re-quires a large number of dispersed sensors to collect information and high-performance computing to control the physical unit to execute tasks. However, by placing MCPS in a network environment a lot of potential uncertainties in the decision-making process should be solved, while respecting the time limits imposed by this process. Therefore, it is necessary to predict the influence of the result of the decision within the time and the physical environment according to the uncertainty factors. MCPS must be able to combine the characteristics of information world and the physical world to build a stable coalition system, and predict the cyber physical system overhead and performance at

The work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU 187/1.5/S/155420.

Razvan Popescu is PhD student at the Doctoral School of the Faculty of Control and Computers, Politehnica University of Bucharest, in the field of Systems Engineering (corresponding author: phone: +40214029105; e-mail: razvan.popescu@subversiv.ro).

Theodora Dumitrescu is PhD student at the Doctoral School of the Faculty of Control and Computers, Politehnica University of Bucharest, in the field of Systems Engineering (e-mail: theodora.dumitrescu@gmail.com).

Radu Dobrescu, PhD, is professor in the Department of Automatic Control and Industrial Informatics of the Faculty of Control and Computers, Politehnica University of Bucharest, (e-mail: rd_dobrescu@yahoo.com).

the same time. In the decisional process, the person is the designed initiator, while the physical unit is an interactive bridge of the digital world and the real environment.

Accepting the decisive role of the person in decision-making, this paper aims to describe a human-centered CPS architecture. Such system needs to collect environmental information, personalized information and human behavior information, then transform the information into useful knowledge for understanding environmental and people. The architecture is a multilevel layered one and includes the concept of data aggregation, knowledge classification and enrichment based on events generated from the physical world.

II. RELATED WORK

Even if at about 9 years after US National Science Foundation (NSF) has identified (in 2006) cyber-physical systems (CPS) as a key area of research [2], the concept of CPS is not unified as a clear definition, one can consider for the distinct class of medical CPS the approach of Lee and Sokolsky [3] which believe that MCPS is an assembly which integrates computing medical devices, communications, and physical processes, embeds computer real-time monitors and controls the physical process is satisfactory. There are three common points of these theories: 1) interconnection by network between the computation system and the physical system; 2) monitoring capabilities; 3) real-time performance. What we intend to bring to your attention is the constantly increasing role of human factor in the decision to be taken as a result of the events that cause changes.

Event based approach for systems' control is becoming a hot topic in different domains. Makedon et al [4] presents an architectural framework to assist human's daily activities using identification mechanisms for events that influence the actuators and alter the human behavior in a closed loop process. The advantages of the framework are dynamic context awareness, adaptiveness, self-repairing and high confidence that couples computational power with physical testbed control. A similar idea can be found in [5], where the authors propose a human-interactive HILS (Hardware-In-the-Loop Simulation) framework developed to support CPS reliability and reusability in a fully distributed operating environment. Yuanyuan et al [6] suggest a dynamic hierarchical model for health monitoring system which adapts fusion of event driven context model. Don et al [7] propose an architectural design framework for remote health care monitoring system. This framework is a dynamic event based sensing and monitoring system designed by integrating the full functionality of cyber

physical system. The system adapts the concept of local and global awareness for intelligent decision making in physical and cyber world.

Efficient communication is needed between physical world and the cyber world. Kyoung et al [8] present a novel information centric to support CPSs.

The architecture is based on a cluster of real time embedded databases which communicate and control wireless sensors networks to extract important collective information. From the above surveyed research papers, the most important is model design frameworks. Such a framework having the potential to manage different events from different sources, aggregate those events and generate meaningful information augmented with those offered by the human perception system will be discussed in the next section.

III. DESIGN PRINCIPLES OF A HUMAN-CENTERED CPS ARCHITECTURE

Medical cyber physical systems combine both physical and cyber world resources. Fig. 1 shows the proposed architecture.

Two types of data analysis were considered in the design

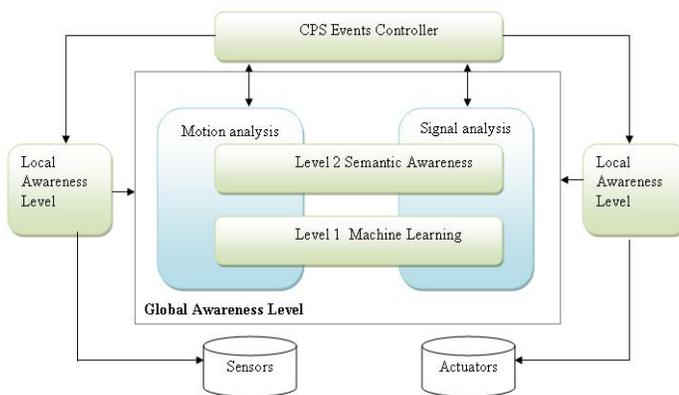


Fig. 1 MCPS architecture

stage. The first one is the motion analysis which includes capturing the body motion of a patient for analyzing the patient behavior. The second one is signal analysis i.e., analysis of the data received from various sensors connected to the patient. This architecture is based on two levels of data processing. The lower (Level 1) is the machine learning which includes different algorithms for event identification. The upper (Level 2) is the semantic awareness level. At this level the system can generate the highest interpretation about the current situation semantically. The system is generic multi-layered architecture with the following components:

A. Sensors and Actuators

Sensors captures motions and signals that provides a communication interface between physical and cyber world. The signal sensor captures various body parameters such as ECG, blood pressure, temperatures etc.

These signal attributes can represent various body conditions of a patient. The motion sensors capture various

body movements. With the help of actuators the systems are able to change the physical attributes such as triggering alarm, wheel chair movement etc.

B. Local and Global Awareness

Awareness is an understanding of the activity of others, which provides a context for your own activity. From the proposed system point of view, Global awareness is the ability to sense and retrieve meaningful information from the physical and cyber world as a single entity, whereas local awareness is the ability to sense the physical world as an individual entity.

C. Local Event Controller

An event can be defined as any occurrence in the physical world at a specific time and location, denoted by E . The type of event considered are the events generated from different biophysical and motion sensors. The local event controller process the events based on local event policies, and only those which are meaningful will be send to the cyber world for further analysis. Each event is characterized by set of features (the event header) $E = \{E_a, E_i, E_t, E_l\}$ where E_a is the event attribute, E_i is the event identification. E_t is the time at which the event is occurred and finally E_l is the location where the event happens. Event policies are set of policies based on the conditional operators CO , each of them having a fixed set of preferential choices by which the event can be further processed or discarded.

The triggered events from the physical world, with the help of communication channel will be received at the cyber space. The received events are then stacked as either motion analysis unit or signal analysis unit depending on the type of event received. The received signals are then processed by the CPS event processing controller. CPS event processing is an inner closed loop controller by which the received events are further processed. The preprocessed events accepted by the policy manager are forwarded to different levels for processing. In our system, we have proposed two levels which can process individually and take action based on the status notifier. Status notifier works as the same concept of lookup tables. It checks the events and provides meaningful information for selecting the levels in-side the controller. Level 1 includes machine learning algorithms for predicting simple events. Level 2 will be considered if the system needs a global awareness of the event happening at the physical world. The corresponding actions are taken in each level in the cyber space and forwarded to the physical world for taking necessary actions in real time.

IV. HUMAN PERCEPTION-BASED DECISIONAL PROCESS

In a MCPS, human always plays an important role, being in the same time the object of the system services and the decision maker. CPS is the communication bridge of the physical world and the digital world. People play a leading role in the physical world; therefore, the systems not only need to perceive the environment, but also perceive the people.

With the aid of the human perception system one can help the system to act in the form the people perceive the world and thus approaching closer to the people's decision-making in order to enhance the intelligent effect of the decisional process.

The human perception system is usually divided into sensation and perception. The sensation is the reflection of individual properties that directly act on the individual sensory stimulation, including external and internal feeling. External feeling is caused by an external stimulus, reflecting the sense of the characteristics of external things, such as vision, hearing, smell, taste and feel of the skin. The receptors located on the body surface, as skin sensory receptors and taste receptors are known as contact receptors, while visual, auditory and olfactory receptors, which can detect external things at a distance, are known as long-distance receptors. Internal feelings, also known as the body's sleep, relative to reflect the feeling of the external environment such as vision, hearing, feeling reflects the internal state of the body and internal changes, including the use of feeling, sense of balance and visceral sensation.

The perception is the procedure that organizes individual sensory information into meaningful perception, object perception and social perception. The CPS system is more concerned about physical perception. Object perception is the perception of material things and external relations. Any matter or thing present spatial characteristics, time features and movement change. Therefore, the perception of the object includes space perception, time perception and motion perception. The spatial perception reflects the perception of the spatial characteristics of the object like shape, size, depth, orientation, etc. The time perception reflects continuity and sequence of the objective phenomenon. Motion perception reflects objects spatial displacement and speed.

All external stimulation perceived by humans must first be received, stored and analyzed by the human perception system, and then can have effect on people. Sensory reacts to appropriate stimulation and process the information, thus forming a meaningful sense of perception, then filter, according to the subjective demand, in order to select only useful information, while ignoring other insignificant things. At this point, we can consider that the human sensory system can complete the data collection in the physical layer of the CPS system, and the organization of the perception system knowledge base and of the CPS system knowledge base are very similar.

It is true that in early CPS system, people and physical units (sensors) perceived the environment, gathered information, analyzed the data collected, made decision, and finally complete the execution of tasks by sending control commands to the remote execution unit. However, with the advance of the intelligence in the CPS system, information analysis and decision-makers gradually shifted from people to computing unit. People only need to perceive the environment and make the demand, and then the computing unit and physical unit deal

with all the subsequent work. In order to allow the computation unit to take decisions that meet people's need, we need to understand and perceive the environment from people's perspective.

Based on the CPS architecture of human perception, the computation unit is divided into three levels: the computation layer, network layer and physical layer (see fig. 2). The computation layer is the core of the CPS system, where data storage, decision-making and the formation of a knowledge base are. In other words, knowledgebase is the brain of CPS. The purpose of the knowledge base is to make the machine to understand better the data meaning, and its role is similar to the perception, ultimately in order to assist decision making. The formation of the knowledge base not only allows the data to fit into the system, but also improves the efficiency of decision-making. The network layer is a communication layer in the CPS system, primarily as a channel of communication exchanges, in which routing, data transmission, data forwarding etc. are completed. The physical layer is the interactive terminal between the CPS system and the physical world, where sensors are used to collect the data of the physical world, and actuators to change the physical world environment. Thus, at the same time of the perception of the physical world, CPS system also needs to perceive people by perceiving what people perceive.

The physical layer of the CPS architecture based on the human perception includes interactive communication function. The communication interaction according to people's exchanges is divided into reading, listening, speaking, writing and observing; the traditional human-computer interaction command is initiated by the way of writing commands input through the keyboard and other external devices. The system then starts to execute commands after reading inputs. However, with speaking and listening added to this behavior, one can improve the performance of a more complex CPS, while simplifying the operation of human behavior, so that the human-computer interaction is more convenient. Observation is a deeper interaction; if the system can observe the behavior of the person and analyzes it to identify the human needs, it can play a very good supporting role on decisions classification and prediction. Moreover, by the system self-learning the system is able to improve the decisional process, mostly by observing the timely receipt of feedback after the task is completed. In general, interactive communication can lead on the one hand to the optimization of information processing, and on the other hand, can improve the convenience of the human-computer interaction, using a variety of means of communication that makes easier to convey the demands to the system, capable of making the formulation of a strategy closer to the people's need.

Unlike the traditional process control systems whose architecture basically centers on machines for the acquisition of environmental information and for the execution of tasks according to specific rules, the CPS architecture based on human perception will understand multi-faceted information

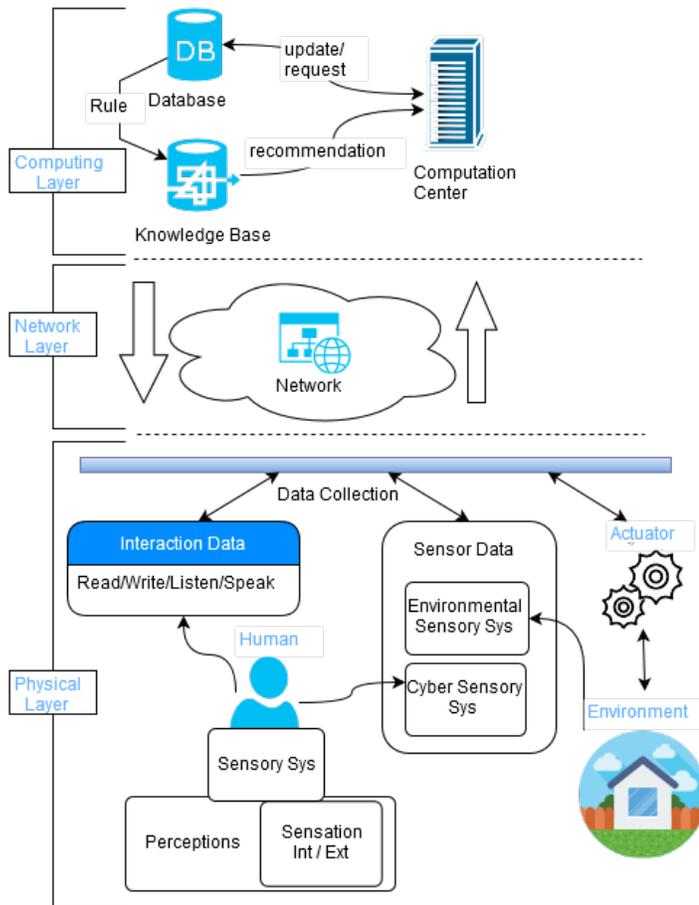


Fig. 2 Human sensory architecture for MCPS

by its own knowledge base on the environment status, and will apply dynamically adjustable rules for performing tasks. The CPS data may be divided into environmental, human, time, positional and historical information. The system unifies the existing rules to formulate the decision-making according to these five types of information. Time and location information is the key relationship between human and environment. Historical information is in order to find the context of data information and rules that can discover a way to deal with the similar issue. Therefore, it is the simplest step to enhance the function of efficiency. Data acquisition is basically completed by sensors, but in a human-centered CPS architecture, the information is separated in two categories: human data and environment data. There is an internal link between human and environment information which offers a certain significance that is the same under time and space and hence, make time and space information as a condition for the human and environment data fusion. After this information fusion, the system could make a reasonable decision by using the knowledge base, and then send the appropriate command to actuators. If the result is not ideal, one can return to the knowledge base, adjust the decisions and proceed to the new execution, until satisfied. To summarize, let mention the steps of the decision making process in a human-centered MCPS :

- 1) Getting of the human and environment information
- 2) Filtering the imperfect or insignificant information
- 3) Storing data to database
- 4) Time and space based human and environment data fusion
- 5) Decision making by using data fusion and knowledge base access Send commands for execution by actuators
- 6) If the result is unsatisfactory, return to step 5
- 7) End the decisional process

V. CONCLUSION

The proposed human-centered architecture for a MCPS work on the principles of event based classification and interpretation. By introducing the concept of local and global awareness the method can achieve higher understanding between the physical world and the cyber world. On the other hand, correlation of environment data from sensors (in particular from the body area networks) and information from the human perception system guarantees an improvement of the decisional process, primarily by reducing the computation time and storage, secondly by enabling performance improvement by iterative feedback adjustments.

REFERENCES

- [1] R. Dobrescu, "Trends in the Development and Use of Medical Cyber-Physical Systems", *First International Workshop on Cyber Physical Systems*, 2012.
- [2] W. Wolf "The Good News and the Bad News", *IEEE Computer* 40 (11): 104, 2007
- [3] L. Insup, O. Sokolsky, "Medical Cyber Physical Systems", *Proc. of the 47th Design Automation Conference*, 2010, pp. 743-748.
- [4] F. Makedon, Z. Le, H. Huang, E. Becker, D. Kosmopoulos, "An event driven frame-work for assistive CPS environments", *SIGBED Revue*, Vol(6), 2009, pp. 1-9.
- [5] M. Kim, S. Kang, W. Kim, I. Chun, "Human-Interactive Hardware-In-the-Loop Simulation Framework for Cyber-Physical Systems", *Second International Conference on Informatics and Applications (ICIA)*, 2013, pp. 198-202.
- [6] C. Yuanyuan, T. Linmi, Xu. Guangyou, "An Event-driven Context Model in Elderly Health Monitoring", *UICATC'09*, 2009, pp. 120-124.
- [7] S. Don, C. Eunmi, M. Dugki, "Event Driven Adaptive Awareness System for Medical Cyber Physical Systems", *4th International Conference on Awareness Science and Technology (iCAST)*, 2012, pp. 238-242.
- [8] K.D. Kang, S.H. Son, "Real-Time Data Services for Cyber Physical Systems", *Proc. of 28th Int.Conf. on Distributed Computing Systems Workshops*, 2008, pp. 1-6

Semi-automated object identification and features extraction of underground faults from tectonic maps

Antonios J. Konstantaras, Nikolaos S. Petrakis, Theofanis S. Frantzeskakis and Emmanouil N. Antonidakis

Abstract—The last few years have seen the eastern Mediterranean sea evolving its status from a popular touristic destination to a petroleum deposit field with multiple oil rigs already established in Cyprus and Israel whilst scanning surveyors are crisscrossing the Greek vicinity. Increased interest in underground structures, such as underground faults discussed in this paper, is now not only due to its geological significance in tectonics and geophysics but also in some cases as potential reserves of hydrocarbons. Extensive thorough surveys of decades have yielded analogue tectonic maps comprised of multiple information encompassing geographical areas' drawings, underground faults, earthquake epicenters, volcanic cones, border lines and other useful information all fused together in a single product. The purpose of this research work is to identify solely underground faults as distinct objects from the tectonic map of Greece and also extract the coordinates of the planar extent of each individual underground fault. The end product is a digital map of the possible planar location and extent of underground faults in the Greek vicinity.

Keywords—data mining, feature extraction, digital image processing, underground faults, tectonic maps, digital mapping

I. INTRODUCTION

UNDERGROUND faults have been objects of significant interest in geology [1] as they are strongly related with the motions of the Earth's crust [2]. In the Greek vicinity, in particular, underground faults are quite extensive [1] and some of them are seismically active [1,3] because of the sinking motion of the African plate beneath the Eurasian plate [2] and the fact that they compose areas of storage of seismic energy [4]. As of late they have also become objects of interest because of their proximity to regions believed to host hydrocarbons [5] both for safety issues regarding oil rigs [6] as well as safety issues regarding the economies of the

neighboring countries. In explaining the latter, countries in the Mediterranean sea rely heavily on tourism [7] and are aware of the fact that the Mediterranean is a closed sea [8] only replenishing its waters once every ninety years.

Over the last few decades several companies and institutions tried to produce tectonic maps of the Greek vicinity using various techniques such as acoustics, electromagnetic, gravity meters, active and passive tomography, etc. [9] but very few and usually older ones are being made openly available and even fewer are trustworthy-concrete. Exceptional and publicly available work in that field is conducted for the Greek vicinity by the IGME institute [10] which has composed a tectonic map of Greece comprising together various types of information such as underground faults, earthquake epicenters, volcanic cones etc. embedded upon a geographical drawing of Greece and surrounding countries including border information. The purpose of this research work is to extract from the overall image only the useful information, which in our case is solely the underground faults' drawing alone as individual distinct objects, and information of the planar extent of each individual underground fault. All individual objects are then merged together to produce a new digital map of the Greek vicinity only containing a planar representation of the underground faults scattered through it.

II. METHODOLOGY

Beginning with the tectonic map and its full context, we want to create and separate only the underground faults as objects. Foremost, we take the image with containing the recorded underground faults and with an image processing tool, eg. Photoshop or Gimp freeware, we ensure that all underground faults are drawn by lines of the same color, eg. black or white in the figures that follow depending on the application. This is a meticulous process that must be done carefully, as many maps containing underground faults also contain additional data such as ground elevation or location of rivers, seismic activity, volcano cones, etc. therefore we must always verify that what is marked as underground fault is indeed that.

A. J. Konstantaras is with the Technological Educational Institute of Crete, Romanou 3, Chania, 73133, Greece, (corresponding author, phone: 0030 28210 23033, fax: 0030 28210 23003, e-mail: akonstantaras@chania.teicrete.gr).

N. S. Petrakis, is with the Technological Educational Institute of Crete, Romanou 3, Chania, 73133, Greece, (e-mail: npet@chania.teicrete.gr).

T. S. Frantzeskakis is with the Technological Educational Institute of Crete, Romanou 3, Chania, 73133, Greece, (e-mail: tfaniz@yahoo.gr).

E. N. Antonidakis is with the Technological Educational Institute of Crete, Romanou 3, Chania, 73133, Greece, (e-mail: ena@chania.teicrete.gr).

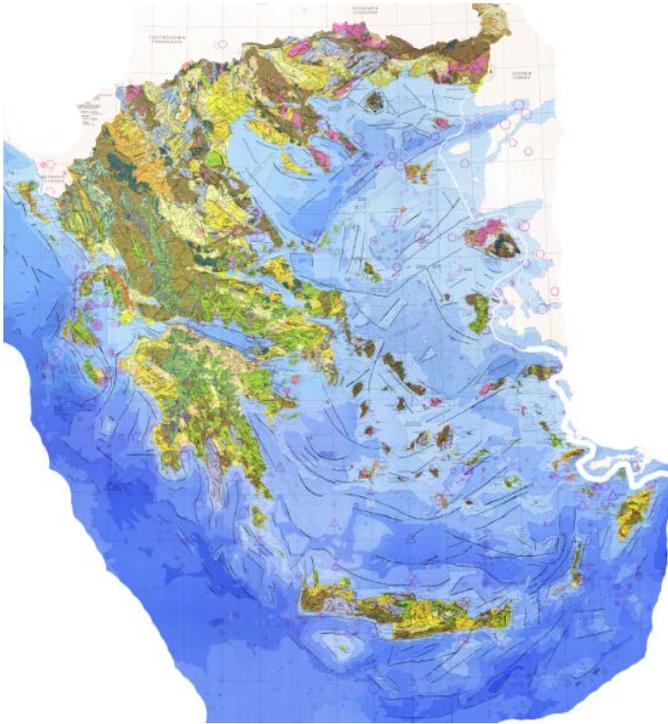


Fig. 1 Processed tectonic map of Greece with all underground faults presented by unique colored lines. The original map prior processing was obtained from IGME [10]

Next, for memory management purposes, we can extract a new image in grayscale with the underground faults marked by black or dark grey stripes, with the next step being the objectification of each individual underground fault.

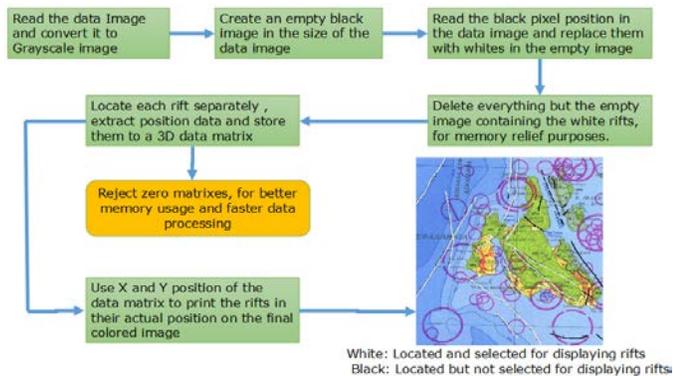


Fig. 2 Schematic flowchart of the of the underground fault objects identification and features extraction algorithm

In order to create objects from images, we have to filter them out of the image. To do so we exploit the narrow underground faults color range, thereby removing all other information represented by other colors. This was done by implementing functions that transform an image into matrix data, with numbers associated to pixel colors, and apply

thresholding to only maintain all the pixels in the selected color range and remove the rest. The resulting image is in effect a black and white figure only demonstrating free pixels positioned within the location range of the multiple present underground faults. The x and y coordinates of each pixel's location is known and recorded but at this stage all pixels are free not belonging in a particular object. Creating objects that actually correspond to individual underground faults was the next task for which several connectivity algorithms were tested. Expert knowledge was also imported to discriminate between separate intercrossing underground faults and single structures that were exhibiting branches.

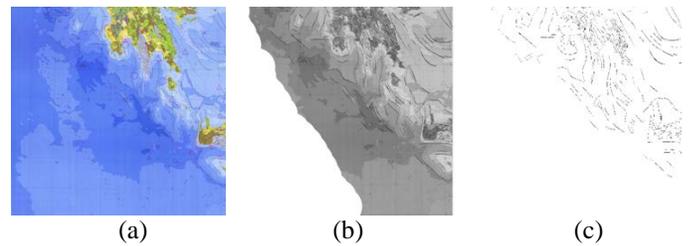


Fig. 3 (a) selected area of interest, (b) interim grayscale image, (c) final image only containing those pixels corresponding to underground faults

The emerging objects are actually matrices containing the horizontal and vertical components of the pixels' positions belonging to a particular distinct underground fault. Each individual object, i.e. underground fault, can be plotted individually or along with any other selected underground for any particular area of interest or the entire tectonic map itself. Furthermore measurements can be derived regarding the proximity of any specific underground fault to areas of increased seismic activity, heavily populated cities, construction sites of large structures, oil rigs, etc. that might be helpful assessing potential hazard risks.

III. RESULTS

One significant problem encountered in processing tectonic maps is their memory size and the fact that lossy compression algorithms yield loss of potential useful information and alter the initial image. As such we resolved to partition the image into smaller figures and ensure pixels' coordinates were organized as such so as to be able to reassemble together the processed figures in produce the processed map as a whole.

The following figure (Fig. 4) demonstrates the application of the above methodology to the north-west quarter of the Greek tectonic map partitioned in four quarters. The upper left and upper right parts of Fig. 4 show the underground faults on the north-west part of the Greek tectonic map in color and in grayscale, respectively. The lower left part shows all the pixels that comprise all the detected underground faults. The lower right part of Fig. 4 zooms in the seismically active area of

Kefalonia and presents all underground faults present as distinct objects. Underground faults are depicted in white and when the user selects to work with one or more of them then those underground faults in particular change their color to black in order to stand out within the overall group of underground faults in that region.

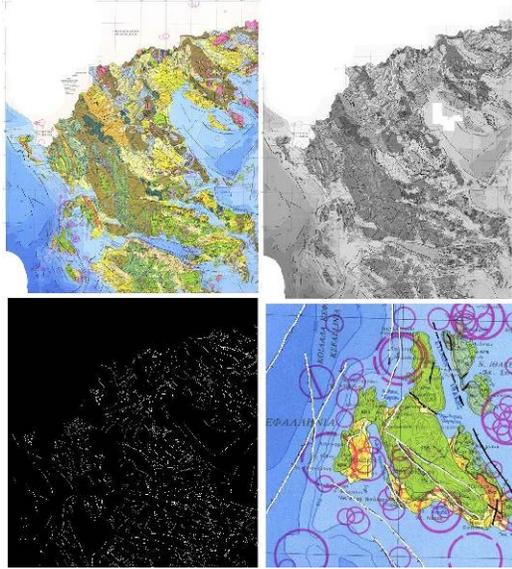


Fig. 4 Partitioning application of the proposed methodology in the north-west part of the Greek vicinity

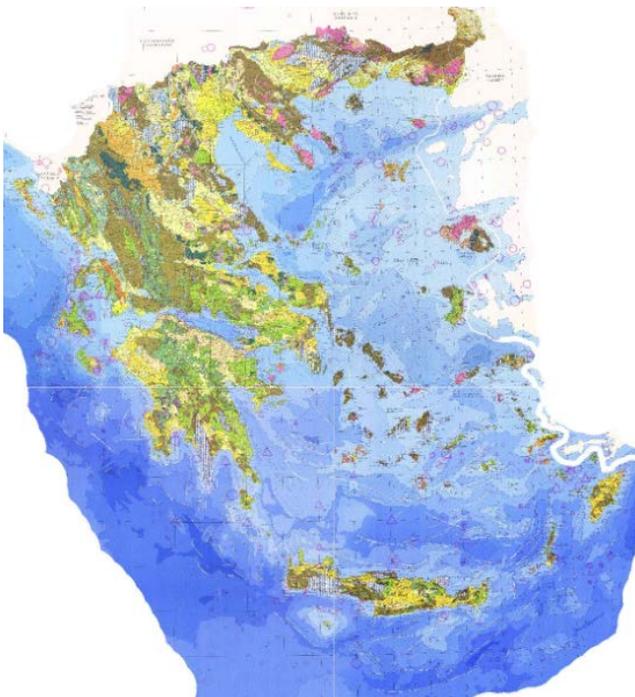


Fig. 5 All processed partitions reassembled together to comprise a digital map with underground faults as single objects depicted by white color lines.

All processed partitions are then reassembled together to comprise an overall processed digital map of the entire country where every single underground fault present is manifested as a distinct object. As objects, all spatial information regarding the extent and planar dimension of any underground fault, are readily available as a set of matrix coordinates.

IV. CONCLUSIONS AND FURTHER WORK

The very nature of underground faults makes their identification and distinction in two dimensional maps a complex process. Apart from orientation, which can be a discriminative factor, there is no apparent way to clarify whether two or more underground faults are intercrossing physically or phenomenally, or whether we are viewing branches of a single structure. Nonetheless, import of expert knowledge [1,11] or additional data, such a depth extent where available, can vastly improve the integrity of the classification process.

The benefits from identifying underground faults as distinct objects are numerous. Foremost they can be studied both individually and in relation to other ones traversing the same region. Features such as vertical coordinates and planar expansion are recorded and documented enabling temporal comparisons [12] with previous or later data recordings. If combined with information regarding their depth expansion then it is possible to derive three dimensional models [13,14] and investigate for actual physical interaction [11,15] amongst two or more underground faults and wider underground faults' networks.

ACKNOWLEDGMENT

The Authors would like to thank the Institute of Geology and Mineral Exploration and the IGME 5000 data compilation in map form that is made available to open access in http://www.bgr.de/karten/igme5000/igme5000_finalversion_2005_g.jpg.

REFERENCES

- [1] B. C. Papazachos, "Large seismic faults in the Hellenic Arc", *Annali di Geofisica*, vol. 39, pp. 891–903, 1996.
- [2] M. Laigle, M. Sachpazi and A. Hirn, "Variation of seismic coupling with slab detachment and upper plate structure along the western Hellenic subduction zone", *Tectonophysics*, vol. 391, pp. 85–95, 2004.
- [3] G. Georgoulas, A. Konstantaras, E. Katsifarakis, C. D. Stylios, E. Maravelakis and G. J. Vachtsevanos, "Seismic-Mass" Density-based Algorithm for Spatio-Temporal Clustering', *Expert Systems with Applications*, vol. 40 (10), pp. 4183–4189, 2013.
- [4] A. Konstantaras, "Expert knowledge-based algorithm for the dynamic discrimination of interactive natural clusters", *Earth Science Informatics*, 2015, DOI 10.1007/s12145-015-0236-0, In Press
- [5] U. S. Allan, "Model for hydrocarbon migration and entrapment within faulted structures", *American Association of Petroleum Geologists Bulletin*, vol. 73 (7), pp. 803-811, 1989.
- [6] D. Shafer, D. "Rig safety and reliability incidents caused by software: How they could have been prevented", *SPE Americas E and P*

- Environmental and Safety Conference*, San Antonio, TX; United States, pp. 109-115, 2009.
- [7] M. D. Sarder, C. Miller and B. Richard, "Modeling Logistical & Economical Impact of Oil Spill on the Gulf Coast", *61st Annual Conference and Expo of the Institute of Industrial Engineers*; Reno, NV; United States, 2011.
- [8] N. Skliris, S. Sofianos, A. Lascaratos, "Hydrological changes in the Mediterranean Sea in relation to changes in the freshwater budget: A numerical modelling study", *Journal of Marine Systems*, vol. 65, pp. 400-416, 2007.
- [9] K. Jacobson, "Petroleum production and exploration", *Acta Universitatis Upsaliensis Uppsala*, 2012.
- [10] IGME 1989, "Seismotectonic map of Greece", Institute of Geological and Mineral Exploration, 1989.
- [11] A. J. Konstantaras, "Classification of Distinct Seismic Regions and Regional Temporal Modelling of Seismicity in the Vicinity of the Hellenic Seismic Arc", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6 (4), pp. 1857-1863, 2013.
- [12] A. Konstantaras, E. Katsifarakis, E. Maravelakis, E. Skounakis, E. Kokkinos, E. Karapidakis, "Intelligent Spatial-Clustering of Seismicity in the Vicinity of the Hellenic Seismic Arc", *Earth Science Research*, vol. 1 (2), pp. 1-10, 2012.
- [13] E. Maravelakis, E. Bilalis, I. Mantzorou, A. Konstantaras and A. Antoniadis, "3D modelling of the oldest olive tree of the world", *International Journal of Computational Engineering Research*, vol. 2 (2), pp. 340-348, 2012.
- [14] E. Maravelakis, A. Konstantaras, A. Kritsotaki, D. Angelakis, M. Xinogalos, "Analysing user needs for a unified 3D metadata recording and exploitation of cultural heritage monuments system", *Advances in Visual Computing*, vol. 8034, pp. 138-147, 2013.
- [15] A. Konstantaras, F. Valianatos, M.R. Varley and J.P. Makris, "Soft computing modelling of seismicity in the Southern Hellenic arc", *IEEE Geosciences and Remote Sensing Letters*, vol. 5 (3), 2008.

Advantages and Disadvantages of Family Entrepreneurship and How to Prevent Distress: Evidence from the Czech Republic

Ondřej Machek, Petra Votavová

Abstract—The relationship between family control and financial performance of firms has been subject to numerous studies in the past literature. However, the academic literature has been particularly silent on family business issues in the countries of the former Eastern Bloc in Europe. This study aims to fill the gap in the past quantitative research focused on Czech family business firms and reflects the need for reasons explaining the differences between family and non-family firms. Based on a review of literature and semi-structured interviews with family firm founders and employees, we summarize the main advantages and disadvantages of family control over firms in the Czech Republic and propose measures to prevent failure of family firms including succession issues. We also identified multiple gaps in contemporary family business research.

Keywords—Family business, advantages, disadvantages, succession, interviews, Czech Republic

I. INTRODUCTION

Family firms, their particularities and differences from non-family businesses have received considerable academic attention especially in the last three decades. According to various estimates, they represent a major part of economies in developed as well as developing countries. Family business research has been focused on family business definition, performance gaps between family and non-family firms, but also on social and psychological aspects of family firms. However, the academic literature has been particularly silent on family business issues in the countries of the former Eastern Bloc in Europe (Czech Republic, Slovakia, Poland, Hungary, among others).

Several quantitative studies have already been carried out, finding differences between family and non-family firms. However, it is also necessary to ask the “why” and “how” questions. Such kind of empirical investigation requires adopting a qualitative approach.

This article is based on semi-structures interviews with family firm founders and employees, including the family and

non-family ones. The research questions we ask are: What are the main advantages and disadvantages of family control over a firm? What measures can be proposed to avoid distress of family firms? At what are the main challenges for future research?

The article is organized in the following manner: first, a review of related literature is carried out. Then, we introduce the methods we use to explore the above-mentioned research questions. Subsequently, a discussion is presented to explain the major findings and provide concluding remarks.

II. LITERATURE REVIEW

The importance of family business as an academic discipline is due to the high number of family firms, which account for a major share of economies around the world [1]. The literature review presented in this article is focused on the following topics: the definition of family firms, differences between family and non-family firms, and current empirical findings on family firms in the Czech Republic.

A. Definition of Family Firms

A whole lot of possible definitions can be found in the past literature. They can be divided into “essence” and “involvement” criteria [2].

The “essence” approach includes the “intention for succession”, self-identification as a family business, or behavioural aspects (“familiness”) as distinguishing factors of family firms ([3]; [4]). Obviously, it is quite difficult to treat such definitions in a quantitative manner. Perhaps this is one of the reasons that the other class of definitions, “involvement criteria”, are prevailing in the academic literature [2]. Such criteria deal with the involvement of family in different areas of control over a company.

Despite the fact that there is no consensus on what defines a family firm, almost all “involvement” criteria include three dimensions of family control [2]:

1. one or several families hold a significant part of the share capital;
2. family members retain significant control over the company, which depends on the distribution of capital and voting rights among nonfamily shareholders, with possible statutory or legal restrictions;
3. family members hold top management and/or supervisory

This work was supported by the Internal Grant Agency of the University of Economics in Prague, project no F3/55/2015 “Family Businesses in the Czech Republic – performance, governance, succession”.

O. Machek is with the Faculty of Business Administration, University of Economics, Prague, Czech Republic (e-mail: ondrej.machek@vse.cz).

P. Votavová is with the Faculty of Business Administration, University of Economics, Prague, Czech Republic (e-mail: xvotp09@vse.cz).

board positions.

The debate on whether “involvement” criteria are sufficient or they should be accompanied by “essence” criteria is still open and will deserve academic attention in the future. It should also be noted that neither the “essence” nor the “involvement” define what actually constitutes a family. It is still not clear whether the family includes the nuclear family, extended family or segment of the extended family. In Table 1, we present the overview of De Massis et al. [2] of recurring criteria used to define a family business in past studies (until 2012). Obviously, the “involvement” criteria have been by far more prevalent than other criteria.

Table 1: Criteria used to define a family business

Definitional criterion	Frequency (%)
Ownership	79%
Management	53%
Directorship	28%
Self-identification	15%
Multiple generations	9%
Intra-family succession intention	7%

Source: Adopted from De Massis et al. [2].

B. Differences between Family and Non-Family Firms

The past research recognizes that firm-value maximization is not the only goal of family companies [5]. There exists a number of other, family-centered goals [6], such as wealth creation, maintaining socio-emotional wealth [7] and family harmony, as well as providing employment to family members. Besides having different goals, family firms have also been found to be different in terms of long-term orientation (the intention of family business owners to preserve the family inheritance for its transmission to following generations [8]).

While a large number of past studies found superior financial performance of family businesses compared to non-family ones (e.g. [9]), other authors, such as O’Boyle et al. [10] found no significant main effects. According to a recent study, there exists an economically weak, albeit statistically significant, superior performance compared to non-family firms [11]. Besides different goals, performance differences are often explained by agency costs reduction. Since the interests of owners and hired managers are different, managers may act in order to maximize their own utilities instead of those of the shareholders [12]. This ambiguity can be mitigated in the case of family firms. However, other authors suggest that with family altruism and conflict between majority and minority shareholders, principal conflict can exist, offsetting advantages.

C. Family Businesses in the Czech Republic

The history of Czech family firms has never been summarized in detail. The industrial revolution resulted in the emergence of many manufacturing firms that remained in the control of family after the death of the founder, such as the manufacturing firms of herbal liquor Becher, furniture

manufacturer Thonet, piano manufacturer Petrof, or shoe manufacturer Baťa. After the collapse of the Austro-Hungarian Empire, Czech family firms started to develop rapidly [13] and became the backbone of the Czech economy. After the World War II, the leadership was taken over by the communist party, which resulted in nationalization of Czech family firms, and liquidation of entrepreneurship and private ownership in general. 40 years later, the fall of the Iron Curtain in 1989 represented an important milestone, since private ownership was re-established and restituted. However, family firms had to deal with issues related to economic transformation: economic crime, bad legal environment, inefficient financial sector, bad work ethic, insufficient competitiveness and obsolete technologies.

Czech family businesses have been recently getting a topic of interest especially due to “succession issues”. While by the beginning of nineties we could hardly speak of any family businesses, more than 25 years later it is quite common that owners already have transferred their businesses to their heirs or have at least started considering it. From this viewpoint, the situation in the Czech Republic is no different from the situation of family businesses in other developed European countries.

Czech family businesses have received academic attention only recently and deeper understanding of their nature and significance is still missing. There are no official statistics and only few educational programs and consulting services for family firms.

Family businesses been addressed especially by non-academic press. In 2008, Czech authors Koráb et al. [14] published a book focused on family business. An overview of 50 largest Czech family firms has been published by Forbes in 2014 [15]. According to a recent survey, Czech family firms are seen positively and are associated with tradition and quality [16].

It is estimated that Czech family firms do not differ significantly from non-family firms in terms of industry affiliation [17]. Most of them operate in the manufacturing sector where they have been found to be better performing than non-family firms [18]. A matched-pair investigation of Czech family and non-family businesses has been carried out with a sample of large and medium-sized companies [19] finding that Czech family firms were performing better in terms of profitability. They have also found to carry less debt and to keep more short-term capital [20]. These recent studies, rather quantitative in nature, however lack the answers to the “why” and “how” questions. This represents a research gap, which is the subject of interest of this article.

III. METHODS

In order to explore the main advantages and disadvantages of family firms, and to propose measures to avoid distress of family firms, we adopted a qualitative approach. We used semi-structured individual and group interviews with family firm founders and employees, including non-family members. Such approach enables searching for explanations in an open and confidential way. The interviews had the format of open

questions, which allows for asking of follow up questions to dig more into detail. The research has been carried out in six Czech family firms [21]. The questions were focused on the following issues:

- What advantages and disadvantages of family firms do the respondents see?
- Will family firm owners prefer family employees over non-family employees?
- What are the main threats of family control?
- Is there any particular need of government support?

While six managers were interviewed individually, the regular employees (14 of them, including family and non-family members) have been interviewed in groups. We decided to choose family firms from various industries (accounting, auditing, tax advisory, wholesale, entertainment, and travel) in order to gain a broader view on the topic of interest. While it is clear that the qualitative research does not allow for generalizing the results, it is necessary to collect and develop ideas and explanations that cannot be provided by quantitative surveys.

IV. RESULTS AND DISCUSSION

Since the record of interviews is long, we do not state the full answers, but instead we provide the main findings from the interviews and confront them with prior research in the domain of family business.

A. Advantages of Family Firms

Based on the interviews, it seems that family firms are long-term oriented and care about the future. This is supported by prior research [1]. This is associated with a possible greater stability: even in the bad times, family ties keep their firm running, which can be seen as a “pillow” that non-family firms lack. Stability of family firms in a more general perspective, especially in terms of income and revenue stability, has been reported by multiple researches such as Lee [23]. However, stability and change are sometimes considered to be in a trade-off relationship [24]. Aversion to change can possibly limit the opportunities of growth, which has been reported by some researchers [25].

Another aspect of stability mentioned by the respondents was a lower fluctuation of employees. This is also consistent with prior literature, which reports lower rotation of employees, positive employment atmosphere and unwillingness to fire employees during the times of crisis. A better working environment was also reported by the respondents in our interviews, even by nonfamily employees.

Terms such as trust and friendship have been frequently mentioned. While they contribute to a better working atmosphere, they are also associated with another advantage of family firms mentioned by the interviewees, which is the fact that family ties reduce delinquency and crime. This is also consistent with prior research [26].

Another advantage of family firms mentioned by the respondents is the fact that family owners are more willing to share knowledge and know-how with their employees. This

can be seen as a better dissemination of knowledge (both formal and tacit knowledge) within family firms. This advantage is obvious but has not received much academic attention so far.

Family firms also enjoy a positive reputation in terms of quality and tradition [16], which can positively affect the demand for their products and services.

B. Disadvantages of Family Firms

Conflicts seem to be one of the major drawbacks of family entrepreneurship. While all companies have to deal with interpersonal dynamics, family involvement introduces an additional source of complexity [27]. Conflicts can emerge between husband and wife, as well as between parents and children, between siblings, or between family and non-family employees.

The first kind of conflicts can emerge between parents and their ancestors, and perhaps even often between fathers and sons. In particular, parents and children can have different opinions about leadership and operational tasks (they are supposed to have different opinions and attitudes in general). According to our interviews, children may consider their parents old-fashioned, while parents must be ready to admit that their children can perform better, have a better knowledge of current trends in technology, fashion, society, etc. The respondents also mentioned that parents may have too high expectations: children will possibly not be as good as their parents expect them to be. Indeed, control by heirs has been often associated by a lower profitability or growth in the past literature [28].

Conflicts may also arise between siblings (especially due to unequal emotional and material treatment), where a possible competition between siblings may occur, and between family and non-family members – in particular, non-family employees may see negatively parents who give preferential treatment to their children, or favoritism granted to relatives (nepotism).

Conflicts between spouses (husband and wife) but also between generations can be due to the lack of separation of work and family: bringing home work-related problems, lack of boundaries between work and family, and working “24h a day”. Such kind of conflicts have been frequently reported by the literature [29]. Another frequently mentioned source of conflicts is having no hiding place at home and no possibility of being alone, too much togetherness [30].

The respondents also mentioned that family firm managers cannot afford to be too strict when dealing with their children or spouses. Unpleasant things are not easy to tell when dealing with own family members. Gustaffson and Norgen [32] mentioned that too strict policies or policies that entirely prevented family ties within the company could harm the company’s way to success in the long run. Such problems can emerge in the case of autocratic leadership style of the family firm founder [33]. However, the question how to balance authority and family ties has not been much discussed in the past research.

Besides conflicts between family members and possible nepotism, there are other disadvantages seen by the interviewees. Because of emotional ties, a distress of a family firm can have a negative causal impact on the whole family.

This, together with the fear of losing control over family firm, may represent one of the sources of the large risk-aversion of family firms frequently reported in the past literature, which is usually accompanied by a reduced level of debt [20] and possibly a greater liquidity. However, according to some authors, such risk-aversion can be easily turned into risk-willing when the risk of losing control of a family over its firm is too high [31].

C. How to Prevent Distress of Family Firms

Possible government support may include, among others, introducing more flexible working hours, job sharing, but also reducing inheritance taxes, supporting start-ups, or providing easier access to long-term financing. Surprisingly, the respondents did not mention any particular need of state support. At the same time, the challenges perceived by family firm managers did not differ at all from challenges that non-family firms are currently facing (globalization, internationalization, slower economic growth, etc.) At the same time, their desires are the same as those of non-family firms (simple and stable regulatory environment, low interest rates, flexible labor market, investment in infrastructure, etc.) From this perspective, it is questionable whether the government should take any measures to favor or support family firms.

However, based on the presented discussion, opinions and suggestions of our respondents, it is possible to formulate certain managerial implications to avoid distress of family firms and improve their performance while keeping the family and company together.

Cooperation on a common goal and willingness to participate is the basic prerequisite of family business success. Parents must be ready to admit that their children can perform better, and find the right time of succession.

Succession is a critical point in a family business lifecycle. This moment should occur before the founder's death, which is a basic prerequisite of knowledge and know-how sharing. Management shouldn't be transferred directly. Instead, children should first start to work in regular positions (accounting, marketing) to learn how the business works. Ancestor have to deserve the management roles and learn enough skills prior to taking over management. Besides obtaining the necessary knowledge and skills, children also must be interested and motivated to work in the family firm. If these conditions are unmet, the performance of a family firm is likely to decrease.

Nepotism has to be avoided. Parents should be neither too tolerant nor too autocratic. Surprisingly, requirements on family members are sometimes stricter than those on non-family members. At the same time, emotional support and knowledge sharing can improve performance of both family and non-family employees.

To avoid conflicts and envy, a clear division of roles is necessary. One possible way to do so it is to divide a company into multiple units, or assign to every successor a precisely defined role. Attributing equal shares on ownership to children can also prevent envy, but requires that all children are actually willing to participate on the company management. If

some children don't want to continue the business, but their siblings do, it is necessary to find a way to compensate them.

V. CONCLUSION

The academic literature has been particularly silent on issues of family business in the countries of the former Eastern Bloc. Despite the fact that the long-term tradition of family business in the Czech lands has been interrupted by the nationalization of private ownership after the World War II, after some 25 years after the fall of the Iron Curtain, many successful family businesses have emerged or been re-established, such as Metalimex, Kofola, Baťa, Koh-i-noor, or Petrof.

A few studies carried out in the past found that Czech family firms are financially different from their non-family counterparts. While quantitative research is suitable for investigating the differences between family and non-family firms, a qualitative approach is necessary to find the answers to "how" and "why" questions. This article was based on individual and group interviews with owners, employees and non-family members of six Czech family firms operating in various industries.

We presented the main advantages and disadvantages of family firms, and the ways how to prevent nepotism, conflicts and to prepare conditions for a successful succession of a family firm to the following generation. The findings have been confronted with existing empirical findings. Surprisingly, family firm managers didn't perceive any particular need for government support.

The research also identified research gaps, which deserve more academic attention. In particular, the future research should focus on the following question: How exactly is knowledge and know-how disseminated in family firms and how does this process differ from non-family firms? How to improve leadership in order to avoid damaging family ties while being strict enough? And how to compensate children who don't want to participate in the family firm in order to prevent envy and conflicts with their siblings?

REFERENCES

- [1] N. Kachaner, G. Stalk, and A. Bloch, "What you can learn from family business," *Harvard Business Review*, vol. 90, no. 11, pp. 102-106, Nov. 2012.
- [2] A. Massis de, P. Sharma, J. H. Chua, and J. J. Chrisman, *Family Business Studies – an annotated bibliography*. Cheltenham, UK: Edward Elgar Publishing, 2012.
- [3] J. H. Chua, J. J. Chrisman, and P. Sharma, "Defining the family business by behavior," *Entrepreneurship theory and practice*, vol. 23, pp. 19-40, Dec. 1999.
- [4] T. G. Habbershon, and M. Williams, "A resource-based framework for assessing the strategic advantages of family firms," *Family Business Review*, vol. 12, no. 1, pp. 1-25, March 1999.
- [5] P. Sharma, J. J. Chrisman, and J. H. Chua, *A Review and Annotated Bibliography of Family Business Studies*. Assinippi Park, Norwell, MA: Kluwer Academic Publishers, 1996.
- [6] K. Stafford, K. A. Duncan, S. Dane, and M. Winter, "A research model of sustainable family businesses," *Family Business Review*, vol. 12, no. 3, pp. 197-208, Sept. 1999.

- [7] P. Berrone, C. Cruz, L. R. Gomez-Mejia, "Socioemotional Wealth in Family Firms: Theoretical Dimensions, Assessment Approaches, and Agenda for Future Research," *Family Business Review*, vol. 25, no. 3, pp. 258-279, Sept. 2012.
- [8] H. S. James, "Owner as manager, extended horizons and the family firm," *International Journal of the Economics of Business*, vol. 6, no. 1, pp. 41-55, 1999.
- [9] R. Anderson, and D. Reeb, "Founding family ownership and firm performance: Evidence from the S&P 500," *Journal of Finance*, vol. 58, no. 3, pp. 1301-1328, June 2003.
- [10] E. H. O'Boyle, J. M. Pollack, and M. W. Rutherford, "Exploring the relation between family involvement and firms' financial performance: A meta-analysis of main and moderator effects," *Journal of Business Venturing*, vol. 27, no. 1, pp. 1-18, Jan. 2012.
- [11] D. Wagner, J. H. Block, D. Miller, Ch. Schwens, G. Xi, "A meta-analysis of the financial performance of family firms: Another attempt," *Journal of Family Business Strategy*, vol. 6, no. 1, pp. 3-13, March 2015.
- [12] E. F. Fama, and M. C. Jensen, "Separation of ownership and control," *Journal of Law and Economic*, vol. 26, pp. 301-326, June 1983.
- [13] A. Hanzelková, "Re-establishing Traditional Czech Family Businesses," Ph.D. dissertation thesis. School of Business and Economics, University of Jyväskylä, 2004.
- [14] V. Koráb, A. Hanzelková, and M. Mihalisko, *Rodinné podnikání*. Brno: Computer Press, 2008.
- [15] Mašek, J. et al. 50 největších rodinných firem. *Forbes*, 05/2014.
- [16] UnicreditBank (2015, April 29). Chut' malých a středních firem investovat roste, průměrný úvěr dosahuje 6 milionů korun. [Online]. Available: <https://www.unicreditbank.cz/web/novinky/chut-malych-a-strednich-firem-investovat-roste-prumerny-uver-dosahuje-6-milionu-korun>.
- [17] J. Hnilica, O. Machek, and M. Hanuška, "The Significance and Profile of Large and Medium- Sized Family Businesses in the Czech Republic," in *Proc. Managing and Modelling of Financial Risks. 7th International Scientific Conference*, 8. - 9. 9. 2014, Ostrava, 2014, pp. 256-264.
- [18] O. Machek, J. Hnilica, and D. Kolouchová, "The Impact of Family Control on Profitability, Leverage and Liquidity: Evidence from the Czech Manufacturing Industry," in *Proc. of the 7th International Scientific Conference Finance and Performance of Firms in Science, Education and Practice*, 23. - 24. 4. 2015. Zlín, 2015, pp. 883-892.
- [19] O. Machek and J. Hnilica, "Evaluating the Impact of Family Presence in Ownership and Management on Financial Performance of Firms Using Matched-Pair Investigation," *Politická ekonomie*, vol. 63, no. 3, pp. 347-362, 2015.
- [20] O. Machek and J. Hnilica, "The Relationship between Capital Structure and Family Control: Evidence from the Czech Republic," *International Journal of Economics and Statistics*, vol. 3, no. 1, pp. 9-14, 2015.
- [21] P. Votavová, "The issue of family business in Czech Republic," Bachelor thesis. Prague: University of Economics, Prague, 2015.
- [22] D. Miller and I. Le Breton-Miller, *Managing for the Long Run: Lessons in Competitive Advantage from Great Family Businesses*. Boston: Harvard Business Press, 2005.
- [23] J. Lee, "Family Firm Performance: Further Evidence," *Family Business Review*, vol. 19, no. 2, pp. 103-114, 2006.
- [24] P. Poutziouris, K. Smyrniotis, K., and S. Klein, *Handbook of research on family business*. Edward Elgar Publishing, 2008.
- [25] C. M. Daily, and M. J. Dollinger, "An Empirical Examination of Ownership Structure in Family and Professionally Managed Firms," *Family Business Review*, vol. 5, no. 2, pp. 117-136, 1992.
- [26] J. P. Wright, F. T. Cullen, and J. T. Miller, "Family social capital and delinquent involvement", *Journal of Criminal Justice*, vol. 29, no. 1, pp. 1-9, 2001.
- [27] R. K. Mitchell, E. A. Morse, and P. Sharma, "The transacting cognitions of nonfamily employees in the family businesses setting", *Journal of Business Venturing*, vol. 18, no. 4, pp. 533-551, 2003.
- [28] M. Bennesen, K. M. Nielsen, F. Perez-Gonzalez, and D. Wolfenzon, "Inside the Family Firm: The Role of Families in Succession Decisions and Performance," *The Quarterly Journal of Economics*, vol. 122, no. 2, pp. 647-691, 2007.
- [29] K. Gersick, J. Davis, M. Hampton, and L. Lansberg, *Generation to Generation: Life Cycles of the Family Business*. Boston, MA: Harvard Business Press, 1997.
- [30] J. A. Cox, K. K. Moore, P. M. Van Auken, "Working Couples in Small Business," *Journal of Small Business Management*, vol. 22, no. 4, pp. 24-30, 1984.
- [31] L. R. Gómez-Mejía, K. T. Haynes, M. Núñez-Nickel, K. J. Jacobson, and K. Moyano-Fuentes, "Socioemotional wealth and business risks in family-controlled firms: Evidence from Spanish olive oil mills," *Administrative science quarterly*, vol. 52, no. 1, pp. 106-137, 2007.
- [32] C. Gustafsson, H. Norgren, *Nepotism perceived by managers in northern Sweden: An explorative study on attitudes towards nepotism and its usage*. Umea: Umeå School of Business and Economics, 2014.
- [33] M. Ram, R. Holliday, "Relative merits: Family culture and kinship in small firms," *Sociology*, vol. 27 no. 4, pp. 629-648, 1993.

Design Sequences over the Finite Field of Order Four with High Linear Complexity and Arbitrary Even Period

Vladimir Edemskiy

Abstract—We present a method of constructing sequences over the finite field of order four with high linear complexity and arbitrary even period. These sequences are obtained using generalized Ding-Helleseth cyclotomy.

Index Terms—Linear complexity, finite field, sequences

I. INTRODUCTION

FOR cryptographic applications, the linear complexity (L) of a sequence is an important merit factor [2], [5]. It may be defined as the length of the shortest linear feedback shift register that is capable of generating the sequence. The feedback function of this shift register can be deduced from the knowledge of just $2L$ consecutive digits of the sequence. Thus, it is reasonable to suggest that "good" sequences have $L > N/2$ (where N denotes the period of the sequence) [8].

Recently, series of papers have examined the sequences over the finite field of order four \mathbb{F}_4 with high linear complexity. In particular, the sequences of even periods ($2p, 2p^n, 2pq$) with high linear complexity were studied in [4], [7], [1] (see also references therein). Authors of the above-mentioned article refer to these sequences as quaternary. At the same time, the number of researches believe that this name can only be used for sequences with terms $0, 1, 2, 4$ ($\pm i, \pm 1$). Also using the Gray map for these sequences we may easily obtain sequences over \mathbb{F}_4 and vice versa.

In this paper we propose a method of constructing sequences over \mathbb{F}_4 with high linear complexity and arbitrary even period. These sequences are obtained using the generalized Ding-Helleseth cyclotomy [3]. The use of classical cyclotomic classes and generalized cyclotomic classes to construct sequences is an important method for sequence design.

II. PRELIMINARIES

In this section we give a partition ring of residue classes. Let N be an even integer and $N = 2^m n$, where $\gcd(n, 2) = 1$. Then $n = p_1^{e_1} \cdots p_t^{e_t}$, when p_1, \dots, p_t are pairwise distinct odd primes. First of all, we briefly recall the definition of Ding-Helleseth generalized cyclotomic classes of order two [3].

Let Z_n be a ring of residue classes modulo n . According to the Chinese Remainder Theorem

$$Z_n \cong Z_{p_1^{e_1}} \times \dots \times Z_{p_t^{e_t}}$$

V. Edemskiy is with the Department of Applied Mathematics and Information Science, Novgorod State University, Veliky Novgorod, Russia, 173003 e-mail: Vladimir.Edemskiy@novsu.ru.

This work was supported by the Ministry of Education and Science of Russia as a part of state-sponsored project no 1.949.2014/K.

relatively to isomorphism $\varphi(x) = (x \bmod p_1^{e_1}, \dots, x \bmod p_t^{e_t})$ [6]. Here and hereafter $x \bmod n$ denotes the least non-negative integer that is congruent to x modulo n .

It is well known that there exists a primitive root g_i modulo $p_i^{e_i}$ [6]. Let $D_0^{(p_i^{e_i})} = \{g_i^{2j} | j \in Z\}$ be the subgroup of $Z_{p_i^{e_i}}^*$, generated by g_i^2 , and $D_1^{(p_i^{e_i})} = g_i D_0^{(p_i^{e_i})}$, where the arithmetic is that of $Z_{p_i^{e_i}}$, $i = 1, 2, \dots, t$.

Let $\mathbf{a} = (a_1, \dots, a_t)$ be a nonzero vector from $(Z_2)^t$ and

$$I_0^{(\mathbf{a}, n)} = \left\{ (i_1, \dots, i_t) \in (Z_2)^t \mid \sum_{k=1}^t i_k a_k = 0 \right\},$$

$$I_1^{(\mathbf{a}, n)} = (Z_2)^t \setminus I_0^{(\mathbf{a}, n)}.$$

By definition, put [3]

$$E_j^{(\mathbf{a}, n)} = \prod_{(i_1, \dots, i_t) \in I_j^{(\mathbf{a}, n)}} D_{i_1}^{(p_1^{e_1})} \times \dots \times D_{i_t}^{(p_t^{e_t})}$$

and

$$D_j^{(\mathbf{a}, n)} = \varphi^{-1} \left(E_j^{(\mathbf{a}, n)} \right), j = 0, 1.$$

From our definition it follows that [3]

$$Z_n^* = D_0^{(\mathbf{a}, n)} \cup D_1^{(\mathbf{a}, n)}, \quad D_0^{(\mathbf{a}, n)} \cap D_1^{(\mathbf{a}, n)} = \emptyset. \quad (1)$$

$D_0^{(\mathbf{a}, n)}$ and $D_1^{(\mathbf{a}, n)}$ are called generalized cyclotomic classes of order 2 with respect to \mathbf{a} and n [3].

Further, by [3] we have a partition

$$Z_n \setminus \{0\} = \bigcup_{d|n, d>1} \frac{n}{d} Z_d^* \quad (2)$$

Let $d > 1$ be a positive integer and $d|n$, and the nonzero vector $\mathbf{a}_d = (a_1^{(d)}, \dots, a_m^{(d)}) \in (Z_2)^m$, where m is a number of different prime numbers participating in the factorization d . By (1) and (2) we obtain

$$Z_n \setminus \{0\} = \bigcup_{d|n, d>1} \frac{n}{d} \left(D_0^{(\mathbf{a}_d, d)} \cup D_1^{(\mathbf{a}_d, d)} \right).$$

Let

$$C_0 = \bigcup_{d|n, d>1} \frac{n}{d} D_0^{(\mathbf{a}_d, d)} \quad \text{and} \quad C_1 = \bigcup_{d|n, d>1} \frac{n}{d} D_1^{(\mathbf{a}_d, d)}.$$

Then $\{C_0, C_1\}$ is a partition of $Z_n \setminus \{0\}$, i.e. $Z_n = C_0 \cup C_1 \cup \{0\}$ and $C_0 \cap C_1 = \emptyset$.

Using Ding-Helleseth cyclotomy, we obtain a partition of Z_N . The ring of residue classes $Z_{2^m n} \cong Z_{2^m} \times Z_n$ relatively

to isomorphism $\phi(x) = (x \bmod 2^m, x \bmod n)$. Put, by definition

$$H_{j,i} = \phi^{-1}(\{j\} \times C_i), j = 0, \dots, 2^m - 1; i = 0, 1.$$

Here we have a partition

$$Z_N = \{0, n, \dots, (2^m - 1)n\} \cup \bigcup_{j=0}^{2^m-1} H_{j,0} \cup H_{j,1},$$

$$H_{j,i} \cap H_{l,k} = \emptyset \text{ for all } j \neq l, i \neq k.$$

In the following section we construct sequences with high linear complexity using this partition.

III. THE DESIGN OF SEQUENCES WITH HIGH LINEAR COMPLEXITY

Let $\mathbb{F}_4 = \{0, 1, \mu, \mu + 1\}$ be a finite field of order four, where μ satisfies $\mu^2 = 1 + \mu$. By assigning the elements of \mathbb{F}_4 to each of generalized cyclotomic classes with respect to Z_N , one obtains a quaternary sequence of length N naturally. However, in order to guarantee that the constructed sequences have high linear complexity, one should do it specially.

It is well known that if $\{s_i\}$ is a sequence with period N , then the minimal polynomial $m(x)$ and the linear complexity L of this sequence is defined by

$$m(x) = (x^N - 1) / \gcd(x^N - 1, S(x)),$$

$$L = N - \deg \gcd(x^N - 1, S(x)), \quad (3)$$

where $S(x) = s_0 + s_1x + \dots + s_{N-1}x^{N-1}$.

In our case $N = 2^m n$, hence over \mathbb{F}_4 we have

$$L = N - \deg \gcd((x^n - 1)^{2^m}, S(x)). \quad (4)$$

Let α be a primitive n th root of unity in the extension of \mathbb{F}_4 . Then, according to (3) and (4), in order to find the minimal polynomial and the linear complexity of $\{s_i\}$ it is sufficient to find the zeros of $S(x)$ in the set $\{\alpha^v, v = 0, 1, \dots, n-1\}$ and determine their multiplicity. In order to investigate the values of $S(\alpha^v)$, let us introduce subsidiary polynomials. Let $S_A(x) = \sum_{i \in A} x^i$, where A is a subset of Z_n or Z_N .

Lemma 1: If $1 \leq v \leq n-1$ then $S_{C_0}(\alpha^v) + S_{C_1}(\alpha^v) = 1$.

Proof: From our definition it follows that $S_{C_0}(\alpha^v) + S_{C_1}(\alpha^v) = \sum_{i=1}^{n-1} \alpha^{vi}$. ■

Lemma 2: If $0 \leq v \leq n-1$ then $S_{H_{j,i}}(\alpha^v) = S_{C_i}(\alpha^v)$.

Proof: By definitions $S_{H_{j,i}}(\alpha^v) = \sum_{i \in H_{j,i}} \alpha^{vi}$ and $H_{j,i} \bmod p = C_i$. This completes the proof of Lemma 2. ■

A. The sequences with a period $2n$

Let a, b, c, d belong to \mathbb{F}_4 and a, b, c, d are pairwise distinct. We construct a quaternary sequence with the first $2n$ terms of sequence $\{s_i\}$ defined as

$$s_i = \begin{cases} 0, & \text{if } i = 0, \\ a, & \text{if } i \in H_{0,0}, \\ b, & \text{if } i \in H_{0,1}, \\ c, & \text{if } i \in H_{1,0}, \\ d, & \text{if } i \in H_{1,1}, \\ e, & \text{if } i = n. \end{cases} \quad (5)$$

for $e \neq c + d$. The sequence defined by (5) is balanced for $e \neq 0$.

Remark 3: If $n = p^k$ then this sequence equals the sequence from [7] for $p \equiv \pm 1 \pmod{8}$ and when replacing $\{c, d\}$ with $\{d, c\}$ for $p \equiv \pm 3 \pmod{8}$.

Theorem 4: Let $\{s_i\}$ be defined by (5) for $e \neq c + d \in \mathbb{F}_4, e \neq 0$. Then $L = N$ and $m(x) = x^N - 1$.

Proof: Let us show that $S(\alpha^v) \neq 0$ for $v = 0, 1, \dots, n-1$. By (5) we have

$$S(x) = ex^N + a \sum_{i \in H_{0,0}} x^i + b \sum_{i \in H_{0,1}} x^i + c \sum_{i \in H_{1,0}} x^i + d \sum_{i \in H_{1,1}} x^i.$$

Let $1 \leq v \leq n-1$. By Lemma 2 we obtain

$$S(\alpha^v) = e + aS_{C_0}(\alpha^v) + bS_{C_1}(\alpha^v) + cS_{C_0}(\alpha^v) + dS_{C_1}(\alpha^v)$$

or by Lemma 1 $S(\alpha^v) = e + (a + b + c + d)S_{C_0}(\alpha^v) + b + d$. By definition $a + b + c + d = 0 + 1 + \mu + \mu^2 = 0$. Thus, the above expression is equivalent to following $S(\alpha^v) = e + b + d$. So, $S(\alpha^v) \neq 0, 1 \leq v \leq n-1$.

To conclude the proof, it remains to note that $S(1) = e + (a + b + c + d)(n-1)/2$. ■

In conclusion of the subsection we say a couple of words about the case when $e = b + d$. Here $\alpha^v S'(\alpha^v) = b + (a + b)S_{C_0}(\alpha^v)$. So, in general case when $n \neq p^k$ it is possible that $|S'(\alpha^v)| = 0, 1 \leq v \leq n-1 > (n-1)/2$, i.e., L can be less than $(n+3)/2$ ($L \geq (n+3)/2$ for $n = p^k$ [7]).

B. The sequences with a period $4n$

Let as earlier, a, b, c, d be a permutation of the elements of \mathbb{F}_4 . We consider a sequence $\{s_i\}$ defined by

$$s_i = \begin{cases} a, & \text{if } i \in H_{0,0} \cup H_{1,0} \cup \{0\}, \\ b, & \text{if } i \in H_{2,0} \cup H_{0,1} \cup \{n\}, \\ c, & \text{if } i \in H_{3,0} \cup H_{1,1} \cup \{2n\}, \\ d, & \text{if } i \in H_{2,1} \cup H_{3,1} \cup \{3n\}. \end{cases} \quad (6)$$

The sequence defined by (6) is balanced.

Theorem 5: Let $\{s_i\}$ be defined by (6). Then $L = N - 1$ and $m(x) = (x^N - 1)/(x - 1)$.

Proof: Similarly as in Theorem 4, by Lemmas 1 and 2 we obtain $S(\alpha^v) = b + c$. Hence, $S(\alpha^v) \neq 0$ for $v = 0, 1, \dots, n-1$. By definition $S(1) = (a + b + c + d)(n+1) = 0$.

Further, $xS'(x) = a \sum_{i \in H_{1,0}} x^i + bx^n + c \sum_{i \in H_{3,0}} x^i + c \sum_{i \in H_{1,1}} x^i + d \sum_{i \in H_{3,1}} x^i + dx^{3n}$. So, $S'(1) = (a + d)(n-1)/2 + b + d$, i.e. $S'(1) \neq 0$. ■

C. The general construction

Let j_1, j_2, j_3, j_4 be pairwise distinct integers between 0 and $2^m - 1$. We consider a subsidiary subsequence $\{t_i\}$ defined as

$$t_i = \begin{cases} a, & \text{if } i \in H_{j_1,0} \cup H_{j_2,1} \cup \{j_1 n\}, \\ b, & \text{if } i \in H_{j_2,0} \cup H_{j_3,1} \cup \{j_2 n\}, \\ c, & \text{if } i \in H_{j_3,0} \cup H_{j_4,1} \cup \{j_3 n\}, \\ d, & \text{if } i \in H_{j_4,0} \cup H_{j_0,1} \cup \{j_4 n\}, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Put, by definition $F_t(x) = \sum_{l=0}^{N-1} t_l x^l$.

Lemma 6: If $0 \leq v \leq n-1$ then $F_t(\alpha^v) = 0$.

Lemma 6 may be proved similarly as Theorem 4.

Now we will give a general definition of sequence with a period $N = 2^m n, m > 2$. If $m > 2$ then we can take the partition $\{4, \dots, 2^m - 1\} = \cup_{k=1}^{2^{m-2}-1} I^{(k)}$, where $I^{(k)} = (j_1^{(k)}, j_2^{(k)}, j_3^{(k)}, j_4^{(k)})$ and $\{j_l^{(k)}, l = 1, 2, 3, 4; k = 1, \dots, 2^{m-2} - 1$ are pairwise distinct integers between 4 and $2^m - 1$. Put, by definition

$$u = s + \sum_{k=1}^{2^{m-2}-1} t^{(k)} \quad (8)$$

where s is defined by (6) and $t^{(k)}$ is defined by (7) for $I^{(k)} = (j_1^{(k)}, j_2^{(k)}, j_3^{(k)}, j_4^{(k)})$. By Theorem 5 and Lemma 6 we obtain the following statement.

Theorem 7: Let $\{u_i\}$ be defined by (8). Then $L = N - 1$ and $m(x) = (x^N - 1)/(x - 1)$.

IV. CONCLUSION

In this paper, we propose a method of constructing sequences over the finite field of order four with high linear complexity and arbitrary even period. These sequences are obtained using generalized Ding-Helleseth cyclotomy. We generalize design of the sequences over the finite field of order four proposed by Ke et al. [7]. The sequences with high linear complexity are significant for cryptographic applications.

July 25, 2015

REFERENCES

- [1] Z. Chang, D. Li. "On the linear complexity of quaternary cyclotomic sequence of length $2pq$ ". *IEICE Transactions on Fundamentals*, vol. E97-A(2), pp. 679–684, 2014.
- [2] T.W. Cusick, C. Ding, A. Renvall. *Stream Ciphers and Number Theory*, North-Holland Publishing Co., Amsterdam (1998)
- [3] C. Ding, T. Helleseht. "New generalized cyclotomy and its applications". *Finite Fields and Their Applications*. vol.4(2), pp. 140–166, 1998.
- [4] X. Du, Z. Chen. "Linear complexity of quaternary sequences generated using generalized cyclotomic classes modulo $2p$ ". *IEICE Trans. Fundamentals*, vol. E94-A(5), pp. 1214–1217, 2011.
- [5] S.W. Golomb, G. Gong. *Signal Design for Good Correlation: For Wireless Communications, Cryptography and Radar Applications*. Cambridge University Press (2005)
- [6] K. Ireland, M. Rosen. *A Classical Introduction to Modern Number Theory*, Springer, Berlin (1982)
- [7] P. Ke, S. Zhang. "New classes of quaternary cyclotomic sequence of length $2p^m$ with high linear complexity". *Information Processing Letters*. vol. 112, pp. 646–650, 2012.
- [8] R. Lidl, H. Niederreiter. *Finite Fields*. Addison-Wesley (1983).

Study on starting the high power induction motors with wounded rotor

Ion Vlad , Sorin Enache, Monica A. Enache, Ionut D. Smarandache

Abstract—The paper analyzes the energetic aspects at the rheostatic startup of the existing three-phase induction motors with wounded rotor, used in conveyor belts actuation, characterized by hard starting. The higher moment of inertia of the system significantly increases the startup time, so it results a higher power consumption and a higher operating cost. The study was made on customer request, which proposes a modernization of the start by using an electronically controlled rheostat at the rotor and a three-phase bridge. The paper makes a technical and an economical comparison between the known classic rheostatic starting and the version proposed by the beneficiary. The proposed modernization of the induction motor with wounded rotor that exists at the conveyor, means a controlled starting with lower current shocks, but it results an increased electricity consumption by 10.7% during the start.

Keywords— high power induction motors, startup optimizing.

I. INTRODUCTION

THE basic components in all power systems are the electrical machines. Therefore, the progress in cutting-edge industries is conditioned by developments in electrical machines, so by their performance. It follows, as a necessity, that high power induction motors to have better operating characteristics, higher specific power, and lower overall dimensions and weight [1-4], [8].

These requirements are achieved through optimal designing of the motors, which have high energy efficiency, and through new experimental trials and product quality certification [10-12]. Designers should lay down in a short time an optimal asynchronous motor variant that meets the customer's requirements, such as the producing company to gain marketplace and the investment to be effective [4], [6-7], [9], [12].

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

Ion Vlad is with the University of Craiova, corresponding author; Romania, e-mail: ivlad@em.ucv.ro.

Sorin Enache is with the University of Craiova, Romania, e-mail: senache@em.ucv.ro.

Monica A. Enache is with the University of Craiova, Romania, e-mail: menache@em.ucv.ro.

Ionut D. Smarandache is with the University of Craiova, Romania, email: smarandescu.ionut@yahoo.com

The paper presents an energy analysis for the startup of the high power induction motors with wounded rotor, used in actuation of the machines with high inertia moment.

II. THE MATHEMATICAL MODEL AND THE OBJECTIVE FUNCTION

On a worldwide scale it is imposed an increase in energy efficiency, so that a rational use of electrical energy consumption in industry to be achieved [1-5], [8]. This results in a reduction of the operating expenses by minimizing losses of the active and reactive power at load operation and during the induction motors with wounded rotor startup.

A. Objective function

In the paper it is exposed a study of the rheostatic starting ($k = 1, 2, \dots, n$ steps) for high power induction motors, considering the minimal startup costs criterion [3], [10-11], by using the *objective function*:

$$f(\bar{x}) = C_{ep} = C_{epa} + C_{epr} = \sum T_{pk} (P_{1k,med} c_{el,a} + Q_{1k,med} c_{el,r}) \quad (1)$$

where: C_{ep} – total startup cost, C_{epa} , C_{epr} – active and reactive electricity costs at startup, T_{pk} – operating time at step k of the resistance, $P_{1k,med}$, $Q_{1k,med}$ – the average values of active and reactive power when running on step k , $c_{el,a}$, $c_{el,r}$ – the cost of a kWh for the active or reactive electricity.

The restrictive conditions are those for current limitation during startup: $I_{pmin} = 0,9 * I_{2N}$ – the lower limit and $I_{pmax} = 1,2 * I_{2N}$ – the upper limit. This limitation determines a sufficient starting torque, even at full load. This provides a lower dynamic torque in order to have a slow, longtime start of the conveyor belt, characterized by a high moment of inertia.

B. Rheostatic starting

In order to reduce the voltage between slip rings to an acceptable value, $U_{e20} = 2745$ V, that doesn't pierce the insulation, the designer opted for the Delta connection at the rotor's winding. Because it is a high power induction motor with wounded rotor, was made a rheostatic starting for limiting the inrush current from the power network [3], [11].

Fig.1.a depicts the Delta connection of the rotor's winding (with the additional resistance on phase included) connected in short-circuit during startup.

Knowing the total resistance for a phase of the rotor - $R_{\Delta p}$ (fig. 1.a), and r_2 - own resistance for a phase of the rotor, one

computed the additional resistance $R_{f\Delta p}$ mounted on a phase at Delta connection:

$$R_{f\Delta p} = R_{\Delta p} - r_2 \quad (2)$$

In fig.1.b is depicted the switching from Delta to Star connection, and was made using the relationship:

$$R_{yp} = \frac{R_{\Delta p}}{3} = \frac{r_2 + R_{f\Delta p}}{3} = \frac{r_2}{3} + \frac{R_{f\Delta p}}{3} \quad (3)$$

In fig.1.c is highlighted the rotor's phase resistance $R_{2y}=r_2/3$, supposing it was in Star connection and assuming the conservation condition of the engine operating mode.

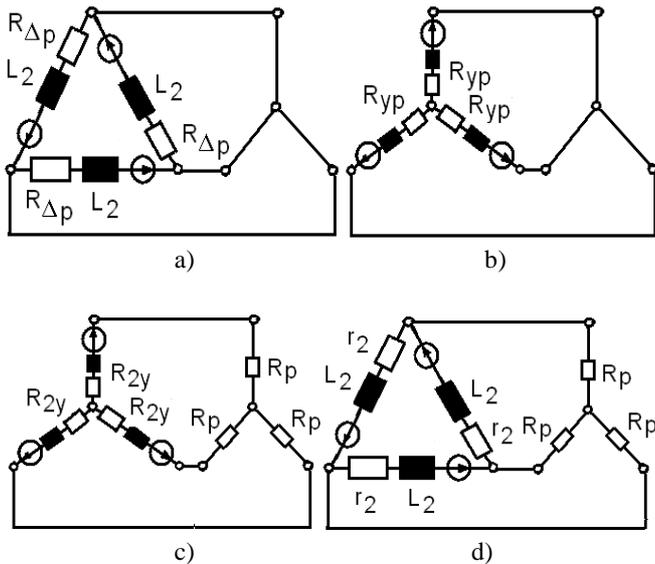


Fig. 1. Connections and transformations necessary to calculate the starting rheostat: a) initial Delta connection with the additional resistance included; b) the transition from Delta to Star connection; c) setting the startup phase resistance R_p ; d) Star-Delta transformation for the rotor's winding.

In this case, for the additional startup resistance in Star connection, results the value:

$$R_p = \frac{R_{f\Delta p}}{3} \quad (4)$$

The rotor passes from Star to Delta connection (fig.1.d), and the receiver is kept in Star connection. With the known transformation relations, results the value:

$$r_2 = 3 \cdot r_{2y} = 3 \cdot \frac{r_2}{3} = r_2 \quad (5)$$

namely own known resistance.

It follows from the results that the total resistance of a rotor's phase $R_{\Delta p}$ is determined according to the literature, and then the receptor's phase resistance R_p in Star connection is determined using (4).

III. RESULTS, SIMULATIONS AND CONCLUSIONS

In the paper are analyzed some aspects regarding the startup optimization of the high power induction motors with wounded rotor, the rheostat sizing, the calculation of energy losses.

Analysis was done on a specific case of a high power and high voltage three-phase induction motor with wounded rotor, used to drive long lengths conveyors.

The rated data of the engine are: $P_N=3800$ kW – rated power, $U_N=6$ kV – rated voltage, $I_{1N}=410$ A, $I_{2N}=1330$ A, $n_1=1000$ rot/min – synchronism speed.

The cost for a rheostatic starting of the engine was made for $c_{el.a}=0.132$ €/kWh - cost of an active electricity kWh, $c_{el.r}=0.036$ €/kVARh - cost of a reactive electricity kVARh.

A redesign of the engine (according to the literature) gave the following results: $\cos\phi_m=0.92$; $\eta_m=0.9665$; $\Sigma p_m=127.3$ kW; $M_{max.m}=2.60 \cdot M_N$; $r_1=0.043$ Ω , $r_2=0.014$ Ω , $X_1=0.762$ Ω , $X_2=0.27$ Ω , $k_r=3.621$ – reference factor.

A. Analysis of the classical rheostatic startup method

The phase current for the rotor [3], [11] is determined by the relationship:

$$I_{2pk} = \frac{s_k U_{e20}}{\sqrt{(R_{\Delta p})^2 + (s_k X_2)^2}} \quad (6)$$

Using (6), the condition $I_{2pk} = I_{pmax}$ and the slipping s_k , corresponding to the transition on the following step of the resistance, it gives the total phase resistance of the rotor $R_{\Delta pk}$. From the condition $I_{2pk} = I_{pmin}$, equation (6) and resistance value $R_{\Delta pk}$, results the slipping s_k , when switching on the next step. In this way the resistance steps of the starting rheostat are calculated.

For the current limits imposed by the beneficiary, I_{pmin} , I_{pmax} , resulted a starting rheostat with 14 steps, the related features being depicted in fig.2.

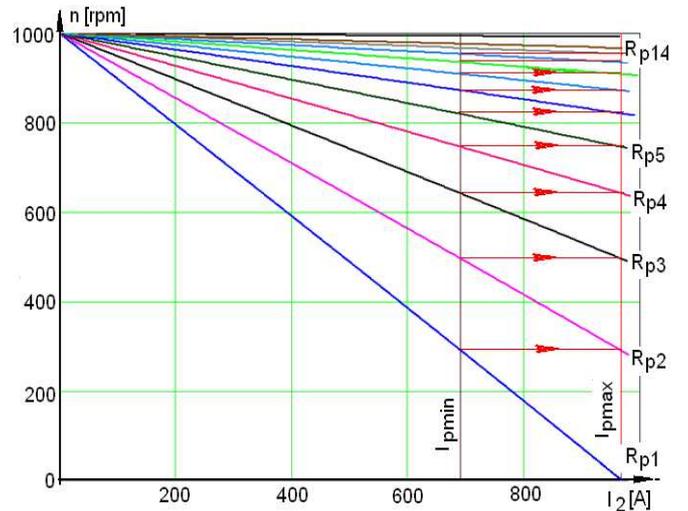


Fig. 2. Rheostatic characteristics necessary for the calculation of the startup rheostat: I_{pmin} and I_{pmax} – limits imposed to the rotor's current.

Further, based on the diagram in fig.3, a detailed study was been made to highlight the startup's specific measurements. It will be noted $R_{pk}=R_k+R_{k+1}+ \dots +R_{14}$ - the phase resistance, corresponding to the startup on step k of the rheostat.

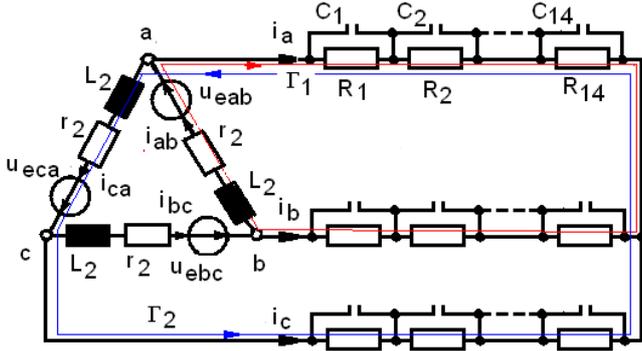


Fig.3. The three-phase electrical scheme of the rotor, the startup step rheostat, and necessary contactors.

The rotor's voltages on all three phases are sine waves. The equations that describe the motor operation on step k of the rheostat, are:

$$\begin{aligned} r_2 i_{ab} + L_2 \frac{di_{ab}}{dt} + R_{pk} i_a - R_{pk} i_b &= u_{eab} \\ r_2 i_{bc} + L_2 \frac{di_{bc}}{dt} + R_{pk} i_b - R_{pk} i_c &= u_{ebc} \\ r_2 i_{ca} + L_2 \frac{di_{ca}}{dt} + R_{pk} i_c - R_{pk} i_a &= u_{eca} \end{aligned} \quad (7)$$

$$i_{ab} - i_{ca} - i_a = 0$$

$$i_{bc} - i_{ab} - i_b = 0$$

$$i_{ca} - i_{bc} - i_c = 0$$

and from these relationships results:

$$\begin{aligned} f_{ab}(u_{eab}, i_{ab}, i_a, i_b) &= \frac{u_{eab} - r_2 i_{ab} - R_{pk} i_a + R_{pk} i_b}{L_2} \\ f_{bc}(u_{ebc}, i_{bc}, i_b, i_c) &= \frac{u_{ebc} - r_2 i_{bc} - R_{pk} i_b + R_{pk} i_c}{L_2} \\ f_{ca}(u_{eca}, i_{ca}, i_c, i_a) &= \frac{u_{eca} - r_2 i_{ca} - R_{pk} i_c + R_{pk} i_a}{L_2} \end{aligned} \quad (8)$$

$$i_{ab} - i_{ca} - i_a = 0$$

$$i_{bc} - i_{ab} - i_b = 0$$

$$i_{ca} - i_{bc} - i_c = 0$$

Using these functions and the 4th order Runge Kutta method [6], [10], it results the numerical solution of the problem (the time variation waveforms of the phase and line currents).

It will be examined the initial startup time $n=0$, (step $k=1$, $s_k=1$), $U_{e20}=1730$ V – the phase emf with the rotor immovable, $r_2= 0.014 \Omega$, $L_2= 0.859$ mH – phase resistance and phase inductance for the rotor.

For current limitation imposed condition $I_{2,p} < I_{pmax}=1.2 \cdot I_{2N}$, results the total phase resistance of the rotor $R_{p1}=R_1+R_2+ \dots +R_{14} = 1.75 \Omega$, (fig.2 and fig.3).

B. Analysis of the startup for the proposed method

Fig. 4 depicts the modernization solution proposed by the beneficiary, where the rotor's phase resistances are replaced with a single electronically controlled resistance, placed at the exit of the rectifier bridge, putting the condition to preserve the value of the line current in the rotor.

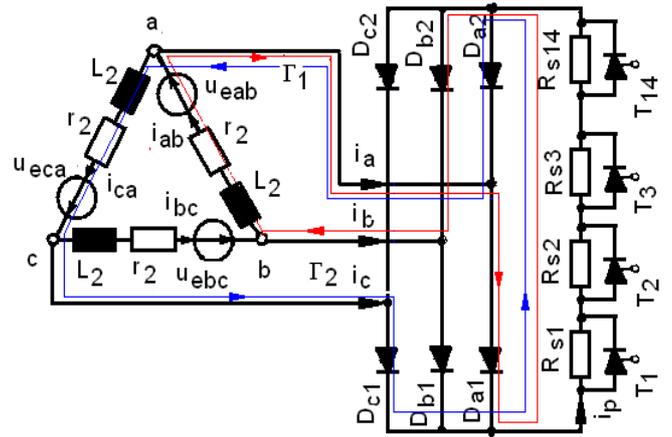


Fig. 4. The three-phase electrical scheme of the rotor, the double bridge, the starting step rheostat and the necessary thyristors.

It will be noted with $R_{ppk}=R_{s_k}+R_{s_{k+1}}+ \dots +R_{s_{14}}$ – the load resistance of the rectifier bridge, corresponding to the starting resistance with step k . The resulting emf, u_{dab} , for the closed circuit containing the rotor's phase ab (fig. 4), when the diodes D_{a1} , D_{b2} respectively D_{a2} , D_{b1} are conducting, is:

$$u_{dab} = \begin{cases} 0 & \omega t = (0 \div 60^\circ) \\ u_{ab} & \omega t = (60^\circ \div 120^\circ) \\ 0 & \omega t = (120^\circ \div 180^\circ) \\ 0 & \omega t = (180^\circ \div 240^\circ) \\ u_{ab} & \omega t = (240^\circ \div 300^\circ) \\ 0 & \omega t = (300^\circ \div 360^\circ) \end{cases} \quad (9)$$

The equations that describe the motor's operation at step k of the rheostat are obtained from (8), where the phase voltages u_{ab} , u_{bc} , u_{ca} are switching with the voltages that occur at the operation of the diode bridge u_{dab} , u_{dbc} , u_{dca} :

$$\begin{aligned} r_2 i_{ab} + L_2 \frac{di_{ab}}{dt} + R_{ppk} i_a &= u_{dab} \\ r_2 i_{bc} + L_2 \frac{di_{bc}}{dt} + R_{ppk} i_b &= u_{dbc} \\ r_2 i_{ca} + L_2 \frac{di_{ca}}{dt} + R_{ppk} i_c &= u_{dca} \end{aligned} \quad (10)$$

$$i_{ab} - i_{ca} - i_a = 0$$

$$i_{bc} - i_{ab} - i_b = 0$$

$$i_{ca} - i_{bc} - i_c = 0$$

B1. The initial moment of the startup

In order to make a fair comparison with the rheostatic startup classical method, it will be analyzed the same starting time $n = 0$, (step $k=1$, $s_k=1$). To comply with the current limitation imposed condition $I_{2,p} < I_{pmax}=1.2 \cdot I_{2N}$, results the total resistance of the diode bridge $R_{pp1}=R_{s1}+R_{s2}+ \dots +R_{s_{14}} = 0.89 \Omega$ (fig. 4).

By solving the differential equations system with the 4th order Runge Kutta method, the numerical solution is obtained and results the time variation waveforms for the phase and line currents. Through numerical methods also, are determined the

effective value of the phase current $I_{ab}=960$ A, the fundamental harmonic value $I_{ab1}=916$ A, the distortion factor $k_{dis}= 30.1\%$ and the harmonic spectrum.

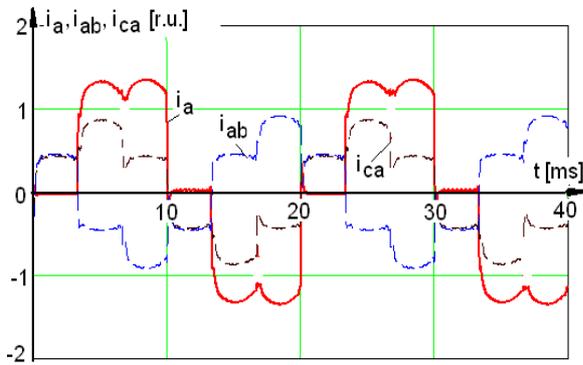


Fig. 5. Phase currents i_{ab} , i_{ca} and line currents i_a waveforms.

The effective value of the line current is $I_a=1670$ A, and for the rectified current it results $I_{red}=4318$ A. Fig. 5 depicts the phase currents i_{ab} , i_{ca} and the resultant line current i_a .

The version proposed by the beneficiary presents a high deformed regime with a rich harmonic spectrum.

C. Energy balance at startup

It will be analyzed, in terms of energy efficiency, the starting rheostatic methods (the classical rheostatic starting, and the one proposed by the beneficiary).

Since the conveyor belt has a high moment of inertia, results a lasting startup with a high consumption of electricity.

C1. Energy balance at the rheostatic startup

In this case all sizes are sinusoidal and a detailed analysis can be made for the motor operation on each step of the starting rheostat R_{pk} ($k=1, 2, 3, \dots, 14$), using the variable $s=(0\div 1)$ and the relationships:

$$I_{I_k}(s) = \sqrt{(I_{10a} + I'_{2ak})^2 + (I_{1\mu} + I'_{2rk})^2} \quad (11)$$

$$\cos \varphi_{I_k}(s) = \frac{I_{10a} + I'_{2ak}}{I_{I_k}} \quad (12)$$

$$P_{I_k}(s) = \sqrt{3} U_N I_{I_k} \cos \varphi_{I_k} \quad (13)$$

where I_{10a} , $I_{1\mu}$, I'_{2ak} , I'_{2rk} represents the active and reactive components for the current at no load operation, respectively of the related rotor current, determined according to the literature [3], [10-11] for the variable $s=(0\div 1)$.

At the rheostatic starting the rotor current must be within the limits $I_{pmin} = 0.9 \cdot I_{2N} = 1197$ A and $I_{pmax} = 1.2 \cdot I_{2N} = 1596$ A. When it reaches the minimum value of the current, is given the command to close the contactor C_i (fig.2, fig.3), corresponding to the operation step.

For the step resistance R_{pk} and the imposed limits (I_{pmax} , I_{pmin}) results the proper slidings s_{k-1} si s_k , respectively the speeds n_{k-1} , n_k . Using the motion equation, the relationship

(14) is obtained which allows us to calculate the operating time on the step resistance R_{pk} , having the speed as variable.

$$TP_k = \sum_{i=n_{k-1}}^{n_k} \frac{2\pi J_t}{M_k(s_i) - M_r} (n_i - n_{i-1}) \quad (14)$$

The total duration of the startup will be:

$$Tp = \sum_{k=1}^{14} TP_k \quad (15)$$

The average value of the active power from the grid when running on step R_{pk} is:

$$P_{I_k med} = \sum_{i=s_{k-1}}^{s_k} \frac{P_{I_k}(s)}{N} \quad (16)$$

N - is the number of points of the range (s_{k-1}, \dots, s_k).

The energy costs during start up will be:

$$C_p = \sum_{k=1}^{14} P_{I_k med} TP_k c_{el,a} \quad (17)$$

C2. Energy balance at the proposed rheostatic startup

Because we have a deformed regime during startup, for the proposed rheostatic starting method, all determinations are made through numerical methods by using the previous relationships.

In fig. 4 we have the rotor winding in Delta connection, the three-phase rectifier bridge, the 14 resistances used at startup and the thyristors that take out of circuit the step resistances.

During the operation at step resistance R_{ppk} , with corresponding slides s_{k-1} and s_k (and speeds n_{k-1} , n_k), the powers P_{I_k} are determined by numerical calculation using the relationships (11), (17).

D. Conclusions

The study performed on customer's demand, aimed to compare functionally and energetically the two methods of rheostatic starting (classical with three-phase rheostat and the proposed method with single-phase electronically controlled rheostat).

In terms of energy, the classical rheostatic startup means a lower cost of energy consumption by 10.7% and a reduced process time by 10.6%. The solution proposed by the beneficiary, with the single-phase electronically controlled rheostat, means a reduction in the amount of the fundamental harmonic for a imposed value of the rotor current. This explains the increased startup time and increased energy consumption for the new method.

REFERENCES

- [1] B.K. Bose, "Power Electronics and Motor Drives Recent Progress and Perspective", *Industrial Electronics, IEEE Transaction on Volume 56, Issue: 2*, 2009, pp.581-588.
- [2] C.U. Brunner, "International Standards for Electric Motors", *Standards for Energy Efficiency of Electric Motor Systems (SEEEM)*, 2007, pp. 6-10.
- [3] A. Campeanu, I. Vlad, S. Enache, L. Augustinov, G. Liuba, I. Cautil, "Optimization of startup characteristics of medium power asynchronous motors with short circuit rotor", *ICEM'2012*, Marseille.
- [4] M. Centner, U. Schäfer, "Machine design software for induction machines," in *Proc. ICEM*, Vilamoura, Portugal, 2008, pp. 1-4.

- [5] J. Faiz, M.B.B. Sharifian, "Optimal design of three-phase Induction Motors and their comparison with a typical industrial motor", *Computers and Electrical Engineering*, vol. 27, 2001, pp. 133-144.
- [6] I. Daniel, I. Munteanu,s.a., *Metode numerice in ingineria electrica*. Bucuresti, Editura Matrix Rom, 1998.
- [7] D. Samarkanov, F. Gillon, P. Brochet, D. Laloy, "Techno-economic Optimization of Induction Machines: an Industrial Application", *ACEMP - Electromotion 2011*, Istanbul –Turkey, 8-10 September 2011, pp. 825-830.
- [8] V.M. Stanciu, M. Cistelean, V. Nitigus, M. Popescu, "Politics for the use of electric machines with high electric efficiency", *Electrical Engineering, Electronics and Automation Journal*, Vol. 53, no. 3, 2005, Bucharest, Romania.
- [9] T. Tudorache, L. Melcescu, "FEM Optimal Design of Energy Efficient Induction Machines", *AECE Journal*, Vol.9, Issue 2, Year 2009, pp.58-64.
- [10] I. Vlad, A. Campeanu, S. Enache, Monica Enache, "Aspects regarding optimal design of high power squirrel cage asynchronous motors", *OPTIM 2012*, pp. 503 – 508.
- [11] I. Vlad, A. Campeanu, S. Enache, *Proiectare asistată a maşinilor asincrone. Probleme de optimizare*, Craiova, Editura Universitaria Craiova, 2011.
- [12] *** CEI 60034-2-1 Standard: "Rotating electrical machines-Part 2-1. Standard methods for determining losses and efficiency from tests", Edition 1.0, 2007.

A Printed Circuit Board Exposure Device Design for Course Projects using LEDs

J. Chatzakis

Abstract— A low cost device for printed circuit board (PCB) exposure is presented. The device uses ultra violet (UV) light emitting diodes (LEDs) and has significant advantages when compared to the conventional devices. The proposed device increases the resolution of the PCB, decreases the PCB exposure time and has higher efficiency. An experimental prototype was constructed and experimental results verify the theoretical expectations.

Keywords—LED, Printed Circuit Board, Exposure Device

I. INTRODUCTION

ELECTRONIC prototyping in universities and small companies almost always requires the construction of a printed circuit board (PCB) [1]. Today, it is easy to produce a high quality mask using a high resolution printer and obtain a photo-resist covered standard copper board (unused PCB) up A4 standard dimensions. The procedure is quite easy and usually involves the following steps:

- 1) Exposure of the photo-resist through the mask using ultraviolet light.
- 2) Development of the photo-resist.
- 3) Etching of the copper.
- 4) Removal of the photo-resist.

Although the removal of the photo-resist can be easily achieved using alcohol, however the resist development and the copper etching use chemicals and are specialized procedures. Engineering laboratories usually do not have the equipment to control accurately the concentration and the temperature of the chemical solutions and this is one of the main reasons for errors in these procedures. The sensitivity on the chemical procedures also depends on how sharp the PCB image is applied to the resist during the resist exposure through the PCB artwork mask.

As well as the mask quality and materials, the artwork transfer depends on the mask to copper contact, the illumination uniformity and the angle of incidence of the UV light onto the resist surface. In the laboratory, the artwork mask is usually a printer transparency. A good quality, laser printer transparency is usually sufficient for the production of the mask. The toner deposited onto the transparency surface

allows for a good contact with the photo-resist and it has high absorption for all types of light. The problem of transparency stretch, caused by heating during the printing process is a problem for some types of transparencies and printers. This problem can have a significant effect on the production of some large and high density double layer PCBs. Furthermore, after some printing, the laser printer drum gets worn and can produce black lines and cuts to the artwork printing. For these reasons some people prefer to use inkjet printers to produce the artwork. Ink is not as opaque as toner and uses a much thicker layer inside the transparency to produce a quality artwork. This degrades the quality of the artwork transfer during the resist exposure since it can affect more the angle of the incident UV light.

A usual method for exposing photo-resist is an UV floodlight positioned a long distance from the PCB. This method has a low efficiency due to reflections and low light fluency and in addition, everything inside the exposure room will be exposed to the UV light. Using this method, it is difficult to expose simultaneously both layers of a double layer PCB.

Many methods for exposing photo-resist exposure have been tested including the use of professional equipment. Even when using professional equipment there are problems such as the warm up time, the equipment cost, the chemicals procedures sensitivity that is caused by the use of printer transparencies under non vertical and non uniform UV light illumination.

II. OPERATION PRINCIPLE

Recent advances in light emitting diode (LED) construction has enabled LEDs to gradually become the preferred choice for many lightning applications [2]. As the production volume of the LEDs has increased their price has reduced and so the cost is no longer an obstacle for their use. LEDs have a lot of advantages in comparison to the conventional lighting methods:

- a) They can easily achieve any color temperature.
- b) They produce stable light almost instantaneously.
- c) They are solid-state (high reliability and vibration resistant).
- d) They can give light within a narrow viewing angle.
- e) They have longer life and higher efficiency than any other competitive light source (especially in violet and UV).

J. Chatzakis is with the Department of Electronics, TEI of Crete, Chania 73133, Crete, Greece (e-mail: xatzakis@chania.teicrete.gr).

In order to use LEDs for photo-resist exposure it is necessary that light is emitted in the range between 350 - 420 nm. A lot of constructors make LEDs in the upper limit of the visible light between 395 – 405 nm in competitive price [3]. These LEDs are produced in large quantities and are intended for vivid car illumination. Their light emittance is suitable for the photo-resist exposure and has other advantages:

- a) A large part of their light is visible and the human eye reacts to it.
- b) The biological effects of their light are not far of those from the visible light.
- c) Normal glass has relatively low absorption in this wavelength.
- d) The viewing angle of “through-hole” LEDs can be narrower than the 20°.

A problem that occurs with the low cost UV LEDs is that they are non-uniform. The problem can be exacerbated if the LED is soldered very close to its body because heat transfer can affect the plastic case. During the development of the experimental prototype was apparent that the vertical incidence of the UV light immunizes sufficiently the photo-resist despite the light intensity variations. Non uniform illumination and exposure time vary significantly without seriously affecting the exposure result. To verify this, a small demonstration was devised. A transparency with various

This experiment, that tests how the angle of the incident UV light affects the exposure result, was not setup with high precision, but the results are very clear. A couple of narrow viewing angle LEDs illuminates each square with a different angle, but with the same distance from the square. On the left of the figure, the angle from the vertical of each couple of LEDs is denoted in degrees. The vertical UV light (on the top of the photograph) in high over-exposure passes from the low dense black ink but still does not affect significantly the thickness of the lines. As the angle from the vertical is increased the lines gradually get thinner and disappear. The line that remains visible is 15 mils wide on the transparency and all the other 10 mils.

In order to design the optical system of the exposure device there are two parameters that have to be minimized: The maximum deflection from the vertical of the effective UV light and the non-uniformity of the illumination.

In reference [2] the definition of the uniformity can be found:

$$Uniformity = \frac{E_{min}}{E_{max}} \tag{1}$$

Where E_{min} and E_{max} are the minimum and the maximum light intensity respectively.

The intensity is ideal uniform when *Uniformity* is equal to 1. The most convenient way to approximate this is to place the LEDs in a grid of squares that are defined by vertical and horizontal lines. This arrangement is in use for uniform illumination and it will be compared to LEDs arranged on the vertices of equilateral triangles. For simplicity in all of the evaluations, LEDs with angle of 30° degrees are assumed, as these that are the type used on the prototype. The LED to LED distance is expressed as d and the design is made to achieve a uniform light in the line between two adjacent LEDs. To achieve this in the middle of the line, the luminosity of each LED is assumed to be half the intensity of the direct, on-axis view. The half intensity angle is denoted $\theta_{1/2}$ and is equal to the one half of the LED viewing angle. In this configuration all the points of the perimeter of the geometrical shape defined by the LEDs have about the same light intensity. The highest intensity variation and the lowest angle of incidence appear in the center of this geometrical shape. The contribution of all the other LEDs is negligible and omitted since the emission angle where the area inside the geometrical shape is reached significant reduces the intensity. In practice, the LED non-uniformity also limits the need for a very accurate expression. For a square geometry we have $l = \frac{d}{\sqrt{2}}$ and for an equilateral triangle we have $l = \frac{d}{\sqrt{3}}$ as the distance between a vertex and the center of the geometrical shape. The distance between the LED and the

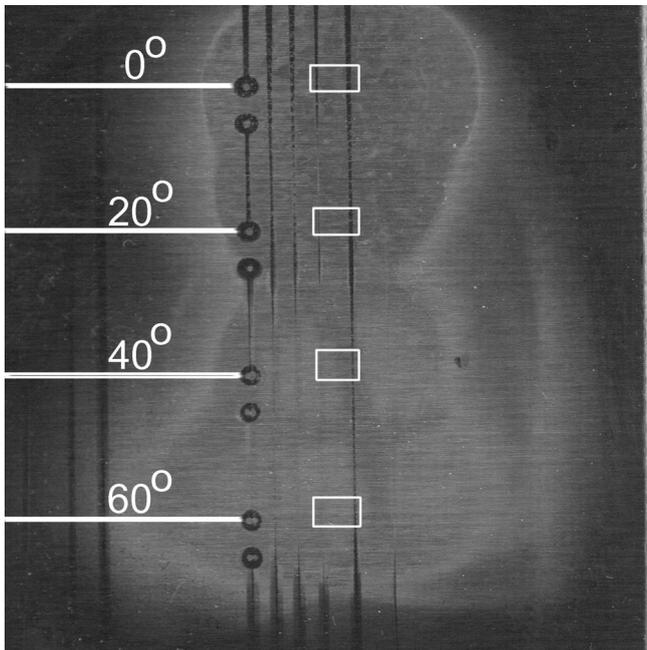


Fig. 1 Sensitivity of the exposure to the angle from the vertical of the incident UV light

parallel lines was printed using an inkjet printer and the transparency overlaid onto the photo-resist was over-exposed in several points with different angles of incidence using two LEDs for each point. The results are illustrated in the photograph of Figure 1.

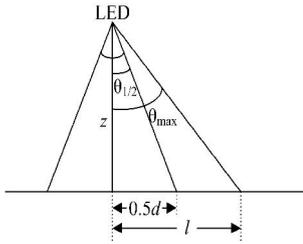


Fig. 2 Optical problem geometry

illuminated surface has to be $z = \frac{d}{2 \tan \theta_{1/2}}$ as is shown in

Figure 2.

The maximum useable angle of the system θ_{max} can be calculated as $\theta_{max} = \tan^{-1} \frac{l}{z}$. The calculations give 20.75° for the square and 17.19° for the equilateral triangle array configuration, for the given LEDs.

The *Uniformity* of the LED array at the center of each shape can be calculated. Since the LED source is assumed to be a Lambertian emitter with irradiance distributed according to a cosine function of the emitting angle. The irradiance distribution (W/m^2) is assumed to be given by:

$$E(r, \theta) = E_0(r) \cos^m \theta, \text{ where } m = \frac{-\ln 2}{\ln(\cos \theta_{1/2})} \quad (2)$$

In this case m is almost equal to 20. Each LED contributes $0.26E_0$ and $0.4E_0$ for the square and the triangular configuration respectively. For the square the four LEDs give *Uniformity* = 0.96 and the three LEDs for the triangular geometry give a *Uniformity* = 0.83. As a result the equilateral triangle array has the desirable lower angle but poorer *Uniformity* that increases the resist exposure time. For the triangular geometry, if the distance between the LED and the illuminated surface is set at a distance $z = \frac{d}{\sqrt{3} \tan \theta_{max}} < \frac{d}{2 \tan \theta_{1/2}}$ with $\theta_{max} = 19.9^\circ$, each LED contributes

in the center of the equilateral triangle $(1/3)E_0$ and better results can be achieved. The most intensive illumination is in front of each LED, the illuminated surface distance from the LEDs is smaller than the previous cases and the maximum angle is still smaller from the case of the square LED array. The exposure time is equal to the square LED array system but the maximum angle is smaller and the whole construction is about 10% thinner. However, *Uniformity* is not improved, but in this case, the most intense points the UV light incidences almost vertical so they are a lot of intense tolerant. Since cheap mass produced UV LEDs have inherent bad *Uniformity* it is important to keep θ_{max} as small as possible in order to make the system tolerant to the exposure time. This configuration will also be favorable in the case of wide viewing angle LEDs because the reflection and the absorption

of the glass close to the triangle centers will correct the *Uniformity* and make it much more advantageous than that of the square LED array.

The arrangement of the LEDs in the equilateral triangular array configuration is shown in Fig. 3. There are lines with (N+1) LEDs that are denoted by l_e and lines with N-LEDs each is denoted by l_d . The board must be constructed wider

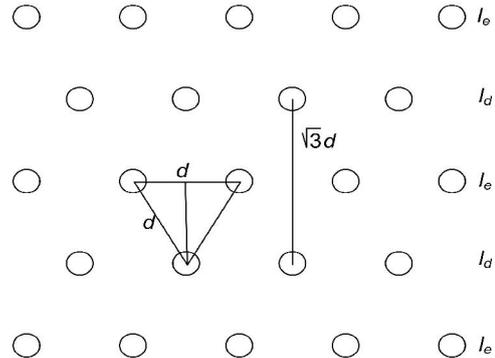


Fig. 3 Placement of the LEDs in the equilateral triangle array configuration

than an A4 PCB board. In order to cover the corners satisfactory the extreme top and bottom lines must have more LEDs. In this case only the extreme left and right vertical lines must be out of the useful exposure area in order to prevent effects from the worst uniformity in the area between their LEDs.

In order to find the number of LEDs required, the number of the excess LED lines has to be found. Let this number be K. To the longest dimension of the maximum PCB area the placement tolerance has to be added. If this sum is divided by K-1 the result represents the distance between two N+1 LED lines. From the Fig. 3 is apparent that if this number is divided by then the result is the distance between two adjacent LEDs “d”. If the other dimension of the maximum PCB area will be divided by the distance between two adjacent LEDs “d” the next bigger integer of this result is the number N. The total LED number that is needed to cover the one PCB side will be $K(N+1)+(K-1)N$.

III. THE ELECTRONIC CIRCUIT

One of the biggest advantages of the proposed exposure device is its very low power consumption. This allows various electronics configurations to be considered. The electronic circuitry that was used on the prototype that is described here was chosen because of its simplicity and ease of construction. A lot of suitable solutions are available today. Voltage of the mains is rectified and an adequate number of LEDs are connected in series driven from a current source. The basic configuration of the LED drive source circuit for a 230V AC system is shown in Fig. 4.

The current applied from the current source is slightly

lower than 30mA. Each UV LED has a forward voltage drop

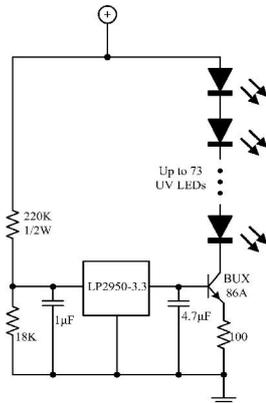


Fig. 4 The proposed basic configuration of the LED drive circuitry of 3.6V and the array of 73 LEDs has a forward voltage drop of about 283V. This is enough low voltage for stable operation of the rectification filter even for the lower limit of the mains voltage specification. The BJT requires a small heatsink. For the case of a lower number of LEDs in the array, the size of the heatsink will be affected. It is better to construct an individual drive to each of 73 LEDs array because otherwise the faults of the one string will affect the current of the other strings.

For some minutes of exposure it is not necessary, but it is very convenient, to have an Operation Control Timer (OCT) to control the illumination circuit. The prototype is equipped

output signal has been placed on P3.5. This signal can control the contact of a relay. An in-circuit programming circuit, that allows the microprocessor to be programmed from a personal computer serial port, has been also added. The clock of the microcontroller has been set to 11.0592MHz for low power operation and to support the in-circuit programming function.

The code of the microcontroller is simple and is based on the use of interrupts. The three timers of the DS89C420 are flexible and more capable than those of the standard 8051. One of them is responsible for controlling the display multiplexing. It produces an interrupt at less than 10msec (higher than 100Hz frequency). The interrupt reads the current digit number from a variable, and reads the respective display value from another variable. A look-up table converts the value from binary to 7-segment and outputs this to the respective display digit. The current digit number is set to the next digit and the interrupt is ended. Each digit flashes faster than 33Hz and this is fast enough for the human eye.

Another interrupt is responsible for the time base for the OCT countdown. For the used operating clock none of the microprocessor timers can produce interrupts at 1 sec. The respective timer is set to produce interrupts with a higher frequency, e.g. 100Hz, and the frequency is divided using a dedicated variable and some code. Care must be taken when interrupts are disabled inside the code, so as not to exceed the multiplexing interrupt time. This can produce temporary display errors.

IV. THE EXPERIMENTAL PROTOTYPE

In order to verify the possibilities of the proposed methodology a prototype was constructed. The case of a

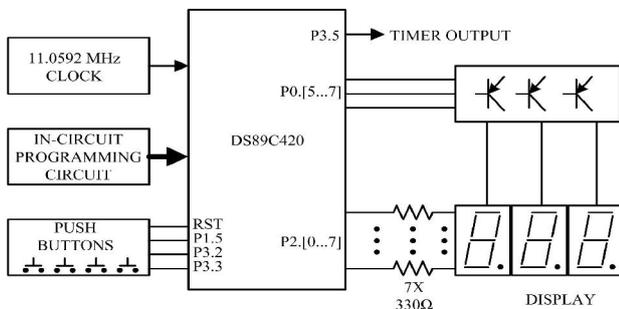


Fig. 5 The OCT (operation control timer) block diagram.

with a specially designed OCT. The OCT is electrically isolated from the other circuit in order to enhance the user protection. The OCT is preloaded with the typical exposure time so that the whole procedure usually requires only a single button press. Two more buttons, with fast moving capability -if kept pressed- allow the user to add or subtract time to the OCT. A fourth reset button loads the default time value again to the OCT. The prototype OCT was based on DS89C420. The block diagram of the OCT is shown in Figure 5.

The display is multiplexed on port 2 and a part of the port 0 is used for the digit selection. The four push buttons use a respective number of active low interrupts and the OCT

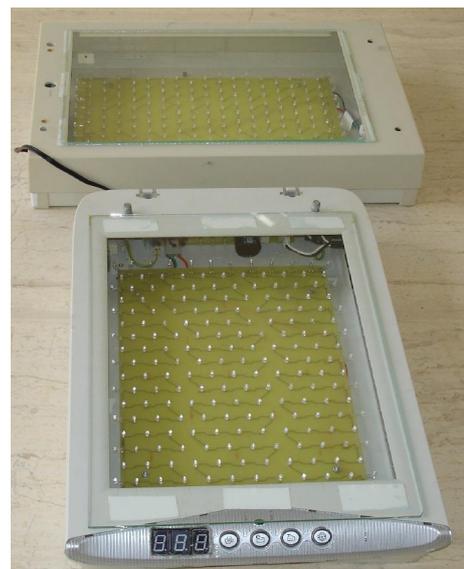


Fig. 6 The experimental prototype.

flatbed scanner found to be a useful enclosure for such an exposure device. For the base, a case of a scanner equipped

with buttons was modified to take in the OCT. For the top an old flatbed SCSI scanner was used. These old scanners have sufficient depth and provide high flexibility for the design. The glass of both scanners is removed and placed in the outer side of their case. This way the glass of each scanner presses the corresponding side transparency on the photosensitive PCB side. Guides were also made in order to center both the exposure surfaces when the cases are placed one over the other. A photo of the experimental prototype is demonstrated in Figure 6.

The prototype uses 218 LEDs with viewing angle of 30° degrees for each side. These are cheap Chinese made LEDs with a typical luminous intensity of about 2000mcd at a continuous current of 30mA. The first observation was that these LEDs have a high non-uniformity. This problem is worse when they are soldered close to their body, probably because of the plastic of their case that is very sensitive to the temperature. Despite this, the experimental results were successful even in the case of very high detail PCBs that were using SMDs (Surface Mounded Devices). This was due to the fact that the incident UV light is very close to the vertical and it immunizes the exposure result from the transparency dense, the photoresist thickness, and the variations of the exposure time and intensity.

The top side case was sufficiently high to allow the examination of the device sensitivity to the construction tolerance. The distance between the LEDs and the exposure surface affects the illumination uniformity but without a noticeable difference in the result. The exposure times were about 2 minutes for the experimental prototype and for good quality printed laser-printer transparencies times up to 10 minutes were used without problems. The power consumption is estimated to be less than 70W. The device is lightweight, does not need pre-heating and it has obvious advantageous when compared with other non-LED devices.

V. CONCLUSION

A methodology for the construction of a LED based PCB exposure device has been presented in this paper. The methodology produces a thin device that incorporates narrow viewing angle LEDs, but can also give better uniformity if wide angle LEDs are implemented. LEDs are very advantageous for such applications because they are efficient, do not need preheating and are solid-state. Also their light can be very close to visible light and so it can be safer in case of accidental exposure of humans in their light. Using the proposed methodology a low cost, lightweight, small and convenient exposure device can be easily constructed for course projects or prototyping.

REFERENCES

- [1] Scott Baldwin, "A Successful Process Approach To Electronic Device Design", *Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition*, Session 2425, 2001.
- [2] Allen Jong-Woei Whang, Yi-Yung Chen, and Yuan-Ting Teng, "Designing Uniform Illumination Systems by Surface-Tailored Lens and Configurations of LED Arrays", *J. Display Technol.* 5, 2009, pp. 94–103.
- [3] Chen-Kang Huang, Jeng-Gang Sung, "The application of UV-LEDs to printed circuit board process", *Microsystems, Packaging, Assembly and Circuits Technology Conference*, 2009, pp. 613–616.

Educational App for Android Mobile Devices

Krzysztof Szklanny, Marcin Wichrowski, and Alicja Wiczorkowska

Abstract— This paper describes an Android application “Apollo” for selected topics in physics for students. The goal of this application (app) is to present physics topics in a simple and pleasant way. The topics selected include: how artificial satellites orbit the Earth, geocentric orbit, gravitational acceleration, and so on. Our application combines virtues of a computer game and educational multimedia presentation. The app is not a pure teaching tool; its goal is to support teaching through play and visualization of physical phenomena. Our app is an RPG game (role playing game), in which the player assumes the role of a pilot-astronaut in a space station. His task is to find hidden hints for each task within a narrative. The narrative introduces the physics problems, and solution methods. The target groups of this app are junior high school students, high school students, and college students. The level of difficulty of the tasks performed by the user corresponds to the high school exams. The app was prepared using Unity 3D 4.5 game engine. Therefore, the app resembles first person perspective games, which are popular among students. The application was inspired by a physics teacher, reporting his difficulties to teach students. We hope our app can be useful, and it will help students to learn physics.

Keywords—mobile learning, Android app, student engagement, teaching physics.

I. INTRODUCTION

THE turn of the first and the second decade of the twenty first century is an era of mobile devices. Smartphones became very popular and according to Parks Associates [1] about 500 million smartphones were sold in 2011. Since the world population exceeds 7 billion, every fourteenth person owns a smartphone since 2011. Moreover, the smartphone sale keeps growing, and increased by 175% in 2012. Other financial agencies also confirm this trend. According to International Data Corporation [2], the smartphone sale increases by 70-80% every year, and since 2013 more smartphones than regular phones are sold.

Our application is designed for Polish students. Smartphones are very popular among them, and the students spend a lot of time using smartphones. The app was written in

This work was partially supported by the Research Center of Polish-Japanese Academy of Information Technology, supported by the Ministry of Science and Higher Education in Poland.

K.. Szklanny is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (corresponding author, phone: +48 22-58-44-500; fax: +48 22-58-44-501; e-mail: kszklanny@pjwstk.edu.pl).

M.. Wichrowski is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (corresponding author, phone: +48 22-58-44-500; fax: +48 22-58-44-501; e-mail: mati@pjwstk.edu.pl).

A. Wiczorkowska., is with the Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland (e-mail: alicja@poljap.edu.pl).

Polish, to facilitate learning of difficult abstract notions in physics.

A. Mobile Technologies in Poland

According to the TNS research performed in 2014 [3], consumer awareness of new mobile technologies increases, and only 7% of smartphone users cannot tell the difference between a regular mobile phone and a smartphone, compared to 42% in 2013. The number of mobile app users also keeps growing, and 31% increase in 2013 was reported, although 1/3 of smartphone users do not buy apps from app stores at all. Still, the number of Android, iOS and Windows Phone users increases every year. Out of all mobile phone users, 44% owned a smartphone in 2014, and since the sell increases by 1/3 every year, the number of smartphone owners is expected to reach 60% of mobile phone users in 2015. The most common activity of smartphone use is using the Internet (64% of Poles) – 23% increase compared to the previous year.

Smartphones are especially popular among young people:

- 74% of 15-19 year-old teenagers own a smartphone,
- 70% of 20-29 year-old young people own a smartphone,
- 60% of 30-39 year-old young people own a smartphone,

Among 40-49 year-old persons, only 52% own a smartphone, and the percentage decreases with the age (28% for 50-59, 9% for older users).

Since smartphones are very popular among young people, it is reasonable to prepare application to be used on smartphone, or another mobile device (tablet etc.).

B. Mobile Apps for Physics Learning

Millions of various applications are available for smartphone users. Google Play offers 1,600,000 apps for Android; App Store has 1,500,000 apps for iOS system; 340,000 apps are available for Windows Phone users [4]. Every month 100,000 more apps are introduced.

Polish educational apps available at Google Play or App Store are mainly prepared for foreign languages teaching. There are just a few apps for physics teaching, compared to several hundred apps for English teaching (apps in Polish). German and other languages are also popular, with about 250 such apps. Not so many apps are available for science, with about 150-170 apps for mathematics, and only 5 for physics. Still, our students are interested in learning physics. Let us examine physics learning apps available for Polish users.

The best app in users’ ranking is „Fizyka na 5” (Physics, grade A), with 4.1 points scored in 5-point scale and 500,000 downloads. To compare with, other apps (in English) that scored 4 points have much lower number of downloads:

“Basic Physics” – 50,000 downloads, “Optics Physics” – 10,000 downloads, “High School Physics” – 10,000 downloads, “Learn Physics via Videos” – 10,000 downloads. The number of downloads of “Fizyka na 5” is especially impressive, as only Polish-speaking persons can be interested in downloading this app. Other apps also have 10,000-100,000 downloads. „Fizyka na 5” is free and simple to use. It contains physical formulas with explanations, presented as text and diagrams. The target groups are junior high school and high school students, and the goal of this app is to help the students to do their homework.

The second ranked app is „Fizyka w szkole LITE” (Physics at school LITE), with 3.7 points scored and 100,000 downloads. It contains a Flash animation collection, combined with formulas and solutions of tasks to do. No explanations of the presented topics are given, so this app is addressed to students having such knowledge. It can also be used as visual illustration of physics classes.

The third ranked app is „Karta wzorów – fizyka” (Formulas – physics), with 4.0 points scored and 10,000 downloads. This app presents physical formulas, as well as solutions of typical physics tasks, presented in clear and aesthetic form. Most of the topics covered in physics classes in junior high school and high school are also covered by this app. The goal of this app is to help students in doing their homework, and prepare for tests.

The fourth ranked app is „Tablice Maturalne Fizyka” (High school exams – Tables - Physics), with 4.0 points scored and 10,000 downloads. This app also contains formulas and tables necessary in doing homework and exams, presented as text with graphics.

The fifth (last) ranked app is „Fizyka w szkole” (Physics at school), not rated and with 5 downloads only. This is the only app in this ranking which is not free. „Fizyka w szkole” is an extended version of „Fizyka w szkole LITE”. The extension includes calculating tasks, improvement of the quality of animations and graphical design of this app.



Fig. 1 Start screen of the app described in this paper

As we can see, there are not many apps available for learning physics for Polish students. Most of the apps basically present the formulas and solving tasks, in rather wearisome

form. It is quite probable that students use these apps to do homework or prepare for tests, and this is why the number of downloads is so big, as the students may expect getting a quick solution of their tasks.

The physics (and generally science) teachers report that the students have difficulties to understand and use abstract ideas. This is why students are often unable to understand physics laws. Therefore, the goal of this work was to create an app that can illustrate hypothetical situations when physics laws manifest themselves, and the students can learn through play. The start screen of our app is shown in Figure 1.

II. PROJECT GUIDELINES

The app is designed for Android operating system version 2.3.1 or higher with Cortex ARMv7 processor.

Graphics processing unit supporting OpenGL ES version 2.0 or higher is assumed, Internet access, at least 1 GB of RAM, and at least 650 MB of disc free space. We assumed that using the app should be easy, and the following guidelines were implemented to assure high User-Experience level [5]:

- The app is playable on big and small screens (tablets, smartphones),
 - Intuitive control through gestures which are easy to learn, 2 joysticks, and a touch-activated interface of high usability,
 - Technical tests were performed to assure the reliability of the app, as well as its high usability,
 - Additional functions: educative movies which can be stopped and started again any time,
 - The app is fast to use and works in real time, as smartphone users are impatient and uninstall slow apps; the users get feedback immediately,
 - Our app uses simple communication, and the content is presented in an engaging way.

A. Narrative

The narrative of the game is based on space travel within our galaxy, in search for necessary minerals and for planets that are suitable to be inhabited by people. The biggest ever space travel project, called Apollo, experiences propulsion breakdown and loses contact with Earth, twelve years after the mission started. The computer must be fixed, so it is necessary to contact Earth and download again the system functions. In order to do this, the pilot must help in determining the parameters necessary to send the message. And then learning comes – the computer gives the user a physics task to solve, and shows the video presentation which explains how to solve this task.

B. User Interface and Interaction in the Application

The user of our app moves in a virtual space station (in 3D space) using virtual joysticks, located at the bottom of the screen, on the left and on the right hand side. Such interaction allows the user choosing the walk direction, moving forward or backward. The interior of the space station is shown in

Figure 2 and Figure 3.

The goal of the spaceship exploration is to find hints and educational materials, which can be found in the screens of the guidance computers. Interactive elements can be easily distinguished in the spaceship, see Figure 4; for example, a chart with physical formulas, see Figure 5. After the interactive element is found, the user can interact with it by touching. This can launch a movie, which can be stopped and resumed any time. Such a form of interaction and a very simple interface are intuitive and easy to use. The narrative discovered by the user encourages him or her to get to know next topics and test his/her knowledge.



Fig. 2 Interior of the space station in Apollo app

III. IMPLEMENTATION

The application is based on Unity 3D game engine [6]. Unity is a game development platform, with IDE (Integrated Development Environment) to design applications for mobile devices, desktop computers, and web pages. Apollo app was implemented using Unity 4.5. Folder structure, game environment, first person perspective controller, materials, animations and triggers were created using Unity, and the Apollo app was built and compiled within Unity platform.

The planet geometry, spaceship and its contents were modeled using 3ds Max [7]. Poly objects were used for this purpose. Geometry was adjusted to real size scale, and then textures and materials were assigned to the objects, using UVW Map and Unwrap UVW modifiers, as well as applying lights and shadows in the scenes. After setting a camera in 3ds Max, animation was created, rendered using V-Ray 2.0, and 3D objects exported to FBX format.

Speech synthesis was also applied in Apollo app. We decided to use Ivona Text to Speech synthesizer [8]. This synthesizer was applied to obtain artificial sounding voice of guidance computer. Text and graphics were edited using Adobe Photoshop [9]. Normal maps for diffuse, specular, normal, and translucency effects were created using Quixel Suite [10]. Intro was created using After Effects [11], and Adobe Audition was applied to edit music and audio files, including speech for narrative. Unity 3D Remote was used to test the app at each stage; the device was connected through

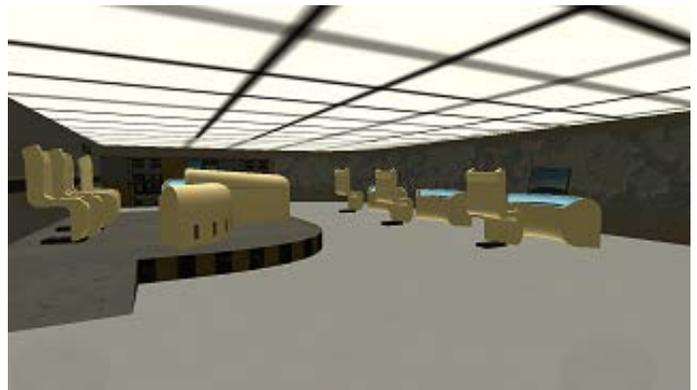


Fig. 3 A room inside the space station in Apollo app



Fig. 4 A room with an interactive element (big table in the back)

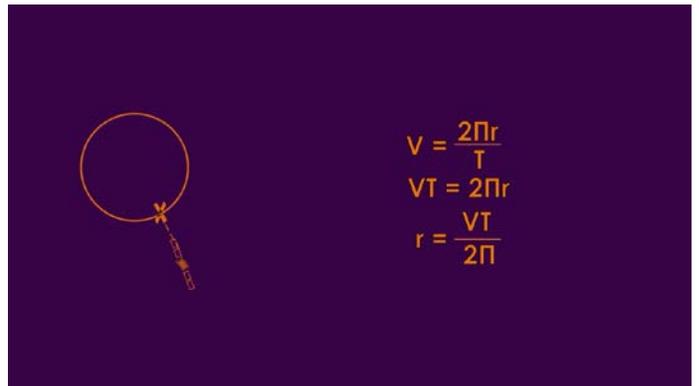


Fig. 5 A chart with physical formulas

USB port.

IV. TESTS

The app was tested on three mobile devices: Nexus 7 tablet computer, Samsung Galaxy S3 smartphone, and Samsung Galaxy S4 smartphone. The following aspects were tested:

- whether the game can be played using various platforms,
- how many errors are indicated by the processor,
- visual quality,
- correct display of movies,

- use of controls,
- usability of the app,
- educational value of the app.

The results of these tests are shown in Tables 1-5.

Tab. 1. Errors indicated by the processor

Device	Processor signaled errors	Number of tests
Nexus 7	0	20
Galaxy S3	0	20
Galaxy S4	0	20

Tab. 2. Game stops and slow-downs

Device	Game stopped because of CPU overload	Game slowed down because of CPU overload	Tests
Nexus 7	0	0	20
Galaxy S3	0	0	20
Galaxy S4	0	0	20

Tab. 3. Visual quality

Device	Very good	Good	Average	Below average	Unsatisfying
Nexus 7	X				
Galaxy S3		X			
Galaxy S4		X			

Tab. 4. Movie playing tests

Device	Movie did not play	Movie interrupted	Tests
Nexus 7	0	0	20
Galaxy S3	0	1	20
Galaxy S4	0	0	20

Tab. 5. Tests of controllers

Device	No reaction to touch	Move other than intended	Tests
Nexus 7	0	0	20
Galaxy S3	0	0	20
Galaxy S4	0	0	20

Usability tests and the tests of educational value of Apollo were performed for two target groups: college students, and high school students. College students represented Polish-Japanese Academy of IT (five students, fourth year), and high school students represented Zespół Szkół Ogólnokształcących in Kamienna Góra, Poland (10 students). The results are shown in Tables 6 and 7.

Tab. 6. Usability tests

Group	Very useful	Useful	Do not know	Not useful	Definitely not useful
PJAIT	0	2	1	1	1
High school	0	5	3	2	0

Tab. 7. Educational value assessment

Group	Very useful	Useful	Do not know	Not useful	Definitely not useful
PJAIT	0	2	1	1	1
High school	1	5	3	1	0

As we can see, the app can be played on various devices practically without errors, and most of the students consider this app useful, as indicated in usability tests and educational value assessment.

V. SUMMARY

In this paper we analyzed the trends in using smartphones, and the conclusions were applied to create an educational app. We observed the tendency to use smartphones rather than regular mobile phones (especially among young people), clearly visible in market research, so we believe that Apollo app supporting learning physics can be helpful. The presented paper describes the project guidelines, narrative, interaction within the app, user interface, and the results of tests with users. Therefore we conclude that the techniques applied to create this app were properly selected, especially Unity platform came very handy. This platform allows fast and effective creation of a multimedia game.

Since many users are interested in such apps, and the results of the tests were encouraging, we believe that more educational apps in physics can be created, and be used by young generation.

ACKNOWLEDGMENT

The authors would like to thank Łukasz Homoncik for his help in implementing the Apollo application.

REFERENCES

- [1] Park Associates [Online]. Available: <http://www.parksassociates.com/index.php>
- [2] International Data Corporation [Online]. Available: <https://www.idc.com/>
- [3] TNS [Online]. Available: <http://www.tnsglobal.com/>
- [4] Statista, “Number of apps available in leading app stores as of July 2015” [Online]. Available: <http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/>
- [5] D. Laufer, “The Best and Worst of Mobile User Experience”, Forrester Cambridge, 2014 [Online]. Available: <https://www.progress.com/docs/default-source/default-document-library/Progress/Documents/Pacific/Whitepaper/The-Best-And-Worst-Of-Mob.pdf>
- [6] Unity [Online]. Available: <https://unity3d.com/>
- [7] 3ds Max [Online]. Available: <http://www.autodesk.com/products/3ds-max/overview>
- [8] Ivona Text to Speech [Online]. Available: <https://www.ivona.com/pl/>
- [9] Adobe Photoshop [Online]. Available: <http://www.adobe.com/uk/products/photoshop.html?promoid=KQLBI>
- [10] Quixel Suite [Online]. Available: <http://www.quixel.se/dev/index>
- [11] Adobe After Effects [Online]. Available: <http://www.adobe.com/uk/products/aftereffects.html>
- [12] Adobe Audition [Online]. Available: <https://www.adobe.com/uk/products/audition.html>

Krzysztof Szklanny obtained his PhD. in 2010. He is the head of the photography laboratory at the Multimedia Department at the Polish-Japanese Academy of Information Technology. He published about 20 papers including conference papers, journal papers, and book chapters. He is also an author of high-speed photographs, including commercial photos, exhibited in the Polish-Japanese Academy of IT.

Marcin Wichrowski is a PhD. candidate at the Multimedia Department at the Polish-Japanese Academy of Information Technology. He published about a dozen papers, including conference papers, journal papers, and book chapters. His research is focused on methodology of designing user interfaces for medical devices and applications. He also collaborates with artists in the creation of projects/installations using new media technologies.

Alicja A. Wiczorkowska obtained her PhD. in 1999 and DSc. in 2012. She is the head of the Multimedia Department at the Polish-Japanese Academy of Information Technology. She published about 90 papers including conference papers, journal papers, and book chapters. She is also the author of the handbook on multimedia (in Polish).

The hybrid methods are 60 years in the scientific works

G.Mehdiyeva, V.Ibrahimov, and M.Imanova

Abstract— In the middle of XX century the scholars began construct the methods which have the best characteristics of the known and multi-step methods with the constant coefficients. And in the 1955 years there appears the works of Gear and Butcher dedicated investigation of the hybrid method. Here we want to show how this theory developed and compared them with the known methods. Constructed the hybrid method with the higher order of accuracy and illustrated them by the model equation. For application of hybrid methods, here proposed the simple algorithms with the order of accuracy $p \leq 10$ in the case, when the amount of the used mesh points is equal to 2 or the order of the difference methods variable satisfies the condition $k = 1$.

Keywords— hybrid methods, Volterra integral equation, accuracy and stability of numerical methods, ODE, multistep methods with the second derivative.

I. INTRODUCTION

MANY problems of natural sciences reduced to solving integral equations with the variable boundaries which in linear case was fundamentally investigated by Vito Volterra. Many scholars have investigated the numerical solution of the Volterra integral equations. But they in basically use the quadrature methods or its modification. But here proposed a new way for construction the numerical methods for solving the integral equation with the variable boundaries.

Let us to consider the following integral equation

$$y(x) = f(x) + \int_{x_0}^x K(x, s, y(s)) ds. \quad (1)$$

Sometimes integral equation (1) called the integral equation of Volterra-Urison type.

This convention is due to the fact that equation (1) in the linear case was thoroughly investigated by Volterra, and at a sufficiently high level was studied such equations occurrence in practical problems (see, e.g. [1], [2]). Note that the singular integral equation with a variable boundary in a particular form

G.Yu.Mehdiyeva - doctor of science, PhD, professor, head of chair of Computational mathematics of Baku State University, Baku, Azerbaijan (corresponding author to provide phone 994125106048 e-mail: imn_bsu@mail.ru)

V.R.Ibrahimov - doctor of science, PhD, professor of the department of Computational mathematics of Baku State University, Baku, Azerbaijan (e-mail: ibvag@mail.com)

M.N.Imanova is with the Baku State University, PhD, teacher of department of Computational mathematics, Baku, Azerbaijan (e-mail: imn_bsu@mail.ru).

was studied for the first time by Abel (see, e.g., [1, p.12]). Given that even in the linear case, the exact solution to equation (1) is not always possible, therefore many experts have used approximation methods to solve it (see, e.g., [3] - [7]).

Here we suppose that the Volterra integral equation (1) has the unique solution determined on the interval $[x_0, X]$, for investigation of the numerical solution of equation of (1), assume that the kernel $K(x, z, y)$ of the integral equation define in the domain $\bar{G} = \{x_0 \leq s \leq x + \varepsilon \leq X, |y| \leq b\}$, where it has partial derivatives up to some order p , inclusively. But the given sufficiently smooth function $f(x)$ has determined on the interval $[x_0, X]$. For determining the approximate values of solution of equation (1), we divide the interval $[x_0, X]$ into N equal parts by the mesh points $x_i = x_0 + ih$ ($i = 0, 1, 2, \dots, N$). Denote by y_i the approximate and $y(x_i)$ exact values of solution of equation (1) at the mesh points x_i ($i = 0, 1, 2, \dots, N$).

As is known one of the popular methods for solving equation (1) is the quadrature method. The quadrature method which applies to solution of equation (1) can write as follows (see [2]):

$$y(x_n) = f(x_n) + h \sum_{j=0}^n \bar{a}_j K(x_n, x_j, y(x_j)) + R_n, \quad (2)$$

where R_n is the remainder term of the quadrature method and \bar{a}_j ($j = 0, 1, \dots, n$) are real numbers; these numbers called the coefficients of the quadrature method. By discarding the remainder term, we obtain the following method:

$$y_n = f_n + h \sum_{j=0}^n \bar{a}_j K(x_n, x_j, y_j) \quad (n = 1, 2, 3, \dots), \quad (3)$$

$$y_0 = f(x_0),$$

which called the quadrature method with the variable boundary.

Remark that the method (3) for $\bar{a}_n \neq 0$ is implicit, but this method for $\bar{a}_n = 0$ is explicit. And in the result of using method of (3), by increasing of the values of the quantity n ,

the amount of computing functions $K(x, s, y)$ is increases, also. Consequently, at each step the amount of calculation work is increases, which is a major disadvantage of quadrature methods. To eliminate this drawback of method (3), some authors have proposed the use of methods such as the following k -step method with constant coefficients (see for example [8] - [11]):

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f(x_{n+i}, y_{n+i}), \quad (4)$$

It is known that if the method (4) is stable then the following holds:

$$p \leq 2\lceil k/2 \rceil + 2. \quad (5)$$

Here the degree of the method (4) quantity p is order of the accuracy and k is the order of the difference method (4).

For the construction the stable methods with the degree $p > k + 2$, the scholars are use modification of the method (4) in with resulted appears the forward-jumping methods etc. Some of these scholars proposed to change the method (4) by the following:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i y'_{n+i} + h^2 \sum_{i=0}^k \gamma_i y''_{n+i} \quad (6)$$

Here the coefficients $\alpha_i, \beta_i, \gamma_i (i = 0, 1, \dots, k)$ are the some real numbers and $\alpha_k \neq 0$.

It is clear that for generalization of the Runge-Kutta and Adams methods one may use the different ways, one of which can be written as the follows:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \gamma_i y'(x_{n+i} + v_i h) \quad (7)$$

$(|v_i| < 1; i = 0, 1, 2, \dots, k)$

Here quantities are the hybrid points.

Remark that for function of the methods Runge-Kutta and Adams. We rewrite the method (6) as the following:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i y'_{n+i} + h \sum_{i=0}^k \gamma_i y'_{n+i+v_i} \quad (8)$$

If we continue the above proposed way, then the method (5) can rewrite as:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \left(\sum_{i=0}^k \beta_i y'_{n+i} + \sum_{i=0}^k \gamma_i y'_{n+i+v_i} \right) + h^2 \left(\sum_{i=0}^k \beta_i y''_{n+i} + \sum_{i=0}^k \gamma_i y''_{n+i+v_i} \right) \quad (9)$$

It is not difficult to understand that the method (9) is more accurate, than the known methods.

Method (4) generalized so that every following method is more accurate than the previous one. For example, in family methods of type (8) there are stable methods that are more accurately than the methods of the type (6) and (7). Therefore research of methods (8) and (9) is more promising

than the above methods of research. For the objectivity, we note that hybrid methods are more accurate than the known. However, using of the hybrid method is a difficult process. Therefore, construction of algorithms for application of hybrid methods is sometimes more difficult process than construction of the method itself. In the scientific literature, some authors such method called fractional steps.

II. APPLICATION OF THE METHOD OF UNDETERMINED COEFFICIENTS TO THE STUDY OF HYBRID METHODS PROCEDURE

For construction of hybrid methods we can use different schemes, in resulting of this, get methods with the different properties. It is known that the properties of numerical methods depend on the values of the coefficients in the formulas (5) - (9). Previously, the above method applies to solution of the following initial value problem:

$$y' = F(x, y), y(0) = 0. \quad (10)$$

In these studies, we used the method of undetermined coefficients for determination of coefficients in the formulas (5) - (9). Therefore, we will try in construction of hybrid methods for solving the equation (1) using the method of undetermined coefficients. Note that one of the most popular methods for finding solutions of Volterra integral equations is quadrature method. But as shown above, when using the quadrature methods the volume of computations increases with the number of quantity of the points used in constructing them. In this regard, here we offer apply to solution of equation (1), the following multi-step method

$$\sum_{i=0}^k \alpha_i y_{n+i} = \sum_{i=0}^k \alpha_i F_{n+i} + h \sum_{j=0}^k \sum_{i=0}^k \beta_i^{(j)} K(x_{n+j}, x_{n+i}, y_{n+i}). \quad (11)$$

Note that from the formula (11) we can obtain implicit and explicit methods, as well as forward-jumping methods. Here we want by using the above schemes construct hybrid methods for solving of the equation (1). The method (11) has a straight connection with the method (5). Indeed, if in the equation (1), the kernel of the integral, the function $K(x, s, y)$ the next: $K(x, s, y) = F(s, y)$, then solution of the equation (1) coincides with solution of the problem (10). Using a similar connection between integral and differential equations, consider construction of hybrid methods.

Here, to solve equation (1), we use the finite-difference method, which can be written as follows:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k (\beta_i y'_{n+i} + \hat{\beta}_i y'_{n+i+m_i}) \quad (12)$$

$(|m_i| < 1; i = 0, 1, \dots, k).$

On the first let us consider any bounders for the coefficients of the method (2).

A: The coefficients $\alpha_i, \beta_i, \gamma_i, v_i (i = 0, 1, 2, \dots, k)$ are

some real numbers, moreover, $\alpha_k \neq 0$.

B: Characteristic polynomials

$$\rho(\lambda) \equiv \sum_{i=0}^k \alpha_i \lambda^i, \quad \delta(\lambda) \equiv \sum_{i=0}^k \beta_i \lambda^i; \quad \gamma(\lambda) \equiv \sum_{i=0}^k \gamma_i \lambda^{i+\nu_i}$$

($\beta_i = \gamma_i$).

have no common multipliers different from the constant.

C: $\sigma(1) + \gamma(1) \neq 0$ and $p \geq 1$.

In particular, for $\hat{\beta}_i = 0$ ($i = 0, 1, 2, \dots, k$), from the method of (12) we obtain an ordinary k -step method with the constant coefficients.

III. ON A WAY TO CONSTRUCTION HYBRID METHOD OF TYPE (12).

Consider the construction of the methods of type (12) for $\hat{\beta}_i = 0$ ($i = 0, 1, 2, \dots, k$). To this end, we construct a formula for calculating the values of the function $y'(x)$. Suppose that by some methods the solution of the equations is found after taking into account that in equation (1) one identity is obtained. Then, from this identity we can write:

$$y'(x) = f'(x) + \int_{x_0}^x K'_x(x, s, y(s)) ds. \tag{13}$$

Finding the values of the quantities y'_{n+k} by means of formulas (13) is reduced to the calculation of the following integral:

$$\int_{x_0}^{x_m} K(x_m, s, y(s)) ds. \tag{14}$$

If quadrature formula (3) is applied to an evaluation of integral (14), then we obtain the integral sum of the variable boundaries. But for the application of the method (12) to solving equation (1), we give the known values y_{n+k-1} , y'_{n+k-1} and we want to find the relationship between these values with the values of y_{n+k} , y'_{n+k} which are unknown.

The computation in the segment $[x_0, x_{n+1}]$ the integral with the variable boundary has been reduced to calculations that in the segment $[x_n, x_{n+1}]$. Moreover, this reduction was accomplished without increasing the computational work by increasing the values of the variable n . Note that this type of scheme is applied to the determination of the relationship between the variables y'_{n+1} and y'_n .

It is known that for a sufficiently smooth function, the following is holds:

$$\sum_{i=0}^k \hat{\alpha}_i z_{n+i} = h z'(\xi_{n+1}), \tag{15}$$

here $\hat{\alpha}_i$ ($i = 0, 1, \dots, k$) are real numbers.

After carrying out same operations, we obtain:

$$\sum_{i=0}^k \alpha_i y'_{n+i} = \sum_{i=0}^k \alpha_i f_{n+i} + \sum_{i=0}^k \alpha_i a_{n+i} + h \sum_{j=0}^k \sum_{i=0}^k \gamma_i^{(j)} K'_x(x_{n+j}, x_{n+i}, y_{n+i}). \tag{16}$$

here $a(x) = K(x, x, y(x))$

Depending on the needs of the user, the coefficients in formulas (14) can be given different meanings. For this purpose, equations (13) can be rewritten as follows:

$$y'(x) = f'(x) + a(x) + \frac{1}{h} \sum_{i=0}^k \hat{\alpha}_i \int_{x_0}^x K(x-ih, s, y(s)) ds + R_1, \tag{17}$$

Here, $\hat{\alpha}_i$ ($i = 0, 1, \dots, k$) are real numbers, but R_1 is the remainder term. By replacing the integrals in formulas (17) with an integral sum and discarding the remaining terms, we obtain methods to calculate the values of the quantity y'_{n+k} . However, in this case, the coefficients impose the following additional conditions. It is not difficult to proof that for the finding of the coefficients of the method (12) may be use the following system of the nonlinear algebraic equation:

$$\sum_{i=0}^k \alpha_i = 0, \quad \sum_{i=0}^k i \alpha_i = \sum_{i=0}^k \beta_i + \sum_{i=0}^k \gamma_i, \tag{18}$$

.....

$$\sum_{i=0}^k \frac{i^p}{p!} \alpha_i = \sum_{i=0}^k \frac{i^{p-1}}{(p-1)!} \beta_i + \sum_{i=0}^k \frac{l_i^{p-1}}{(p-1)!} \gamma_i.$$

By using the solution of algebraic equations can construct methods having different properties. From the viewpoint of the application of numerical methods, the most significant property is the stability and accuracy. The accuracy of these methods depends on the number of equations in the system (18). However, when the number of equations in the system (18), there is some difficulty in finding the solution of the system (18). The above mentioned difficulty is not finding the solution of linear algebraic equations. Note that we constructed methods maximum accuracy only in some particular cases, the following are some of the mood of them:

$$y_{n+1} = y_n + h(y'_{n+1} + y'_n)/12 + 5h(y'_{n+1/2-\alpha} + y'_{n+1/2+\alpha})/12, \tag{19}$$

$$(\alpha = \sqrt{5}/10)$$

$$y_{n+1} = (11y_n + 8y_{n+1/2})/19 + h(10y'_n + 57y'_{n+1/2} + 24y'_{n+1} - y_{n+3/2})/114 \tag{20}$$

$$y_{n+1} = y_n + h(64y'_{n+1} + 98y'_{n+1/2} + 18y'_n)/360 + h(18y'_{n+1/2+\beta/2} + 98y'_{n+1/2} + 64y'_{n+1/2-\beta/2})/360. \tag{21}$$

$$y_{n+2} = y_n + h(5y'_{n+1+\alpha} + 8y'_{n+1} + 5y'_{n+1-\alpha})/9, \tag{22}$$

$$(\alpha = \sqrt{15}/5),$$

As can be seen from the above stable hybrid methods are more accurate than corresponding methods of Runge-Kutta and Adams.

In a specific example, we show that if we use the Simpson method to solve it, the result obtained by step $h/2$ is better than the result in the step h .

To this end, consider the following tasks:

$$y' = \cos x, y(0) = 1, x \in [0,1],$$

exact solution is written as: $y = \sin x$.

The result at the step $h = 0,1$ to place in the following:

Table 1.

Value of the variable x	Error for the Simpson method by the step $h/2$	Error for the Simpson method by the step h
0.10	0.34E-08	0.11E-06
0.20	0.69E-08	0.10E-06
0.30	0.10E-07	0.21E-06
0.40	0.13E-07	0.21E-06
0.50	0.16E-07	0.31E-06
0.60	0.19E-07	0.30E-06
0.70	0.22E-07	0.39E-06
0.80	0.24E-07	0.38E-06
0.90	0.27E-07	0.46E-06
1.00	0.29E-07	0.44E-06

As seen from Table 1, the results obtained by the following method of Simpson

$$y_{n+1} = y_n + h(y'_{n+1} + 4y'_{n+1/2} + y'_n) / 6$$

are accurate.

Now consider the application of the next hybrid method of type (7):

$$y_{n+1} = y_n + h(f_{n+l_0} + f_{n+1+l_1}) / 2, (y'(x) = f(x, y(x))),$$

$$(l_1 = 1/2 + \sqrt{3}/6; l_0 = 1/2 - \sqrt{3}/6)$$

To solving following examples:

1. $y' = \cos x, y(0) = 0, x \in [0,1]$. Exact solution: $y(x) = \sin x$.
2. $y' = \lambda y, y(0) = 1, x \in [0,1], \lambda = \pm 1$. Exact solution: $y(x) = \exp(\lambda x)$.

For the calculating of the values y_{n+1} and y_{n+1-l} may be used the methods:

$$y_{n+l} = y_n + lh f_n,$$

$$\hat{y}_{n+l} = y_n + lh(f_n + \bar{f}_{n+l}) / 2,$$

$$(\bar{f}_m = f(x_m, \bar{y}_m), m = 0,1,2,\dots).$$

Repeat these schemes for $l := 1-l$.

The results of calculations accommodated in Table 2 and Table 3.

Table 2

Step size	Variable x	Error for the example 1	Error for the example 2 ($\lambda = 1$)
$h = 0.10$	0.20	0.11E-06	0.61E-05
	0.30	0.10E-06	0.78E-05
	0.40	0.21E-06	0.15E-04
	0.50	0.21E-06	0.19E-04
	0.60	0.31E-06	0.28E-04
	0.70	0.30E-06	0.34E-04
	0.80	0.39E-06	0.46E-04
	0.90	0.38E-06	0.56E-04
	1.00	0.46E-06	0.71E-04

Table 3

Step size	Variable x	Error for the example 1	Error for the example 2 ($\lambda = 1$)
$h = 0.05$	0.10	0.34E-08	0.36E-06
	0.20	0.69E-08	0.81E-06
	0.30	0.10E-07	0.13E-05
	0.40	0.13E-07	0.19E-05
	0.50	0.16E-07	0.27E-05
	0.60	0.19E-07	0.36E-05
	0.70	0.22E-07	0.47E-05
	0.80	0.24E-07	0.59E-05
	0.90	0.27E-07	0.74E-05
	1.00	0.29E-07	0.91E-05

IV. CONCLUSION

Considering that the solution of ordinary differential equations is fundamentally investigated. Using the examples above, we have shown some of the benefits of hybrid methods. As we know the order of accuracy of the method of Simpson coincides with the above given order of accuracy of the hybrid method, but the comparison of the results set forth in these tables routinely comes that the stability region for the hybrid method is more advanced. Several authors have given an advantage of hybrid methods applied them to the solution of some problems in mechanics and called them by fractional steps size.

It is not difficult to show that the use of hybrid methods to the solution of equation (1) gives better results than the corresponding known methods (see e.g. [8]-[10]).

Consequently, the use of hybrid methods in the scientific literature for 60 years shows that these methods are more promising. From the foregoing, it follows that in the theory of hybrid methods, there are many unsolved problems.

ACKNOWLEDGMENT

The authors wish to express their thanks to academician Ali Abbasov for his suggestion to investigate the computational aspects of our problem and for his frequent valuable suggestion.

REFERENCES

- [1] E.M. Polishuk Vito Volterra. Leningrad, Nauka, 1977, 114p.
- [2] V. Volterra Theory of functional and of integral and integro-differential equations, Dover publications. Ing, New York, Nauka, Moscow, 1982 p.304 (in Russian).
- [3] A.V. Manzhirov, A.D. Polyaniin Handbook of Integral Equations: Methods of solutions. Moscow: Publishing House of the "Factorial Press", 2000, 384 p.
- [4] A.F. Verlan, V.S. Sizikov. Integral equations: methods, algorithms, programs. Kiev, Naukovo Dumka, 1986, 384 p.
- [5] Y.D Mamedov. Methods of calculation, 1978.
- [6] A.Quarteroni, R. Sacco, F. Saleri. Numerical Mathematics, Second Edition, Springer, 656p.
- [7] R.L.Burden, J.D.Faires Numerical analysis. Cengege Learning, 2001, 850 pp, 7-th edition.
- [8] G.Yu. Mehdiyeva, M.N. Imanova, V.R. Ibrahimov On one application of forward jumping methods. Applied Numerical Mathematics, Volume 72, October 2013, p. 234–245.
- [9] P.Linz Linear Multistep methods for Volterra Integro-Differential equations, Journal of the Association for Computing Machinery, Vol.16, No.2, April 1969, p. 295-301.
- [10] G.Yu Mehdiyeva, M.N. Imanova, V.R. Ibrahimov Application of the hybrid method with constant coefficients to solving the integro-differential equations of first order. 9th International conference on mathematical problems in engineering, aerospace and sciences, AIP, Vienna, Austria, 10-14 July 2012, p. 506-510.
- [11] G.Yu Mehdiyeva, M.N. Imanova, V.R. Ibrahimov On a Research of Hybrid Methods. Numerical Analysis and Its Applications, Springer, 2013, p. 395-402.
- [12] V.R.Ibrahimov V.D.Aliyeva The construction of the finite-difference method and application Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014 (ICNAAM-2014) AIP Conf. Proc. 1648, 850049-1–850049-5.
- [13] A.A.Makroglou Block - by-block method for the numerical solution of Volterra delay integro-differential equations, Computing 3, 1983, 30, №1, p.49-62.
- [14] A.Feldstein, J.R. Sopka Numerical methods for nonlinear Volterra integro-differential equations. SIAM J. Numer. Anal. 1974, V 11, p. 826-846.
- [15] V.R. Ibrahimov, M.N. Imanova On a Research of Symmetric Equations of Volterra Type. International journal of mathematical models and methods in Applied sciences Volume 8, 2014, p.434-440.
- [16] M.N. Imanova One the multistep method of numerical solution for the Volterra integral equation. Transactions issue mathematics and mechanics series of physical -technical and mathematical science, XXBI, 2006, №1.
- [17] A.Makroglou Hybrid methods in the numerical solution of Volterra integro-differential equations. Journal of Numerical Analysis 2, 1982, p.21-35.
- [18] V.R.Ibrahimov V.D.Aliyeva The construction of the finite-difference method and application Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014 (ICNAAM-2014) AIP Conf. Proc. 1648, 850049-1–850049-5.
- [19] E. Hairier, S.P.Norsett, G.Wanner Solving ordinary differential equations. (Russian) M., Mir, 1990, 512 p.
- [20] Modern numerical methods for solving ordinary differential equations. Editors J.Holl and J.Watt Publishing House "Mir", Moscow, 1979.
- [21] G.Dahlquist Convergence and stability in the numerical integration of ordinary differential equations. Math. Scand. 1956, №4, p.33-53.
- [22] G. Dahlquist, Stability and Error Bounds in the Numerical Integration of Ordinary Differential Equations. Trans. Of the Royal Inst. Of Techn. Stockholm, Sweden, 1959, №130, p.3-87.
- [23] G.Yu Mehdiyeva, V.R. Ibrahimov On the research of multistep methods with constant coefficients, LAP LAMBERT Academic Publishing, 2013, 314 p. (Russian).
- [24] G. Yu. Mehdiyeva, M. N. Imanova, V. R. Ibrahimov An application of the hybrid method of multistep type. Advances in Applied and Pure mathematics, Proceedings of 2 Intern.Conf. on Math.Comp and Aqtatist. Science (MCSS), 2014, p. 270-276.
- [25] G.K. Gupta. A polynomial representation of hybrid methods for solving ordinary differential equations.Mathematics of comp., volume 33, number 148, 1979, p.1251-1256.
- [26] L.M. Skvortsov Explicit two-step Runge-Kutta methods. Math. modeling, 21, 2009, №9, p. 54-65.
- [27] G.Yu. Mehdiyeva, M.N. Imanova, V.R. Ibrahimov Some application of the hybrid methods to solving Volterra integral equations Advances in Applied and Pure mathematics, Proceedings of 2 Intern.Conf. on Math.Comp and Aqtatist.Science (MCSS), 2014, p. 352-356.
- [28] E.A.Areo, R.A. Ademiluyi, P.O. Babatola Accurate collocation multistep method for integration of first order ordinary differential equations. J.of Modern Math.and Statistics, 2(1): 1-6, 2008, p.1-6.
- [29] V. Ibrahimov On the maximal degree of the k-step Obrechhoff's method. Bulletin of Iranian Mathematical Society, Vol.28, №1, 2002, p. 1-28.
- [30] V.R. Ibrahimov On a relation between order and degree for stable forward jumping formula. Zh.Vychis.Mat., № 7, 1990, p.1045-1056.

Hybrid Control System for the Hyper Redundant Arm HHR in Creeping Movements

Ionel Cristian Vladu, Viorel Stoian, and Ileana Vladu

Abstract— This paper presents creeping movements application of the hyper-redundant arm with an electro-pneumatically system for position control. A system of cables actuated by DC motors is used for robotic arm bending. Control system is based on a PIC microcontroller. The position of the robotic arm can be obtained by bending it with the cables and by blocking the position of the elements we need, using the electro-pneumatically system. The major advantage of this type of actuation consists in the fact that the robot can be actuated using a boundary control by cables, the position blocking system for any element being relatively simple. A new control algorithm for a hyper-redundant manipulator, which moves in a labyrinth and can touch the boundaries of the niche, is presented. The collisions or contacts are part of the task. The support-points from the walls help with the moving process. The dynamic model by Lagrange method of the hyper-redundant robotic structure is established. Afterwards, a hybrid position/force controller for the hyper-redundant locomotion is proposed. Finally, the simulations of the hyper-redundant locomotion by climbing are presented.

Keywords— hyper-redundant robot, robotic arm, hybrid control algorithm, hyper-redundant arm, creeping movements, dynamic model.

I. INTRODUCTION

THE hyper-redundant arms are a class of manipulators that can reach any position and orientation in space. Inspired by biological structures, like trunks (Fig. 1), tentacles and snake backbones, these arms are the mechanical structures with continuum elements described by distributed parameter



Fig. 1. Example of a trunks animal vs manipulator

Ionel Cristian Vladu is with the Faculty of Electric Engineering, University of Craiova, Craiova, Romania (corresponding author to provide phone: +40728148258; e-mail: cristian.vladu@ie.ucv.ro).

Viorel Stoian, was with Faculty of Automation, Computers and Electronics, University of Craiova, Craiova, Romania (e-mail: stoian@robotics.ucv.ro).

Ileana Vladu is with the Faculty of Automation, Computers and Electronics, University of Craiova, Craiova, Romania, (e-mail: ileana.vladu@gmail.com).

models. In comparison with classic manipulators, hyper-redundant robots can operate in restricted areas.

The robotic arm is composed of identical elements with series connections. The final position of the robotic arm is obtained by changing relative position of these elements. The structure is the „backbone” type one, actuated by a set of cables. Robot’s body is structured in sections composed of successive elements. By modifying their angular position, one can obtain section curvature, robot’s curvature, respectively (Fig. 2).

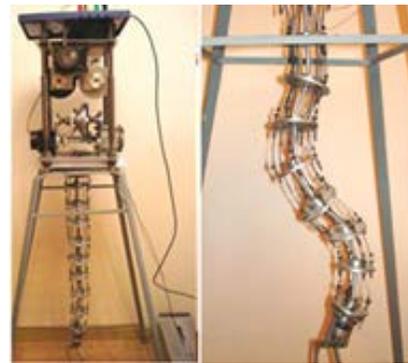


Fig. 2. Hyper-redundant robot elephant trunk – HipRob I

The robot can be considered as composed of groups of elements (modules), every such group can have its own bending. At least three actuation cables are necessary for every manipulator section bending. By actuating this cables section curvature is obtained. The number of cables for each section represents a major disadvantage because of the actuation system complexity, from the constructive and control point of view [1-12].

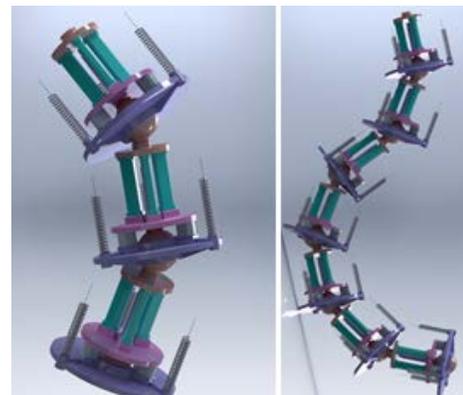


Fig. 3. Hyper-redundant robot HHR – Solidworks model

One of the cardinal goals of the robotic researchers is to provide control algorithms that allow robots to move in an unknown environment without collisions with the unstructured obstacles. Generally, in specific literature this problem is addressed as the collision avoidance problem [6], [9], [10], [11]. In this paper the hyper-redundant manipulator is analyzed in following situation: his job is to infiltrate, to creep through dropped walls after a cataclysm in a humanitarian action or for inspection in highly constrained environments (Fig.4).

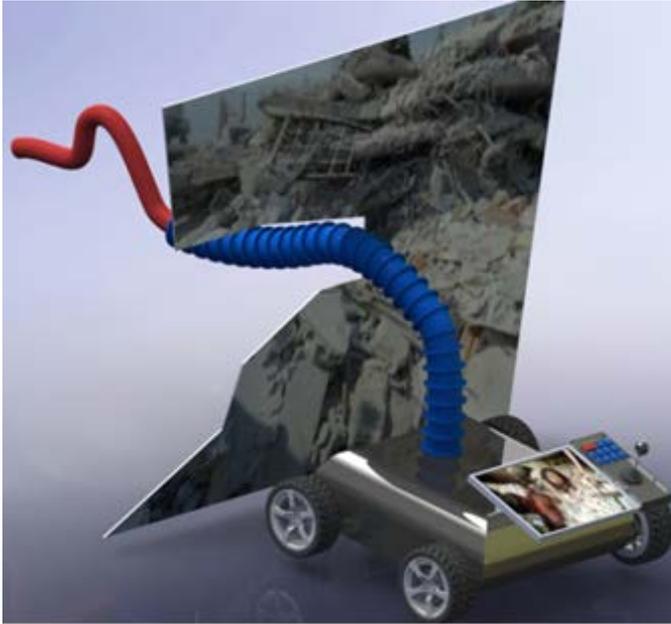


Fig. 4. Hyper-redundant robot in a humanitarian action

The mechanics are for a snake-like structure but with serial rigid links. In this case is necessary a compliant motion by touching of objects in many points of the manipulator like a human been in a very narrow cave which press with his shoulder the wall for his stability and with his hand push up a block (gravestone, manhole) or for climbing. The collisions or contacts are a part of the task.

The support points from the walls help to the moving process. For example, in fig. 1 (for simplicity, a planar structure) the robot fixes its i -link on the walls pressing the link's ends by F and F forces. In this situation the upper part (links $i+1, i+2 \dots n$) can be relaxed and inferior part (links $1, 2 \dots i-1$) can be constricted. This process allows a creeping in horizontal plane or a climbing in vertical plane.

This applications type consist in possibility of carry a conduct with rescue fluid, or another, in creeping spaces.

HHR it has an ideal architecture for this purpose, imposed by hybrid actuating system. Whatever the length of the robot only three actuating cables are required. We consider the arm a series module, each module contains k segments or k elements. Acting one module to one time moment, with control creeping algorithm, obtain movement on robotic arm in desire position, module by module, whatever the the length of them.

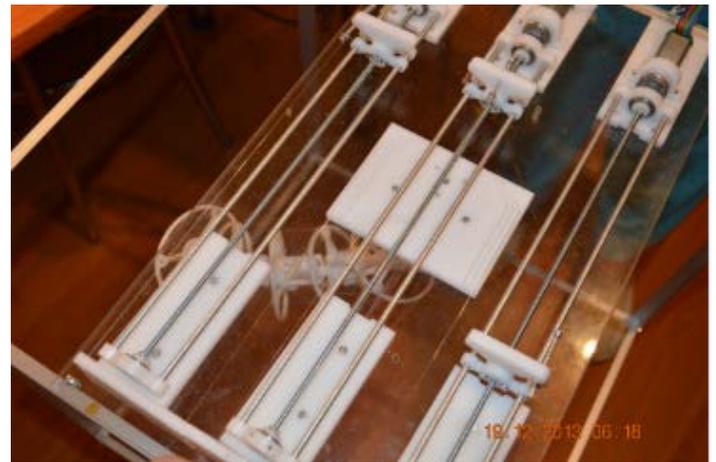
II. THE ROBOT ARCHITECTURE

A. General presentation

The system proposed considers the robotic arm as a single section, from the actuating point of view. Desired robot curvature is obtained using just three cables for actuating together with electro-pneumatically system. The actuating system is simplified by this fact. The position of every element can be blocked, independently of the others, using an electro-pneumatically system (Fig. 5).

The following operations are repeated for obtaining the desired position:

- The electro-pneumatically blocking system is used for all the components, except those for which one desires to change position.
- The position of unblocked elements is modified, using cable actuation system.
- The process is repeated for every element/group of elements until the desired position is reached.



(a)



(b)

Fig. 5. Backbone hyper-redundant arm actuated by a set of cables, (a)-acting part, (b)-robot trunk part

B. Element description

From the constructive point of view, a robot element consists in two disks connected by a rigid rod. One of the disks is considered the base of the element and is fixed rigidly on the rod, perpendicularly on it (Fig. 6 - Solidworks model, Fig. 7 - physical robot).

The second disk is attached to the rod through a spherical articulation. One end of the rod is composed of a sphere representing the internal part of the spherical articulation. The core of the disk is composed of a spherical cake representing the external part of the spherical articulation. By the restriction imposed by actuation system, this can be considered a four class coupling; the rotation around the axis is cancelled. Referring to the element construction this one represents the mobile disk (Fig. 7 b).

Constructively, the mobile disk can change its angular position with respect to the fixed disk, or the rigid rod, respectively, realizing partial rotation around axis OX and OY.

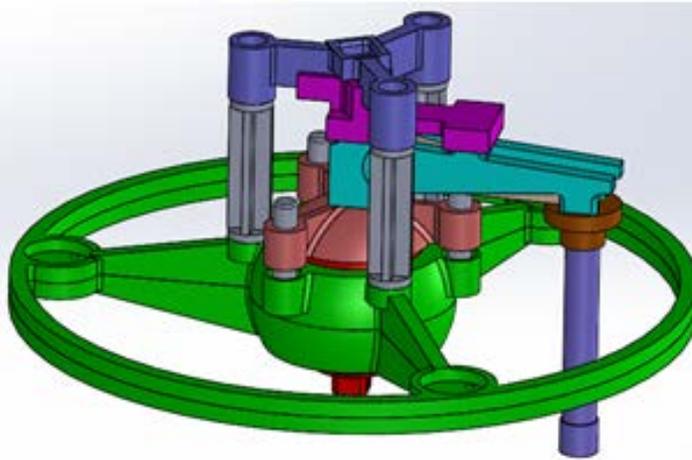


Fig. 6. Element of robotic arm – Solidworks model

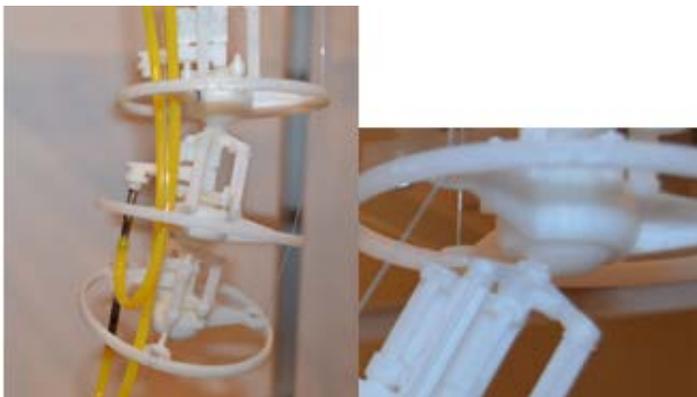


Fig. 7. Element of robotic arm, a-robot ansamble, b-spherical articulation detail

Every string is crossed by a cable, which also passes by the two disks. Cables are attached to the mobile disk and cross the fixed disk up to their actuation elements. By actuating cables, mobile disk rotates around spherical articulation as against axis OX, OY or both of them.

C. The electro-pneumatically blocking system

A simple action pneumatic piston is implemented inside of the rod. For acting, the mobile element of the piston presses on the internal sphere of spherical articulation (Fig. 8, Fig. 9).

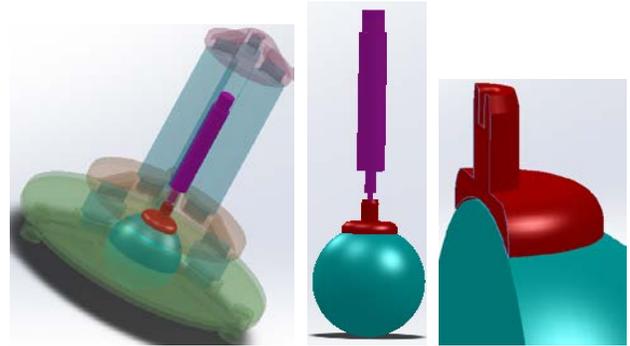


Fig. 8. Element of robotic arm – electro-pneumatically system – Solidworks model

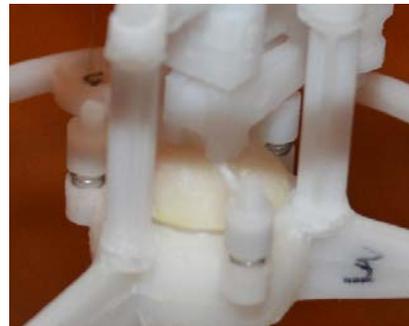


Fig. 9. Element of robotic arm – electro-pneumatically system detail

Both of the parts of the sphere that interact with the piston are properly processed in order to ensure a great rubbing coefficient. The pressure of the system is controlled through an electro valve.

The manipulator is committed from a series element connection. Two elements have a common disk, the mobile disk of an element representing the fixed disk for the following disk. The cables are connected at the terminal point of the arm, at the last mobile disk.

The electro-pneumatically system ensures the blocking-on or blocking-off of the elements. The elements that are not blocked (blocking-off state) can be bended by this cable-based actuation system.

III. DYNAMIC MODEL

In Figure 10 is presented a hyper-redundant robotic structure for establish a dynamic model by Lagrange method.

With the notations from that figure are passed get through all phases of the algorithm. The coordinates $x(q)$ and $y(q)$ are:

$$x_k = \sum_{i=1}^n l_i \cos\left(\sum_{j=1}^i \theta_j\right) \tag{1}$$

$$y_k = \sum_{i=1}^n l_i \sin\left(\sum_{j=1}^i \theta_j\right) \tag{2}$$

The velocities of the x and y are:

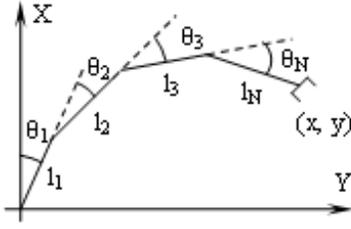


Fig. 10. The backbone architecture

$$\dot{x}_k = -\sum_{i=1}^k \left\{ l_i \left[\sin \left(\sum_{j=1}^i \theta_j \right) \right] \left(\sum_{j=1}^i \dot{\theta}_j \right) \right\} \quad (3)$$

$$\dot{y}_k = -\sum_{i=1}^k \left\{ l_i \left[\cos \left(\sum_{j=1}^i \theta_j \right) \right] \left(\sum_{j=1}^i \dot{\theta}_j \right) \right\} \quad (4)$$

The kinetic energy and the potential energy are:

$$E_c(q, \dot{q}) = \frac{1}{2} \dot{q}^T M(q) \dot{q} \quad (5)$$

$$E_p(q) = \sum_{i=1}^n E_{p_i} = g \sum_{i=1}^n m_i \left[\sum_{j=1}^i l_j \sin \left(\sum_{k=1}^j \theta_k \right) \right] \quad (6)$$

where $M(q)$ is the inertial matrix and

$$q = [\theta_1, \theta_2, \dots, \theta_n]^T \quad (7)$$

The Lagrange function is:

$$L(q, \dot{q}) = E_c(q, \dot{q}) - E_p(q) = \frac{1}{2} \sum_{i,j=1}^n M_{ij}(q) \dot{q}_i \dot{q}_j - E_p(q) \quad (8)$$

Now, we can apply Lagrange's equation are:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = T_i \quad (9)$$

The two terms of the Lagrange's equation are:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) = \frac{d}{dt} \left(\sum_{j=1}^n M_{ij} \dot{q}_j \right) = \sum_{j=1}^n (M_{ij} \ddot{q}_j + \dot{M}_{ij} \dot{q}_j) \quad (10)$$

$$\frac{\partial L}{\partial q_i} = \frac{1}{2} \sum_{j,k=1}^n \frac{\partial M_{kj}}{\partial q_i} \dot{q}_k \dot{q}_j - \frac{\partial E_p}{\partial q_i} \quad (11)$$

Now, the Lagrange's equation is:

$$\sum_{j=1}^n M_{ij}(q) \ddot{q}_j + \sum_{j,k=1}^n \Gamma_{ijk} \dot{q}_j \dot{q}_k + \frac{\partial E_p}{\partial q_i}(q) = T_i \quad (12)$$

where:

$$\Gamma_{ijk} = \frac{1}{2} \left(\frac{\partial M_{ij}(q)}{\partial q_k} + \frac{\partial M_{ik}(q)}{\partial q_j} + \frac{\partial M_{jk}(q)}{\partial q_i} \right) \quad (13)$$

If we define the Coriolis matrix as:

$$C_{ij}(q, \dot{q}) = \sum_{k=1}^n \Gamma_{ijk} \dot{q}_k = \frac{1}{2} \sum_{k=1}^n \left(\frac{\partial M_{ij}(q)}{\partial q_k} + \frac{\partial M_{ik}(q)}{\partial q_j} + \frac{\partial M_{jk}(q)}{\partial q_i} \right) \dot{q}_k \quad (14)$$

Then $C(q, \dot{q})\dot{q}$ is the Coriolis and centrifugal forces vector.

The third term is a position term representing loading due to gravity.

$$\frac{\partial E_p}{\partial q_i}(q) = G(q) \quad (15)$$

T_i represents the outer forces:

$$T_i = T_i - B(\dot{q}) - \tau \quad (16)$$

where T_i are the actuators forces (inner generalized forces), B represents the friction forces and τ represents the forces from joints due to the forces acting on the tool tip or on the another points from the robotic structure due to a load or contact with the environment. With the above-mentioned notations we have the dynamic model of the hyper-redundant robotic structure from Figure 10:

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + B(\dot{q}) + G(q) + \tau = T \quad (17)$$

In this paper, by "position" we mean "position and orientation", and by "force" we mean "force and torque".

IV. HYBRID CONTROL

During creeping in tunnels or labyrinths the robot comes in contact with environment; therefore, interaction forces develop between the robot and environment. Consequently, these interaction forces, as well as the position of the points from the robot, must be controlled. A number of methods for obtaining force information exist: motor currents may be measured or programmed, motor output torques may be measured and joint-mounted sensors may be used. There are two primary methods for producing compliant motion: a passive mechanical compliance built into the manipulator, or an active compliance implemented in the software control loop, force control. Passive compliance offers some performance advantages, but the force control method offers the advantage of programmability. This allows the manipulator to use the particular form of compliance necessary for the particular application. Throughout this paper is assumed that the desired velocity and force trajectories, which are commanded by the controllers, are consistent with the model of the environment [19]. If this is not the case, it may be possible to modify the desired velocity and force trajectories to be consistent with the model of the environment. The compliant motion occurs when the position of a robot is constrained by the task. The so-called hybrid position/force controller [20], [21] can be used for tracking position and force trajectories simultaneously. The basic concept of the hybrid position/force controller is to decouple the position and force/control problems into subtasks via a task space formulation because the task space formulation is valuable in determining which directions should be force or position controlled. Figure 11 presents a hybrid position/force controller.

$$x_d = \begin{bmatrix} x_d \\ y_d \\ z_d \end{bmatrix}, x = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \Delta x = x_d - x, \quad (18)$$

$$F_d = \begin{bmatrix} F_{xd} \\ F_{yd} \\ F_{zd} \end{bmatrix}, F = \begin{bmatrix} F_{xd} \\ F_{yd} \\ F_{zd} \end{bmatrix}, \Delta F = F_d - F \quad (19)$$

$$q = [q_1, q_2, \dots, q_n]^T, f = [f_1, f_2, \dots, f_n]^T \quad (19)$$

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, S = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & S_z \end{bmatrix} \quad (20)$$

where S defines the compliance selection matrix.

$S_i=0$ for position control

$S_i=1$ for force control

$$T_{Fj} = J_j, T^{-1}_{Fj} = (J^{-1})^T \quad (21)$$

where $J(q)$ is the Jacobian matrix and T_{Fj} is a metrical operator which make the transformation between the forces F and the forces F . $x=f(q)$ is a relation which represent the forward kinematics. For position controller is advantageously to use a PID controller and force a PD controller.

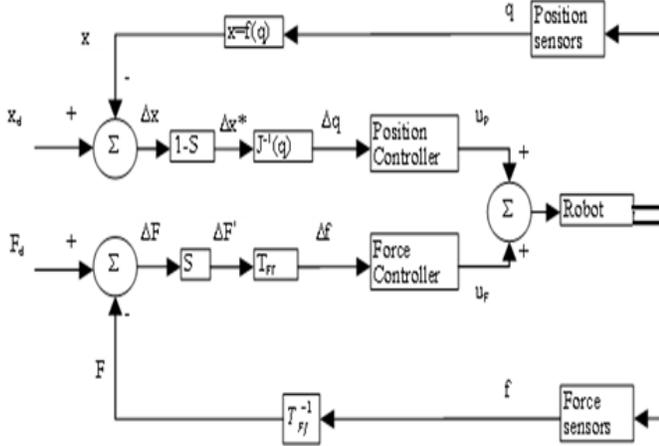


Fig. 11. The Hybrid Position / Force Controller

$$u_p = k_1 \Delta q + k_2 \Delta \dot{q} + k_3 \int_0^t \Delta q dt \quad (22)$$

$$u_f = k_4 \Delta f + k_5 \Delta \dot{f} \quad (23)$$

The general control law u_i is

$$u_i = \sum_j [\Psi_{ij} S_i \Delta_j i + \Xi_{ij} (1 - S_j) \Delta q_i] \quad (24)$$

where Ψ_{ij} is a force control law and Ξ_{ij} is a position control law. Figure 12 presents a simulation for the climbing operation through a vertical labyrinth for one module with k_1, \dots, k_n elements.

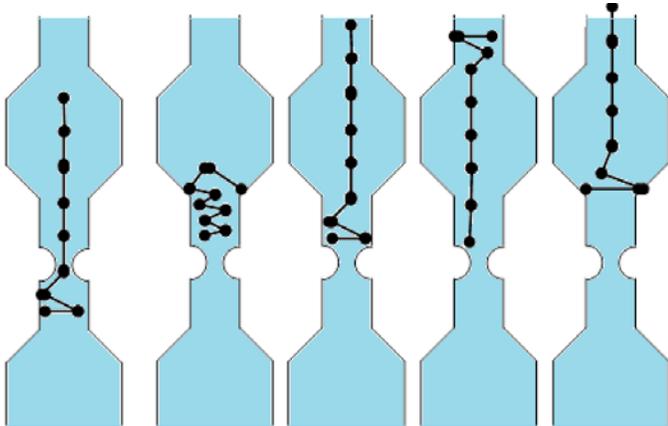


Fig. 12. Simulation for a climbing action for one module, $k_1 \dots k_n$ elements (evolution in time from left to right).

Any segments of the hyper-redundant robot are used for the anchorage of the structure on the walls of the vertical labyrinth while the others segments are moving up. The hyper-redundant locomotion by creeping is analogous.

V. CONCLUSIONS

The cable drive system and sensorial system presented in the paper simplifies the mechanical structure, the drive and controls a hyper-redundant robotic arm.

Control algorithms specially adapted to this structure confirmed the stability, robustness and reliability of the system. This is presented in a conference paper [13, 14].

A new control algorithm for a hyper-redundant manipulator, which moves in a labyrinth and can touch the boundaries of the niche, is presented in this paper. The collisions or contacts are part of the task. The support-points from the walls help with the moving process. The dynamic model by Lagrange method of the hyper-redundant arm is established. Afterwards, a hybrid position/force controller for the hyper-redundant locomotion is proposed. Finally, the simulations of the creeping of the hyper-redundant arm are presented.

REFERENCES

- [1] Kapadia, I. Walker, D. Dawson, "A model - based sliding mode controller for extensible continuum robots", Recent Advances in Signal Processing, Robotics and Automation, ISPRa Conf., 2009, pp. 103 - 120
- [2] A. Khopalov, "Source localization and sensor placement in environmental monitoring", Int. Journal Appl. Math. Computer Science, 2010, vol. 20, no. 3, pp. 445 - 458
- [3] B. Jones, I.D. Walker, "Practical kinematics for real-time implementation of continuum robots", IEEE Trans. Robotics, vol. 22, no. 6, Dec. 2006, pp. 1087 - 1099
- [4] Camarillo, D., Milne, C., Mechanics Modeling of Tendon - Driven Continuum Manipulators, IEEE Trans. on Robotics, vol. 24, no. 6, December 2008, pp. 1262 - 1273
- [5] G. Robinson and J.B.C. Davies, "Continuum Robots - A state of the art", Proc. IEEE Int. Conf. on Robotics and Automation, Detroit, May 1999, pp. 2849 - 2854
- [6] Gravage, Ian A., Walker, Ian D., On the kinematics of remotely - actuated continuum robots, Proc. 2000 IEEE Int. Conf. on Robotics and Automation, San Francisco, April 2000, pp. 2544-2550
- [7] Gravage, Ian A., Walker, Ian D., Kinematic Transformations for Remotely-Actuated Planar Continuum Robots, Proc. 2000 IEEE Int. Conf. on Rob. and Aut., San Francisco, April 2000, pp. 19-26
- [8] Ivanescu, M., Florescu, M.C., Popescu, N., Popescu, D., Position and Force Control of the Grasping Function for a Hyperredundant Arm, Proc. of IEEE Int. Conf. on Rob. and Aut., Pasadena, California, 2008, pp. 2599-2604
- [9] Ivanescu, M., Bizdoaca, N., Florescu, M., Popescu, N., Popescu, D., Frequency Criteria for the Grasping Control of a Hyper-redundant Robot, Proc. of IEEE International Conference on Robotics and Automation, Anchorage, Alaska (ICRA 2010), May 3 - 8, 2010, pp. 1542-1549
- [10] Krstic, M., Smyshlyayev, A., Boundary Control of PDEs: A Short Course on Backstepping Design, VCSB, 2006. I. Walker, M. Hannan, "A novel elephant's trunk robot", AIM '99, pp. 410 - 415
- [11] Vladu, I., Strimbeanu, D., Ivanescu, M., Bizdoacă, N., Vladu, C., Florescu, M., - Control System for a Hyper-Redundant Robot - INCOM '12 - Bucharest, may 2012
- [12] Vladu, I., Strimbeanu, D., Ivanescu, M., Bizdoacă, N., Vladu, C., - "Robotics Structure Control with Rheological Actuators" - National Conference of PhD Students "Romanian PhD Research with UE Support" - Timișoara, feb. 2013

- [13] Nirvana Popescu, Mircea Ivanescu, Dorin Popescu, Cristian Vladu, Ileana Vladu, "Force Observer-Based Control for a Rehabilitation Hand Exoskeleton System," în Asian Control Conference ASCC 2013, Istanbul, Turcia, 2013
- [14] Ileana Vladu, Mircea Ivănescu, Ionel Cristian Vladu, "HYPER-REDUNDANT ARM actuation with electro-pneumatic actuation system – part I - Construction", în ICSTCC, Sinaia, 2013.
- [15] Chirikjian, G.S. and J.W. Burdick, 1993. "Kinematically Optimal Hyper-Redundant Manipulator Configurations", IEEE Transactions on Robotics and Automat., vol II, No.6, pp. 794-805.
- [16] Colbaugh, R.; H. Seraji; and K.L. Glass. 1989. "Obstacle Avoidance for Redundant Robots Using Configuration Control", J. of Robotic Systems, vol. 6, pp 721 – 744.
- [17] Khatib, O. 1986. "Real-Time Obstacle Avoidance for Manipulators and Mobile Robots", Int. J. of Robotic Research, vol. 5, pp. 90 – 98.
- [18] Zghal, H.; R.V. Dubey; and J.A. Euler. 1992. "Collision Avoidance of a Multiple Degree of Redundancy Manipulator Operating Through a Window", Transaction of the ASME, Journal of Dynamic Systems, Measur., and Control, vol 114, pp. 717-721
- [19] Lipkin, H. and J. Duffy. 1988. "Hybrid Twist and Wrench Control for a Robot Manipulator", Trans. ASME J. Mechan. Transmissions Autom. Design, vol 110, pp. 138-144.
- [20] Chae, A.; C. Atkenson; and Hollerbach J. 1988. "Model-Based Control of a Robot Manipulator", Cambridge, MA, MIT Press.
- [21] Raibert, M. and J. Craig. 1981. "Hybrid Position/Force Control of Manipulators", J. Dyn. Syst. Meas. Control, vol. 102, pp. 126-133.

Implementation of Type Wheels to Design Caterpillar Prototype Chair with Automated Features for People with Disabilities in Paraplegia

Maribel Aguilar Echeverri, Arnold Romero Chaparro, Milena Alejandra Pirajan Zamora, Juan Diego López Vargas

Abstract—at present, there is large variety of conditions in the health that prevents the autonomous movement of the people, in some severe cases, in which patients are totally motor of its powers, it becomes necessary to use transport aircraft, such as the use of wheelchairs. The purpose of the following article, is the design of a prototype wheelchair, that through the application of type wheels, track suits dissimilar land and that by means of an automated control system, the operator can adjust the speed of acceleration and deceleration of the same. The results show that is a viable project, that meets the needs of the population with disabilities paraplegia and that has a high social projection, because it is a design of wheelchair to low acquisition cost, with a management control simple and effective.

Keywords—Autonomous; Design; paraplegia; track type wheels; social projection.

I. INTRODUCTION

The manual wheelchairs, since 1595 have been a useful aid to the mobility of persons with disabilities, but they have some shortcomings in its design which produce transportation difficulties to those who use them; they can even cause a disability even greater, as determined by the World Health Organization (WHO) in her article on manual wheelchairs drafted in 2008.

One of the most noticeable difficulties of the wheelchairs current, is the shift in areas that are unstable or who possess steps, because the design of the wheels does not allow the promotion by these and the person must make greater effort to be able to scroll; also if the person does not have total movement of the arms and legs is of greater utility a automatic wheelchair that allows their mobility. The following research is conducted in the wake of the next question: How can you improve the problem of wheeled mobility in different areas for people with disabilities in paraplegia? The arguments described on this problem previously and a thorough investigation supported by verifiable sources, underpin the approach to the problem of the present investigation. People with paraplegia have difficulty daily to be transported to their places of work or at home, according to a report of the WHO, as shown in the source [1].

In the users whose disability involves pain (patients with AR and EA), were able to observe changes in their physical structure which have a direct impact on their bodily functions, and, therefore, in its activities, in their accessibility to the environment and their social participation.

"When you walk on the pavement completely new [...] this soft. But when you're on the road very bad [...] ratchets much the wheelchair and greatly affects the wear of the hip" (L. , Colombian user, 59 years, AR, housewife) [2].

Fig. 1. Problems of the Conventional Wheelchair



Source: [1]

The track type wheels, consists of a set of link modular allowing a displacement stable even on rough terrain, the majority of the tracks are part of a flexible belt with a rigid set of links to other strongly.

In order to enable people with disabilities to live independently and participate fully in all aspects of life, States Parties shall take appropriate measures to ensure the access of persons with disabilities, on an equal footing with the other, the physical environment, transport, information and communications, including the systems and information and communications technologies, and other services and facilities open to the public or for public use, in both urban and rural areas. These measures, which will include the identification and elimination of obstacles and barriers to access, shall apply to [3]; hence this project is intended to be a part of these measures of inclusion, designing a wheelchair that allows access to several areas, in order that people with disabilities are treated equally, pretending to remove as many barriers in the displacement by areas of the country [4].

II. METHODOLOGICAL DESIGN

The methodology in the process of research, is based on the study of population, materials, techniques and procedures, aimed at ensuring compliance with the objectives. The innovation of this project is based on the projection that has social and economic design that looks for, as this is not a project for high-tech to high costs, but a viable project affordable for all types of people:

-Population: the type of population group to which you are targeted by the project, are people who suffer from the disease of paraplegia, due to the fact that if they have other conditions such as immobility of arms may not activate the wheelchair; although the automated wheelchair of this project is intended to mitigate the problem of mobility of people with paraplegia, there are other people that they can also make use of it, as are people who are in recovery for any other type of illness or those who do not may be transported by themselves by dissimilar reasons.

-Materials, techniques and procedures: emphasis is on the following methodology:

A. In Software Development

For the realization of drawings in the design of the wheelchair, was used in a computer program with the rhinoceros, which allows the modeling in 3D; the program for the ARDUINO programming of the card that commanded the instructions toward the gear motor and the MATLAB program for the modeling of the PID control [5].

B. Development of Hardware

The development of the prototype is made in rods CR (chrome) of 8 mm for the chassis, upholstery and foam to the

polyurethanes FMG coated, wheels of carts of 16 cm in diameter, toothed belt ½ inch, plastic and velcro for water,

infrared sensors, sprockets and gearmotor with encoder for the operation. For the design to scale would require real rods in stainless steel, upholstery for body assembly type track wheels so Queta, belts for automobile and electronic instrumentation as it is: infrared sensors, gear motor high-end, wiring THW AWG No. 12, Three-position switches of shank card ARDUINO, H-bridge 1298, terminals, connectors, halogen lights, etc.

C. Development of the Automation

The development of the automated portion is linked in controlling the acceleration and deceleration of the chair and to efficiently control the speed by means of a PID controller that will be programd into the ARDUINO card; it should be noted that in the automated portion of the wheelchair is has in mind the protections and emergency electrical necessary to give the user greater peace of mind [6].

It is endeavoring to comply with the following objectives:

➤ General Objective

Designing a chair with wheels type track that allow the movement of people with disabilities of paraplegia, adaptable to all sorts of terrain.

➤ Specific Objectives

- Analyze the state of the art of the wheelchairs for persons with disabilities of paraplegia.

- Analyze the types of wheelchair, using a characterization.

- Implementing and designing a wheelchair of automatic handling, adaptable to different land, taking into account specialized hardware and software, at low cost.

- Look for strategic alliances, participate in calls external or internal to the University for the financing of the second phase of the project.

III. RESULTS

A. Mechanical Description of the Prototype of the Wheelchair

Automated wheelchair can be mobilised by all sorts of terrain by means of wheels type track and gear motor controlled by card programd ARDUINO in computation; in addition to mobilize to any side, is designed under standard to actual scale and ergonomic, with security features that give the user reliability to be transported in it.

The design of the wheelchair was carried out in automated program rhinoceros version 4.0 as shown in Figure 2:

Fig. 2. Design 3D Prototype of the Wheelchair

Source: Authors

B. Development of Mathematical Calculations.

- Tangential velocity of the wheel:

$$v = w * r \text{ m/s}$$

$$v = \text{wheel speed}$$

$$w = \text{angular frequency}$$

$$r = \text{radio} = 8\text{cm}$$

$$w = 2\pi f$$

$$f = 200 \text{ rpm}$$

$$f = \frac{200}{60 \text{ s}} = 3.33 \text{ rps}$$

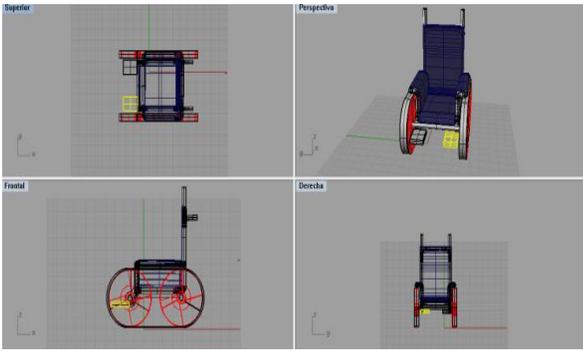
$$w = 2\pi * (3.33)$$

$$w = 20.92 \text{ rad/s}$$

$$v = 20.92 * 0.08$$

$$V=1.6736 \text{ m/s}$$

Resistivity of a chrome rod:



$$R = \frac{\rho * l}{A}$$

$R = \text{resistivity of the rod}$
 $\rho = \text{chrome Resistivity} = 12.5 \times 10^{-8}$
 $l = \text{length} = 3\text{m}$

- Across – sectional area depending on the diameter:

$$A = \frac{\pi d^2}{4}$$

$$A = \frac{\pi (8 \times 10^{-3})^2}{4}$$

$$A = 0.05026 \times 10^{-3} \text{ m}^2$$

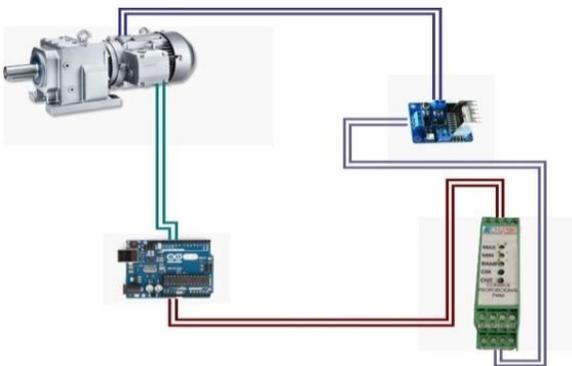
$$R = \frac{(12.5 \times 10^{-8}) * (3)}{0.05026 \times 10^{-3}}$$

$$R = 7.4612 \times 10^{-3} \Omega$$

C. Description of the Automated Wheelchair to Actual Scale

The wheelchair to actual scale will be according to the design a card with ARDUINO, a/c gear motor with encoder, a PWM, a H-bridge l298 and a switch style shank for start, as shown in Figure 3:

Fig. 3. Control System Design



Source: Authors

The design of control of the wheelchair is based on a c gearmotor with DC power from the NDCMG series, with up to 250 watts, which has the following features mechanical; it should be borne in mind that the prototype was used a motor reducer of lesser range.

It requires a card program ARDUINO and version 6, to schedule the following lines of code and through the PWM

internal card simulate the operation of the control of the speed of the gear motor [7].

It should be borne in mind that the error must be less than 5 %, furthermore must contain the following control variables:

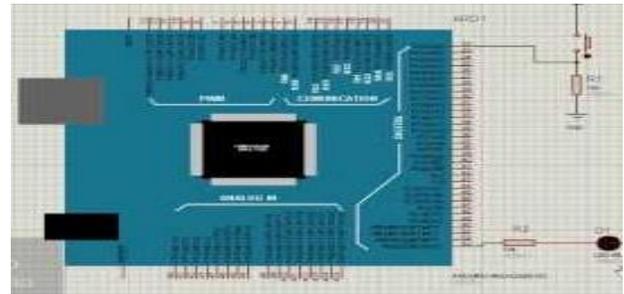
KP= proportional constant

Ki= cash integral

Derivative Kd= constant

In turn, is shape in MATLAB the PID controller, in order to find values that are approximated to have a minimum response error in the output and enter it in the Arduino to control the passage of the PWM and so set the speed of the engine, the start-up time, stop and stabilization at rated speed, it should be noted that the margins of error must be minimal due to the fact that this is the responses that the user is going to get the wheelchair in operation and you cannot put in danger the same; the simulation of the ARDUINO will be as in Figure 4:

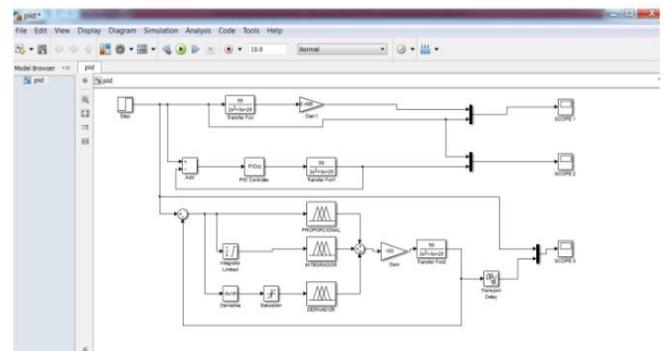
Fig. 4. Simulation in ARDUINO

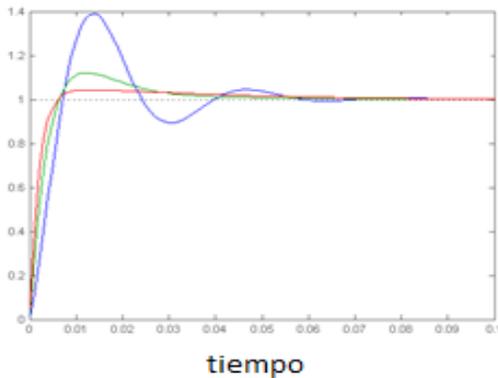


Source: Authors

The plant is mounted in MATLAB as a dynamic model of a motor load and controls the starter until it is stable:

Fig. 5. Simulation in MATLAB





Source: Authors

Once you get the values of the constants of the PID, digital on the agenda of the ARDUINO, which controls the engine speed, because the input of the same is measured by means of the internal encoder of the motor reducer that by means of drivers entering the ARDUINO give the number of revolutions per minute and by means of the output of the L 298 at the time of the signal of the ARDUINO calculates the voltage that will enter the engine.

It should be noted that the automated system will be powered by battery type car of 24 V.D.C. with electric charger connection to ac power [8].

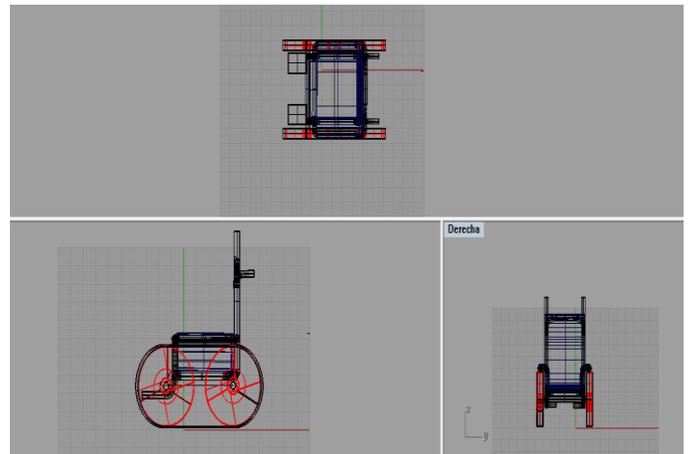
D. Implementation of the Prototype

Complying with the aim of implementing and designing a wheelchair of automatic handling, adaptable to different land, taking into account software and specialized hardware, at a low cost, is carried out the assembly below:

For the design of the wheelchair for people with disabilities of paraplegia, has been taken into account the aforementioned investigations, to generate the design that best suits the needs of users, in order to meet the needs of the mobility in all sorts of terrain with an automatic control, taking into account each of the following steps:

1) Design a scale of 1:3 of the Chair, the gender program design in the Rhinoceros.

Fig. 6. Simulation in MATLAB



Source: Own, developed by the project developers, rhinoceros 4.0 .

2) Assemble the Chassis; figure 7 shows the design that is intended to establish to meet the proposed objectives of displacement of the wheelchair in disparate fields.

Fig. 7. The Prototype Chassis



Source: Authors

According to the design of the levels mentioned above, was the implementation of the prototype. The materials used and the results obtained are described below:

- CR 8 mm rod: The chassis of the prototype of the wheelchair was made with 8 mm rod CR, this has a high resistance to corrosion, the metals that contains have great affinity for oxygen and reacts with the forming a layer passivating, thus avoiding the corrosion of iron. Introducing high resistance to corrosion is an optimum material that has a high durability and long life that is what you are looking for when choosing a wheelchair, in the figure 8 shows the chassis soldier.

Fig. 8. Chassis Soldier



Source: Authors

- **Toothed Belts:** The toothed belts of the prototype were used in the implementation of the type wheel track, these belts are high intensity, high flexibility, good durability, possess excellent resistance to heat and wear resistant, additionally their slots have a frictional force optimum to adapt to the different land, in the figure 9 shows the belts used.

Fig. 9. Toothed Belts



Source: Authors

- **Wheel of 16 cm in diameter :** required the use of four wheels according to the design of the plane, the wheels are plastic and the rubber cover that achieves adjustment with belts, in figure 10 shows the wheels used.
- The mount was conducted and presented in the following manner, along with the implementation of safety belts:

Fig. 10. Wheels



Source: Authors

- **Polyurethanes FMG coated:** fabric was used for auto upholstery color and blue foam king of 1 cm of thickness for chair, in order to shape it in the chassis.

Fig. 11. Wheels



Source: Authors

- **Transparent plastic:** plastic and velcro used 0.5 mm thickness for the top of the chair, with the purpose of preventing the user and electronic system will get wet.

Fig. 12. For waters of the Chair



Source: Authors

Fig. 13. Prototype of the Wheelchair



Source: Authors

IV. CONCLUSIONS

In the document signed is reflected the work of the project, in which the design of a prototype of a wheel chair with implementation of type wheels and track through the simulation program is able to control the speed system.

It was assumed by the design of the wheelchair wheels by means of the Rhinoceros software, this turns out to be very useful when it comes to choosing a design program, allowed us to analyze the different mechanical variables that would involve the chair to the prototype and the actual scale. Mathematical calculations were necessary for the selection of the transmission elements of movement. By means of the relevant simulations of the automation program for the motor reducer of the chair with the ARDUINO program generated the lines of code of the PID controller.

Simulating the start and deceleration of the gear motor in MATLAB, was chosen from the constants required for the program in ARDUINO, verifying the stabilization of the plant such as a dynamic model of a motor load. Once you have verified that the project was not feasible, the prototype was

implemented at a scale of 1:3, as realistic as possible, this in order to qualify as an option of degree of specialists in the ECCI University. With the approval of the project is intended to be attested by a legal entity for its functioning and launch a real scale.

REFERENCES

- [1] A new dynamic model of the manual wheelchair for straight and curvilinear propulsion. Publication, Rehabilitation Robotics (ICORR), IEEE International Conference on 2011.
- [2] Patricia Herrera-Saraya , Ingris Peláez-Ballestasb, Luciana Ramos-Lirac , David Sánchez-Monroyd, Ruben Burgos-Vargas , Colombia - Mexico, 2012.
Orthopedics Guzman. Types of components of a wheelchair. Interview Date 16 April 2015.
- [3] A new dynamic model of the manual wheelchair for straight and curvilinear propulsion. Publication, Rehabilitation Robotics (ICORR), IEEE International Conference on 2011.
- [4] Arduino - Arduino.cc/in/Guide/Introduction - consultation date of 16 April 2015.
- [5] Multidodal intelligent wheelchair control based on fuzzy algorithm. Publication Information and Automation (ICIA), International Conference on 2012.
- [6] A new dynamic model of the manual wheelchair for straight and curvilinear propulsion. Publication, Rehabilitation Robotics (ICORR), IEEE International Conference on 2011.
- [7] Orthopedics Guzman. Types of components of a wheelchair. Interview Date 16 April 2015.
- [8] ICONTEC. NTC 6047. Colombian technical standard for accessibility to the physical environment . Spaces for citizen service in public administration. Requirements. Consulted in August of 2014.

First author, Maribel Aguilar Echeverri, born in Bogota, Colombia on 15 December 1988. He graduated as a technologist in Industrial Electricity in the SENA in 2010; He completed his undergraduate studies at the Colombian School of Industrial Races. He graduated in Media Telecommunications technologist at the Electronic Engineering in 2013 and 2014; He is attending the second module Eng. In Industrial Automation, 2015. He is currently working as a teacher of physics and technology at the British Institute of Bogota.

Second Author, Milena Pirajan Alejandra Zamora, born in Bogota, Colombia on February 4, 1992. He completed his undergraduate studies at the Colombian School of Industrial Races technologist Graduated in 2012 in Industrial Electronics and Electronic Engineering in 2014. It is located in his second module Eng. In Industrial Automation, 2015. He is currently working in the area of support for company audits the Ministry of Education.

Third Author, Arnold Yesid Chaparro Romero, born in Bogota, Colombia on 29 March 1992. He graduated as a technician in assembly and maintenance of computers in 2009; He completed his undergraduate studies at the Colombian School of Industrial Races, winning the title of Technologist in Industrial Electronics in 2013 and Electronic Engineering in 2014; He is attending the second module Eng. In Industrial Automation, 2015. He served for five years in the area of maintenance and support of computers, 3 years in the administrative area of the Civil Aeronautics (one year in telecommunications), currently Huawei is

working as an engineer in support to customers massive failures clear

Author quarter, Juan Diego Lopez Vargas, born in Tulua Valle, Colombia January 25, 1981. Telecommunications Engineer, Specialist, Master and PhD in Integration of IT in Organizations, of the Polytechnic

University of Valencia. He has been a research faculty of the Universidad Santo Tomas and Manuela Beltran University, about 8 years. He has also been Vice-President for Research and Investogaciones Coordinator, is currently Research Coordinator Research Masters of the ECCI University and Director of Business Innovation Thinki

Metaheuristics based on the Variables integration Method applied to Reactive Power Compensation in Multi-objective Optimization.

Iliana González Palau, Secundino Marrero Ramírez, Arístides Legrá Lobaina, Daniel Mendiola Ellis

Abstract—Compensation of reactive power is widely used in industrial energy supply networks. To guarantee an efficient operation and adequate amount of devices several technical and/or economic criteria should be considered. This paper proposes a method to solve an optimization problem with two objective functions, and with the use of the Tchebycheff's distance achieve the reduction of the distance between calculated values and desired objectives values. A metaheuristic method is employed as technique and the initial population is obtained and refined by Random Conditional Search Method (RCS) with stop criteria adjusted over a satisfying and bounded solution set. The procedure and algorithm are tested with theoretical and real cases in order to measure efficiency, speed and tradeoffs.

Keywords—Multi-objective Optimization, Reactive Power Compensation, Evolutionary Algorithms.

I. INTRODUCTION

Commonly, the solutions for compensation studies in industrial systems are the location of capacitors/filters to achieve voltage profiles within the pre-established ranges, energy loss reduction and the reduction of harmonic contamination. These studies carries an optimization problem with one correlated objective associated to reduce the compensation costs and the other parameters included in constrains[2]. In a broader way, the reactive power optimization problem solutions has been focused from conventional optimization techniques including the gradient method, quadratic programming, nonlinear programming, linear programming and interior point method, but the recent and actual trends targets to a multi-objective context through evolutionary optimization algorithms [3-7]. In general, these techniques are still associated with some difficulties such as handling of the multi-model characteristic of the problems and

also non differential, nonlinear and non-convex nature of the reactive power optimization problem.

Genetic Algorithms (GA) method is the mostly used heuristic method among others. The objective of GA is to look for the best solution within a series of solutions; by "best solution" we refer those that optimize to a predetermined metrics for a given problem, meaning the value that comes closest to this numerical value, once evaluated for a particular evaluation of a function [8].

The traditional optimization algorithms generally obtain an optimal solution, however the application of evolutionary optimization methods permits several simultaneous and independent objectives originating a set of so-called optimally efficient solutions or Pareto solutions of which those that satisfy the reviewer's set of preferences will be selected.

The method formulated in this paper is built upon the Integration Variables Method [9] as a metaheuristic variant or genetic algorithms generalization. The functioning bases rest on a not necessarily use of a variable code to describe the possible solutions, but on any quantity of them and whatever set of operators to update the members of the population.

This variation is because during the application of genetic algorithms to optimization problems the combined and generated solutions can generate bad blocks of solutions depending on the objective function nature and producing scarcity of genetic diversity. This deficiency could be overcome by metaheuristic methods deficiency could be overcome by metaheuristic methods [10], where a modification of Variables Integration Methods (VIM) was developed and named Random Conditioned Search (RCS), and giving solution to convergence problems founded when the VIM was applied to very big populations. The modification permits:

1. The enhancement of the population by subintervals aimed to achieve a greater diversity over the solutions set in the extended population.
2. Search in the near neighborhood of the efficient calculated solutions with target in better performance member addition.
3. Search in adequate samples of population. Starting with the better characteristics founded in the population already calculated, it is selected code-dependent a subpopulation in the uncalculated population for prioritized evaluation.

This work has been supported in part by ISMM-MES, Cuba and in part by RENET- EduLink, Contract Number: FED/2013/320-289.

Iliana González Palau, Cotopaxy Technical University, Ecuador,
palauiliana@gmail.com

Secundino Marrero Ramírez, Cotopaxy Technical University, Ecuador,
smarreroram@gmail.com

Arístides Legrá Lobaina at ISMM, Cuba, alegre@ismm.edu.cu

Daniel Mendiola Ellis at ISMM, Cuba, dmendiola@ismm.edu.cu

II. REACTIVE POWER COMPENSATION AS A MULTI-CRITERIA DECISION MAKING SYSTEM.

The characterization of reactive power compensation as a decision-making process lead to an external analysis, which should be realized in order to include the selection of implied variables (see fig 1)

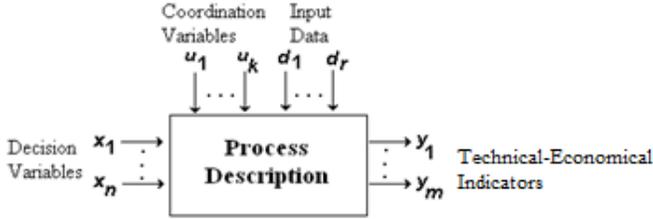


Fig. 1 Information Classification in External Analysis.

The installed load in the supply network was considered as coordination variables, variables u_i and d_i are related to every load data.

The x_i variables represent decisions variables.

The technical and economic indicators y_i , with pre-weights according to the criteria of experts, are as follows: Voltage in nodes, power factor in nodes, total harmonic distortion, the energy losses of the network (energy loss reduction obtained in the process of compensation) and the economic variable (VAN). To obtain these indicators, proper procedures of calculations in engineering systems are used as a model.

III. PROBLEM FORMULATION.

The Chebyshev's distance weight method permits the reduction of the distance of the calculated value to the desired value of every object included in the function by means of the expression;

$$\max_i \left\{ \omega_i \frac{|y_{ci} - y_{di}|}{|y_{di}|} \right\} \quad (1)$$

The minimization of the equation (1) over a set of different combinations of ω_i values results in an efficient solutions space; that is, solutions that are not worse than the remaining ones belonging to the set of possible solutions, for at least one of the objectives.

The proposition of two objective functions is associated with a partial one applied in nodes and another one to the network as a whole. It is assumed as the partial Objective Function (OF) in each node $i = 1 \dots m$ at the system.

$$Z_i = \max_j \left\{ w_{i,j} \left| \frac{Z_{cij} - Z_{dij}}{Z_{dij}} \right| \right\} \quad (2)$$

it accomplished that $0 \leq Z_i \leq 1$,

$$\text{since } \left| \frac{Z_{cij} - Z_{dij}}{Z_{dij}} \right| \leq 1$$

The W_{ij} coefficient can take different values depending on the evaluated objective; but $0 \leq W_{ij} \leq 1$ and also bounded

by $\sum_{j=1}^m W_{ij} = 1$. It also permits taking the solutions search process, from a multi-objective focus to a mono-objective one, when the undesired objectives are cancelled out in the function.

The desired value of each objective follows the each decisor considerations, and it will be used as information directly associated to the overall search, with restrictions for iteration. Equation 2 includes three objectives analysed at node level: Z_1 (voltage in nodes), Z_2 (power factor in the nodes) and Z_3 (total voltage harmonic distortion). Finally for each node i , it has:

$$Z_i = \text{Max}_i \{ \lambda_i Z_{\max i} \} \quad (3)$$

Then, the objective function to minimize for an entire network is formed from (3) and the objectives analysed on a global system level, Z_4 and Z_5 , correspond respectively to the active losses associated with the compensation and the economic indicator examined through VAN.

$$Z_T = \max_i \{ \lambda_1 Z_1, \dots, \lambda_n Z_n, \lambda_{n+1} Z_4, \lambda_{n+2} Z_5 \}$$

restricted by

$$\sum_{i=1}^n \lambda_i = 1 \quad \text{and} \quad 0 \leq \lambda_i \leq 1.$$

The values of the variables within the objective function are normalized, so that, these values have the same order of magnitude to those pre-weights λ that assume its role.

The weight constant for nodes i , correspond to the type of node and the indicators will depend on the importance of each particular case. Besides, a set of constraints can be considered to guarantee a feasible region of existing solutions. These restrictions are given as:

- $Q^{\inf} \leq Q(x) \leq Q^{\sup}$ the capacitive reactive power of the node that must be between the allowed limits
- $\text{Cos} \varphi_{\text{entrada}} \geq 0.9$ power factor of the input node
- $\text{Cos} \varphi_{\text{carga}} \geq 0.7$ power factor of load nodes
- $\text{VAN} \geq \text{VAN}_{\min}$

The mathematical model allows the calculation of all the indicators and necessary intermediate variables from the decision variables and the input data, and it is given by the load flow calculation procedures at fundamental frequency and harmonics.

IV. WORKAROUND OF THE METHOD.

A. VARIABLES INTEGRATION METHOD (VIM).

The principal characteristic of the method is the creation of a population, container of a set of feasible solutions (individuals),

The population's structure is determined by a set of decision variables. In order to process possible solutions, there is a mechanism of encoding, which permits the setting of only one value for each individual of the population, which is representative to its quality as a solution. From this very point, the process, which is iterative, enable the production of new generations of individuals, bettering characteristics each time closer to the solution of the settled problem.

The stop criteria of the algorithm is activated in different ways, by fixing a number of iterations, when a percent of the entire population in evaluation has been calculated or, if there is no significant improvement of the aptitude of the population in a calculation process after a certain number of iterations.

The algorithm of the Variables Integration Method is showed in the Fig 2.

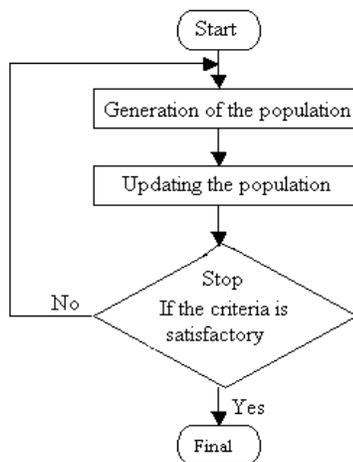


Fig.2. Algorithm of Variables Integration Method

B. RANDOM CONDITIONAL SEARCH ALGORITHM

The proposed general method is called **Random Conditional Search** (RCS) and responds to the particularities of the objective function and the restrictions of the problem.

The RCS method follows the basic ideas of Variables Integration Method as described in Fig. 2, in particular, it brings the built-in characteristics of random search of the extreme of the Variable Code Method [3, 4]. RCS is composed by several algorithms, with handlers functions for search in the initial population, also the improvement of the initial population and the selective reduction of the population be accomplished.

C. SEARCH ALGORITHM OF THE INITIAL POPULATION.

The initial population (IP) consists of K elements, K is an investigator-defined number. For each exploration step, the best solution founded is included in the population, and this is

done until the population's size matches the pre-assigned or it exceeds the initial established one.

There are two obtaining ways for this IP ranging from a set of T possible configurations of the system:

1. Take k different random integers between 1 and T
2. Divide the interval in K subintervals [11] and get a solution of the IP in each sub-interval.

D. INITIAL POPULATION IMPROVEMENT

Afterward the IP has a fixed size, follows its improvements or updating; updating means, the comparison of the obtained solutions in the objective function; select the best among them all, i.e. the one that has a less value of Z; then this new solution is compared with the already calculated worst solution of the population and determining a possible replacement. At the moment which the foreseen accuracy δ is attained, the process of generation of random values of the population is restarted.

For each iteration, two variables codes are generated in an interval [A, B], overcomes a subdivision in three subintervals [A, x_1], [x_1 , x_2] and [x_2 , B], after that; the subinterval containing a bigger value of Z, which is the worst results is eliminated. In this point, two other random values are generated within the obtained subinterval [A, B]. This process is repeated while the length of the interval $x_1 - x_2$ stays superior to a given precision. While, the population size remains inferior to the established size, the solution with the smallest value of Z amongst 2 generated codes is added to the population, in each step. Once the required size of the population is reached, for a new calculated Z value, this solution is compared with the worst solution, and in case of a smaller value of Z it becomes in substitute of the worst solution already found in the actual population.

This algorithm leads to the search of individuals with every time lesser Z's values; in other words, it means convergence, but an obvious deficiency can be detected around Z_{X_1} or Z_{X_2} ; there are possibilities in some internal steps, of smaller Z_{X_h} for X_h , before the ending of the principal algorithm. To remediate this inconvenience, truncating possibilities have been implemented in the general algorithm, it means, when values of Z_{X_1} or Z_{X_2} less than Z_{X_p} are detected, it is assumed that X_h is Z_{X_1} or Z_{X_2} as appropriate. This alternative has demonstrated to be more efficient in time when 5 % of the entire population is evaluated. Nevertheless, there is a preconceived saving action of all the calculated values of

x and Z_x during the execution of the principal algorithm and its ordering over the finalization, taking as a consequence, the smallest value of Z as a solution.

E. VARIABLES CODIFICATION

The candidate variables to be encoded belong to the pre-declared decision variables. The decision variables could take different connections or configurations, defined as the X chromosomes. It can be assumed that an X chromosome is a string character, where each character represents a transformer, a bank of capacitors or a filter amongst others and every

possible value of X depends on the definition of the analysed element.

e.g. Let be, an electric system constituted by:

A transformer T , with two taps 1 and 2

A capacitor's bank C , with two switchable sections numbered 1 and 2

A filter F , with two positions (connected or not).

Then, the number of independent combinations C_p , of the configurations, taken by these decision variables, would be;

$$C_p = 2 \times 3 \times 2 = 12$$

It is necessary to state out, the proposed codification is decimal, i.e., each digit of the code accepts decimal values. Because this code, a special but small algorithm is required to achieve the implemented codification in two functions providing knowledge about X_A .

With these two functions, it could be established that knowing sub-code X warrants and univocal form of the corresponding codification for each electrical system and the reciprocal statement is also true; since the order of every element (transformer, filter and bank of capacitors) of the electrical system is invariant during each calculation.

F. STOP CRITERIA

The stop criteria used during optimization is a mixed condition, in which a number of solution codes are calculated without changes in the composition of the population; and the difference between the Z calculated values in the objective function should be less than a predetermined value (the worst and best solution).

V. VALIDATION OF THE RANDOM CONDITIONAL SEARCH OPTIMIZATION ALGORITHM

Several experiments were realized with 40 samples in each case. The results show off the behaviour of the relation Z/Z_{exa} (calculated value of Z and the smallest value of Z_{exa} found in the exhaustive search), for different percentages of population respect to the total.

Overall cases due to the search process randomness, values of good solutions were always attained from 7 % evaluation of the population, in spite of different initial populations; (Fig. 3). it can be deduced for search values around 10 % of population, that, it is possible achieve solutions for the demand during the searching for efficient solutions and post-improving process, an aggregate reinforce and consistent with other previous results [9]. In all running, the standard of quality used to evaluate the best solutions, corresponds to best solution found in the exhaustive search represented through Z_{exa} .

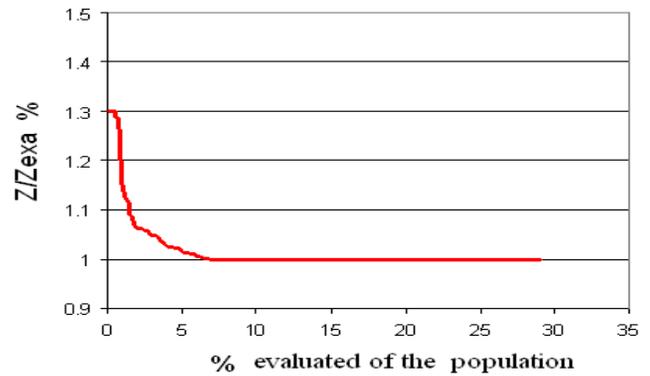


Fig.3. Average relation of the calculated Z and the evaluated percentage of the population

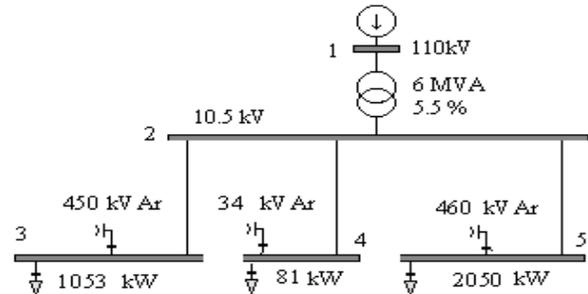


Fig.4. Electric diagram

The Fig. 4 shows the graphical evolution for solution of the proposed algorithm is shown from the electric scheme. Every each point corresponds to a solution.

Table 1 shows the results with three solutions 1, 2 and 3; where the solution1 has a big value, solution 2 has an average value and solution 3 a small value (see Fig.5).

TABLE I
VALUE OF LOSSES AND COS Φ FOR DIFFERENT SOLUTIONS

Calculated (Z) in the Objective Function	ΔE (kWh)	Cos ϕ			
		Node 3	Node 4	Node 5	
Solution 1	0.025	298015	0.90	0.76	0.92
Solution 2	0.019	283824	0.94	0.89	0.92
Solution 3	0.009	276728	0.93	0.92	0.93

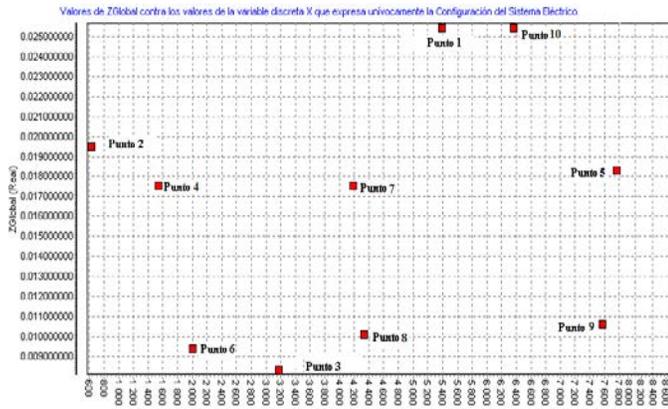


Fig.5. Behaviour of Z value for 10 configurations (IP). The variability of taps in the transformer and capacitors are shown in table II.

TABLE II
TRANSFORMERS TAP POSITION AND QC IN CAPACITORS IN KVAR

Calculated (Z) in the Objective Function	Transformer position and value of QC in kVAr				
	T1	C1	C2	C3	
Point 1	0.025	2	472	17	234
Point 2	0.019	3	472	17	469
Point 3	0.009	6	225	17	234

Table III shows four of the selected solutions as optimal over the 10% of all the possible solutions were calculated. This show off possible to finding optimal solutions (so-called Pareto solutions) from calculating between 7% and 10 % of the total population, if the proposed algorithm is used and also effectiveness warranties without exhaustive search, all of this, with no doubts permits the reduction of calculation time.

TABLE III.
CALCULATED VALUES OF FOUR OF THE SELECTED OPTIMAL SOLUTIONS

Sol.	(Z)	ΔE (kWh)	Cos ϕ in load nodes			Tap s	Qc in capacitors (kVAr)		
			4	5	6		T1	C1	C2
1	0.007	27247	0.9	0.9	0.9	6	472	27	37
2	0.007	27459	0.9	0.9	0.9	6	450	34	37
3	0.007	27034	0.9	0.9	0.9	6	450	34	46
4	0.009	27672	0.9	0.9	0.9	6	225	17	23

VI. CONCLUSIONS

1. The compensation of reactive power and reduction of the distortion of harmonics, can be evaluated as a problem of preparation and decision making under multiple criteria in discreet variables where different technical-economic indicators which are associated to the quality of energy

supplied are conciliated, and whereby including Tchebycheff's weight distance in the objective function allows the evaluation of correspondent system configurations for efficient solutions.

- The character of the formulated task for compensation of reactive power permit the use of procedures for the generation of solutions based in codes evolution, that is, any one of the proper algorithms of Variables Integration Method.
- The application of the Random Conditional Search Method generate a number of correlated solutions to the problem of optimization in a relatively fast and effective form, also it allows the evaluation of multiple connection options of the elements, in a manner that the investigator elects the best-suited one according to the studied situation.
- The use of weights in the objective function is possible, for the best selection of reserved indicators. In very large systems, the values must be determined appealing to experimental tests, in concordance with the importance of the analysed indicator in order to reduce the quantity of weight coefficients in the objective function.
- The validation process shows off that given a number of generated solutions by the RCS method; 10 % of the total possible options over the quality of the obtained solutions could be considered efficient respect of those obtained with Exhaustive Search.

REFERENCES

- B. S. Prajapati and L. Srivastava, "Multi-Objective Reactive Power Optimization Using Artificial Bee Colony Algorithm," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, 2012.
- J. Arzola, *Sistemas de Ingeniería*. La Habana: Editorial "Felix Varela", 2000.
- I. González, A. Legrá, S. Marrero, and J. Arzola, "Optimización de la Compensación de la Potencia Reactiva en Redes de Suministro Eléctrico Industriales con el empleo de un algoritmo de criterios múltiples I.," *Revista Energética*, vol. XXVII, 2006.
- I. González, A. Legrá, S. Marrero, and J. Arzola, "Optimización de la Compensación de la Potencia Reactiva en Redes de Suministro Eléctrico Industriales con el empleo de un algoritmo de criterios múltiples. II.," *Revista Energética*, vol. XXVII, 2006.
- I. González, S. Marrero, and A. Legrá, "Metodología para el control de la calidad en sistemas de suministro eléctricos en redes industriales," FIE, Santiago de Cuba, 2002.
- J. Hernández Galicia and R. Gómez Nieva, "Planificación de la compensación reactiva mediante programación evolutiva," presented at the Reunión de Verano de Potencia 2000, IEEE Sección., Acapulco, México., 2000.
- D. Pires, A. Gomes, and A. Henggeler, "Multi-objective Model for VAR Planning in Radial Distribution Networks Based on Tabu Search," *IEEE Transactions On Power Systems*, vol. 20, May 2005.
- Huang Bao, Zhuang Jian, and Y. De-Hong, "A Novel and Accelerated Genetic Algorithm " *WSEAS Transactions on Systems and Control*, vol. 3, p. 10, April 2008 2008.
- J. Arzola, M. Simeón, and A. Maceo, "The integration of variables method: a generalization of genetic algorithms," in *Intensive Workshop in Optimal Design of Material and Structure*, Paris, 2003.
- I. González and S. Marrero, "Convergence of the Variables Integration Method Applied to Multiobjective Optimization of

Reactive Power Compensation in Supply Industrial Networks",
CINAREM 2013, ISMM, Moa, Cuba.

- [11] I. Pérez A. and J. A. González Q., "VAR Compensation by Sequential Quadratic Programming," *IEEE Transactions on Power Systems*, vol. 18, 2003.

Edge Detection Based Nearest Neighbor Linear Cellular Automata Rules

Nima Aberomand

Abstract—The goal of this paper is to survey linear cellular automata rules to detect edge in both monochromatic and RGB images. Transition matrix representation use as uniform cellular automata to detect edges. Rules are based on linear cellular automata for edge detection by using matrix representation that in some states they are strong and some other rules are not useful for practical edge detection. Result are compared to classical methods like Canny, Sobel, Prewitt and Robert. To evaluate this project, MSE, PSNR, SNR are used in proposed method.

Keywords—edge detection, image processing, linear cellular automata, transition matrix

I. INTRODUCTION

In order to extract the contour of an object in an image detecting edges give complete information about region of image and edge detection is an important part of image processing [1]. These edges can be used as a final result or as intermediate information such as segmentation, object recognition, tracking, face recognition, image retrieval, corner detection, tumor identification, breast cancer and so on. An important property of the edge detection method is its ability to extract the accurate edges with good orientation in the considered image, and a lot of papers at this field had been published in the past two decade.

There are many methods for edge detection, and most of them use the computed gradient magnitude of the pixel value as the measure of edge strength [2]. A different edge detection method i.e. Prewitt, Laplacian, Roberts, Canny [3] and Sobel [4] use different discrete approximation of the derivation function. Surface fitting approach for edge detection is adopted by several authors [5], [6], [7], and [8].

There are several methods for image edge detection that divided into two groups: 1) gradient: this method search for minimum and maximum in the first of image derivative. Methods like Robert, Prewitt and Sobel are in Gradient family. 2) Laplacian: this methods search zero crossing in second derivative of image. This method is independence from any direction.

Some optimization based detectors are presented in [9] and [10]. Use of statically techniques are represented in [11], [12], and [13]. Other approach shows the usage of Genetic algorithm and other meta-heuristic algorithms like PSO, BCO, ACO and others [14] and [15]. Use of neural network in edge

detection specially MLP, SOM, MSOM, BP [16], the Bayesian approach [17], and residual analysis-based techniques [18] and [19]. Some papers have tried to study of an effect of noise in images on the performance of image detectors and noise reduction in the image detection level by thresholding and median filter [20] and [21].

Cellular Automata (CA) were introduced by Ulam and Von Neumann. A cellular automata is a computer algorithm and operates on area of location (e.g. pixels in image processing, peak in signal processing, and in other area is different). CA have been used for many different applications because of property and simple rules generating complex behavior. CA have some advantages like fast in time, parallel computation and due to these advantages used in image processing, too.

II. METHODOLOGY

At first, introducing CA over the filed Z^2 by using the uniform linear local rules is written due to importance of it. The definition of CA is recall, then considering the 2D integer lattice Z^2 and the configuration space $\Omega = \{0,1\}^{Z^2}$ with elements and show in formula (1).

$$\{0,1\} \rightarrow \sigma: Z^2 \quad (1)$$

The value of σ at a point $v \in Z^2$ will be denoted by σ , let $u_1, \dots, u_s \in Z^2$ be a finite set of distinct vectors and $f: \{0,1\}^s \rightarrow \{0,1\}$ be a function.

CA with local rule f is defined as a pair (Ω, T_f) where the global transition map $T_f: \Omega \rightarrow \Omega$ is given by formula (2).

$$(T_f \sigma)_v = f(\sigma_{v+u_1}, \dots, \sigma_{v+u_s}), v \in Z^2 \quad (2)$$

The function f is called local rule. The space Ω is assumed to be equipped with a Tychonoff topology and it is easily seen that the global transmission map T_f introduced (2) and the shift operator U_y are continuous. The 2D finite CA consists of $m \times n$ cells arranged in m rows and n columns, where each cell takes one of the values of 0 or 1. A configuration of the system is an assignment of states to all the cells.

There are nine cells arranged in 3×3 matrix for 2D CA nearest neighbors. Table (1) shows the nearest neighborhood comprises for eight cells with surrounded center X_{ij} . In that case, the rules are extracted from these matrix as CA rules.

Table 1 - nearest neighborhood comprises for eight cells with surrounded center X_{ij}

$X\{i-1,j+1\}$ (c_64)	$X\{i,j+1\}$ (c_128)	$X\{i+1,j+1\}$ (c_256)
$X\{i-1,j\}$ (c_32)	$X\{i,j\}$ (c_1)	$X\{i+1,j\}$ (c_2)
$X\{i-1,j-1\}$ (c_16)	$X\{i,j-1\}$ (c_8)	$X\{i+1,j-1\}$ (c_4)

III. IMPLEMENTATION

In this part, the result of the experimentation is represent. Both RGB and monochromatic images are used for edge detection. Classical methods applied like Canny, Sobel, Prewitt, and Robert operators to selected images. At end to evaluate the proposed technique, best known standard test images are selected and transformed from gray level to the monochromatic structure. This approach for designing rule-changing CA uses two-state CA to deal with monochromatic and RGB images. The proposed method represents a transition rule in a matrix form which can be easily applied directly to the images by multiplication. This matrix representation scheme enables the number of state to be increased which will consider at the future for grey or RGB images as an expanded study. In this paper evaluation the effectiveness of proposed method for image edge detection is done by CA rules.

When the program run, the input image has RGB color mode that will be converted into gray-scale. This program will be run even choosing gray-scale images as input. Fig.1 show this steps.

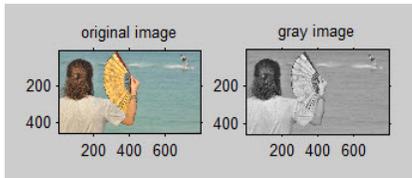


Fig.1

After this step, edge detection will be start with classical methods like Sobel, Canny, Prewitt and Robert as Fig.2.

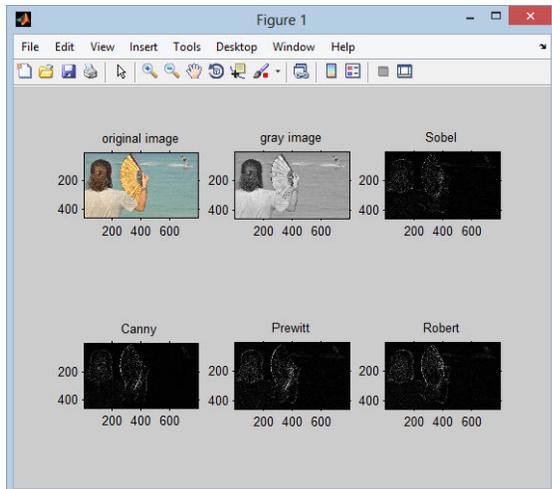


Fig.2

After classical methods take effected, the main method means edge detetion by CA rules will be appeared in 3 level as Fig.3, Fig.4 and Fig.5.

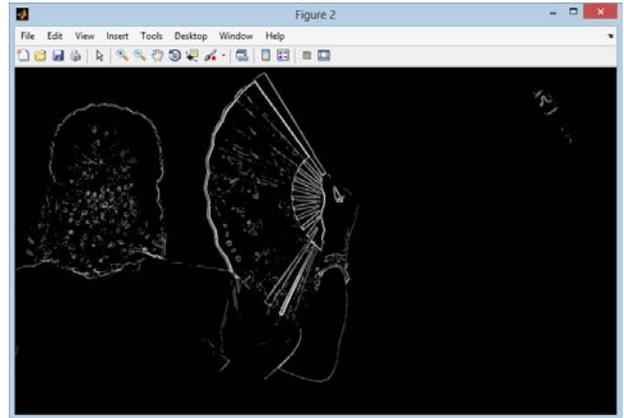


Fig.3

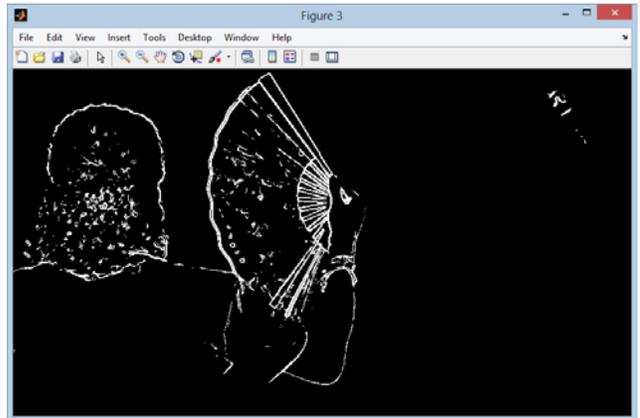


Fig.4

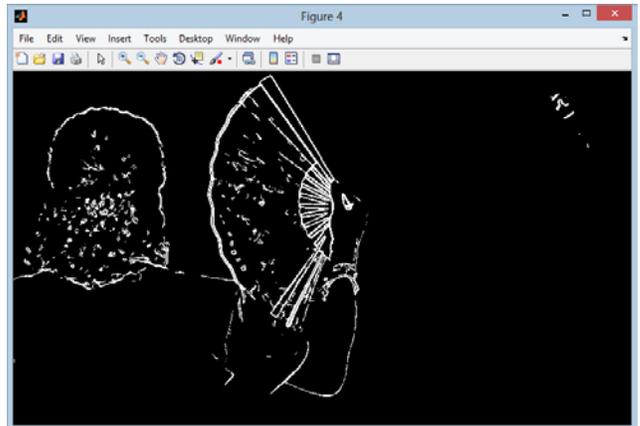


Fig.5

For evaluation classic methods, MSE (Mean Square Error) have been applied as Fig.6.

```

result_sobel_MSE =
    251.2000

result_canny_MSE =
    315.2338

result_prewitt_MSE =
    391.6482

result_robert_MSE =
    669.6406

```

Fig.6

The last output is like Prewitt but better detected edges. The three level of CA rules, evaluated by MSE as Fig.7 for same tested image. It's obvious that in final level, MSE is reduced as 2 level and of course in comparison of classical methods.

```

Command Window

first_level_CA =
    10.6248

second_level_CA =
    8.1462

final_level_CA =
    8.1200

```

Fig.7

The result of PSNR for proposed method is shown at Fig.8 and SNR value is shown at Fig.9.

```

result_PSNR =
    14.9698

```

Fig.8

```

result_SNR =
    33.1610

```

Fig.9

IV. CONCLUSION

Edge detection by transition matrix representation used as uniform cellular automata for both monochromatic and RGB images is proposed in this paper. Rules are based on linear cellular automata for edge detection that in some states they are strong and some other rules are not useful. For practical edge detection. Results are compared to classical methods and evaluation parameters like MSE, PSNR and SNR, showed that this approach have better performance than others.

REFERENCES

- [1] C. C. Kang, W. J. Wang, "A novel edge detection method based on the maximizing objective function", Pattern Recognition, vol. 40, no. 2, pp. 609-618, February 2007.
- [2] A. Bovik, "Handbook of Image and Video Processing. New York, Academic, 2000.
- [3] J. Canny, "A Computational Approach To Edge Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8(6), pp. 679-698, 1986.
- [4] E. Sobel, "Camera Models and Machine Perception". Ph.D thesis. Stanford University, Stanford, California, 1970.
- [5] G. Chen and Y. H. H. Yang, "Edge detection by regularized cubic B-spline fitting", IEEE Trans. Syst., Man, Cybern., vol. 25, pp.636-643, 1995
- [6] V. S. Nalwa and T. O. Binford, "On detecting edges", IEEE Trans. Pattern Anal. Machine Intell. vol. PAMI-8, pp.699-714, 1986.
- [7] S. S. Sinha and B. G. Schunk, "A two stage algorithm for discontinuity-preserving surface reconstruction", IEEE Trans. Pattern Anal. Machine Intel, vol. 14, pp.36-55, 1992.
- [8] J. Canny, "A computational approach to edge detection", IEEE Trans. Pattern Anal. Machine Intel., vol. PAMI-8, pp.679-697, 1986
- [9] S. Y. Sarkar and K. L. Boyer, "Optimal infinite impulse response zero-crossing based edge detectors", Comput. Vis. Graph. Image Process: Image Understanding, vol. 54, no. 9, pp.224-243, 1991.
- [10] J. Shen and S. Castan, "An optimal linear operator for step edge detection", Graph. Models Image Process, vol. 54, no. 1, pp.112-133, 1992.
- [11] P. deSouza, "Edge detection using sliding statistical tests", Comput. Vis., Graph. Image Process., vol. 23, no. 1, pp.1-14, 1983.
- [12] E. Chuang and D. Sher, "Chi-square test for feature extraction", Pattern Recognit., vol. 26, no. 11, pp.1673-1683, 1993.
- [13] P. Qie and S. M. Bhandarkar, "An edge detection technique using local smoothing and statistical hypothesis testing", Pattern Recognit. Lett., vol. 17, no. 8, pp.849-872, 1996.
- [14] S. M. Bhandarkar, Y. Zhang, and W. D. Potter, "An edge detection technique using genetic algorithm based optimization", Pattern Recognit, vol. 27, no. 9, pp.1159-1180, 1994.
- [15] L. Caponetti, N. Abbattisti, and G. Carapella, "A genetic approach to edge detection", Int. Conf. Image Processing, vol. 94, pp.318-322, 1994.
- [16] V. Srinivasan, P. Bhatia, and S. H. Ong, "Edge detection using neural network", Pattern Recognit., vol. 27, no. 12, pp.1653-1662, 1995.
- [17] M. H. Chen, D. Lee, and T. Pavlidis, "Residual analysis for feature detection", IEEE Trans. Pattern Anal. Machine Intell., vol. 13, pp.30-40, 1991.
- [18] T. J. Hebert and D. Malagre, "Edge detection using a priori model", Int. Conf. Image Processing, vol. 94, pp.303-307, 1994.
- [19] C. Spinu, C. Garbay, and J. M. Chassery, "Edge detection by estimation and minimization of errors", Proc. Int. Conf. Image Processing, vol. 94, pp.303-307, 1997.
- [20] F. L. Valverde, N. Guil, J. Munoz, R. Nishikawa, and K. Doi, "An evaluation criterion for edge detection techniques in noisy images", Int. Conf. Image Processing, pp.766-769, 2001.
- [21] R.R. Rakesh, P. Chaudhuri, C.A. Murthy, "Thresholding in edge detection: a statistical approach", Image Processing, IEEE Transactions on , vol.13, no.7, pp.927-936, 2004 doi: 10.1109/TIP.2004.828404.

Automatically Diagnosis of Suspicious Lesions in Mammograms

A. Elmoufidi Member IEEE, K. El Fahssi, S. Jai-Andaloussi, A. Sekkaki, G. Quellec, M. Lamard, G. Cazuguel.

Abstract—Computer-aided detection/diagnosis (CAD) is usually used as a second opinion by the radiologists and the mammography represents the most effective tool for the early detection of breast cancer. The main objective of this study is to introduce a new approach to extract and select the features of suspicious lesions in mammograms and classifying them, in order to build a Computer-aided diagnosis (CADx) model to discriminate between malignant and benign parenchyma. Our method contains two phases: The first one is the generation and selection of features of Regions of Interest (ROIs). The second one is the diagnosis the ROIs detected. Our method has been verified with the well-known Mammographic Image Analysis Society (MIAS) database and we have used the Receiver Operating Characteristics (ROC) to measure the performance of our method. The experimental results show that our method achieved an overall classification accuracy of 94.29%, with 94.11% sensitivity and 94.44% specificity.

Index Terms: Mammography, computer aided diagnosis, feature extraction, Support Vector Machine, Image classification, ROC analysis.

I. INTRODUCTION

Breast cancer is one of the frequent types of cancer contributing to the increase in mortality among women worldwide. Recent statistics have shown that one in ten women in Europe and one in eight in the United States develop breast cancer during their lifetime [1],[2]. Early detection and diagnosis of breast cancer is the most important factors affecting the possibility of recovery from the disease. For that, the mammography represents the best and most accurate tool in detecting breast cancer [3],[4],[5]. In order to improve the accuracy of interpreting mammograms, a variety of CAD systems that perform computerized mammogram analysis have been proposed. These systems are usually employed as a supplement to the radiologists' assessment. Thus, their role in modern medical practice is considered to be significant and important in the early detection and diagnosis of breast cancer. Generally, the procedure to develop a CAD system for the detection and the diagnosis of suspicious regions in mammograms takes place in two phases: The first one is a Computer-aided detection (CADE) contains two steps: 1)

preprocessing step, 2) Image Analysis. And the second one is a Computer-aided diagnosis (CADx) also contains two steps: 1) Extraction and selection of features of ROIs, 2) the Classification of ROIs detected in the first phase [6],[7].

1) Pre-Processing: the purpose of this stage is to prepare the image for the next stage of operations; 2) Image Analysis: the purpose of this stage is to analyze the image and extract the necessary information; 3) Features Extraction and selection of ROIs: In this stage, we can find, match, and identify specific patterns, shapes, density and texture; 4) Classification of ROIs: The purpose of this stage is to classify the mammogram to normal, malign or benign class [7].

In this paper, we have proposed fully automatic and robust CADx for diagnosis of suspicious lesions in a mammogram. We have started by detecting and extracted the features of ROIs, and we have finished by classified the ROIs extracted to malignant or benign parenchyma, so the classification of mammograms to malignant or benign mammogram. The proposed algorithm is a very accurate technique for diagnosing breast cancer by using mammography images. The obtained quantitative and qualitative results demonstrate the efficiency of this method and confirm the possibility of using it in improving the CADx system.

Paper organization : The setup of the paper is organized as follows: An introduction is given in section I; Section II discusses related work; Section III presents materials and method; Section IV describes our proposed research; The results and performance are presented in section V; Section VI includes a conclusion; References are given at the end.

II. RELATED WORK

Many methods have been proposed for the diagnosis of abnormalities in mammography images. i.e, K. Ganesan, et al. [8] provided an overview about recent developments and advances in the field of Computer-Aided Diagnosis (CAD) of breast cancer using mammograms. M. Veta, et al [20] Published a review titled "Breast cancer histopathology image analysis" introduce the steps of image analyses. A.Jalalian, et al. [6] presented the approaches which are applied to develop CAD systems on mammography and ultrasound images. The diagnosis of regions of interest (ROIs) is a capital step in a development CAD system. Hence, a number of methods have been used to feature extraction and classification. For example, Nasseer et al. [9] developed an algorithm for Classification of Breast Masses in Digital Mammograms using Support Vector Machines. Cascio D. et al.[10] Used an approach for Mammogram Segmentation by Contour Searching and

A. Elmoufidi, K. El Fahssi, S. Jai-Andaloussi, A. Sekkaki are with Department of Mathematics and Computer Sciences, Faculty of sciences, Hassan II University, Casablanca, Morocco. and G. Quellec, M. Lamard, G. Cazuguel are with Inserm, UMR 1101, Brest, F-29200 France. M. Lamard is with Univ Bretagne Occidentale, Brest, F-29200 France. G. Cazuguel is with Institut Mines-Telecom; Telecom Bretagne; UEB; Dpt ITI, Brest, F-29200 France. (E-mail: Abdelali.Elmoufidi09@univcasa.ma; Elfahsi@etude.univcasa.ma; Said.jai-andaloussi@univcasa.ma; Abderahim.sekkaki@univcasa.ma; gwenole.quellec@insrm.fr; mathieu.lamard@univbrest.fr; guy.cazuguel@telecombretagne.eu

Massive Lesion Classification with Neural Network. Jacob Levman et al. [18] proposed a method titled "Classification of Dynamic Contrast-Enhanced Magnetic Resonance Breast Lesions by Support Vector Machines (SVM)" for classified the breast lesions using SVM. Stylianos.D et al [21] proposed a fully automated scheme for mammographies segmentation and classification based on breast density and asymmetry. The CAD system proved to be powerful tools that could assist medical staff in hospitals and lead to better results in diagnosing a patient.

III. MATERIALS AND METHOD

To develop and evaluate our proposed method we have used the Mammographic Image Analysis Society (MIAS) database [12], and Support Vector Machine (SVM) for classifying the suspicious regions to benign or malignant parenchyma.

A. Database

In this work, to develop and evaluate the proposed method we have used the Mammographic Image Analysis Society (MIAS) database [12]. The mammograms have a size of 1024×1024 pixels in Portable Greymap (PGM) format, and resolution of 200 micron. Each pixel in the images is represented as an 8-bit word with a pixel intensity of range [0, 255], where the images are in grayscale format. This database is composed of 322 mammography images of right and left breast, from 161 patients, where 54 mammograms were diagnosed as malignant, 69 benign and 207 normal. The details about MIAS were discussed in [1].

B. Support Vector Machine (SVM)

Support vector machine (SVM) classification algorithm, developed from the machine learning community is a discriminative classifier formally defined by a separating hyperplane. The hyperplane is determined in such a way that the distance from this hyperplane to the nearest data points on each side, called support vectors, is maximal [11]. SVM was used to diagnose breast cancer. For example, an approach with wavelet SVM was discussed in [13]. The details about SVM and its application to breast cancer diagnosis were discussed in [14],[15], which uses similar kernel.

IV. FEATURES GENERATION AND EXTRACTION

Below a list of eighteen features selected for using as input parameters of SVM for training and testing our CADx system.

1) Mean Value:

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N I(i, j) \quad (1)$$

Where: $I(i, j)$ is the pixel value at point (i, j) in a ROI of size $M \times N$.

2) Standard Deviation:

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - \mu)^2} \quad (2)$$

3) Entropy:

$$H = - \sum_{k=1}^{L-1} P_k * \log_2(P_k) \quad (3)$$

Where: P_k is the probability of the k^{th} grey level, L is the total number of grey levels.

4) Skewness:

$$S = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left[\frac{I(i, j) - \mu}{\sigma} \right]^3 \quad (4)$$

Where: μ is the mean and σ is the standard deviation.

5) Kurtosis:

$$K = \left\{ \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left[\frac{I(i, j) - \mu}{\sigma} \right]^4 \right\} - 3 \quad (5)$$

6) **Uniformity (U)**: The Uniformity (U) is a texture measure based on histogram:

$$U = \sum_{k=0}^{L-1} P_k^2 \quad (6)$$

Where: P_k is the probability of the k^{th} grey level. Because the k^{th} have values in the range [0,1] and their sum equals 1.

7) Sum Entropy (SE):

$$SE = - \sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\}. \quad (7)$$

8) Sum Average (SA):

$$SA = \sum_{i=2}^{2N_g} i p_{x+y}(i) \quad (8)$$

9) Difference variance (DV):

$$DV = \sum_{i=2}^{2N_g} (i - SE)^2 p_{x-y}(i) \quad (9)$$

10) **Difference entropy**: The Difference Entropy (DE) is an entropy measure which provides a measure of no uniformity while taking into consideration a different measure obtained from the original image:

$$DE = - \sum_{i=2}^{2N_g} p_{x-y}(i) \log\{p_{x-y}(i)\}. \quad (10)$$

11) Inverse Difference Moments (IDM) :

$$IDM = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j). \quad (11)$$

12) **Area (A)**: The area (A) is calculated as the sum of the number of all pixels (x) of the segmented ROI.

$$A = \sum_{x \in ROI} 1. \quad (12)$$

13) **Perimeter:** The perimeter (P) is the length of a polygonal approximation of the boundary (B) of ROI:

$$P = \sum_{x \in B} 1. \quad (13)$$

14) **Convexity:**

$$C(S) = \frac{A}{Area(CH(S))}. \quad (14)$$

Where: S is a ROI, CH(S) is its convex hull and A is the ROI's area.

15) **Compactness:**

$$C = \frac{P^2}{4\pi A}. \quad (15)$$

Where : P is the ROI's perimeter, A is the area of ROI.

16) **Aspect Ratio (AR):** The AR corresponds to the aspect ratio of the smallest window fully enclosing the ROI in both directions (see Fig.1.):

$$AR = \frac{D_y}{D_x}. \quad (16)$$

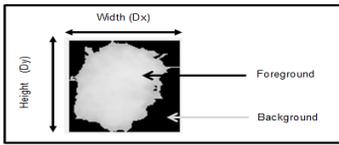


Fig. 1. An example of ROI's window from which features will be extracted.

Where: D_x , D_y are the height and width of the previously mentioned window (see Fig.1.).

17) **Area to background percentage:** The Area Ratio (R_Area) is specified by dividing the area of the segmented ROI in pixels by the area of the same window given in Fig.1:

$$R_Area = \frac{Area_ROI(in\ pixels)}{Area_window(in\ pixels)}. \quad (17)$$

Where: $Area_window = D_x * D_y$, D_x is the width's ROI and D_y is the height's ROI.

18) **Perimeter Ratio:** The Perimeter Ratio (R_Perim) presents the ratio between the perimeter of the segmented ROI to the perimeter of the same rectangular window of fig.1:

$$R_Perim = \frac{Perimeter_ROI(in\ pixels)}{Perimeter_window(in\ pixels)}. \quad (18)$$

V. OUR PROPOSED RESEARCH

In this paper, we have implemented a new method for automatically diagnosis of suspicious lesions in mammograms. Our proposed method is divided into two major blocks, namely: (1) Extraction and selection of technical features for each region of interest, and (2) classification of ROIs extracted to benign or malignant parenchyma.

One among the novelties of our algorithm, that in the case of detection of multiple regions of interest, we are going to separate the ROIs detected one by one and extracted the features of each one separately, and then the diagnosing.

In the end, if all ROIs belong in the same mammogram are benign, then the mammogram is benign. Otherwise, the mammogram is malignant. In addition, our algorithm is capable to diagnosing the different objects in the mammogram: the masses, the calcifications and the micro-calcifications. The obtained quantitative and qualitative results demonstrate the efficiency of this method and confirm the possibility of its use in improving the computer-aided diagnosis (CADx).

A. Diagnosis of Regions of Interest (ROIs)

After the detection of regions of interest and selected their features, the next step is to classify them to benign or malignant mass. In the literature the different classifiers such as: Support Vector Machine(SVM), Artificial Neural Network (ANN), Nearest Mean Classifier(NMC), Linear Discriminant Classifier (LDC) etc. have been used. But, the Support Vector Machine (SVM) has given the best results for diagnosis the regions of interest detected [16], the details about (SVM) are discussed above. For that, we have used the SVM for diagnosis the ROIs detected in the detection phase of our system.

1) **Experimental results:** The next three figures display the details of the discussed method for two cases. In the first case, the suspicious lesion was correctly located without any false positive and exact diagnosis (see fig.2 and fig.3). In the second case, we have detected two suspicious regions, which one correct and the second one is a false positive warning, but in the end, the mammogram was correctly diagnosed (fig.4).

a) **Example 1:** A lesion malignant correctly detected/diagnosis without false positive

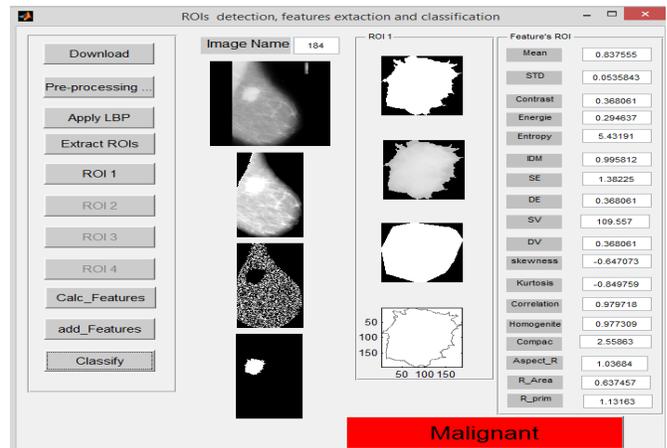


Fig. 2. Example1: Mammography image mdb184.pgm : Exactly suspicious lesion detected, and correctly diagnosis.

b) **Example 2:** A lesion benign correctly detected/diagnosis without false positive

c) **Example 3:** A lesion malignant detected with another one as false positive, but correctly diagnosis

VI. RESULTS AND PERFORMANCE

The global diagnostic methods were tested on 89 images from the online available MIAS database. The detail about MIAS database is given above. Each segmentation

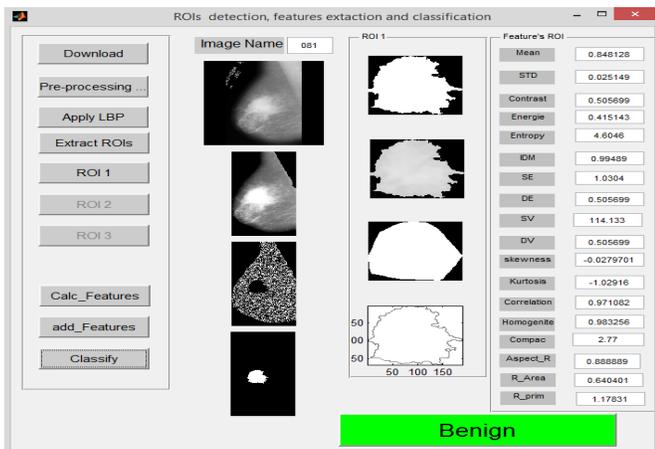


Fig. 3. Example2: Mammography image mdb081.pgm : Exactly suspicious lesion detected, and correctly classified.

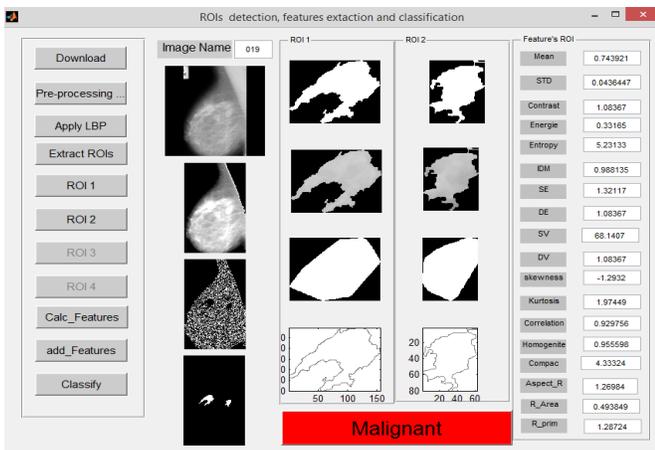


Fig. 4. Example3: Mammography image mdb019.pgm :Suspicious lesion detected with another ROI as false positive, and correctly diagnosis.

and classification result needs evaluation of its performance. Performance evaluation for researches on classification of breast density involves comparison of research result with a density class that has been given by radiologist.

A. Performance evaluation

In the classification of ROIs to benign or malignant mass, a positive case means correct classification of ROIs to benign or malignant while a negative case means incorrect classification of ROIs as such a type. In this work, we have used 89 mammograms, representing the number of mammography images for evaluating our algorithm, these mammograms are divided into 29 mammography images (15 benign and 14 malignant) for training and 70 mammography images (36 benign and 34 malignant) for testing our proposed algorithm.

Fig.5 shows the ROC curve of the proposed diagnosis method. ROC analysis is based on statistical decision. The diagnosis stage achieved an overall classification accuracy of 94.29%, with 94.11% sensitivity and 94.44% specificity.

TABLE I
PERFORMANCE EVALUATION OF RESULT ANALYSIS.

	Malignant	Benign
Malignant	32	2
Benign	2	34

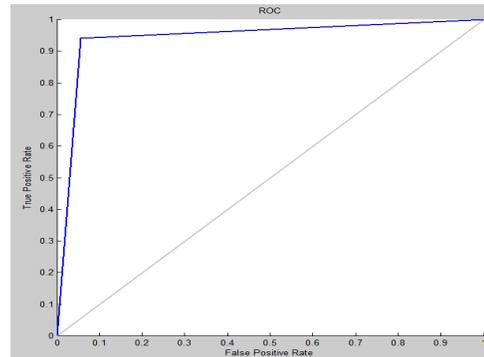


Fig. 5. Plot illustrating ROC curve.

B. The Comparison of our proposed algorithm with existing papers.

TABLE II
THE COMPARISON OF THE PERFORMANCE'S OUR METHOD WITH PAPERS PUBLISHED RECENTLY.

Authors	Method used	Accuracy
Veena et al.[17]	CAD Based System for Automatic Detection & Classification of Suspicious Lesions in Mammograms	92.13%
Nasseer et al.[9]	Classification of Breast Masses in Digital Mammograms Using Support Vector Machines	93.069%
K. Ganesan et al.[11]	One-Class Classification of Mammograms Using Trace Transform Functionals	92.48 %
Our method	Automatic Detection and diagnosis , of Suspicious Lesions in Mammograms	94.29%

VII. CONCLUSION

In this paper, an algorithm for breast mass diagnosis has been implemented under the MATLAB environment for automatic diagnosis of suspicious regions in mammogram by using SVM as a classifier. First, we have extracted and selected the technical features of regions of interest (ROIs). Secondly, we have classified the ROIs to benign or malignant mass. The performance of our algorithm has been evaluated by using Receiver Operating Characteristics (ROC), and the experimental results show that our method achieved an overall classification accuracy of 94.29%, with 94.11% sensitivity and 94.44% specificity. The obtained results demonstrate the efficiency of this method and comparable to other methods. Our proposed algorithm can contribute to solving the main problem in mammography image processing such as the detection and the diagnosis of masses, of calcifications also of micro-calcifications. The efficiency of the proposed method confirms the possibility of its use in improving the CADx system.

REFERENCES

- [1] A. ELMOUFIDI Member IEEE et al., "Automatically Density Based Breast Segmentation for Mammograms by using Dynamic K-means Algorithm and Seed Based Region Growing," I2MTC 2015 - International Instrumentation and Measurement Technology Conference , PISA, ITALY, MAY 11-14, **2015**.
- [2] K. Hu et al., "Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms," IEEE Trans on Instrumentation and Measurement, vol. 60, no. 2, pp. 462-472, **2010**.
- [3] A. Ferrero Fellow IEEE et al., "Uncertainty evaluation in a fuzzy classifier for microcalcifications in digital mammography," I2MTC 2010 - International Instrumentation and Measurement Technology Conference Austin, TX, 3-6 May **2010**.
- [4] Abdelali Elmoufidi et al., "Detection of Regions of Interest in Mammograms by Using Local Binary Pattern, Dynamic K-Means Algorithm and Gray Level Co-occurrence Matrix," 2014 Fifth International Conference on Next Generation Networks and Services (NGNS'14) 28-30 May **2014**, Casablanca, Morocco.
- [5] A. Elmoufidi et al., "Detection of Regions of Interest in Mammograms by Using Local Binary Pattern and Dynamic K-Means Algorithm," International Journal of Image and Video Processing: Theory and Application Vol. 1, No. 1, 30 April **2014** ISSN: 2336-0992.
- [6] A. Jalalian et al., "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound," Clinical Imaging, 37 **2013** 420-426.
- [7] S. Shirmohammadi and A. Ferrero, "Camera as the Instrument: The Rising Trend of Vision Based Measurement," IEEE Instrumentation and Measurement Magazine, Vol. 17, No. 3, June **2014**, pp. 41-47. DOI: 10.1109/MIM.2014.6825388.
- [8] Karthikeyan Ganesan et al., "Computer-Aided Breast Cancer Detection Using Mammograms," IEEE Reviews in biomedical engineering, vol. 6, **2013**.
- [9] Nasseer M. Basheer et al., "Classification of Breast Masses in Digital Mammograms Using Support Vector Machines," International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277 128X, Volume 3, Issue 10, October **2013**.
- [10] Cascio D. et al., "Mammogram Segmentation by Contour Searching and Massive Lesion Classification with Neural Network," Institute of Electrical and Electronic Engineering (IEEE), **2006**.
- [11] Karthikeyan Ganesan et al., "One-Class Classification of Mammograms Using Trace Transform Functionals", IEEE Transactions on Instrumentation and Measurement, Vol. 63, No. 2, February **2014**.
- [12] J. Suckling et al., "The Mammographic Image Analysis Society digital mammogram database," Excerpta Medica, International Congress Series 1069 pp. 375-378., **1994**.
- [13] M. Shen et al., "A prediction approach for multichannel EEG signals modeling using local wavelet SVM", IEEE Trans. Instrum. Meas., vol. 59, no. 5, pp. 1485-1492, May **2010**.
- [14] H. X. Liu, et al., "Diagnosing Breast Cancer Based on Support Vector Machines", J. Chem. Inf. Comput. Sci. **2003**, 43, 900-907.
- [15] L. Wei, et al., "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," IEEE Trans. Med. Imag., vol. 24, no. 3, pp. 371-380, Mar. **2005**.
- [16] M. Arfan Jaffar et al. "DCT Features Based Malignancy and Abnormality Type Detection Method for Mammograms," International Journal of Innovative Computing, Information and Control ICIC, Volume 7, Number 9, September **2011** ISSN 1349-4198.
- [17] Veena, et al., "CAD Based System for Automatic Detection et Classification of Suspicious Lesions in Mammograms," International Journal of Emerging Trends et Technology in Computer Science (IJETTCS) ISSN 2278-6856 , Volume 3, Issue 4 July-August **2014**.
- [18] Jacob Levman, "Classification of Dynamic Contrast-Enhanced Magnetic Resonance Breast Lesions by Support Vector Machines", IEEE Transactions On Medical Imaging, Vol. 27, No. 5, May **2008**.
- [19] S. Timp et al., "Computer-aided diagnosis with temporal analysis to improve radiologists' interpretation of mammographic mass lesions," IEEE Trans. Inform. Technol. Biomedicine, vol. 14, no. 3, pp. 803-808, May **2010**.
- [20] M. Veta, et al. "Breast cancer histopathology image analysis: a review," IEEE transactions on bio-medical engineering, vol. 61, no. 5, pp. 1400-11, May **2014**.
- [21] Stylianos.D et al., "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry," computer methods and programs in biomedicine **2011**, 47-63.

Exploitation of Chaotic and Synchronization Properties of Logistic Maps for Application in Wireless Communication

Bijoy Kamal Bhattacharyya, Hemanta Kumar Sarmah, Kandarpa Kumar Sarma and Nikos Mastorakis

Abstract—Logistic maps have been preferred for generating chaotic sequences to study dynamical systems. In wireless communication, the widest range of chaotic behavior is observed in the channel. This is due to the fact that the propagation conditions show fluctuations in magnitude and phase with change in time which is known as fading. In this paper, we use two important properties of logistic maps to minimize shortcomings of the rapid fluctuations observed in the wireless channel. These are the chaotic behaviour and the synchronization property of the logistic map. Chaotic sequences of varied lengths are generated using logistic map for use as spreading factor of a spread spectrum modulation system. Next, the synchronization property of the logistic map is used to provide proper recovery of a sequence of bits in the identified modulation system. Experimental results show that the proposed method is effective in wireless channels and can provide proper quality of service during recovery of signals.

Keywords—*logistic, channel, chaotic, synchronization, spread, spectrum.*

I. INTRODUCTION

Fading is a significant aspect in wireless communications and occurs due to multiple reflections, refraction, scattering and diffraction of the transmitted signal while propagating through the medium [1]. Fading generates rapid variations in the magnitude and phase of the original signal. Also, channel and background noise corrupts the signal considerably. To ensure proper quality of service (QoS) of any system while executing a communication process through such a medium needs proper modeling, understanding and design of appropriate mechanisms for such a situation [2].

Spread spectrum modulation (SSM) is an important class of techniques used extensively in wireless systems for obtaining proper QoS. The primary advantage of SSM is the fact that it

uses a noise like rapidly fluctuating signal to spread the bandwidth of the signal much more than the normal requirement during transmission and just the reverse process during recovery. As a result, signal components are spread throughout the channel and specific portions of the signal cannot be corrupted by propagation related variations and noise. There are two major techniques to spread the spectrum; one is the frequency hopping (FH) technique, which makes the narrow band signal jump in random narrow bands within a larger bandwidth and the other one is the direct sequence (DS) technique which introduces rapid phase transition to the data to make it larger in bandwidth. The spreading factor is obtained by using a pseudo-noise (PN) generator. In some cases Gold codes are also used. One of the critical disadvantages of PN and Gold code generators are their limited sequence lengths. It is directly dependent on the physical size of the generator. But practical communication conditions may demand varying or increasing sequence lengths. It has been observed that increasing sequence lengths in wireless communication serves as an effective aid to fight propagation related variations and contributes towards better QoS [3].

The chaotic behavior observed in specific operations region of the logistic map has been effectively used for such a purpose [3] [4]. The logistic map provides non-linear recursion generating iterative values which exhibit deterministic chaos [5]. Another property of the logistic map namely synchronization can be effectively used in wireless communication to provide better QoS while used as part of SSM systems. In this paper, we use the chaotic behaviour and the synchronization property of the logistic map as an aid to SSM in a wireless setup. Chaotic sequences of varied lengths are generated using logistic map for use as spreading factor of a spread spectrum modulation system. Next, the synchronization property of the logistic map is used to provide proper recovery of a sequence of bits in the identified modulation system. Experimental results show that the proposed method is effective in wireless channels and can provide proper quality of service during recovery of signals.

The rest of the paper is organized as follows. Section 2 gives a description of the theoretical considerations involved, Section 3 shows the experimental details and results obtained and Section 4 concludes.

B. K. Bhattacharyya is with the Department of Mathematics, L C B College, Guwahati-781011, Assam, India (e-mail:).

H. K. Sarmah is with the Department of Mathematics, Gauhati University, Guwahati, 781014, Assam, India (e-mail:).

K. K. Sarma is with the Electronics and Communication Technology Department, Gauhati University, Guwahati, 781014, Assam, India (e-mail: kandarpaks@gmail.com).

N. Mastorakis is with the Military Institutions of University Education, Hellenic Naval Academy, Terma Hatzikyriakou, 18539, Piraeus, Greece (e-mail: mastor@ieee.org).

Table 1. Behaviour dependent on r

Range of r	Behaviour of population
Between 0 and 1	Independent of the initial population
Between 1 and 2	Independent of the initial population
Between 2 and 3	Fluctuate around the value $r - 1/r$ for some time
Greater than 3	Dependent of the initial population

II. LOGISTIC MAP AS CHAOS GENERATOR

A logistic map is a polynomial mapping having a degree of 2. It gives the idea of how a very complex, chaotic behaviour can occur from very simple non-linear dynamical equations. Prediction thus becomes impossible, and then the system behaves randomly [5], [6]. Mathematically, the logistic map is written as

$$x_{n+1} = r * x_n(1 - x_n) \tag{1}$$

where x_n denotes a number between zero and one, population (at year 0). The logistic map behaviour is totally dependent on r , which is clearly seen in the Table 1.

An example of 1-dimensional map is $x_{n+1} = \cos(x_n)$ where, the sequence x_0, x_0, x_0, \dots is called the orbit starting from x_0 .

Maps arise in various ways:

- As tools for analysing differential equations (e.g. Poincare and Lorenz)
- As models of natural phenomena (in economics and finance)
- As simple examples of chaos

Maps are capable of much wilder behaviour than differential equations because the points x_n hop discontinuously along their orbits rather than flow continuously.

For fixed point, an equation $x_{n+1} = f(x_n)$ is considered, where $f()$ is a smooth function from the real line onto itself.

Suppose x^* satisfies $f(x^*) = x^* \Rightarrow x^*$ is a fixed point of the map. Its stability is determined by considering a nearby orbit $x_n = x^* + \eta_n$. Thus

$$x^* + \eta_{n+1} = x_{n+1} = f(x^*)f'(x^*)\eta_n + O(\eta_n^2) \tag{2}$$

Since $f(x^*) = x^*$,

$\eta_{n+1} = f'(x^*)\eta_n$ is the linearized map and $\lambda = f'(x^*)$ is the eigen value or multiplier.

- If $|\lambda| = |f'(x^*)| < 1$ then $\eta_n \rightarrow 0$ as $n \rightarrow \infty \Rightarrow x^*$ is linearly stable.
- If $|\lambda| = |f'(x^*)| > 1$ then x^* becomes unstable.
- If $|\lambda| = |f'(x^*)| = 1$ then the terms $O(\eta_n^2)$ have to be considered.

Similarly, if we consider $x_{n+1} = x_n^2$, now for fixed points at $x^* = (x^*)^2 \Rightarrow x^* = 0, 1$.

$\lambda = f'(x^*) = 2(x^*) \Rightarrow x^* = 0$ is stable and $x^* = 1$ is unstable. Cobwebs allow us to see global behavior at a glance.

Considering the equation of logistic map, let $0 \leq r \leq 4, 0 \leq x \leq 1 \Rightarrow$ map is a parabola with maximum value of $r/4$ at $x = 0.5$.

- For $r < 1, x_n \rightarrow 0$ as $n \rightarrow \infty$
- For $1 < r < 3, x_n$ grows as n increases, reaching a non-zero steady state.
- For larger r (e.g. $r = 3:3$) x_n eventually oscillates about the former steady state \Rightarrow period 2 cycle.
- At still larger r (e.g. $r = 3:5$), x_n approaches a cycle which repeats every 4 generations

\Rightarrow period 4 cycle.

- Further period doublings to cycles of period 8, 16, 32 etc occur when r increases. Hence, $r = 3$ (period 2 is generated); $r = 3.449$ (period 4 is created); $r = 3.54409$ (period 8 is obtained), $r = 3.5644$ (period 16 is found); ...; $r_c = 3.569946$ (period r is born);
- Successive bifurcations become faster and faster as r increases.
- The r_n converge to a limiting value r_c .
- For large n , the distance between successive transitions shrinks by a constant factor

$$\delta = \lim_{n \rightarrow \infty} \frac{r_n - r_{n-1}}{r_{n+1} - r_n} = 4.669. \tag{3}$$

When $r > r_c$, for many values of r , the sequence $\{x_n\}$ never settles down to a fixed point or a periodic orbit, the long term behavior is aperiodic. It is expected that the system would become more and more chaotic as r increases, but in fact the dynamics are more subtle.

At $r = 3.4$ the attractor is a period 2 cycle. As r increases, both branches split, giving a period 4 cycle i.e. a period-doubling bifurcation has occurred. A cascade of further period-doublings occurs as r increases, until at $r = r_c \cong 3.57$, the map becomes chaotic and the attractor changes from a finite to an infinite set of points. When $r > r_c$, the orbit reveals a mixture of order and chaos, with periodic windows interspersed with chaotic clouds of dots. When, $r \cong 3.83$, then there is a stable period 3 cycle.

Now, considering the logistic map equation $x_{n+1} = r * x_n(1 - x_n); 0 \leq r \leq 4, 0 \leq x_n \leq 1$

For fixed point, $x^* = f(x^*) = rx^*(1 - x^*)$

$\Rightarrow x^* = 0$ or $1 - 1/r$, where

$x^* = 0$ is a fixed for all r , and $x^* = 1 - 1/r$ only if $r \geq 1$; (since $0 \leq x_n \leq 1$)

Stability depends on $f'(x^*) = r - 2rx^*$

$x^* = 0$ is stable for $r < 1$ and unstable for $r > 1$

$x^* = 1 - 1/r$ is stable for $-1 < (2 - r) < 1$, i.e. $1 < r < 3$ and unstable for $r > 3$;

Therefore, at $r = 1, x^*$ bifurcates from the origin in a transcritical bifurcation. As r increases beyond 1, the slope at x^* gets increasingly steep. The critical slope $f'(x^*) = -1$ is attained when $r = 3$, the resulting bifurcation is called a flip bifurcation that means two cycle.

Now, logistic map has a 2-cycle for all $r > 3$, which is clear from the following description :

A 2 cycle exists if and only if there are two points p and q such that $f(p) = q$ and $f(q) = p$. Equivalently, such a p must satisfy $f(f(p)) = p$ where $f(x) = rx(1 - x)$. Hence, p is a fixed point of the second iterate map $f^2(x) = f(f(x))$. Since $f(x)$ is a quadratic map and $f^2(x)$ is a quadratic polynomial.

Now, to solve $f^2(x) = x, x^* = 0$ and $x^* = 1 - 1/r$ are trivial solutions. The other two solutions are

$$p, q = \frac{r+1 \pm \sqrt{(r-3)(r+1)}}{2r} \tag{4}$$

Which are real for $r > 3$. Hence a 2 cycle exists for all $r > 3$ for logistic map [4].

III. SYNCHRONIZATION IN LOGISTIC MAPS

In logistic maps, synchronization is related to coupled maps. Chaotic behavior of logistic maps is considered as random oscillation generations. In a coupled system, if amplitude, frequency and phase of two logistic maps share compatible attributes, these are called to be synchronized maps [5] [6]. Two logistic maps can be shown to be synchronized as below:

$$y_{n+1} = x_n \quad (5)$$

$$x_{n+1} = y_n \quad (6)$$

Replacing the above with respective logistic map gives

$$y_n = rx_n(1 - x_n) \quad (7)$$

$$x_n = rq(1 - q) \quad (8)$$

where

$$q = ax_n + (1 - a)y_n \quad (9).$$

Taking $r = 4$ for x and y gives three solutions which further yields $a = \frac{1}{2}$ and $a = \frac{5}{4}$ with the later discarded for realistic cases [5] [6]. For a symmetrically coupled case $= \frac{1}{4}$.

IV. PROPOSED METHOD OF USING VARYING LENGTH CHAOTIC SEQUENCES AND SYNCHRONIZED LOGISTIC MAPS AS PART OF DS-SSM

Here, we discuss the generation of binary chaos codes of varying length for use in a DS-SSM system in certain wireless communication setup and show how synchronization can be achieved to recover transmitted data bits.

A. Binary spreading sequence generation using logistic map

Two methods are adopted to generate the binary spreading sequence using logistic chaotic map. These are

- Binary sequence generation using thresholding method.
- Binary sequence generation using floating point to bit conversion method.
- The generated sequences for three different values of r are shown in the Table 2.

Table 2. Sequences generated by logistic map for $r=3.61, 3.65$ and 3.69

Value of r	Binary sequence
3.61	00111000011010010100
3.65	01001001010010010101
3.69	01101001001110000110

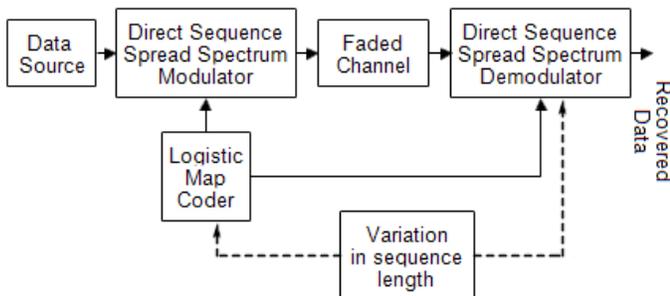


Figure 1: System model

B. System Model of using varying length chaos code as part of a DS-SSM

Certain experiments are performed as per the logical flow outlined in Figure 1. For fixed sized binary data blocks generated by following the process logic outline in Figure 1, chaotic codes of varying length are generated using a logistic map generator. The basic system is a direct sequence SSM (DS-SSM). In traditional cases, PN sequences and Gold codes are used as spreading/ de-spreading factors. Here, the chaotic code acts as a spreading factor in the transmitter and is considered for de-spreading in the receiver. Depending upon the fading observed in the channel, chaotic sequences of varying lengths are generated. Fading increasing with rise in speed of transmitters and receivers and /or either one or both. So depending upon the channel condition, the logistic map generator produces chaos codes of varying length. In case of traditional coding techniques like PN sequence or Gold code generators, the size of the spreading factor is dependent on the physical length of the register used for the purpose which is not the case with the chaos code. As varying sequence lengths of the spreading factor provides better QoS in a SSM [3], hence certain experiments are performed to check the suitability of the chaos code in faded channels generated due to varying speeds of transmitter, receiver and/or either one or both of the two.

Inside the DS-SSM, data blocks are passed through binary phase shift keying (BPSK) where the chaotic code provides spectrum spreading and de-spreading. Unlike PN sequence and Gold code, the chaotic sequence length variation is achieved by changing the iteration count within the logistic map generator. This is a major advantage which removes the dependence on the physical devices. For a given data block, as per the instantaneous channel state, a unique chaotic code is generated during transmission which is also made available at the receiver side during recovery of the content.

C. Achieving synchronization between coupled logistic maps for data recovery in a DS-SSM

Another set of experiments are performed as per the logical flow shown in Figure 2. Two coupled logistic maps are considered to execute the spread factor generation for bits 1 and 0 as part of a DS-SSM. Data coming from a source are bifurcated into streams of 1 and 0 bits after passing through a bit selector block.

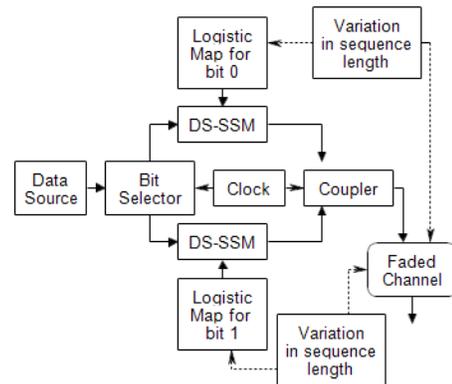


Figure 2: Coupled logistic map chaotic code generator for DS-SSM

Two coupled logistic maps are used to provide the spreading factors for generated DS-SSMed forms for both the streams. A coupler combines the output of both the DS-SSM blocks as per a clocking cycle used during bit separation. The coupled output is passes through the faded channel.

At the receiver side, two logistic maps attempt to achieve synchronization for recovery of the bits 1 and 0. The received signal is given as

$$y_{m,n}(x, \theta) = f_1(x(n), c(n)).S_n * H_{m,n}(x, \theta) + N \quad (10)$$

where,

- $f_1(x(n), c(n))$ is a modulation process with spreading factor $c(n)$ and bit sequence $x(n)$,
- $H_{m,n}(x, \theta) = Rayleigh, Rician_{m,n}$ is the channel matrix for x sample value and θ phase,
- $S_n = Step(l_n)$ with $l_{n+1} = r * l_n(1 - l_n)$ is a binary chaos sequence obtain logistic map with $r = 3.582$.

Two different logistic maps try to attain synchronization using the received signal y , and a feedback as per the following arrangement:

$$x(n + 1) = f\{(1 - \epsilon)x(n) + \epsilon y(n)\} \quad (11)$$

with

$$\epsilon \in \left[\frac{1+a}{2}, \frac{3-a}{2} \right] \quad (12)$$

Synchronization is achieved when

$$e(n) = |x(n) - y(n)| \quad (13)$$

is minimized.

Table 3. Simulation parameters.

Sl. No.	Item	Description
1	Modulation type	DS-SSM with BPSK
2	Data block size	Between 1000 to 10,000
3	Coders	PN Sequence, Gold code, Logistic Map.
4	Channel types	Rayleigh, Rician
5	No. of trials per sequence length	at least 10

V. EXPERIMENTAL DETAILS AND RESULTS

Extensive experiments have been performed to ascertain the performance of the proposed approaches under different wireless communication conditions. Simulation parameters are summarized in Table 3.

For the DS-SSM system, experiments are performed using PN, Gold code and chaos code sequences. Rayleigh channel with Doppler shift due to vehicular movement between 10 to 100 Km/h have been considered. Sequence lengths of 4, 8, 16 and 32 are considered for a range of channel conditions including vehicular and pedestrian. A range of 0 to 10 dB signal to noise ratio (SNR) is considered. The effect of sequence length variation on the QoS of the system is noted down in terms of bit error rate (BER) v/s SNR.

Table 4. BER v/s SNR plot of varying sequence lengths of PN, Gold and Chaos codes

Sl no	Sequence Length	SNR in dB	PN Sequence	Gold Code	Logistic Map
1	4	0	0.1	0.1	0.1
		2	0.07	0.05	0.05
		4	0.06	0.04	0.034
		6	0.05	0.03	0.03
		8	0.04	0.018	0.015
		10	0.01	0.008	0.007
2	8	0	0.091	0.085	0.083
		2	0.0637	0.0425	0.0415
		4	0.0546	0.034	0.02822
		6	0.0455	0.0255	0.0249
		8	0.0364	0.0153	0.01245
		10	0.0091	0.0068	0.00581
3	16	0	0.078	0.0752	0.059926
		2	0.0546	0.0376	0.030005
		4	0.0468	0.03024	0.020347
		6	0.039	0.02259	0.017928
		8	0.0312	0.013536	0.008964
		10	0.0078	0.006016	0.004183
4	32	0	0.065	0.0623	0.061
		2	0.04585	0.0311	0.0305
		4	0.03924	0.02492	0.020808
		6	0.0326	0.0186	0.01839
		8	0.02608	0.01116	0.0093
		10	0.0063	0.00488	0.00434

An average of ten trials of reading under fading conditions is shown in Table 4. Compared to the PN sequence, for sequence length 4, the chaos code shows a -3dB gain at 10 dB SNR. Similarly, for sequence length 8, the gain at 8dB is 9.3dB, for sequence length 16 the gain is 10.8dB and the 32 the corresponding SNR gain is 8.95dB which is significant. It indicates increase in QoS as established by falling values of BER with rise in SNR due to the use of chaos codes. One noticeable advantage is that for PN and Gold codes, the sequence lengths require registers of specific size unlike that in case of the chaos code. Here, with rise in the number of iterations, the sequence lengths increase. It leads to the saving of the hardware layout and power.

In case of the coupled logistic map case, the recovery of the data involves iterative processing. It requires a few cycles to establish the synchronization between the received signal and the expected output as described in Section III C. Figure 3 shows a plot between error and iteration obtained during synchronization of the two logistic maps used for transmission and recovery of 1 and 0 bits respectively while the fading is less severe. .

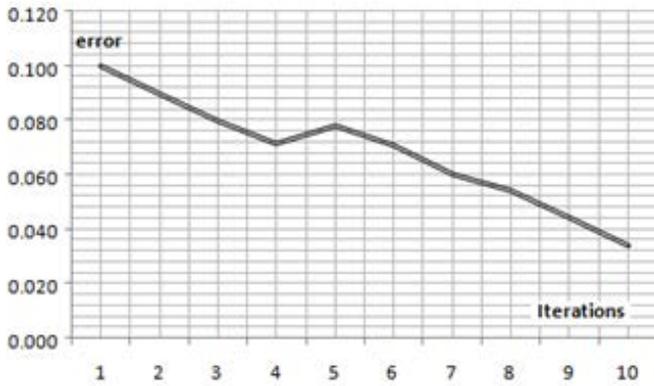


Figure 3: Error v/s iteration during synchronization between coupled maps

Table 5: Average computational time due to varying sequence lengths

Sl no	Sequence Length	Time is sec.s		
		PN Sequence	Gold Code	Logistic Map
1	4	1.35	1.42	1.4
2	8	2.13	2.3	3.1
3	16	2.52	2.55	3.6
4	32	3.1	3.3	4.4

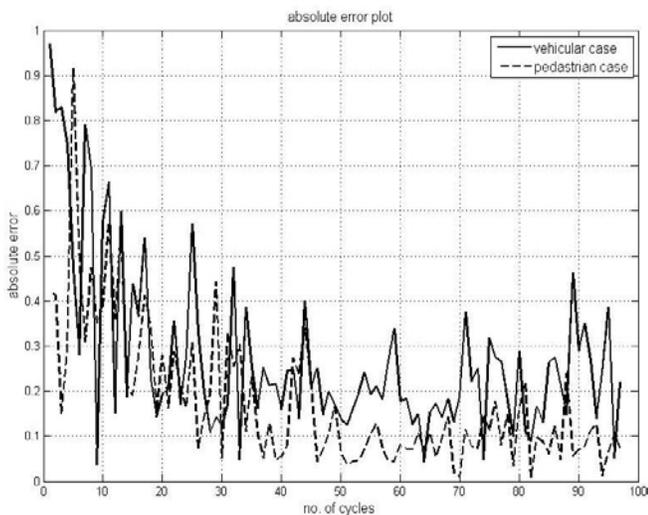


Figure 4: Absolute error plot from LMS Algorithm

Table 6: Average BER at 10dB SNR between logistic map showing chaotic behavior and in synchronized state

Sl no	Sequence Length	Logistic Map	Synchronous Logistic Map
1	4	0.007	0.0061
2	8	0.00581	0.0037
3	16	0.004183	0.002
4	32	0.00434	0.0013

The error plot shown in Figure 3 is obtained using a chaos code of length 32 which adds to the computational complexity of the system. Table 5 shows the average computational complexity of PN, Gold and chaos codes under fading conditions due to varying sequence lengths. It indicates that the chaos code is computationally expensive.

Another set of results are shown in Figure 4 which is obtained while establishing synchronization in vehicular and pedestrian conditions of the channel. There are distinct fluctuations which are due to the variations and dynamic behavior of the channel. Increasing the sequence length however reduces the number of iterations required to minimize the error.

The synchronization between two coupled logistic maps improves the performance of the system. In case of the DS-SSM, the synchronized logistic maps recover the bits 1 and 0 separately and generate the BER. The comparative values noted at 10 dB SNR for ten different trials are summarized in Table 6. It shows the clear advantage that the coupled synchronized maps provide. It, however, adds to the computational complexity.

VI. CONCLUSION

In this paper, we have discussed the use of chaotic behavior and synchronization property of logistic maps for improvement of QoS in wireless communication systems. We have explored the performance variations derived in case of chaos codes while used as a spreading factor in DS-SSM compared to PN and Gold code sequences. Next, we have shown how synchronization can be achieved in coupled logistic maps for improving QoS while transmitting bits separately. Though the chaotic behavior and the synchronized state of the logistic map add to the computational complexity, it improves efficiency of the design by not having any dependence on physical size of the registers. It reduces resign cost and saves power.

REFERENCES

- [1] T. S. Rappaport, *Wireless Communications-Principles And Practice*, 2nd ed.ition, PHI, New Delhi, 2008.
- [2] W. Turin, R. Jana, C. Martin and J. Winters J., "Modeling Wireless Channel Fading", *Proceedings of Vehicular Technology Conference*, Atlantic City, NJ, October 2001.
- [3] D. S, Swami and K. K. Sarma, "A Chaos based PN Sequence Generator for Direct-Sequence Spread Spectrum Communication System", *International Journal of Circuits, Systems and Signal Processing*, vol.8, pp. 351-360, 2014.
- [4] K. Kashyap, M. P. Sarma, K. K. Sarma and N. Mastorakis, "Generation of Orthogonal Logistic Map Sequences for Application in Wireless Channel and Implementation using a Multiplierless Technique", *Latest Trends in Circuits, Systems, Signal Processing and Automatic Control*, pp. 292-296, Jun., 2014.
- [5] S. N. Rasband, *Chaotic Dynamics of Nonlinear Systems*, Wiley, New York, 1990.
- [6] R. M. May, G. F. Oster, "Chaos from Maps". *Journal of Physics Letters A*, pp. 1 - 124, Vol. 78, Jul. 7, 1980.

Reliability Polynomials: Obtaining and Usage

Alexey S. Rodionov
 ICM&MG SB RAS
 Novosibirsk, Russia, 630090,
 Email: alrod@sscc.ru

Olga Rodionova
 Novosibirsk State University
 Higher College of Informatics
 Novosibirsk, Russia, 630058,
 Email: rolcon@mail.ru

—General problems of obtaining coefficients of reliability polynomials for different reliability indices, such as all-terminal reliability and average pairwise reliability are discussed along with discussion of using these polynomials for solving different theoretical and practical problems. It is shown that one of possible forms of a reliability polynomial presentation helps in considerable speeding up of its obtaining by using an intensional meaning of coefficients.

Index Terms—network reliability, reliability polynomials

NOTATION

G	Undirected probabilistic network
N	Set of n nodes
E	Set of m edges
e_i, e_{ij}	i -th edge or edge that connects i -th and j -th nodes, depending on context
p_j	Operating probability of j -th edge
$w_i = w(v_i)$	weight of node v_i , $WT = (w_1, \dots, w_n)$
$W(G)$	total weight of nodes of G
$R(G)$	all-terminal reliability of G
$r_{ij}(G)$ ($a_{ij}(G)$)	probability of v_i and v_j be connected (disconnected) in G . If G is known from context, then simply r_{ij} (a_{ij}) is used
$\mathbb{C}(G, s)$	mathematical expectation of a number of nodes in a connected subgraph that contains marked node v_s ; if v_1 is marked, then simply $\mathbb{C}(G)$
$N(G)$ ($M(G)$)	mathematical expectation of a number of disconnected (connected) pairs of nodes in G
$\bar{R}(G)$	average pairwise reliability of G
C	Edge chain composed of k edges e_1, \dots, e_k
G/C (G/e)	Network obtained from G by contracting chain C (edge e)
$G \setminus C$ ($G \setminus e$)	Subnetwork of G obtained by deleting chain C (edge e)
$G'(x, y)$	Network obtained from G by contracting nodes v_x and v_y
$G_C(n)$	Complete network with n nodes
$G_I(n_1, n_2)$	Grid network with $n_1 \cdot n_2$ nodes

I. Introduction

This paper is generalization and extension of authors' results in reliability polynomials of random graphs obtaining and usage [1]–[4]. Reliability polynomial of random graphs with unreliable elements are investigated

for a long time and by many authors [5]–[9]. Most often polynomials for all-terminal reliability (ATR) of a random graph with unreliable edge are considered. In [10] different forms of presentation of these polynomials are presented and usage of coefficients for graph's properties analysis is discussed. In [4] one of the authors of current paper had shown how the meaning of coefficients of a reliability polynomial for ATR in one of its forms can be used for considerable speeding up of their obtaining. Hereafter we show how meaning of coefficients can be used for their faster obtaining in the case of some other reliability indices.

In [8] reliability polynomials are used for estimating an “edge's” impotence in a graph's structure: edge, whose removal leads to a maximal decreasing of a polynomial's plot, is considered as contributed maximal income into graph's reliability. At the same time it is known that for some indices reliability polynomials for graphs with the same number of nodes and edges may intersect (see [3], [6], [7], for example). Thus as choice of optimal structure, as finding edges' “significance” depends on edges' reliability. Some examples of such structures are presented in the paper.

The rest of the paper is organized as follows. In Section II the factoring method is described in connection with polynomial's representation, in Section III some results on polynomial foe ATR are given

II. Factorization Method and Polynomial Representation

As it is noted in [10], there are various ways of a reliability polynomial representations. In some cases coefficients have intensional meaning. In [4] the presentation

$$R(G, p) = \sum_{i=0}^m a_i (1-p)^i p^{m-i}, \quad (1)$$

had been efficiently used for speeding up coefficients' obtaining. It is known that a_i is equal to a number of connected sugraphs (subgraphs on complete set of nodes), that may be obtained by removing exactly i edges. Coefficients of polynomial in its classic presentation

$$R(G, p) = \sum_{i=0}^m b_i p^i \quad (2)$$

are connected with those in (1) by the following equations:

$$b_0 = a_m; \quad b_{m-i} = \sum_{j=i}^m (-1)^{i+j} C_j^i a_j, \quad i = \overline{0, m-1}; \quad (3)$$

$$a_m = b_0; \quad a_i = b_{m-i} + \sum_{j=i+1}^m (-1)^{i+j-1} C_j^i a_j, \quad i = \overline{m-1, 0}. \quad (4)$$

Note that the meaning of coefficients in (1) not allows obtaining some coefficients directly and is used for derivation of finite expressions for graphs of small dimension (2-5 nodes) or of some special kind (trees, cycles) only, but it well corresponds for the factorization method also [12], [13]:

$$R(G, p) = pR(G/e_{ij}) + (1-p)R(G \setminus \{e_{ij}\}). \quad (5)$$

We use this method as basic one hereafter.

General scheme of the factorization method

General recursive scheme of the method is as follows:

- 1) Check if a graph allows direct obtaining of coefficients. If YES, then calculate them and exit, else go to the next step.
- 2) Check if graph's reduction is possible. If YES, then do it and go back to the step 1, else go to the next step.
- 3) Choose a pivot element and execute factorization, that is prepare pair of graphs and make recursive calls of the procedure.

Note that in some cases reduction leads to obtaining more than one graph of smaller dimension and, consequently, several calls of basic procedure are needed.

III. Reliability Polynomial for All-Terminal Connectivity

As it is mentioned above, the paper [4] is devoted to the subject. Here we cite some main points that allows significantly speed up the calculation process.

First to all, it is obvious that all a_i for $i > m - n + 1$ in (1) are zeros: it is impossible connect n nodes by less than $n - 1$ edges. Thus, when manipulating with intermediate polynomials during calculations, we need not consider coefficients that lay in this zone.

Next, when summarize independently obtained polynomials we need reduce them to a same power. For this purpose polynomials with lesser power are multiplied by the special polynomial of the kind (1) that is identically equal to one (its power is equal to a difference of powers of polynomials and maximal one):

$$I(n) \equiv 1 \equiv 1^n \equiv (1-p+p)^n = \sum_{i=0}^n C_n^i p^i (1-p)^{n-i}. \quad (6)$$

This trick allows reduce obtaining the vector of coefficients to summarizing vectors of binomial coefficients, may be shifted or truncated. Truncation allows omitting calculation of coefficients that are known beforehand.

A. Chain reduction

Main way of a graph dimension reduction is "branching by chain" [2]. Let us cite the theorem from [4].

"Let graph G have a chain Ch , that consists of k multi-edges e_1, e_2, \dots, e_k with multiplicities $\lambda_1, \lambda_2, \dots, \lambda_k$, correspondingly, that connects nodes v_u and v_t . Then

$$R(G, p) = \left[I(\lambda_{ut}, p) \prod_{i=1}^m M(\lambda_i, p) + M(\lambda_{ut}, p) \times \sum_{i=1}^k N(\lambda_i, p) \prod_{j \neq i} M(\lambda_j, p) \right] R(G/Ch, p) + N(\lambda_{ut}, p) \sum_{i=1}^k N(\lambda_i, p) \prod_{j \neq i} M(\lambda_j, p) R(G \setminus Ch, p), \quad (7)$$

where G/Ch – graph, obtained by contracting G by chain Ch , and $G \setminus Ch$ – graph, obtained from G by removing edges of this chain, λ_{ut} – multiplicity of edge that connects ending nodes of Ch directly (zero if no such edge). $N(\lambda, p) = (1-p)^\lambda$; $M(\lambda, p) = 1 - N(\lambda, p)$."

The theorem is a corollary of the corresponding theorem for calculation of all-terminal reliability [12].

IV. Polynomial for Average Pairwise Reliability

Task of obtaining this reliability index (APNC) is equivalent to one of obtaining mathematical expectation of a number of disconnected pairs of nodes in a random graph (EDP), that was examined in [14], [15], and for the case of equally reliable edges in [3]. Indeed, the following equations are clear:

$$\bar{R}(G) = \frac{C_n^2 - N(G)}{C_n^2}, \quad (8)$$

$$N(G) = C_n^2 (1 - \bar{R}(G)). \quad (9)$$

From this we have that if EDP-polynomial is

$$N(G, p) = \sum_{i=0}^m n_i (1-p)^i p^{m-i},$$

then polynomial for APNC is

$$\bar{R}(G, p) = \sum_{i=0}^m \frac{C_n^2 - n_i}{C_n^2} (1-p)^i p^{m-i}. \quad (10)$$

While obtaining EDP-polynomial in this form, we also can use meaning of its coefficients: n_i is equal to a total number of disconnected pairs of nodes by all possible variants of deleting exactly i edges. We use it as for obtaining equations for graphs of small dimension, as for calculating some coefficients beforehand.

Let all edges fail. This means that all connections are broken and their total number is $n_m = C_n^2 = n(n-1)/2$.

If only one edge remains, then we have m variants of such event. In all these events only one connection exists and $C_n^2 - 1$ are broken, thus

$$n_{m-1} = m(C_n^2 - 1) = m[n(n-1) - 2]/2. \quad (11)$$

In the case of two remaining edges we need knowledge about nodes' degrees in G . Let D_2 be a set of nodes with degrees exceeding one. Then total number of chains with length two is

$$K_2 = \sum_{v \in D_2} C_{deg(v)}^2, \quad (12)$$

and for each such case we have $C_{n-3}^2 + 3(n-3)$ broken pairwise connections, while for non-adjacent pair of edges — $C_{n-4}^2 + 4(n-4) + 4$. From this we obtain

$$\begin{aligned} n_{m-2} &= K_2 [C_{n-3}^2 + 3(n-3)] + (C_m^2 - K_2) [C_{n-4}^2 + 4(n-4) + 4] \\ &= K_2 (C_n^2 - 3) + (C_m^2 - K_2) (C_n^2 - 2) = C_m^2 (C_n^2 - 2) - K_2. \end{aligned} \quad (13)$$

Derivation of the next from the end coefficient (n_{m-3}) requires additional knowledge about structure of graph G , that means about existence and number of triangles (their finding has square complicity). Complete analysis of variants requires too much space, let us present final equations.

Let D_3 be a set of nodes with degrees not less than three. Let us denote a set of edges that connects nodes from D_2 as U^+ and consider a number of triangles T as known. In this case total number of chains with length three is equal to

$$K_3^{\sim} = \sum_{e_{ij} \in U^+} [(deg(v_i) - 1)(deg(v_j) - 1)] - 3T = S_1 - 3T. \quad (14)$$

Using these information and considering cases of non-adjacent remaining edges and of chain with length two and separate edge, we obtain final

$$n_{m-3} = C_m^3 (C_n^2 - 3) - K_2 (m - 2) + 6T - S_1. \quad (15)$$

Similar to the case of ATR-polynomial, existence of pre-defined coefficients allows significantly reduce number of operations at calculations and thus speeding up them.

When obtaining EDP-polynomial, one must take into account that when reducing graph's dimension, for example by removing dangling nodes, polynomially expressed weight of one or more nodes may occur (refer to [?]). In this case we need equating degrees of intermediate polynomials using(20). For example, if a graph G has one dangling node v_x with numerical value w_x of weight, that is connected to a main body by a single edge e_{sx} , then

$$N(G, p) = N(G^o, p) + (1 - p)w_x [W(G) - w_x], \quad (16)$$

where G^o has a structure of $G \setminus e$ with new weight of a node v_s , recalculated by equation $WT_s(G^o) = w_s + pw_x$. If we use form of a polynomial representation (1), then we transform these equations to $WT_s(G^o) = w_s(1 + p - p) + pw_x = (w_s + w_x)p + w_s(1 - p)$ and

$$N(G, p) = N(G^o, p) + (1 - p)w_x [W(G) - w_x] I(m - 1). \quad (17)$$

V. Polynomial for Average Size of a Connected Subgraph Containing Dedicated Node

Mathematical expectation of a number of nodes in a connected subgraph (MENC) that contains some dedicated node corresponds to an average size of monitoring part of a network. Without loss of generality we usually assume that this dedicated node has number one.

There are several ways of obtaining this polynomial. Most obvious is through obtaining all polynomials for two-terminal probabilistic connectivity $R_{1i}(G, p)$:

$$\mathbb{C}(G, p) = w_1 + \sum_{i=2}^n w_i R_{1i}(G, p). \quad (18)$$

Another way is also quit obvious and is based on the definition of mathematical expectation:

$$\mathbb{C}(G) = \sum_{H \in \mathcal{G}} W(H) P(H \text{ is connected and contains } v_1), \quad (19)$$

where \mathcal{G} – set of all possible subgraphs of G .

Modification of this way allows to decrease enumeration of variants:

$$\mathbb{C}(G) = \sum_{H \in \mathcal{G}} W(H) P(H \text{ is cut off}) P(H \text{ is connected}), \quad (20)$$

where \mathcal{G} is now a set of all subgraphs of G , that contains v_1 . We suppose that G itself is cut off with probability 1.

Last method that we consider is enumeration of connections of a dedicated node (v_1) with all other nodes:

$$\mathbb{C}(G) = w_1 + \sum_{i=2}^n w_i r_{1i}. \quad (21)$$

All these methods can be used for obtaining finite expressions for graphs of small dimension, that are needed for factorization method.

Polynomials for graphs with 2 or 3 nodes are obvious.

A. Case of 2-node graph

In this case m is a multiplicity of edge e_{12} .

$$\mathbb{C}(G(2, m), p) = w_1 I(m) + [I(m) - (1 - p)^m] w_2. \quad (22)$$

B. Case of 3-node graph

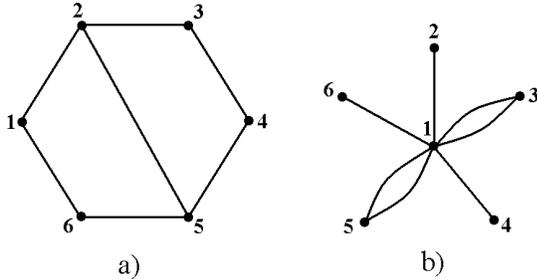
Let λ_{12} , λ_{13} and λ_{23} be multiplicities of corresponding edges ($\lambda_{12} + \lambda_{13} + \lambda_{23} = m$). Then

$$\begin{aligned} \mathbb{C}(G(3, m), p) &= w_1 + \{ [1 - (1 - p)^{\lambda_{12}}] + [1 - (1 - p)^{\lambda_{13}}] \times \\ &\quad [1 - (1 - p)^{\lambda_{23}}] - \\ &\quad [1 - (1 - p)^{\lambda_{12}}] [1 - (1 - p)^{\lambda_{13}}] [1 - (1 - p)^{\lambda_{23}}] \} w_2 + \\ &\quad \{ [1 - (1 - p)^{\lambda_{13}}] + [1 - (1 - p)^{\lambda_{12}}] [1 - (1 - p)^{\lambda_{23}}] - \\ &\quad [1 - (1 - p)^{\lambda_{13}}] [1 - (1 - p)^{\lambda_{12}}] [1 - (1 - p)^{\lambda_{23}}] \} w_3. \end{aligned} \quad (23)$$

In the case of numerical values of all weights we have

$$\begin{aligned} \mathbb{C}(G(3, m)) &= (w_1+w_2+w_3)I(m) - \\ &w_2I(\lambda_{13})(1-p)^{m-\lambda_{13}} - w_3I(\lambda_{12})(1-p)^{m-\lambda_{12}} - \\ &(w_2+w_3)[I(\lambda_{23})(1-p)^{m-\lambda_{23}} - (1-p)^m]. \end{aligned} \quad (24)$$

VI. Ambiguous Choice of Best Structure



1. Test 6-node 7-edge graphs

Different criteria may lead to different choosing of “best” graph’s structure. Let us examine graphs in the Fig. 1. For graph a) we obtain the following polynomials:

$$R(G_a, p) = p^7 + 7p^6(1-p) + 15p^5(1-p)^2, \quad (25)$$

$$\begin{aligned} N(G_a, p) &= 36p^5(1-p)^2 + 231p^4(1-p)^3 + 356p^3(1-p)^4 + \\ &263p^2(1-p)^5 + 98p(1-p)^6 + 15(1-p)^7, \end{aligned} \quad (26)$$

$$\begin{aligned} \mathbb{C}(G_a, p) &= 6p^7 + 42p^6(1-p) + 112p^5(1-p)^2 + 126p^4(1-p)^3 + \\ &85p^3(1-p)^4 + 34p^2(1-p)^5 + 9p(1-p)^6 + (1-p)^7, \end{aligned} \quad (27)$$

and for graph b):

$$R(G_b, p) = p^7 + 4p^6(1-p) + 4p^5(1-p)^2, \quad (28)$$

$$\begin{aligned} N(G_b, p) &= 15p^6(1-p) + 97p^5(1-p)^2 + 254p^4(1-p)^3 + \\ &345p^3(1-p)^4 + 256p^2(1-p)^5 + 98p(1-p)^6 + 15(1-p)^7, \end{aligned} \quad (29)$$

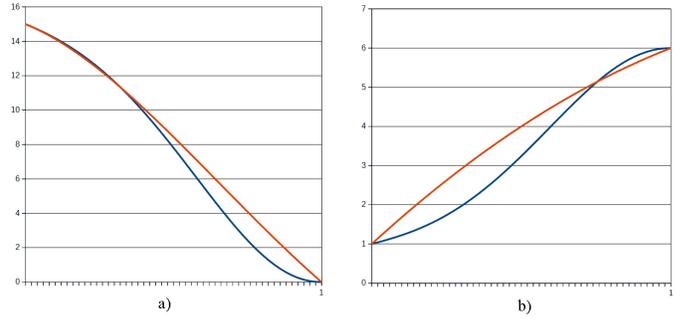
$$\begin{aligned} \mathbb{C}(G_b, p) &= 6p^7 + 39p^6(1-p) + 106p^5(1-p)^2 + 155p^4(1-p)^3 + \\ &130p^3(1-p)^4 + 61p^2(1-p)^5 + 14p(1-p)^6 + (1-p)^7, \end{aligned} \quad (30)$$

Comparison of these polynomials gives the following results. Graph a) is more reliable than b) for all $p \in (0, 1]$ by ATR-criterion, on the interval $(0, 0.319923)$ by the APNC criterion and on interval $(0, 0.681982)$ by the MENC-criterion¹. Graphics of $N(G, p)$ and $\mathbb{C}(G, p)$ of these graphs are presented in the Fig 2, a) and b), correspondingly. Thus, first graph is better by all three criteria on $(0, 0.319923)$, which corresponds to low-reliable edges. In the case of high-reliable edges choice depends on a criterion that is more impotent for concrete purpose of a modelled network.

VII. Conclusion

In the paper we show how organize process of obtaining polynomials for different criteria of a graph’s reliability in the case of independent edge’s failures. We show also how these polynomials can be used for choosing best

¹Matlab R2014b was used for finding polynomials’ roots



2. Graphics of $N(G, p)$ and $\mathbb{C}(G, p)$ for graphs in 1

(for given criterion) structure at the stage of a network design. Our further researches in the area are connected with other reliability criteria and consideration of different conditions applied to a network, its diameter (number of admitted commutations in connection) or number of possible connections through an edge, for example.

- [1] Rodionova O.K., Rodionov A.S., Choo H., Network Probabilistic Connectivity: Optimal Structures, ICCSA-2004, Springer LNCS. Vol. 3047 (2004) 431–440.
- [2] Rodionov A.S., Rodionova O.K., Choo H., Network Probabilistic Connectivity: Reliability Polynomial, in: ISPC Communication-2004. Proc. 8th Intern. Conf. “Problems of Operation of Information Networks”, Vol. 1 (2004) 321–327.
- [3] Gadyatskaya O., Rodionov A., Rodionova O., Using EDP-polynomials for network structure optimization, ICCSA-2008, Springer LNCS, Vol.5073, (2008) 1061–1077.
- [4] Rodionov A.S., Speeding up computation of the reliability polynomial coefficients for a random graph, Automation and remote control, n.7 (2011) 134–146.
- [5] Chari M. and Colbourn C.J., Reliability polynomials: a survey, Combin. Inform. System Sci, Vol. 22 (1997) 177–193.
- [6] Kelmans A.K., Crossing properties of graph reliability functions, Journal of Graph Theory, Vol. 35, Iss. 3 (2000) 206–221.
- [7] Kelmans A.K., Multiple crossings of network reliability functions, RUTCOR Research Report, Rutgers University (1994) 43–94.
- [8] Page L.B., Perry J.E., Reliability polynomials and link importance in networks, IEEE Transactions on Reliability, Vol. 43 Iss. 1 (1994) 51–58.
- [9] Camarda P., Bounds Evaluation of Coefficients in the Reliability Polynomial, Microelectronics and Reliability, Vol. 30, Iss. 2. (1990) 1699-1110.
- [10] Colbourn C.J., Some open problems on reliability polynomials, in: Congr. Numer. 93 (1993) 187–202.
- [11] Ayanoglu E., Cpih-Lin, A Method of Computing the Coefficients of the Network Reliability Polynomial, in: Proc. GLOBECOM’89. IEEE, Vol. 1. (1989) 331–337
- [12] Rodionova O.K., Rodionov A.S., Choo H., Network Probabilistic Connectivity: Exact Calculation with Use of Chains, ICCSA-2004, Springer LNCS, Vol. 3046. (2004) 315–324
- [13] Satyanarayana A. and Chang M.K., Network reliability and the factoring theorem, Networks, Vol. 13. (1983) 107–120.
- [14] Rodionov, A.S., Rodionova, O.K., Network Probabilistic Connectivity: Expectation of a Number of Disconnected Pairs of Nodes, in: HPCC 2006, Springer LNCS, Vol. 4208 (2006) 101–109.
- [15] Rodionov, A.S., Rodionova, O.K., Choo, H., On the Expected Value of a Number of Disconnected Pairs of Nodes in Unreliable Network, ICCSA-2007, Springer LNCS, Vol. 4707 (2007) 534–543.

MESH REFINEMENT WITH FINITE ELEMENTS AND ARTIFICIAL NEURAL NETWORKS

Fatima BELHABI¹, Mohamed ETTAOUIL²

Modelling and Scientific Computing Laboratory

University Sidi Mohamed Ben Abdelah
Faculty of science and Technology
Box 2202, Fez
MOROCCO

¹F.belhabib@yahoo.fr , ²MohamedEttaouil@yahoo.fr

Abstract: - In this paper, we present a new modeling for Mesh-size refinement with finite elements and artificial neural networks adopted by standards actual videos based on the SOM for image one domain, in the form of a structure. We developed in this study a mesh based on the object of interest by finite elements method and reduce the effort required to apply finite element analysis to image, this presentation that allows the identification of edges is a good representation of the movement of network nodes, and then we approach the follow-up of objects on sequences of Mesh-size refinement images. The algorithm of SOM the Kohonen is one of the important methods; it is a biologically inspired data clustering technique. It is a question of determining the Mesh adapte of an object nets, from one image to another. For that we used the algorithm allowing following a deformable plane object. On the one hand, we improve its performance, and then we study the optimization of the error function by error the Mesh-size refinement object simplification of our model, among the different meshes associated with images references. At the end of this work, we present simulation results.

Key-Words: - Refinement mesh-size, Learning Kohonen SOM, Mesh-size by Finite Elements, deformation the Mesh-size refinement, interpolation.

1 Introduction

Finite element analysis is a powerful computational tool for modeling the deformation objects. Getting place on a domain that corresponds to the representation space of the physical problem. To simplify the presentation, we will constraint to the case where is a differential variety of dimension 2D, and the Kohonen algorithm is an automatic classification method which is the origin of Self-Organizing Maps (SOM)[9]. This famous method falls within the framework of algorithms quantification vector and the method of k-means algorithm. More precisely, the SOM can be seen as an extension of the algorithm of pattern recognition and the automatic classification algorithms [5]. The latter algorithms detail the possible methodologies

for introducing expertise that follows the learning unsupervised stage [1].

Because its interesting and specific quality, Kohonen is efficient tool in the important domain which is the unsupervised classification . Indeed this latter is considered as a useful tool in data mining because it accelerates the research of relevant information. In fact the access to this latter is in suitable time has become a daily need and difficult task. This is due to the huge amount of information available in the website. An unsupervised classifier groups the similar information that refer to the same topic in same cluster and these which are dissimilar in the distinct ones. This avoid the search of the desired information in a lot of clusters, consequently an important time is economized. The Kohonen algorithm is unsupervised partitioned classifier i.e.

it treat with unlabeled inputs and provides classes with no overlap. Beside its robustness i.e. its ability to resist to the noise, the Kohonen algorithm possesses other interesting properties. Indeed the self-organizing map is an unsupervised neural network which projects high-dimensional data onto a low-dimensional grid which called a topological map [9]. This projection must preserve the topology of inputs (more details of this point is given after in this paper). This lead to an organization and representation in map. This latter provides a visualization of the data. Furthermore the distances between classes are identified and visualized. This propriety is the most attracting quality of this algorithm. For this reason many huge efforts are performed aiming to overcome the shortcoming of this algorithm and improve the performance of kohonen algorithm.

A mesh is therefore determined by geometry and topology [3]. A mesh size is said to be consistent if the intersection of two of these distinct triangles is either an arc or a node, or null. In other words it prevents a summit to be specified only at the end of an edge. Property compliance is critical to ensure the continuity of function interpolating thereof. We distinguish different types of meshes. A priori triangles can be, but it often imposes constraints on the number of edges of these triangles. It often sets the number of edges by triangles (triangular meshes) or (quadrilateral meshes). Triangular meshes are however a number of advantages. On the one hand, triangles can more accurately model the border of any domain. Indeed, any area polygon border is triangularly. One the other hand triangular meshes interest to provide continuous representation modeled objects and the associated interpolation functions are often easier [5][32].

Generally, the size of the topological map is randomly chosen. Indeed such choice effect the performance of the Kohonen algorithm. Consequently, the labeling phase becomes difficult. The weight vectors are also randomly selected, hence the result is affected too. To facilitate the selection phase, we propose in this work a learning method which allows selecting the size of the topological map. In order to attempt this goal, we add to learning Kohonen algorithm a phase called selection stage. This phase is based on a function called selection function [17][15]. To construct this latter, we use a sub-set of the data set. In addition, we divide the topological map on two parts. The first contains the used neurons; the second part is formed by the unused ones. It should be noted that the size of the first and the second parts are

modified by the use of the selection function. This method can be also used as a mean of neural architecture optimization.

This paper is organized as follow: The first paragraph relates of the Mesh-Size Finite Elements Method, that is to say, approximation by piecewise, the second consists introduced Mesh –size refinement objects by models of the SOM, the third consists in modeling a model of the Optimum Adaptation of a Mesh Size by finite Elements [6] and SOM. Simulation results are also available.

2 Mesh-Size Finite Elements Method

2.1 Finite Element Method (FEM)

The principle of finite elements solution is to show values calculated from variable approach taken in a number of end points properly chosen. We then rebuilt solution across the field by interpolation [10]. Extends in the method discretization involves setting up a domain partition considers in simple elements. This partition is the mesh of the domain and must possess properties that depend on the application. The finite element methods provide approximations of exact solutions and we try to get the error associated with the smallest possible [7].

We present generally how can be approximated by piecewise that are to say mesh-size by finite elements. The method finite element (FEM) uses a representation based nodes by estimating the values in the facet by functions the nodal values weighting. Polynomial representation is also widely used in the field of image deformation or coefficients are obtained by identifying the nodal values to a polynomial approximation of piecewise. Mesh-size is used for modeling by discretization of partial differential equations. It is also possible, by a judicious choice of weighting functions of the FEM, these two performances are the same, and however, the nodes based representations have several advantages:

- Numerical analysis more efficient and more stable.
- The definition of positions and values of the nodes require less accuracy that polynomial coefficients estimation (this last point is especially important in applications of coding that all these parameters are subject to the quantification).
- It is easier to visualize a function or deformity interpolated from the positions of the nodes from polynomial coefficients.

2.2 Mesh size by models of surface finite elements

This method is to evaluate values approached the field looked only at certain points in the domain, and then deduct these values by interpolation, the solution at any point. A first step will be to achieve a Mesh size of the field that will define, among other things, points. This pre-treatment is an essential phase of the method, the mesh nodes, will determine the convergence of the method to a good solution[34]. Also, the domain can have a complex geometry, which implies that the mesh is not trivial. We focus only, in this section triangular mesh [20]. The algorithms presented here for automating finite element analysis are potentially applicable to a variety of objects videos that require deformation modeling. In this paper, the conception of the finite elements method of the mesh can be realized by three phases:

- The first phase is the analysis of the problems, which consist the studying of domain geometry that is very complicated problem will be broken down into simple forms problems.
- The second phase is the formal construction the mesh: Takes into account the analysis results, and defines simple objects allowing division the total work into phase.
- The third phase is the realization of the mesh, which includes the previous phases.

3 Refinement Mesh-Sizes with Maps SOM AND FINITE ELEMENTS

3.1 Self-Organizing Map (SOM) and Finite Elements.

The Self-Organizing Map (SOM), proposed by Kohonen, consists projecting high dimensional data onto a low-dimensional grid [9]. The projected data preserves the topological relationship of the original data; therefore, this ordered grid can be used as a convenient visualization surface for showing various features of the training data, for example, cluster structures [8].The Kohonen network has one single layer, let name this one the output layer. The additional input layer just distributes the inputs to output layer. The neurons of this latter are arranged in a matrix. We consider, in this work, that the map is in two dimensions. The goal of self-organizing maps is to associate with each neuron a referent of a vector space data; see figure 1. The number of neurons on input layer is equal to the dimension of input vector. Kohonen has proposed various

alternatives for the automatic classification, and presented the Kohonen topological map [9]. Basing on the structure graph defined on the topological map, Kohonen has defined the discrete distance δ . For any pair of neurons $(c; r)$, it calculates a discrete distance $\delta_{c,r}$, where $\delta_{c,r}$ is defined as the length of the shortest path between c and r on the graph. For each neuron c , this discrete distance can define the concept of a neighborhood as follows: $V_c = \{r / \delta_{c,r} \leq d\}$, where d represents the ray of the neighborhood of the neuron c .

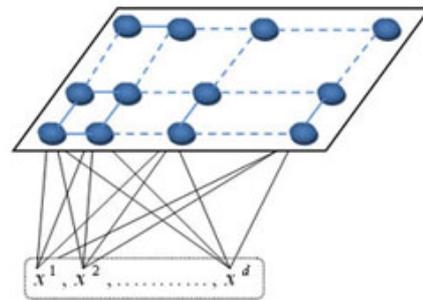


Fig. 1 Kohonen topological map

3.2 The Proposed Method

In this part, we describe with some detail the main component of our approach.

It should be noted that this method consists integrating a new phase in the Kohonen algorithm.

This later is called Interpolation learning phase and has as role to facilitate the refinement Mesh.

Given a sequence of images, we seek to refine a Mesh-size each pixel based on the principle of conservation of intensity [13]. Our method is based on a specific architecture of neural networks this architecture allows to refine the mesh in two steps. The first is to refine the mesh in each pixel using the algorithm of the SOM [20] winner. The second step is used to adapt the mesh according to the movement range of each pixel, the results of the previous phase, in which the displacement of a pixel configuration takes into account the movement in the vicinity [19]. The approach we propose is based on two phases: an initialization phase is to assign the weight of each neuron-image value of the gray level of the pixel corresponding to the images of the first part. During the second phase, after building the learning that includes all the images of all cycles together With interpolation finite elements, we present the elements of this set iteratively to

neuron-pixel to deduce what the neuron-image of the input values. For updating the weights of the neuron, called "winner neuron" we can not change is that the weights of the winner without changing the weights of the neighboring neuron-images (SOM algorithm) or modify the weights of the neuron-picture winner and the weight of the neighboring neuron-images (SOM algorithm.) the purpose of the change is to allow different weights converge to values that represent the corresponding image appropriately according to measurements from the acquisition of several cycles. At the end of the process, the neuron-map images form a sequence of images.

The network we propose is inspired by Kohonen maps in the principle of self-organization. It consists two layers. An input the number of neurons which is equal to the number of pixels of a block layer of the mesh of predetermined size, a self-organizing map having the same size as an image where each pixel represents a neuron of an image of the sequence and a self-organizing is a Kohonen map of input picture size. The images of the sequence are considered one by one by presenting, after cutting into blocks of fixed size, to the input of the network. Figure 3 shows the network topology that we propose for the mesh refinement.

- o The input layer E: composed of $n * m$ neurons, with $m * n$ representing the size of the vicinity.
- o The self-organizing map consists of N neurons having the size of an image sequence in which each neuron is connected to all the neurons of the input layer.

The solution of finite elements of type winner is interpolated on the structured cloud of mesh builds in the steps. To do it, we proceed as follows:

- The network of points mesh calculated in the previous step is planned on the mesh of the face not deformable.
- These coordinate are used as functions to calculate the point surface of interpolation to deform; by using the functions of interpolation; calculated the solution finite elements by type movements with the algorithm of the learning a method [5].

3.3 Algorithm

The algorithm we propose for our method consists of two stages: the first allows initializing the mesh of each pixel based on the principle of the Kohonen SOM algorithm. The second which may be recursive

or non-recursive can refine the mesh in the case of displacement obtained in the previous phase.

Initialization phase:

The initializing phase for the Kohonen maps generally involves initializing the weight vectors with random values or, when these are known a priori. In the approach we propose, the weight vectors are initialized with mesh images, supposedly close to the other images of the periodic and orderly sequence, which requires the sequence of neural network (image 1 mesh -> neuron 1 , mesh image 2 -> neuron 2 ...). This phase is therefore to present sequentially input to the network, T-meshes corresponding image (Equation 3.1). The network is initialized by writing to each neuron-mesh t an image of a CAD wt weight vector each of whose components $n, i = 1 \dots n$, is equal to the grayscale noted $w_n^t = I_n^t, i = 1 \dots n$, each of the n pixels in the input image. The operation is repeated for T-meshes neuron-images using pictures successively T cycle used for initialization.

$$w_1^1 = I_1^1 \quad w_1^t = I_1^t \quad w_1^T = I_1^T \quad (1)$$

$$w_n^1 = I_n^1 \quad w_n^t = I_n^t \quad w_n^T = I_n^T$$

With the vector associated with each neuron-cell-picture:

$$w^1 = (w_1^1, \dots, w_n^1), \quad w^t = (w_1^t, \dots, w_n^t) \quad \text{AND} \quad (2)$$

$$w^T = (w_1^T, \dots, w_n^T) \quad (3)$$

Learning phase

The sequence we have is composed of L cycles, the first cycle is used to initialize the network, other cycles form our learning together. During the learning phase, we present successively the elements of the training set, $L-1$ cycles of the input image grid. The first values, presented to neuron-pixels of the input layer are the intensity of the first image of the second mesh-cycle values, and then the second-grid image and so on until the meshing of the picture of the last cycle $T L$. The entrance is mesh mesh image image I_i amplitude of pixel i of the image being analyzed is assigned to the input neuron-pixel e_i or $e_i = I_i$. We are looking for each entry, wt component corresponding to the neuron-mesh-ct picture of CAD that minimizes the gap with the values shown in entry. This neuron-mesh-image is called neuron-mesh-picture "winner" or<< simply "winner."

A measure of the difference between T-neuron

meshes images of the self-organizing map, construction and meshes the input image (the values of which are carried by the neuron-pixels of the input layer) is defined as equation 1:

$$Em = \text{Ming}(w^t, e) \tag{4}$$

With $t=1..T$.

Em is the minimum value for the neuron-mesh-image t^* of the self-organizing map, considered the winner among other neurons-mesh-map images. Then we proceed to the modification of the weights of neuron-mesh-picture winner and the weight of neighboring neurons-images according to the following formula:

$$w^j(t) = w^j(t-1) + \beta_{t,k} (e - w^j(t-1)) \tag{5}$$

$$\beta_{t,k} = \exp\left(\frac{\|r_i - r_k\|}{(\sigma(t)^2)}\right) \tag{6}$$

Where $\|r_i - r_k\| \approx \|w^j - w^k\|$ and r_i the vector represents the coordinates of weights and functions are decreasing.

The main important parameter in our method is the selection set. In this context, the choice of the later must be done carefully, taking into account the quality of learning and reduced complexity of the proposed method.

Interpolation phase

The image I of the surface that we seek to approximate \hat{I} by a surface defined a polynomial in two variables: $\hat{I}(x, y) = P^n(x, y)$ we know that it is preferable to perform interpolation piecewise rather than global phenomena that lead to instability, but this involves studying between the different areas in order to ensure the consistency of the interpolate[3].

- We restrict ourselves by triangular elementary domains, also called elements. To triangulate and cover domain Ω image elementary domains support, triangle offers flexibility
- This is particularly useful for approaching the boundaries whatsoever such as the edges of object. Image border Ω is a polygon and in this case Ω is fully and accurately covered by triangles Basic fields defining a partition Ω , assembled them so that they have:
 - An intersection is empty.
 - Either an intersection reduced to a common vertex.

- Either an intersection reduced to one side common.
- Should therefore define a global numbering of elementary domain e_i , global numbering for all S_i summits and a relationship that each elementary domain associates list heights.

Given a triangulation $T_\Omega = \cup_{i=1}^N e_i$ formed by triangles of a closed polygon domain Ω . Knowing the value S_i to the surface to approximate in each vertex of a triangle, I_{S_i} , it seeks to build \hat{I} such as:

Restricting Court an e triangle is a polynomial of degree $P_T \leq 1$.

- Determining P_T on e triangle vertices $S1, S2$ and $S3$ knowing $I(S1), I(S2)$, and $I(S3)$, conditions are imposed:

$$P_T(S_i) = I(S_i) \quad i \in \{1,2,3\} \tag{7}$$

To do this, define 3 basic functions are commanded local ψ_i checking for an e triangle vertices $S1, S2$ and $S3$:

$$\psi_i(S_j) = \delta_{ij} \quad i, j = 1,2,3 \tag{8}$$

- Interpolated point value expression (x, y) owned by an Element therefore corresponds Fig. 2 and Fig. 3.

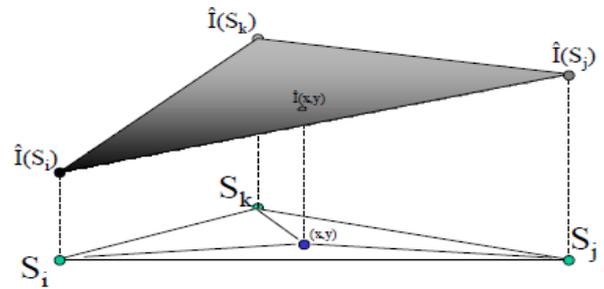


Fig2. Interpolation of value depending on the model of Lagrange, equivalent to the projection of the point (x, y) on the plan.

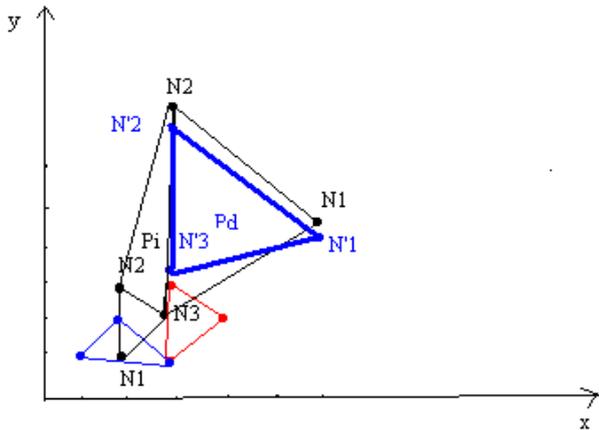


Fig.3: nodes the mesh deformable

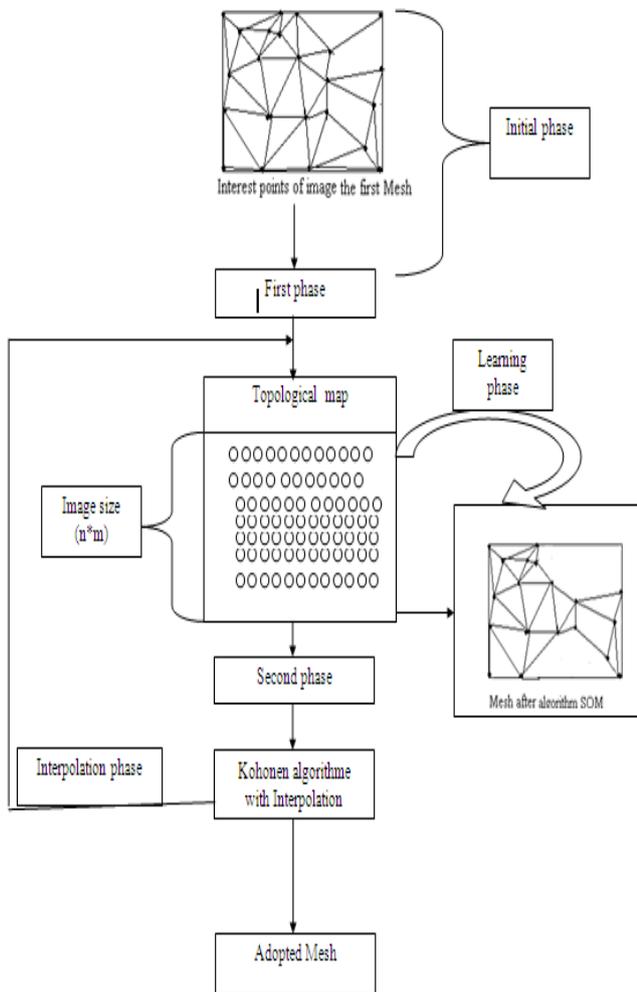


Fig2. Adapted Mesh Diagram

The diagram shows the different steps of our method.

- The choice of the selection set has great impact on the algorithm performance and on its complexity.
- The selection of neuron must reflect the most information of the population space. Since the interpolation function is defined on P, a big size of this set increases the complexity of the system. So, suitable choice is required [17][15].
- The dimension of the investigated inputs is typically huge. The neural network of kohonen is commonly used as an efficient mean to reduce this dimension. In the beginning no information about the class is available. So the choice of map size and the active neuron is made image size. Hence the performance will be affected and some problem will arise in the labeling phase. The update can be considered as a correction of initialization phase and also as a way for searching a suitable architecture map. Thereby the performance of the system will be improved.
- The proposed method modelizes a behaviour used naturally by the human. Indeed, this latter when he wants accomplishes a task., at first he choose randomly a set of persons. After, this set is updated by liberating some persons and adding others basing on their competency. This behavior is also used by certain animals as ants [15].

4 Computational experiments

Experimental results are provided to compare the performance of the proposed refinement Mesh maps with finite elements by occlusion adaptive forward mesh tracking versus test sequence. The sequence starts with slow movements of the head of the Sequences Miss America, Foreman, Salesman, and Table Tennis were selected for the simulation and results can be seen in Table 1. However, the part between frames 150 and 151 is especially challenging and well suited to demonstrate the occlusion-adaptive mesh tracking concept. The processed video, the overall performance is improved in all test sequences at a fixed bitrate (0.2 bpp). We identified two results: refinement Mesh maps and refinement Mesh finite elements.

- 1) a new uniform mesh at each frame, and 2) a new content-based mesh at each frame. An algorithmic description of the refinement-Mesh by finite elements mesh approach is given in the Appendix.

The case of redesigning a uniform mesh is expected to be a lower bound on the performance of mesh maps by the proposed forward tracking content-base mesh, since the structure of the mesh may not fit the motion boundaries well, leading to multiple motions within a single patch. However, it requires transmission of no overhead information about the mesh structure.

The proposed forward tracking content-based mesh is a compromise between these two, since it yields a mesh structure that fits the scene content without too much overhead transmission. To this effect, the efficacy of the proposed method has been evaluated based on how it compares against these benchmarks in terms of motion-compensation PSNR and the number of node points whose coordinates need to be transmitted at each frame. The PSNR values refer to the prediction PSNR of each frame based on the original of the previous frame, using the affine motion field interpolated from the node-point motion vectors.

We have implemented the algorithms described in this paper in Java and employed them in a cloth simulation system. The tests were run on a standard Pentium 4 with 2.8GHz and 512 MB RAM, using the Java Runtime Environment 1.6.0.

Table 1: optimization of neural architecture

Error after using SOM maps with finite elements "Foreman frame 150"	Error after using SOM maps finite elements « Foreman : frame 151 »	Error after using SOM maps finite elements « Miss »
148.477301	166.729442	79.275862
158.679901	147.208811	75.773006
136.577586	144.424079	75.486486
134.106460	143.673505	74.850433
132.891856	141.990975	74.814356
140.155492	141.960921	85.000000

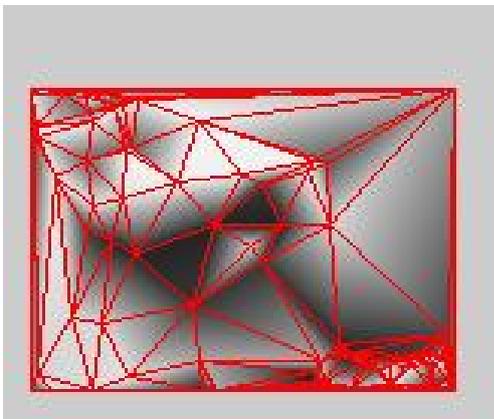
145.939469	141.951198	78.180812
138.193152	141.191111	73.644295
132.405510	140.778883	53.027019
143.106682	140.680321	52.146341
132.343431	140.111111	51.931646
132.515564	150.050847	52.792683
132.242105	139.948010	51.703582
131.612766	139.846154	51.388170
142.120000	139.283898	89.142857
135.786184	139.254237	51.317073
130.721947	146.205882	78.285714
131.190590	138.242820	50.319564
130.398340	138.120346	50.312821
134.925926	137.996516	57.581395
130.240469	137.703812	61.425000
146.929174	138.513292	49.628319
131.159716	135.872483	65.285968
129.073077	140.619495	49.790000
128.806897	139.276257	49.574820
134.695989	136.625000	49.500000



Original Image the Forman



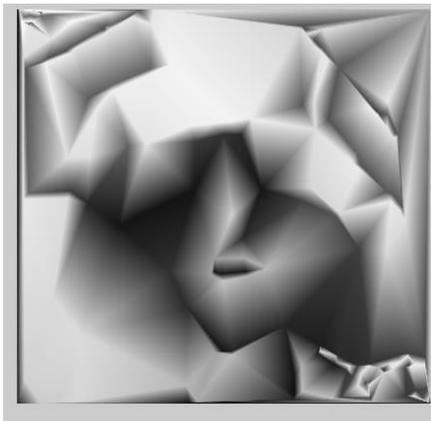
Original Image the Miss



Refinement the mesh SOM maps with finite elements

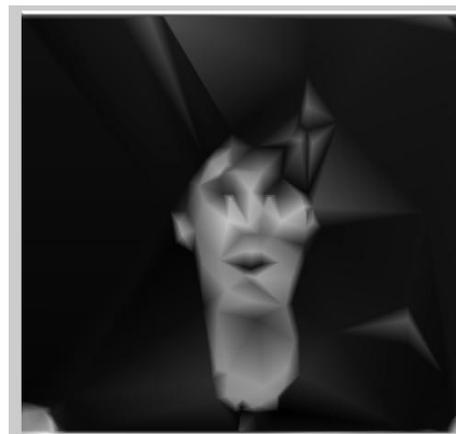


Refinement the mesh SOM maps with finite elements



Error = 136.625000

Table 1 presents the error de PSNR of the the Kohonen map with Finite Elements. It shows that only with 100 iterations, the proposed method lead to a reduction of the card.



Error = 49.500000

From the Table1, we see that the proposed method gives satisfying results, Kohonen algorithm will be well appreciated. The obtained amelioration can be

explained as follows:

- The first term of the objective hybridation of our proposed model controls the geometric error on a given refinement.

The proposed method completes the Kohonen learning method. In fact, our approach realizes two tasks at the same time: the learning task and the regroup of the objet one which consists minimizing the size of the map.

4 Conclusion

In this paper, we have interested in Finite Elements with Artificial Neural Networks. We have chosen as a method of mesh refinement by self-organizing map of Kohonen SOM algorithm with Finite elements , in which we presented and its main techniques we have developed and implemented a method aims supervised learning, error minimization and mesh adaptation in either refine or looping mode. We have applied our algorithm to a widely used dataset, objects videos for refinement Mesh size. In this respect, our method produces good results in reasonable time in comparison with the recent.

References:

- [1] M. Ayache, M. Khalil and F. Tranquart. "Artificial Neural Network for Transfer Function Placental Development: DCT and DWT Approach". *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 5, No 3, September 2011.
- [2] G. Dryfus, J.M.Martinez, M.Samuelides, M.B.Gordan, "Reseaux de neurons Mthodologie et applications". EYRLLES, SC 924, 2000.
- [3] M. Ettaouil, and Y.Ghanou, and K. Elmoutaouakil, M. Lazaar. "A New Architecture Optimization Model for the Kohonen Networks and Clustering", *Journal of Advanced Research in Computer Science (JARCS)*, Volume 3, Issue 1, 2011, pp. 14 - 32.
- [4] M. Ettaouil, and Y. Ghanou, and K. El Moutaouakil, M. Lazaar, "Image Medical Compression by A new Architecture Optimization Model for the Kohonen Networks". *International Journal of Computer Theory and Engineering* vol. 3, no. 2, 2011, pp. 204-210.
- [5] M. Ettaouil, and K. Elmoutaouakil, and Y.Ghanou. "The continuous hopfield networks (CHN) for the placement of the electronic circuits problem ", *WSEAS Transactions on Computer*, Volume 8 Issue 12, December 2009.
- [6] M. Ettaouil, and E.Abdlatif and F.Belhabib and K. Elmoutaouakil. "Learning Algorithm of Kohonen Network With Selection Phase ", *WSEAS Transactions on Computer*, Volume 11 Issue 11, December 2012.
- [7] M. ETTAOUI and M. LAZAAR, "Improved Self-Organizing Maps and Speech Compression", *International Journal of Computer Science Issues (IJCSI)*, Volume 9, Issue 2, No 1, pp. 197-205, 2012.
- [8] A. Ghosh, and S.K. Pal, Object Background classification using Hopfield type neural networks. *International Journal of Patten Recognition and artificial Intelligence*, 1992, pp. 989-1008.
- [9] S. Ghorbel, M. Ben Jemaa and M. Chtourou. "Object-based Video compression using neural networks". *IJCSI International Journal of Comput er Science Issues*, Vol. 8, Issue 4, No 1, July 2011.
- [10] J.J. Hopfield, Neurons with graded response have collective computational properties like those of two-states neurons. *proceedings of the National academy of sciences of the USA* 81 ,1984,pp. 3088-3092.
- [11] C. C. Hsu, Generalizing Self-Organizing Map for Categorical Data. *IEEE Transactions on neural networks*, Vol. 17, No. 2, 2006.,pp. 294-304.
- [12] S .Marsland, S.U .Nehmzow,, and Shapiro, J., "A self-organizing network that grows when required", in *Neural Networks*, Volume 15, 2002, Issue 8-9, pp.1041-1058.
- [13] N. Nasrabadi and Y. Feng, "Vector quantization of images based upon the Kohonen self-organizing feature maps", in *IEEE Int. Conf. Neural Networks*, San Diego, CA, vol. 1, pp. 101-108, 1988.
- [14] T. Kohonen. *Self Organizing Maps*. Springer, 3e edition, 2001.
- [15] H. Shah-Hosseini, and R. Safabakhsh. *TASOM: The Time Adaptive Self-Organizing Map*. The International Conference on Information Technology: Coding and Computing (ITCC'00), 2000, 422.
- [16] H. Yin. "ViSOMA Novel Method for Multivariate Data Projection and Structure

- Visualization“. IEEE Transactions on Neural Networks, Vol 13, 2002, 1, pp. 237-243.
- [17] www.ics.uci.edu/mlearn/MLRepository.html.
- [18] V. Selvi, R.Umarani “Comparative analysis of ant colony and Particle Swarm optimization techniques “ International journal of Computer Application Volume 5-No.4, August 2010, pp.0975-8887
- [19] D. Wang, ‘Fast Constructive-Covering Algorithm for neural networks and its implement in classification’, Applied Soft Computing 8 ,2008, pp. 166-173.
- [20] D. Wang, N.S. Chaudhari, ‘A constructive unsupervised learning algorithm for Boolean neural networks based on multi-level geometrical expansion’, Neuro computing 57C, 2004, pp.455-461.
- [21] M.C. Rivara, Local modification of meshes for adaptive and/or Multigrid finite-element methods, J. Comput. T. Meinders, Developments in numerical simulations of the real life deep drawing process, Ph.D. Thesis, U Ponsen & Looijen Wageningen (publ.), ISBN 90-36514002, 2000.
- [22] G. ROBERT, Représentation et codage De séquences vidéo Par hybridation de fractales et d éléments finis. PhD thesis, Université Joseph Fourier-Grenoble 1 Sciences Géographie, 2000.
- [23] C. Toklu. Object-based Digital Video Processing using 2-D Meshes. PhD Thesis University of Rochester, USA, 1998.
- [24] Y. Wang and O. Lee. Active mesh : “A feature seeking and tracking image sequence representation scheme”. IEEE Transactions on Image Processing, 3(5) :, 1994, pp 610–624.
- [25] O.C. ZienKiewicz, ” La méthode des éléments finis ”, 1973, Edi-science, Paris.
- [26] P.L George. “Génération Automatique de maillage- Applications aux méthodes d’éléments Finis“. Masson 1991.
- [27] K.HO-Le, “Finite elemnt mesh generation methods: Areview and classification l comput. Aided Deseign 20(1), January 1988, pp 27-38.
- [28] ISO/IEC 14496-10 and ITU-T Rec. “H.264. Advanced Video Coding”. 2003.
- [29] H. B. Jung and K. Kim, “A New Parameterisation Method for NURBS Surface Interpolation”, The International journal of Advanced Manufacturing Technology, Vol. 16, 2000, p. 784-790.
- [30] F. Jurie and M. Dhome. “Hyperplane approximation for template matching”. In IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 24(7), 2002, pp996 1000,.
- [31] Iain E. G. Richardson. H.264 and MPEG-4 “Video Compression: Video Coding for Next-generation Multimedia”. Ed. John Wiley & Sons, 2003.
- [32] J.A. Robinson and M.A. Ren. Data-dependant sampling of twodimensional signals. Multidimensional Systems and Signal Processing, 6, 1995, 89–111.
- [33] A. Perera, C-L. Tsai, R. Flatland et C. Stewart. Maintaining valid topology with active contours : theory and application. IEEE International Conference on Computer Vision .
- [34] G. Turk. “Re-tiling polygonal surfaces”. Computer, Graphics SIGGRAPH ’92 Proceedings), 26(2):, July 1992, pp 55–64.

Determination the Coefficient of Regenerative Losses in Stirling

Prof.eng.Traian FLOREA Ph.D, Stud. Catalin OITA, eng. Traian Vasile FLOREA Ph.D
 Department of Engineering Marine
 Naval Academy "Mircea cel Bătrân" of Constanța

Abstract: The coefficient of regenerative losses, X , is the term that includes all the losses due to incomplete heat transfer in the regenerator of a Stirling machine. This parameter clearly depends on a large number of variables. The objectives of this paper are (1) to describe a method for evaluating the coefficient of regenerative losses, X , and (2) to demonstrate the application of this coefficient in calculating the efficiency and the power output of the Stirling engine. The relationship expressing the regenerative losses as a function of all parameters has been evaluated using first law and heat transfer principles applied to both the regenerator and the gas. The analysis resulted in differential equations that were then integrated. This integration is based on (1) a lump sum analysis, which tends to predict higher values of the regenerative losses, X_1 ; or (2) on a linear distribution of the temperature in the regenerator matrix and gas, which tends to predict lower value of the regenerate losses X_2 . The results of computations of regenerator heat losses, efficiency and power output based on this analysis are compared to the performance, data from the twelve operating Stirling engines. A high degree of correlation is found over a wide range of operating conditions for all the engines.

1. Introduction

This paper presents a method for calculating the coefficient of regenerative losses in Stirling engines and it includes the derivation of the equation for the calculation of those losses.

The P-V/P-x diagram has been presented previously and has been shown to be an effective tool for analyzing the Stirling Cycle, both theoretical and actual (Petrescu et al. 1999, 2000a, 2000b, 2000c; Florea 1999). In this paper, a T-x diagram has been added to the P-V/P-x diagram in order to more clearly present and explain the processes that take place in the regenerator.

An equation is presented for calculating the regenerative losses, X , in Stirling engines based on using the P-V/P-x/T-x diagram for insight and using the First Law of Thermodynamics as a basis for calculations. This equation for X has been found to accurately predict the

coefficient of regenerative losses for twelve engines under sixteen conditions of operation (Allen 1987, Fujii 1990, Fared 1988, Geng 1987, Stine and Diver 1994).

2. Using the P-V/P-x/T-x diagram as an aid in understanding the Stirling Engine Cycle

The PV diagram in Fig. 1 illustrates how the isothermal compression process 1-2 and the isothermal expansion process 3-4 occur in the Stirling Engine Cycle. The P-x and T-x diagrams in Fig. 1 illustrate the operation of the regenerator during the isometric processes 2-3 and 4-1. The T-x diagram serves to clarify the operation of the regenerator.

Referring now to the P-V diagram in Fig. 1, the isothermal compression process 1-2

takes place at temperature (T_L) with the regenerator stationary at R_2 while the working piston, P moves from P_t to P_2 . Heat, Q_L is rejected to the cooler during this process.

The isometric process 2-3 occurs next. During this process, the working piston remains stationary at (P_2) while the regenerator moves from R_2 to R_1 in the displacer cylinder. During this process, the pressure variation, assuming no losses, is shown on P-x coordinates as 2'-3', while the actual pressure is shown by the process line 2'-3'_R - The gas is heated from 2 to only 3_R, not 3, due to incomplete regeneration in the regenerator. This incomplete regeneration or heat loss is accounted for through use of the regenerator heat loss coefficient X, defined by Feidt(1987)as:

$$X = \frac{Q_{R,id} - Q_{R,lost}}{Q_{R,id}} = \frac{Q_{23} - Q_{3R3}}{Q_{23}} = 1 - \frac{mc_v(T_3 - T_{3R})}{mc_v(T_3 - T_2)} \tag{1}$$

$$X = 1 - \frac{T_3 - T_{3R}}{T_3 - T_2} = 1 - \frac{T_4 - T_{1R}}{T_4 - T_1} \tag{2}$$

The **T-x** diagram in Fig. 1 includes curves showing tire ideal (complete regeneration) and the real (incomplete regeneration) temperature distributions in the gas and the real temperature distribution in the regenerator with the regenerator at R_1 and also with the regenerator at R_2 . On the **T-x** diagram:

$T_{R,ideal,g}$ is the ideal temperature distribution of the gas contained in the regenerator porous matrix, $T_{R,real,g}$ is the real temperature distribution of the gas contained in the regenerator porous matrix, and $T_{R,real}$ is the real temperature distribution in the regenerator matrix.

With this distribution in mind, the **T-x** diagram aids in clarifying the processes taking place in the regenerator during the isometric process 4-1 (ideal) and 4-1_R (real) as shown on the **P-V** diagram. This process takes place with

the working piston stationary at P, while the regenerator moves from R_1 to R_2 . When the regenerator is at R_1 the temperature distribution in the regenerator is shown on the **T-x** diagram. During the process 4-1, the curve CDA traces the ideal temperature variation of the gas, the line CBA traces the real temperature in the gas and the line CR traces die real temperature variation of the matrix material in the regenerator.

After the regenerator moves from R_1 to R_2 the ideal temperature distribution in the gas is shown by curve AD'C, the real temperature in the gas is shown by curve AB'C and the real temperature in the regenerator matrix is shown by curve AR'.

The isothermal expansion process 3-4 take place at temperature (T_H) with the regenerator stationary at R_1 while the working piston moves from P_2 to P_1 .

The **T-x** diagram shows how the temperature of the gas in the regenerator oscillates between the temperatures CBA and AB'C during each complete cycle and how the temperature of the matrix oscillates between temperatures CR and AR'. Using the temperatures illustrated on this diagram and based on a lump sum analysis of the thermal interaction between the gas and the regenerator matrix, equations for X will be formulated. These equations for X, in one case, include worst-case assumptions and hence tend to overestimate losses. In the other case, the equations for X include best-case assumptions arid tend to underestimate these losses. These two equations eventually will be combined into a single equation for X by bringing the losses calculated analytically into consistency with experimental data (Allen 1987, Fujii, 1990; Farrell, 1988; Geng, 1987, Stine and Diver, 1994).

3. Regenerative Loss, X, Based on High Loss Assumptions

In a lump sum analysis of the heat transfer processes between the gas and the metal matrix in the regenerator, the First Law may be written as follows:

Matrix Internal Energy Variation = Gas Internal Energy Variation = Heat transfer between the Gas and Matrix

Upon substitution for the general terms in the above relationship:

$$m_R c_R (T_R - T_{R,i}) = m_g c_{v,g} (T_g - T_{g,i}) = \int h A_R (T_g - T_R) dt \quad (3)$$

Solving for T_R and substituting for Min Eq. (3), gives:

$$T_R = T_{R,i} + M (T_{g,i} - T_g) \quad (4)$$

with:

$$M = \frac{m_g c_{v,g}}{m_R c_R} \quad (5)$$

where m_g is the mass of the gas passing through the regenerator, m_R is the mass of the porous matrix in the regenerator; and A_R is the heat transfer surface area of the wires in the regenerator:

$$m_R = \frac{\pi \cdot 2D_R \cdot 2dL\rho_R}{16(b+d)} \quad (6)$$

$$A_R = \frac{\pi \cdot 2D_R \cdot 2L}{4(b+d)} \quad (7)$$

Upon substitution of the above into Eq. (3) and Eq. (4) and after rearrangement

$$\frac{-dT_g}{T_g(1+M) - T_{R,i} - MT_{g,i}} = \frac{hA_R}{m_g c_{v,g}} dt \quad (8)$$

This differential equation is then integrated and solved for the temperature at 1_R :

$$T_{1_R} = \frac{(T_H - T_L) \exp\left(\frac{-B}{2}\right) + \frac{T_H + T_L}{2} + MT_H}{1+M} \quad (9)$$

where

$$B = (1+M) \frac{hA_R}{m_g c_{v,g}} \frac{S}{w} \quad (10)$$

The temperature T_{1_R} is then substituted in Eq. (2), the expression for X, This expression for X, based on a lump sum analysis of the regenerator processes, predicts relatively high losses and is designated X_1 . It may be referred to as a "pessimistic" expression for X.

$$X_1 = \frac{1+2M+e^{-B}}{2(1+M)} \quad (11)$$

The values of X_1 were found to be less than 0.5 in a first approximation.

4. Regenerative loss, X, Based on Low Loss Assumptions

In an analysis predicting low losses, the gas is hypothesized to be permanently in contact with the heater at point A and also in permanent contact with the cooler at point B on the **T-x** diagram of Fig. 1. Under these conditions and with the regenerator at R_1 , the real temperature of the gas in the regenerator is represented by the line ABC. After the regenerator moves to R_2 , the temperature profile in the regenerator is represented by the line CBA. During operation of the engine, the temperature of the gas in the regenerator oscillates between ABC and CBA with each cycle. By comparison, the ideal temperature of the gas would oscillate between lines ADC and CD A if there were no losses.

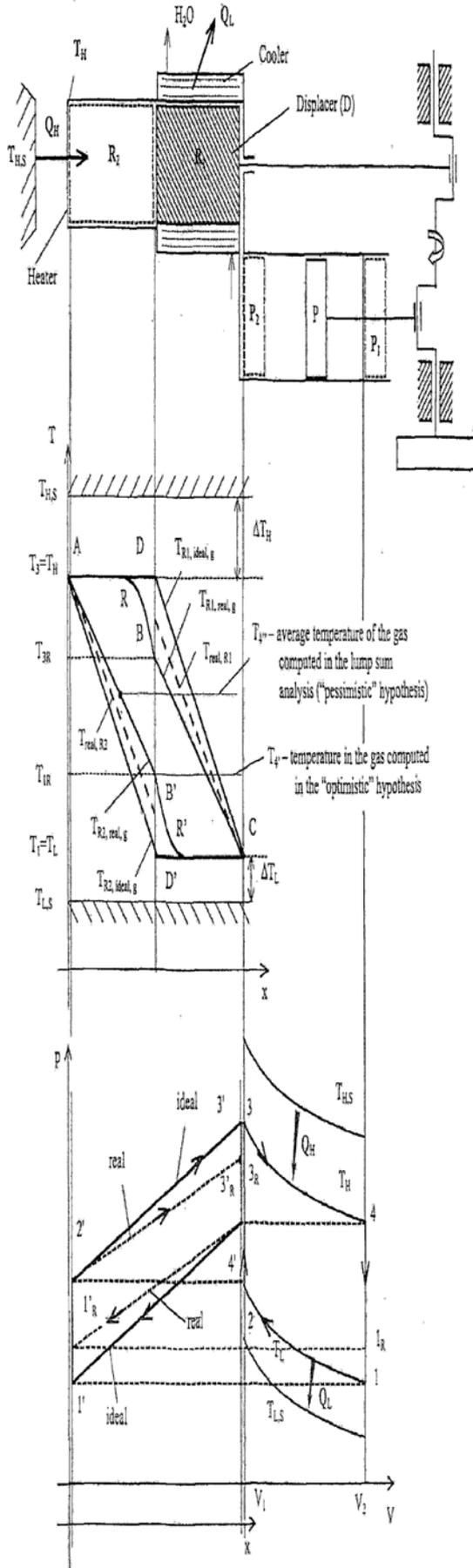


Figure 1. P-V/P-x/T-x diagram of a Stirling engine

In formulating the low loss equation for X, the temperature T_4'' shown on the T-x diagram of Fig.1 is considered equal to the temperature T_{1R} which was obtained in the "high loss" or "pessimistic" analysis given in Eq. (9). Based on the above hypothesis the temperature T_4 is as follows:

$$T_4 = T_H - 2(T_H - T_4'') = \frac{(T_H - T_L)\exp(-B) + T_L + MT_H}{1 + M} \tag{12}$$

With die temperature T_4' assumed equal with T_{1R} , substitution into Eq. (2) for X and letting X_2 equal X for condition of low regenerator losses, we get

$$X_2 = \frac{M + e^{-B}}{1 + M} \tag{13}$$

Eq. (13) represents the low losses or "optimistic" equation for X.

5. The Convection Coefficient in the Regenerator

The Incropera and De Witt correlation formula (Incropera and De Wirt, 1996) for analysis of the heat transfer processes between the gas and porous matrix in the regenerator results in the following equations:

$$St \cdot Pr^{2/3} = \frac{0.79}{p \cdot R_e^{0.576}} \tag{14}$$

where:

$$R_e = \frac{wD}{\nu} \tag{15}$$

$$St = \frac{h}{\rho w c_p} \tag{16}$$

$$Pr = \frac{\nu}{a} = \frac{\rho c_p \nu}{k} \tag{17}$$

The porosity of the regenerator matrix, p , is given by:

$$p = \frac{\text{volume of the gas in pores}}{\text{total volume of the regenerator}} \quad (18)$$

A regenerator made up from N pressed screens having d , diameter of the wire and b , distance between wires has a porosity:

$$p = 1 - \frac{\pi d}{4(d+b)} \quad (19)$$

The average density of the gas is

$$\rho_m = \frac{P_m}{RT_m} \quad (20)$$

where

$$T_m = \frac{T_1(\tau+1)}{2}, P_m = \frac{(\varepsilon+1)(\tau+1)P_1}{4}, \tau = \frac{T_{H,g}}{T_{L,g}} \quad (21)$$

Replacing St from Eq. (16), Re from Eq. (15), the average gas density from Eq. (20) and the porosity from Eq. (19) in Eq. (14) results in an expression for h , the convection heat transfer coefficient in the porous medium in the regenerator:

$$h = \frac{0.395(4P_m / RT_L)w_g^{0.424} c_p(T_m) \nu(T_m)^{0.576}}{(1+\tau) \left[1 - \frac{\pi}{4[(b/d)+1]} \right] D_R^{0.576} \cdot Pr^{2/3}} \quad (22)$$

where c_p , ν and Pr are computed at the average temperature T_m .

Eq. (22) shows that the convection coefficient is influenced by the average pressure of the gas in the engine P_m , low temperature in the gas T_L , ratio of temperatures τ , speed of the gas w_g , gas properties, c_p , ν and Pr , and also by constructive parameters (b , d , D_R). This in turn affects the engine efficiency and the power output through X .

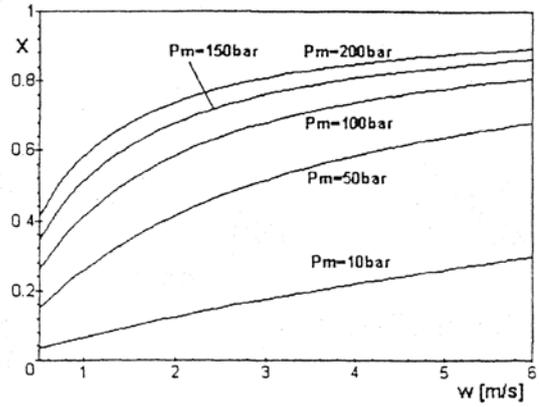


Figure 2. The coefficient of regenerative losses as a function of piston speed for different average working gas pressures ($D_C=80\text{mm}$, $D_R=70\text{mm}$, $b/d=1.5$, $d=0.05\text{mm}$, $S=30\text{mm}$, $\tau=2$,

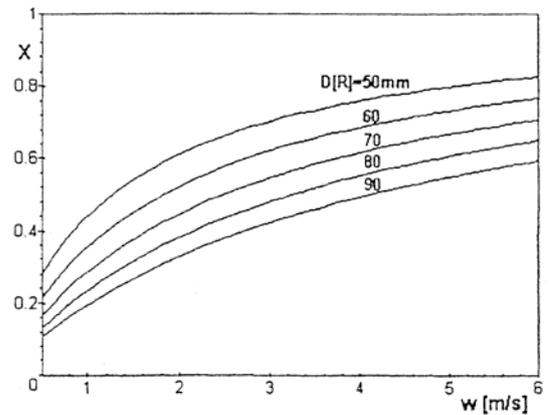


Figure 3. The coefficient of regenerative losses as a function of piston speed for different values of the regenerator diameter ($D_C=80\text{mm}$, $P_m=50\text{bar}$, $b/d=1.5$, $d=0.05\text{mm}$, $S=30\text{mm}$, $\tau=2$, $N=700$).

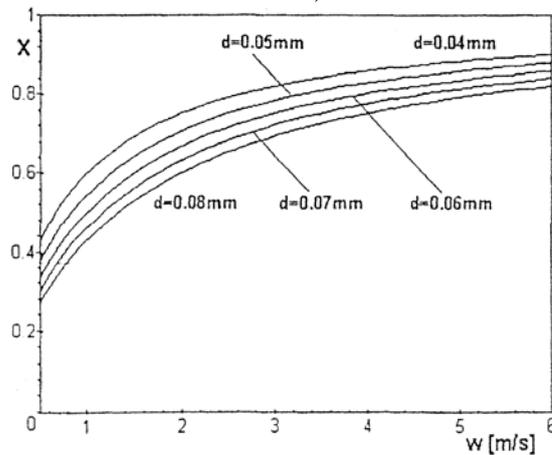


Figure 4. The coefficient of regenerative losses as a function of piston speed for different values of the regenerator wire diameter ($D_C=80\text{mm}$, $D_R=70\text{mm}$,

$P_m=50\text{bar}$, $b/d=1.5$, $S=30\text{mm}$, $\tau=2$,
 $N=700$).

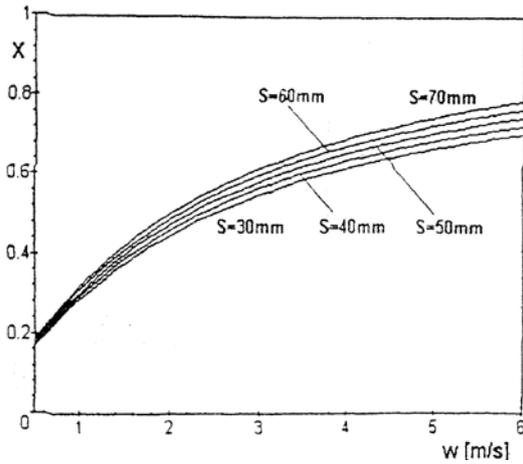


Figure 5. The coefficient of regenerative losses as a function of piston speed for piston strokes ($D_C=80\text{mm}$, $D_R=70\text{mm}$, $b/d=1.5$, $d=0.05\text{mm}$, $P_m=50\text{bar}$, $\tau=2$, $N=700$).

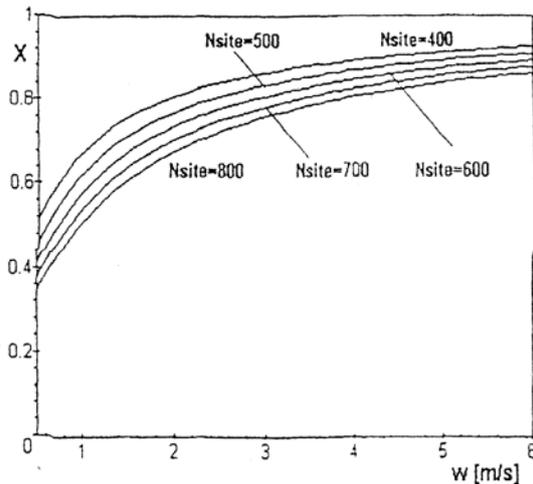


Figure 6. The coefficient of regenerative losses as a function of piston speed for various number of matrix wires ($D_C=80\text{mm}$, $D_R=70\text{mm}$, $b/d=1.5$, $P_m=150\text{bar}$, $S=50\text{mm}$, $\tau=2$, $N=700$).

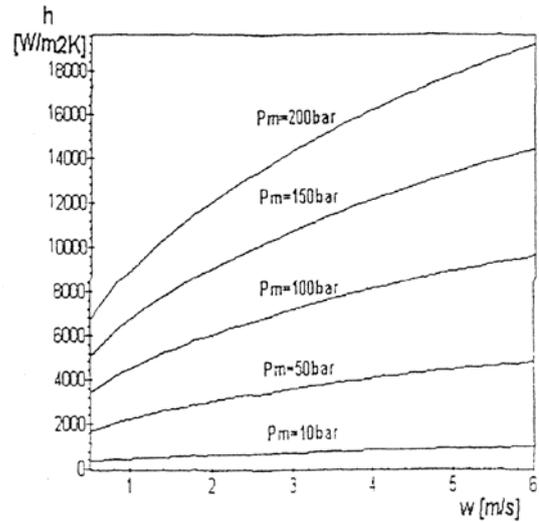


Figure 7. The convective heat transfer coefficient in the regenerator as a function of piston speed for several values of the medium pressure of the working gas ($D_R=70\text{mm}$, $b/d=1.5$, $\tau=2$).

6. The Validation of the Final Formula for X, based on comparison with the experimental data

The effect on X_1 (Eq. 11) and X_2 (Eq. 13) of the operating variables such as piston speed and all the other parameters and properties of the gas, cycle and regenerator was determined. The computed values of X_1 and X_2 were found to accurately predict the values of X determined from experimental data available in the literature (Allen and Tomazic 1987, Farrell 1988, Fujii 1990, Geng 1987, Stine and Diver 1994) using the following equation:

$$X = yX_1 + (1 - y)X_2 \quad (23)$$

where the adjusting parameter y is equal to 0.72.

The losses due to incomplete regeneration X as indicated in Eq. (23) cause a decrease in the efficiency of the Stirling engine through The Second Law Efficiency factor (Petrescu et al. 1999, 2000a, 2000b, 2000c, Florea 1999):

$$\eta_{II,ir,X} = \left[1 + \frac{[X_1 y + X_2 (1-y)] (1 - \sqrt{T_L/T_{H,S}})}{R/c_v(T) \ln \varepsilon} \right]^{-1} \quad (24)$$

which enters in the total efficiency of the Stirling engine:

$$\eta_{SE} = \eta_{CC} \cdot \eta_{II,ir} = \left(1 - \frac{T_L}{T_{H,S}} \right) \left[1 + \frac{T_L}{T_{H,S}} \right]^{-1} \left[1 + \frac{X(1 - \sqrt{T_L/T_{H,S}})}{R/c_v(T) \ln \varepsilon} \right]^{-1} \cdot \eta_{II,ir,\Sigma \Delta P_i} \quad (25)$$

where:

$$\eta_{II,ir,\Sigma \Delta P_i} = 1 - \left[\frac{w}{w_{S,L}} \right] \gamma \cdot (1 + \sqrt{\tau}) \ln \varepsilon + 5 \left[\frac{w}{w_{S,L}} \right]^2 N + \frac{3(0.94 + 0.045w) 10^5}{4} \cdot (\tau \eta \ln \varepsilon) \quad (26)$$

Fig. 2 shows the variation of the regenerative loss coefficient, X , versus the piston speed for several values of the average pressure of the working gas. Figs. 3-6 present the variation of the coefficient of regenerative losses as a function of the piston speed for several values of the analysis parameters (D_R -regenerator diameter, d -wire diameter, S -stroke, N -number of gauzes).

Fig. 7 illustrates the convection heat transfer coefficient dependence on piston speed, for several values of the average pressure of the working gas.

The final analytic expression for the power output of actual Stirling engines is:

$$Power_{SE} = \eta_{SE} \cdot z m R T_{H,g} (w/2S) \ln \varepsilon \quad (27)$$

This expression accurately predicts the power output of twelve Stirling engines under 16 operating conditions (Allen 1987, Fujii 1990, Farell 1988, Geng 1987, Stine and Diver 1994) when the adjusting parameter z equals 0.8.

7. Discussion

The operating processes of the Stirling engines are presented in a comprehensive yet intuitive manner using a **P-V** and **P-x** coordinate system (Petrescu et al. 1999, 2000a, 2000b, 2000c, - Florea 1999) to which the **T-x** diagram has been added. The addition of the **T-x** diagram to Fig. 1 aids in understanding and explaining the processes in the regenerator. An analysis and a technique for calculating of the regenerator losses that occur in Stirling engines and a method of predicting the effect of regenerative losses on engine performance is presented. This analysis provides insight into the processes involved since it utilizes the First Law of Thermodynamics for Processes with Finite

Speed (Stoicescu and Petrescu 1964a, 1964b, 1965a, 1965b, 1965c, Petrescu 1969a, 1971, 1974, 1991, Petrescu et al. 1992, Petrescu and Stanescu 1993, Petrescu and Harman 1994, Petrescu et al. 1996b) in conjunction with the Direct Method. (Stoicescu and Petrescu 1964a, 1964b, 1965c, Petrescu 1969, 1991, Stanescu 1992, Petrescu and Stanescu 1993, Petrescu et al. 1993a, 1993b, Petrescu and Harman 1994, Petrescu et al. 1994, 1996a, 1996b, Costea 1997, Costea et al. 1999, Florea 1999, Petrescu et al. 1999, 2000a, 2000b, 2000c, 2000d). The method simulates the operation of actual Stirling engines while intuitively suggesting the mechanisms that generate the irreversibilities. A method for calculating these losses based on actual regenerator components is presented in Eq. (23) for the regenerator loss coefficient X . This relationship was validated by comparison with experimental data on known components and engine configurations (Allen 1987, Fujii 1990, Farell 1988, Geng 1987, Stine and Diver 1994). The results of computations based on this analysis are compared to performance data taken from

a number of operating Stirling engines in Figs. (8-9) and Table 1.

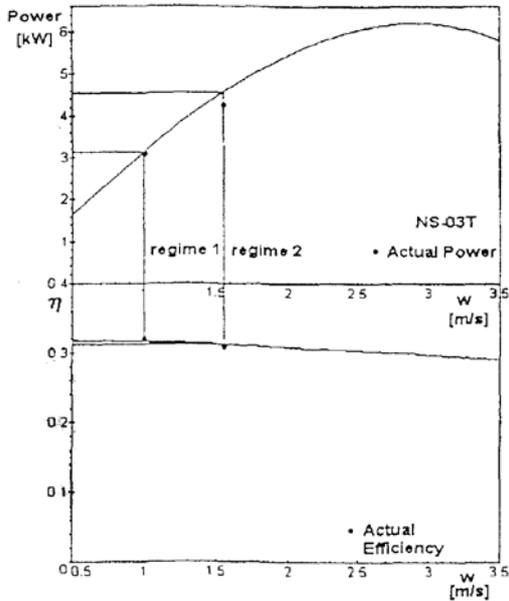


Figure 8. Comparison of the analysis results with actual performance data for the NS-03T Stirling engine.

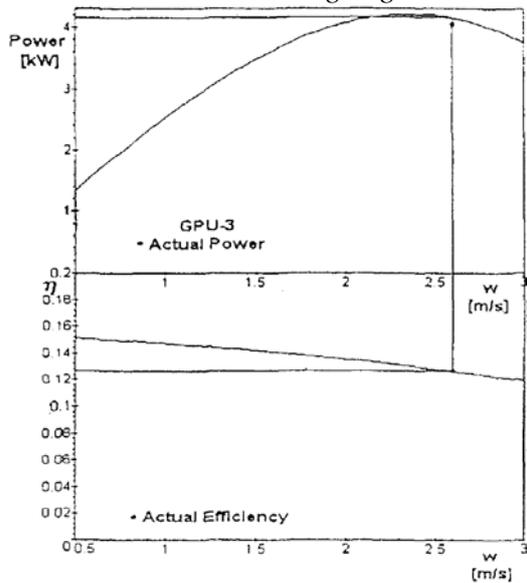


Figure 9. Comparison of the analysis results with actual performance data for the GPU-3 Stirling engine.

Table 1 Comparison between analytical results and actual engine performance data (Fujii, 1990; Allen and Tomazic, 1987; Geng, 1987; Stine and Diver, 1994; Farrell et al., 1988; Organ, 1992)

Stirling engine	Actual Power [kW]	Calculated Power [kW]	Actual Efficiency [-]	Calculated Efficiency [-]
NS-03M, regime 1 (max. efficiency)	2.03	2.182	0.359	0.339
NS-03M, regime 2 (max. power)	3.81	4.196	0.310	0.329
NS-30A, regime 1 (max. efficiency)	23.20	29.450	0.375	0.357
NS-30A, regime 2 (max. power)	30.40	33.820	0.330	CB-36
NS-30S, regime 1 (max. efficiency)	30.90	33.780	0.372	0.366
NS-30S, regime 2 (max. power)	45.60	45.620	0.352	0.352
STM4-120	25	26.360	0.400	0.401

Y-160	9	8.825	0.300	0.30S
4-95 MKII	25	28.400	0.294	0.289
4-275	50	48.610	0.420	0.412
MP 1002 CA	200 W	193.900 W	0.156	0.153
Free Piston Stirling engine	9	9.165	0.330	0.331
RE- 1000	0.93 9	1.005	0.258	0.228

8. Conclusion

The high degree of correlation between the analytic and the operational data shown in Figures (8-9) and Table 1 indicates that the analysis presented using the concept of a coefficient of regenerator losses X accurately predicts the performance in terms of power and efficiency for Stirling engines over the range of conditions. This capability has value in the design of new Stirling engines and in predicting the power and efficiency of a particular Stirling engine.

Nomenclature

A	area, m^2
b	distance between the regenerator matrix wire, m
CC	Carnot Cycle
c_p, c_v	specific heats, $J\ kg^{-1}\ K^{-1}$
D	diameter, m
d	wire diameter, m
h	heat transfer coefficient, $W\ m^{-2}\ K^{-1}$
k	gas thermal conductivity, $W\ m^{-1}\ K^{-1}$
m	mass, kg
N	number of gauzes
P	pressure, Pa
p	porosity
Pr	Prandtl number
Q	heat, J
R	gas constant, $J\ kg^{-1}\ K^{-1}$
Re	Reynolds number

S	stroke, m
St	Stanton number
T	temperature, K
T	time, s
U	internal energy, J
V	volume, m^3
W	work, J
w	piston speed, $m\ s^{-1}$
w_R	gas speed in the regenerator, $m\ s^{-1}$
w_{SL}	speed of the sound at temperature T_L , ($m\ s^{-1}$)
X	regenerative losses coefficient
y	adjusting parameter
z	adjusting parameter

Greek symbols

ϵ	compression ratio (V_1/V_2)
γ	ratio of the specific heats
η	first law efficiency
η_{II}	second law efficiency
ν	viscosity of the gas, $m^2\ s^{-1}$
τ	ratio of the extreme temperatures

Subscripts

g	gas
H,g	gas, at the hot-end of the engine
H,S	source, at the hot-end of the engine
i	instantaneous
L	sink
m	average
R	regenerator
S,L	sink, at the cold-end of the engine
SE	Stirling engine

References

1. Allen, D.I. and Tomazic W.A., 1937, "Hot Piston Ring Tests", *NASA TM-100256*.
2. Farell, R.A. et al., 1988, "Automotive Stirling engine Development Program", *NASA CR-180839*.

3. Florea, T., 1999, "Grapho-Analytical Method for the study of the operating processes irreversibility in Stirling "engines", Ph.D. Thesis, P.U. Bucharest.
4. Fujii, I., 1990, *From Solar Energy to Mechanical Power*, Harwood Academic Publishers, New York.
5. Geng, S.M., 1987, "Calibration and Comparison of the NASA Lewis Free-Pistons Stirling engine Model Prediction with RE-1000 Test Data", *NASA TM-89853*.
6. Organ, J.A., 1992, *Thermodynamics and Gas Dynamics of Stirling Cycle Machine*, Cambridge University Press, Cambridge.
7. Petrescu, S., Florea, T, Costea, M., Harman, C, 1999, "Finite Speed Thermodynamics Applied to Stirling engines", *Proceedings of BIRAC Conference*, Bucharest, Romania, pp.35-43.
8. Petrescu, S., Harman, C, Costea, M., Florea, T., 2000a, "A Method for Determination of the Performances of Stirling Machines Based on a P-V/P-x Diagram and First Law for Processes with Finite Speed", *Proceedings of the Fifth World Conference on Integrated Design & Process Technology*, Dallas, USA.
9. Petrescu, S., Florea, T., Harman, C, Costea, M., 2000b, "A Method for Calculating the Coefficient for the Regenerative Losses in Stirling Machines", *Proceedings of the 5th European Stirling Forum 2000*, Osnabriick, Germany, pp. 121-129.
10. Petrescu, S., Costea, M., Harman, C, Florea, T., 2000c, "Detenmnation of the Pressure Losses in a Stirling Cycle through Use of a P-V/P-x Diagram", *Proceedings of ECOS'2000*, GG. Hirs, ed., Enschede, Netherlands, pp. 659-670.
11. Petrescu, S., Feidt, M., Costea, M., Florea, T., 2000d, "Stirling Refrigerator and Heat Pumps Cycles with Finite Speed", *Proceedings of the 10th National Conference on Thermodynamics*, Sibiu, Romania, Vol. II, pp. 178-183.

Authors Index

Aberomand, N.	213	Grisel, Y.	26	Quellec, G.	216
Albrand, M.	26	Iakovakis, V.	151	Ramírez, S. M.	207
Antonidakis, E. N.	162	Ibrahimov, V.	189	Revesz, P. Z.	21, 40, 53
Batzias, D.	58, 80	Imanova, M.	189	Revesz, P. Z.	75, 89, 101
Batzias, F. A.	31, 58, 80	Jai-Andaloussi, S.	216	Rmanan, J.	75
Belhabi, F.	230	Jakóbczak, D. J.	69	Rodionov, A. S.	226
Bhattacharyya, B. K.	221	Kamarinopoulos, L.	80	Rodionova, O.	226
Brünger, J.	45	Kechagias, J.	110, 151	Safarik, I.	58
Cazuguel, G.	216	Kitsakis, K.	110, 151	Sarma, K. K.	221
Cervantes, R. F. L.	95	Klysz, G.	26	Sarmah, H. K.	221
Chaparro, A. R.	200	Koch, R.	45	Sauer, P.	95
Chatzakis, J.	179	Konstantaras, A. J.	162	Sekkaki, A.	216
Coteli, R.	143	Kyratsis, P.	110	Sidiras, D.	58, 80
Daggumati, S.	89	Lamad, M.	216	Siontorou, C.	58, 80
Dobrescu, R.	128, 158	László, I.	147	Smarandache, I. D.	174
Dumitrescu, T.	128, 158	Li, Z.	40	Stoian, V.	194
Echeverri, M. A.	200	Lietaer, B.	115	Svehla, C.	89
Edemskiy, V.	171	Lobaina, A. L.	207	Szklanny, K.	106, 124, 184
El Fahssi, K.	216	Machek, O.	166	Traulsen, I.	45
Ellis, D. M.	207	Mastorakis, N. E.	110, 151, 221	Tsapatsis, M.	58
Elmoufidi, A.	216	Mehdiyeva, G.	189	Vargas, J. D. L.	200
Enache, M. A.	174	Moza, Z.	151	Vlad, I.	174
Enache, S.	174	Muradova, A. D.	138	Vladu, II.	194
Ettaouil, M.	230	Oita, C.	240	Vladu, Io. C.	194
Ferrieres, X.	26	Palau, I. G.	207	Votavová, P.	166
Florea, T.	240	Petrakis, N. S.	162	Wichrowski, M.	124, 184
Florea, T. V.	240	Pokorny, P.	134	Wieczorkowska, A.	106, 124, 184
Frantzeskakis, T. S.	162	Pollalis, Y.	80	Zamora, M. A. P.	200
Fülep, D.	147	Popescu, R.	128, 158	Zsoldos, I.	147