MATHEMATICAL METHODS in ENGINEERING and ECONOMICS

Proceedings of the 2014 International Conference on Applied Mathematics and Computational Methods in Engineering II (AMCME '14)

Proceedings of the 2014 International Conference on Economics and Business Administration II (EBA '14)

Prague, Czech Republic
April 2-4, 2014

MATHEMATICAL METHODS in ENGINEERING and ECONOMICS

Proceedings of the 2014 International Conference on Applied Mathematics and Computational Methods in Engineering II (AMCME '14)

Proceedings of the 2014 International Conference on Economics and Business Administration II (EBA '14)

Prague, Czech Republic
April 2-4, 2014

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.

ISBN: 978-1-61804-230-9

MATHEMATICAL METHODS in ENGINEERING and ECONOMICS

Proceedings of the 2014 International Conference on Applied Mathematics and Computational Methods in Engineering II (AMCME '14)

Proceedings of the 2014 International Conference on Economics and Business Administration II (EBA '14)

Prague, Czech Republic
April 2-4, 2014

Organizing Committee

General Chairs (EDITORS)

- Professor H. M. Srivastava
  Professor Emeritus
  Department of Mathematics and Statistics
  University of Victoria
  Victoria, British Columbia V8W 3R4
  Canada
- Professor Martin Bohner,
  Missouri University of Science and
  Technology, Rolla, Missouri, USA
- Professor Ivan G. Avramidi,
  New Mexico Tech, Socorro,
  New Mexico, USA
- Professor Martin Schechter,
  University of California, Irvine, USA
- Prof. Morris Adelman,
  Professor of Economics,
  Emeritus, MIT, USA

Senior Program Chairs

- Professor Michel Chipot,
  University of Zurich, Zurich,
  Switzerland
- Professor Gen Qi Xu
  Department of Mathematics,
  Tianjin University, Tianjin, China
- Professor Ji Gao
  Community College of Philadelphia
  1700 Spring Garden Street
  Philadelphia, PA 19130-3991
  USA
- Professor S. P. Yung
  Department of Mathematics
  The University of Hong Kong
  Hong Kong

Program Chairs

- Professor Xiaodong Yan,
  University of Connecticut,
  Connecticut USA
- Professor Yushun Wang,
  Nanjing Normal university,
  Nanjing, China
  Idaho State University, USA
- Professor Ravi P. Agarwal,
  Texas A&M University - Kingsville,
  Kingsville, TX, USA

Tutorials Chair

- Professor Detlev Buchholz,
  Universitaet Goettingen,
  Goettingen, Germany

Special Session Chair

- Professor Patricia J. Y. Wong,
  Nanyang Technological University,
  Singapore

Workshops Chair

- Professor Jim Zhu,
  Western Michigan University,
  Kalamazoo, MI, USA

Local Organizing Chair

- Assistant Prof. Klimis Ntalianis,
  Tech. Educ. Inst. of Athens (TEI),
  Athens, Greece

Publication Chair

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

Publicity Committee

- Professor Gerd Teschke,
  Institute for Computational
  Mathematics in Science and Technology,
  Neubrandenburg, Berlin-Dahlem, Germany
- Professor Lucio Boccardo,
  Universita degli Studi di
  Roma "La Sapienza",
  Roma, Italy

International Liaisons

- Professor Shanhe Wu,
  Longyan University,
  Longyan, Fujian, China
- Professor Natig M. Atakishiyev,
  National Autonomous University of Mexico,
  Mexico
- Professor Imre Rudas,
  Obuda University,
  Hungary
- Professor Jianming Zhan,
  Hubei University for Nationalities,
  Enshi, Hubei Province, China

Steering Committee

- Prof. Ferhan M. Atici, Western Kentucky University, Bowling Green, KY 42101, USA
- Prof. Ravi P. Agarwal, Texas A&M University - Kingsville, Kingsville, TX, USA
- Prof. Martin Bohner, Missouri University of Science and Technology, Rolla, Missouri, USA
- Prof. M. Affan Badar, Indiana State University, Terre Haute, Indiana, USA
- Prof. Dashan Fan, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
- Prof. Martin Bohner, Missouri University of Science and Technology, Rolla, Missouri, USA
- Prof. Ravi P. Agarwal, Texas A&M University-Kingsville, Kingsville, Texas, USA

Program Committee

Prof. Ming Mei, McGill University, Montreal, Quebec, Canada
Prof. Andrew Pickering, Universidad Rey Juan Carlos, Spain
Prof. Jiri Hrebicek, Masaryk University, Brno, Czech Republic
Prof. Angelo Favini, Universita di Bologna, Bologna, Italy
Prof. Yuriy Rogovchenko, University of Agder, Kristiansand and Grimstad, Norway
Prof. Maria Alessandra Ragusa, Universita di Catania, Catania, Italy
Prof. Feliz Minhos, Universidade de Evora, Evora, Portugal
Prof. Julian Lopez-Gomez, Universidad Complutense de Madrid, Madrid, Spain
Prof. Stanislaw Migorski, Jagiellonian University in Krakow, Krakow, Poland
Prof. Simeon Reich, The Technion - Israel Institute of Technology, Haifa, Israel
Prof. Kevin Kam Fung Yuen, Xi?an Jiaotong-Liverpool University, China
Prof. Yansheng Liu, Shandong Normal University, Jinan, Shandong, China
Prof. Kailash C. Patidar, University of the Western Cape, Bellville, South Africa
Prof. Wei-Shih Du, National Kaohsiung Normal University, Kaohsiung City, Taiwan
Prof. Ahmed El-Sayed, Alexandria University, Alexandria, Egypt
Prof. Valery Y. Glizer, Department of Mathematics, ORT Braude College, Karmiel, Israel
Prof. Ivan Ganchev Ivanov, Sofia University "St. Kl. Ohridski", Sofia, Bulgaria
Prof. Lucas Jodar, Universitat Politecnica de Valencia, Valencia, Spain
Prof. Ming-Yi Lee, National Central University, Taiwan
Prof. Carlos Lizama, Universidad de Santiago de Chile, Santiago, Chile
Prof. Juan Carlos Cortes Lopez, Universidad Politecnica de Valencia, Spain
Prof. Khalil Ezzinbi, Universite Cadi Ayyad, Marrakesh, Morocco
Prof. Mitsuharu Otani, Waseda University, Japan
Prof. Luigi Rodino, University of Torino, Torino, Italy
Prof. Narcisa C. Apreutesei, Technical University of Iasi, Iasi, Romania
Prof. Sining Zheng, Dalian University of Technology, Dalian, China
Prof. Stevo Stevic, Mathematical Institute Serbian Academy of Sciences and Arts, Beograd, Serbia
Prof. Daoyi Xu, Sichuan University, Chengdu, China
Prof. Junmin Wang, Beijing Institute of Technology, Beijing, China
Prof. Elsayed M. E. Zayed, Faculty of Science, Zagazig University, Zagazig, Egypt
Prof. Jinde Cao, Southeast University/ King Abdulaziz University, China
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. Satit Saejung, Khon Kaen University, Muang District, Khon Kaen, Thailand
Prof. Juan J. Trujillo, Universidad de La Laguna, La Laguna, Tenerife, Spain
Prof. Tiecheng Xia, Department of Mathematics, Shanghai University, China
Prof. Noemi Wolanski, Universidad de Buenos Aires, Buenos Aires, Argentina
Prof. Zhenya Yan, Chinese Academy of Sciences, Beijing, China
Prof. Juan Carlos Cortes Lopez, Universidad Politecnica de Valencia, Spain
Prof. Zili Wu, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China
Prof. Wei-Shih Du, National Kaohsiung Normal University, Kaohsiung City, Taiwan
Prof. Chun-Gang Zhu, Dalian University of Technology, Dalian, China

Additional Reviewers

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

Table of Contents

ISBN: 978-1-61804-230-9

Keynote Lecture 1

Interpolation and Projective Representation in Computer Graphics, Visualization and Games


Vaclav Skala
University of West Bohemia
Plzen, Czech Republic
E-mail: skala@kiv.zcu.cz

Rongjiang Pan
Shandong University
Jinan, China
E-mail: panrj@sdu.edu.cn

Abstract: Today's engineering problem solutions are based mostly on computational packages. However the computational power doubles in 18 months. In 15 years perspective the computational power will be of 2^10 = 1024 of today's computational power. Engineering problems solved will be more complicated, complex and will lead to a numerically ill conditioned problems especially in the perspective of today available floating point representation and formulation in the Euclidean space.
Homogeneous coordinates and projective geometry are mostly connected with geometric transformations only. However the projective extension of the Euclidean system allows reformulation of geometrical problems which can be easily solved. In many cases quite complicated formulae are becoming simple from the geometrical and computational point of view. In addition they lead to simple parallelization and to matrix-vector operations which are convenient for matrix-vector hardware architecture like GPU.
In this short tutorial we will introduce "practical theory" of the projective space and homogeneous coordinates. We will show that a solution of linear system of equations is equivalent to generalized cross product and how this influences basic geometrical algorithms. The projective formulation is also convenient for computation of barycentric coordinates, as it is actually one cross-product implemented as one clock instruction on GPU. Selected examples of engineering disasters caused by non-robust computations will be presented as well.

Brief Biography of the Speaker: Prof.Vaclav Skala is a Full professor of Computer Science at the University of West Bohemia, Plzen, Czech Republic. He received his Ing. (equivalent of MSc.) degree in 1975 from the Institute of Technology in Plzen and CSc. (equivalent of Ph.D.) degree from the Czech Technical University in Prague in 1981. In 1996 he became a full professor in Computer Science. He is the Head of the Center of Computer Graphics and Visualization at the University of West Bohemia in Plzen (http://Graphics.zcu.cz) since 1996.
Prof.Vaclav Skala is a member of editorial board of The Visual Computer (Springer), Computers and Graphics (Elsevier), Machine Graphics and Vision (Polish Academy of Sciences), The International Journal of Virtual Reality (IPI Press, USA) and the Editor in Chief of the Journal of WSCG. He has been a member of several international program committees of prestigious conferences and workshops. He is a member of ACM SIGGRAPH, IEEE and Eurographics Association. He became a Fellow of the Eurographics Association in 2010.

ISBN: 978-1-61804-230-9

Prof.Vaclav Skala has published over 200 research papers in scientific journal and at international research conferences. His current research interests are computer graphics, visualization and mathematics, especially geometrical algebra, algorithms and data structures. Details can be found at http://www.VaclavSkala.eu

Prof. Rongjiang Pan is a professor in the School of Computer Science and Technology, Shandong University, China. He received a BSc in computer science, a Msc in computer science, a PhD in computer science from Shandong University, China in 1996, 2001 and 2005, respectively. During 2006 and 2007, he was a visiting scholar at the University of West Bohemia in Plzen under a program supported by the international exchange scholarship between China and Czech governments. He is now a visiting professor at the School of Engineering, Brown University from 2014 to 2105 under the support of China Scholarship Council.

He is a Member of the ACM. His research interests include 3D shape modeling and analysis, computer graphics and vision, image processing. He has published over 20 research papers in journal and at conferences

ISBN: 978-1-61804-230-9

Some further results on weak proximal contractions including the case of iteration-dependent image sets

M. De la Sen

***Abstract*—** This paper presents some further results on weak proximal contractions in metric spaces by discussing some restrictions of parametrical type on the parameters which define the contractive conditions which allow guaranteeing that the implying condition of the proximal contraction characterization holds for the studied classes of proximal sequences so that it can be removed from the analysis. Some related generalizations are also given for non self-mappings of the form $T : A \to (\cup B_n)$, such that $D_n = d(A, B_n)$, where A and B_n are nonempty subsets of a metric space (X, d) for $n \in \mathbf{Z}_{0+}$, provided that the set theoretic limit of the sequence of sets $\{B_n\}$ exists as $n \to \infty$. The convergence of the sequences in the domain and the image sets of the non self-mapping as well as the existence and uniqueness of the best proximity points are also investigated.

***Keywords*—**Best proximity point, proximal contraction, weak proximal contractionset theoretic limit.

I. INTRODUCTION

THIS paper is devoted to formulate and prove some further results for a general class of weak proximal contractions [3] in metric spaces which are subject to certain parametrical restrictions on the contractive conditions which allow guaranteeing that the implying condition of the contraction holds for the proximal sequences so that it can be removed from the analysis, [1-4]. See also [5-8] on fixed point theory and some of its applications to stability problems and [9-10] on characterizations of stability of dynamic systems. Some related generalizations are also given for non self-mappings $T : A \to (\cup B_n)$, subject to a set distance $D_n = d(A, B_n)$, where A and B_n are nonempty and closed subsets of a metric space (X, d) for $n \in \mathbf{Z}_{0+}$, provided that the set theoretic limit of the sequence of sets $\{B_n\}$ exists as $n \to \infty$. The convergence of the sequences in the domain and the image sets of the non self-mapping as well as the existence and uniqueness of the best proximity points are also investigated for the different restrictions and the given extension.

This work was supported in part by Spanish Government for its support of this research trough Grant DPI2012-30 651, and to the Basque Government for its support of this research trough Grants IT378-10 and SAIOTEK S-PE12UN015. He is also grateful to the University of Basque Country for its financial support through Grant UFI 2011/07.

M. De la Sen is with Institute of Research and Development of Processes IIDP, Barrio Sarriena, Leioa, Spain (e-mail: manuel.delasen@ehu.es).

II. RESULTS

Firstly, the definition of a so-called generalized class of proximal contractions is recalled for further discussions:

Definition 1. Let (A, B) be nonempty closed subsets of a metric space (X, d). A non self-mapping $T : A \to B$ with $D = d(A, B)$ is a generalized weak proximal contraction if there are real constants $\alpha \in [0, 1)$ and $\beta \in [0, \infty)$ such that for all $x, y, , u, v \in A$ with

$$d(u, Tx) = d(v, Ty) = D \tag{1}$$

one has the following logic implication:

$$d(x, Tx) \le (1 + \alpha + \beta) d(x, y) + D$$
$$\Rightarrow \; d(u, v) \le \alpha d(x, y) + \beta (d(Tx, y) - D) \tag{2}$$

The sets A_0, of best proximity points of $T : A \to B$, and B_0 are defined by.

$$A_0 := \{ x \in A : d(x, y) = D \text{ for some } y \in B \}$$
$$B_0 := \{ y \in B : d(x, y) = D \text{ for some } x \in A \}$$

If (A, B) is a nonempty, weakly compact and convex pair of a Banach space X then A_0 and B_0 are nonempty. The set A is approximatively compact with respect to B if any

ISBN: 978-1-61804-230-9

sequence $\{x_n\} \subset A$ such that $d(x_n , y) \to d(A, y)$ as $n \to \infty$ for some $y \in B$ has a convergent subsequence. If the property holds for $y \in B_0$, and since

$$d(A, y) = d(A, B_0) = d(A_0 , B_0) = D$$

then a sequence $\{x_n\} \subset A$ has a convergent sequence if A is approximatively compact with respect to B and $d(x_n , y) \to D$ as $n \to \infty$ for some $y \in B_0$.

The following result holds from Definition 1:

Proposition 1. Let (A, B) be a pair of nonempty closed subsets of a metric space (X, d). Assume that $T : A \to B$ with $D = d(A,B)$ is a generalized weak proximal contraction such that A_0 is nonempty and $T(A_0) \subseteq B_0$. Construct any sequence $\{x_n\} \subset A$ satisfying for any $x_0 \in A$ and $d(x_{n+1}, Tx_n) = d(Tx_{n+1}, x_{n+2}) = D$; $\forall n \in \mathbf{Z}_{0+}$ and the condition:

$$d(x_{n+1} , x_{n+2}) \le \alpha d(x_n , x_{n-1}) + \beta[d(Tx_n , x_{n+1}) - D]; \forall n \in \mathbf{Z}_{0+} \quad (3)$$

Then, $\{x_n\}$ is bounded, if x_0 is bounded, $d(x_n , x_{n+1}) \to 0$ and $d(x_n, Tx_n) \to D$ as $n \to \infty$ and $\sum_{n=0}^{\infty} d(x_{n+1} , x_n) < \infty$. The condition (3.a) can be removed and the result still holds.

Proof: Note that the given condition (3) follows from (2) with the first condition of the logic implication being true; $\forall n \in \mathbf{Z}_{0+}$ with the replacements $x \to x_n$, $y = u \to x_{n+1}$, $v \to x_{n+2}$. Note that the implying inequality of (2) holds since:

$$\begin{aligned}(1+\alpha+\beta)d(x_n, x_{n+1}) + D &= (1+\alpha+\beta)d(x_n, x_{n+1}) + d(Tx_n , x_{n+1}) \\ &\ge d(x_n, x_{n+1}) + d(Tx_n , x_{n+1}) \ge d(x_n, Tx_n);\end{aligned}$$

$\forall n \in \mathbf{Z}_{0+}$ and the use of the constraint $d(Tx_n , x_{n+1}) = D$; $\forall n \in \mathbf{Z}_{0+}$ in

$$d(x_{n+1} , x_n) \le \alpha d(x_n , x_{n-1}) + \beta[d(Tx_n , x_{n+1}) - D]; \forall n \in \mathbf{Z}_{0+} \quad (4)$$

Note also that $d(x_{n+1} , x_n) \to 0$, then $\{x_n\}$ is bounded if $x_0 \in A$ is bounded, and $d(Tx_n , x_n) \to D$ as $n \to \infty$, since $d(Tx_n , x_{n+1}) = D$; $\forall n \in \mathbf{Z}_{0+}$, and $d(x_{n+1} , x_n) \to 0$ as $n \to \infty$ since

$$\begin{aligned}d(x_{n+1} , x_n) &\le \alpha d(x_n , x_{n-1}) + \beta[d(Tx_n , x_{n+1}) - D] = \alpha d(x_n , x_{n-1}) \\ &\le \alpha^n d(x_1 , x_0) \ ; \ \forall n \in \mathbf{Z}_{+}\end{aligned}$$

$$D \le d(Tx_n , x_n) \le D + d(x_n , x_{n+1}) \ ; \ \forall n \in \mathbf{Z}_{0+}$$

$$\begin{aligned}\sum_{k=0}^{m} d(x_{k+n+1} , x_{k+n}) &= \sum_{k=0}^{\infty} d(x_{k+n+1} , x_{k+n}) - \sum_{k=m+1}^{\infty} d(x_{k+n+1} , x_{k+n}) \\ &\le \frac{1-\alpha^m}{1-\alpha} d(x_n , x_{n+1}) \ ; \ \forall n, m \in \mathbf{Z}_{0+}\end{aligned} \quad (5)$$

$\sum_{n=0}^{\infty} d(x_{n+1} , x_n) \le \dfrac{1}{1-\alpha} d(x_0 , x_1)$. On the other hand since $d(x_{n+1} , Tx_n) = D$,

$$d(x_n, Tx_n) \le d(x_{n+1} , Tx_n) + d(x_n , x_{n+1}) \le (1+\alpha+\beta)d(x_n , x_{n+1}) + D$$

so that the implying condition of (2) always holds and then (3.a) can be removed from this result which directly holds under(3.b). Also, $\{x_n\} \subset A$ with $x_n \in A_0$; $\forall n \in \mathbf{Z}_{+}$ and $\{Tx_n\} \subset A_0 \subseteq A$.

Theorem 1. Under all the assumptions of Proposition 1, assume also that (X, d) is complete and that B is approximatively compact with respect to A, or the weaker condition that A_0 is closed. Then, $T : A \to B$ has a unique best proximity point.

Proof: Since $d(x_{n+1} , x_n) \to 0$ as $n \to \infty$ $\{x_n\} \to x$ from Proposition 1. Since $\{x_n\} \subset A$ and A is closed then $x \in A$. Since $d(x_{n+1} , Tx_n) = d(x, Tx) = D$ and B is approximatively compact with respect to A, the sequence $\{Tx_n\} \subset B_0 \subset B$ has a convergent subsequence $\{Tx_{nk}\} \to z \in B$, since B is also closed. Also, $z \in cl\, B_0$ with $d(x , Tx_{n_k}) \to D$ as $k \to \infty$.

Then, a) $d(x, z) = D$ and $z = Tx$ with $x \in A_0$, since A_0 is closed if B is approximatively compact with respect to A and both A and B are closed and nonempty, and $Tx \in B_0$. Assume not. Then, $\liminf_{n\to\infty} d(x_n , Tx_n) > D$ and taking limits as $n \to \infty$ in the equation:

$$d(x_n , Tx_n) \le d(x_{n+1} , Tx_n) + d(x_n , x_{n+1}) = D + d(x_n , x_{n+1})$$

would imply, since $d(x_{n+1} , x_n) \to 0$ as $n \to \infty$, the contradiction $D > D$. Assume now that there are $x, y (\ne x) \in A_0$ such that $d(x, Tx) = d(y, Ty) = D$. Thus, one gets from (2) for $u = x$ and $y = v$ that for any real constant $\beta \in [0, \infty)$:

$$D = d(x,Tx) = d(y,Ty) \le (1+\alpha+\beta)d(x, y) + D$$

$$\begin{aligned}(1-\alpha)d(x, y) &\le \beta(d(Tx, y) - D) \le \beta d(Tx, x) + \beta d(x, y) - \beta D \\ &= \beta d(x, y)\end{aligned}$$

which fails for $0 \le \beta < 1-\alpha$ if $x \ne y$. Thus, $x = y$.

Also, the following weakened result holds under conditions which guarantee that the implied logic proposition (2) always holds:

Lemma 1. Let (A, B) be a of nonempty closed subsets of a metric space (X, d). Assume that $T : A \to B$ with $D = d(A,B)$ is a generalized weak proximal contraction, such that A_0 is nonempty and $T(A_0) \subseteq B_0$, with $d(u,Tx) = d(v,Ty) = D$. Then,

$$d(u,v) \le (\alpha + \beta(\alpha+\beta))d(x,y) \;\Rightarrow\; d(u,v) \le \alpha d(x,y) + \beta(d(Tx,y) - D) \tag{6}$$

Proof: Since $T : A \to B$ with $D = d(A,B)$ is a generalized weak proximal contraction with $d(u,Tx) = d(v,Ty) = D$ then

$$d(x,Tx) \le (1+\alpha+\beta)d(x,y) + D \;\Rightarrow\; d(u,v) \le \alpha d(x,y) + \beta(d(Tx,y) - D) \tag{7}$$

It remains to be proved that

$$d(x,Tx) > (1+\alpha+\beta)d(x,y) + D \;\Rightarrow\; d(u,v) \le \alpha d(x,y) + \beta(d(Tx,y) - D) \tag{8}$$

Assume that $d(u,v) > \alpha d(x,y) + \beta(d(Tx,y) - D)$ for some $x, y \in A$ such that $d(x,Tx) > (1+\alpha+\beta)d(x,y) + D$. Note that $d(x,Tx) > (1+\alpha+\beta)d(x,y) + D$ is identical to $d(x,y) < \dfrac{d(x,Tx) - D}{1+\alpha+\beta}$ which implies, after using $d(x,Tx) \le d(x,y) + d(y,Tx)$, that $d(y,Tx) > (\alpha+\beta)d(x,y) + D$. Then, we get that the logic implication proposition (6) holds since its equivalent contrapositive logic proposition holds, since:

$$d(u,v) > \alpha d(x,y) - \beta D + \beta d(Tx,y) > (\alpha + \beta(\alpha+\beta))d(x,y) \tag{9}$$

Note that $d(u,Tx) = d(v,Ty) = D$ and $d(x,Tx) \le (1+\alpha+\beta)d(x,y) + D$ imply:

$$\begin{aligned} d(u,v) &\le \alpha d(x,y) + \beta(d(Tx,y) - D) \\ &\le (\alpha+\beta)d(x,y) + \beta(d(Tx,x) - D) \end{aligned} \tag{10}$$

Lemma 1 establishes that the contractive weak proximal condition can be satisfied under stronger contractive conditions irrespective of the amount $(d(x,Tx) - (1+\alpha+\beta)d(x,y) - D)$ being positive, negative or null. Thus, Lemma 1 adopts the following equivalent form:

Lemma 1´. Let (A, B) and, $T : A \to B$ be a pair of nonempty closed subsets of a metric space (X, d), such that $D = d(A,B)$, and, respectively, genmeralized weak proximal contraction, such that A_0 is nonempty, $T(A_0) \subseteq B_0$, and $\alpha + \beta(\alpha+\beta) \in [0,1)$, and let $x, y, u, v \in A$ be such that $d(u,Tx) = d(v,Ty) = D$. Then,

$$d(u,v) \le \min\left[(\alpha + \beta(\alpha+\beta))d(x,y),\, \alpha d(x,y) + \beta(d(Tx,y) - D)\right] \tag{11}$$

In summary, note from triangle inequality that the condition

$$d(y,Tx) - D \le (\alpha+\beta)d(x,y)$$

guarantees that the implying condition of (2) holds since

$$d(x,Tx) \le d(x,y) + d(y,Tx) - D \le (1+\alpha+\beta)d(x,y) + D$$

while the implied one adopts the form

$$d(u,v) \le [\alpha + \beta(\alpha+\beta)]d(x,y)$$

Lemma 1 leads to the following distance convergence result:

Proposition 2. Let (A, B) be a pair of nonempty closed subsets of a metric space (X, d). Assume that $T : A \to B$ with $D = d(A,B)$ is a geenralized weak proximal contraction, such that A_0 is nonempty $T(A_0) \subseteq B_0$ and $\beta \in [0, 1-\alpha]$. Construct a sequence $\{x_n\} \subset A$ satisfying $d(x_{n+1}, Tx_n) = d(Tx_{n+1}, x_{n+2}) = D$; $\forall n \in \mathbf{Z}_{0+}$ and the contractive condition:

$$d(x_{n+1}, x_{n+2}) \le [\alpha + \beta(\alpha+\beta)]d(x_n, x_{n+1}),\ \forall n \in \mathbf{Z}_{0+} \tag{12}$$

Then, $\{x_n\}$ is bounded, if x_0 is bounded, $d(x_n, x_{n+1}) \to 0$ and $d(x_n, Tx_n) \to D$ as $n \to \infty$ and $\sum_{n=0}^{\infty} d(x_{n+1}, x_n) < \infty$.

Proof: Note that the zeros of $p(\beta) = \alpha + \beta(\alpha+\beta) - 1$ are $\beta_1 = -1$, $\beta_2 = 1-\alpha$. Since the parabola $p(\beta)$ is convex then $\rho(\beta) < 0$, equivalently $\alpha + \beta(\alpha+\beta) < 1$, if $\beta \in [\beta_1, \beta_2)$, and since $\alpha \in [0, 1)$ and $\beta \in [0, \infty)$ are constraints associated with $T : A \to B$ being a generalized weak proximal contraction, $0 \le \theta = \alpha + \beta(\alpha+\beta) < 1$ requires $\beta \in [0, 1-\alpha]$ since $\alpha \in [0,1)$. Thus, since $d(x_{n+1}, x_n) \le \theta\, d(x_n, x_{n-1}) \le \theta^n d(x_1, x_0)$; $\forall n \in \mathbf{Z}_+$ then, $d(x_{n+1}, x_n) \to 0$ as $n \to \infty$ and $\{x_n\}$ is bounded if x_0 is bounded. On the other hand,

$$D \le d(Tx_n, x_n) \le d(Tx_n, x_{n+1}) + d(x_{n+1}, x_n) \le D + d(x_{n+1}, x_n)\ ;\ \forall n \in \mathbf{Z}_{0+}$$

so that $d(x_n, Tx_n) \to D$ as $n \to \infty$ since $d(x_{n+1}, x_n) \to 0$ as $n \to \infty$. Also, since $d(Tx_n, x_{n+1}) = D$; $\forall n \in \mathbf{Z}_{0+}$, one gets:

$$\begin{aligned} d(x_{n+1}, x_n) &\le \alpha d(x_n, x_{n-1}) + \beta[d(Tx_n, x_{n+1}) - D] = \alpha d(x_n, x_{n-1}) \\ &\le \alpha^n d(x_1, x_0) \le \theta^n d(x_1, x_0)\ ;\ \forall n \in \mathbf{Z}_+ \end{aligned}$$

which implies

$$\sum_{n=0}^{\infty} d(x_{n+1}, x_n) \le \frac{d(x_0, x_1)}{1-\alpha} \le \frac{d(x_0, x_1)}{1-\theta} < \infty$$

A further arrangement which guarantees that the implying logic proposition of (2) always holds for certain approximation sequences, then guaranteeing that the implied logic condition, can be used for contraction purposes follows below. Its proof is based on an induction argument combined with an associate strict contraction for sequences of distances. The price to be paid is a stronger constraint on the β-

ISBN: 978-1-61804-230-9

parameter depending on the value of $\alpha \in [0, 1)$ instead of its wider constraint $\beta \in [0, \infty)$ of Definition 1.

Theorem 2. Under all the assumptions of Proposition 2, assume also that (X, d) is complete and that B is approximatively compact with respect to A, or the weaker condition that A_0 is closed. Then, $T : A \to B$ has a unique best proximity point.

Proof: The convergence to a best proximity point is proved as in Theorem 1. The uniqueness of such a point follows from the contradiction of the form $0 < 0$ arising from contractive condition with $\alpha + \beta(\alpha + \beta) < 1$, $u = x$, $y = v$ and $x \neq y$:

$$d(x, y) = d(u, v) \leq [\alpha + \beta(\alpha + \beta)] d(x, y)$$

Proposition 3. Let (A, B) be a pair of nonempty closed subsets of a metric space (X, d). Assume that $T : A \to B$ with $D = d(A, B)$ is a generalized weak proximal contraction, such that A_0 is nonempty and $T(A_0) \subseteq B_0$, with

$$\beta \in \left[0, max\left(\frac{\sqrt{(2\alpha+3)^2 + 8(1-\alpha)} - (2\alpha+3)}{4}, 1-\alpha \right) \right),$$

and construct any sequence $\{x_n\} \subset A$ satisfying with given $x_0, x_1 \in A$, such that $d(x_{n+2}, Tx_n) = d(Tx_{n+3}, Tx_{n+1}) = D$; $\forall n \in \mathbf{Z}_{0+}$, and

$$d(x_2, x_3) \leq \frac{1}{2}(1 + \alpha + \beta) d(x_0, x_1), \forall n \in \mathbf{Z}_{0+} \tag{13}$$

Then, $d(x_n, x_{n+1}) \to 0$ and $d(x_n, Tx_n) \to 0$ as $n \to \infty$, $\{x_n\}$ is bounded, if x_0 is bounded, and $\sum_{n=0}^{\infty} d(x_{n+1}, x_n)$ is also bounded.

Proof: Note from (5) that the implying condition of (2) always holds provided that $d(x_{n+2}, x_{n+1}) \leq \frac{1}{2}(1 + \alpha + \beta) d(x_n, x_{n+1})$; $\forall n \in \mathbf{Z}_{0+}$ since then one gets:

$$\begin{aligned} d(x_n, Tx_n) &\leq d(x_{n+2}, Tx_n) + d(x_{n+2}, x_n) \leq d(x_{n+2}, x_n) + D \\ &\leq 2\, max\left(d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2})\right) + D \\ &\leq (1 + \alpha + \beta) d(x_n, x_{n+1}) + D \end{aligned} \tag{14}$$

Assume that (14) holds for $n = 0$, that is $d(x_2, x_1) \leq \frac{1}{2}(1 + \alpha + \beta) d(x_0, x_1)$ with $x_0, x_1 \in A$, x_0 finite and $x_2 \in A$ being such that $d(x_2, Tx_0) = D$. Then, proceed by complete induction by assuming that it also holds for any fixed $n \in \mathbf{Z}_{0+}$

Then, since

$$\beta \in \left[0, max\left(\frac{\sqrt{(2\alpha+3)^2 + 8(1-\alpha)} - (2\alpha+3)}{4}, 1-\alpha \right) \right)$$

which is equivalent to $\alpha + \beta + \beta(1 + \alpha + \beta) \leq \frac{1 + \alpha + \beta}{2} < 1$ and which is guaranteed if $\rho_1 = \alpha + \beta + 2\beta(1 + \alpha + \beta) \leq 1$, so that the maximum zero of the convex parabola $q(\beta) = \alpha + \beta + 2\beta(1 + \alpha + \beta) - 1$ is reached at $\beta_2 = \frac{\sqrt{(2\alpha+3)^2 + 8(1-\alpha)} - (2\alpha+3)}{4}$, so guaranteeing $q(\beta) = \alpha + \beta + 2\beta(1 + \alpha + \beta) - 1 \leq 0$ on $[0, \beta_2]$, which together with $\rho = \frac{1 + \alpha + \beta}{2} < 1$ leads to:

$$\begin{aligned} d(x_{n+3}, x_{n+2}) &\leq \alpha d(x_n, x_{n+1}) - \beta D + \beta d(Tx_n, x_{n+1}) \\ &\leq (\alpha + \beta) d(x_n, x_{n+1}) - \beta D + \beta d(Tx_n, x_n) \end{aligned}$$

$$\begin{aligned} &\leq (\alpha + \beta) d(x_n, x_{n+1}) - \beta D + \beta\left[(1 + \alpha + \beta) d(x_n, x_{n+1}) + D\right] \\ &\leq \rho\, d(x_n, x_{n+1}) < d(x_n, x_{n+1}); \forall n \in \mathbf{Z}_{0+} \end{aligned} \tag{15}$$

if $d(x_n, x_{n+1}) \neq 0$, provided that

$$\beta \in \left[0, max\left(\frac{\sqrt{(2\alpha+3)^2 + 8(1-\alpha)} - (2\alpha+3)}{4}, 1-\alpha \right) \right)$$

since

$$(\rho_1 = \alpha + \beta + 2\beta(1 + \alpha + \beta) < 2) \wedge \left(\rho = \frac{1 + \alpha + \beta}{2} < 1\right) \tag{16}$$

with the second above constraint being equivalent to $\alpha + \beta < 1$. Then (14) also holds for the replacement $n \to n+1$ by the complete induction method, then for any $n \in \mathbf{Z}_{0+}$, since $\frac{1 + \alpha + \beta}{2} \leq (1 + \alpha + \beta)$. Thus, one gets from (15) after replacing $n \to 2n$ that $d(x_{2n+3}, x_{2n+2}) \leq \rho^n d(x_0, x_1)$ so that $d(x_{2n+3}, x_{2n+2}) \to 0$ as $n \to \infty$. As a result, $d(x_n, x_{n+1}) \to 0$ as $n \to \infty$ from (15), $d(x_n, Tx_n) \to D$ as $n \to \infty$ from (14), $\{x_n\}$ is bounded, if x_0 is finite, and $\sum_{n=0}^{\infty} d(x_{n+1}, x_n)$ is also bounded from a closed proof to that of its counterpart of Proposition 1. The following result, whose proof is omitted, can be proved in a similar way as those of Theorems 1-2.

Theorem 3. Under all the assumptions of Proposition 3, assume also that (X, d) is complete and that B is approximatively compact with respect to A, or the weaker condition that A_0 is closed. Then, $T : A \to B$ has a unique best proximity point.

Another further result, which generalizes Proposition 3, while it guarantees that the implying logic proposition of (2) always holds for certain approximation sequences, then guaranteeing that the implied logic condition now follows by using a sequence of distances $d(A, B_n)$ for $n \in \mathbf{Z}_{0+}$. In this context, we can define the non-self mapping $T: A \to (\cup B_n)$ in such a way the (weak) proximal sequences are constructed with the auxiliary $T: A \to B_n$; $\forall n \in \mathbf{Z}_{0+}$. The next result is con¡cerned with booindedness and convergence of (weak) proximal sequences. If, furthermore, the set theoretic limit $B := \lim_{n\to\infty} B_n$ exists and it is closed then the existence and uniqueness of best proximity points is also investigated in a later result

Proposition 4. Let A and B_n; $\forall n \in \mathbf{Z}_{0+}$ be nonempty subsets of a metric space (X, d). Assume that $T: A \to (\cup B_n)$, such that A_0 is nonempty and $T(A_0) \subseteq B_{0n}$; $\forall n \in \mathbf{Z}_{0+}$, with $D_n = d(A, B_n)$; $\forall n \in \mathbf{Z}_{0+}$ satisfies the contractive condition:

$$d(x_{n+1}, x_{n+2}) \le \alpha d(x_{n+1}, x_n) + \beta(d(Tx_n, x_{n+1}) - D_n) \quad ; \forall n \in \mathbf{Z}_+ \quad (17)$$

Then, the following properties hold:

(i) $d(x_{n+1}, x_{n+2}) \le \mu^{n+1} d(x_0, x_1) + \beta \sum_{k=0}^{n} \mu^{n-k}\left(D_k^0 - D_k\right)$; $\forall n \in \mathbf{Z}_+$ (18)

which is bounded for any finite $x_0 \in A$ if $\mu \le 1$ and $\sum_{k=0}^{n} \mu^{n-k}\left(\left|D_k - D_k^0\right|\right) < \infty$, where $\mu = (\alpha+\beta)(1+\beta)+\beta$ and the sequence $\left\{D_n^0\right\}$ satisfies

$$D_n^0 \ge d(x_n, Tx_n) - (1+\alpha+\beta) d(x_n, x_{n+1}); \ \forall n \in \mathbf{Z}_{0+} \quad (19)$$

(ii) If (17) holds, subject to (19), with $\mu < 1$ then

$$d(x_n, x_{n+1}) \le \mu^n d(x_0, x_1) + \frac{1-\mu^n}{1-\mu} \beta \sup_{0\le k\le n+1} \left|D_k - D_k^0\right| \quad (20)$$

$$\limsup_{n\to\infty} \left(d(x_n, x_{n+1}) - \frac{\beta}{1-\mu} \left(\sup_{0\le k\le n} \left|D_k - D_k^0\right| \right)\right) \le 0 \quad (21)$$

$$\limsup_{n\to\infty} \left(\left| d(x_{n+1}, Tx_n) - d(x_n, Tx_n) \right| - d(x_n, x_{n+1}) \right) \le 0 \quad (22)$$

$$\limsup_{n\to\infty} \left(\left| d(x_{n+1}, Tx_n) - d(x_n, Tx_n) \right| - \frac{\beta}{1-\mu} \sup_{0\le k\le n} \left|D_k - D_k^0\right| \right) \le 0 \quad (23)$$

$$\lim_{n\to\infty} d(x_n, x_{n+1}) = 0 \ ; \ \lim_{n\to\infty} \left(\left| d(x_{n+1}, Tx_n) - d(x_n, Tx_n) \right| \right) = 0 \quad (24)$$

if $\left|D_n - D_n^0\right| \to 0$ as $n \to \infty$

(iii) If (17) holds subject to (19) with $\mu < 1$ then

$$\liminf_{n\to\infty} \left[\min\left(d(x_n, Tx_n) - D_n, d(x_{n+1}, Tx_n) - D_{n+1} \right) \right] \ge 0 \quad (25)$$

$$\limsup_{n\to\infty} \left(d(x_n, Tx_n) - D_n^0 - \frac{\beta(1+\alpha+\beta)}{1-\mu} \left(\sup_{0\le k\le n} \left|D_k - D_k^0\right| \right)\right) \le 0 \quad (26)$$

$$\limsup_{n\to\infty} \left(d(x_{n+1}, Tx_n) - D_n^0 - \frac{\beta(2+\alpha+\beta)}{1-\mu} \left(\sup_{0\le k\le n} \left|D_k - D_k^0\right| \right)\right) \le 0 \quad (27)$$

(iv) If $\mu < 1$ and $\left|D_n - D_n^0\right| \to 0$ as $n \to \infty$ then the limits below exist and are identical:

$$\lim_{n\to\infty} \left(d(x_n, Tx_n) - D_n^0\right) = \lim_{n\to\infty} \left(d(x_n, Tx_n) - D_n\right) = \lim_{n\to\infty} \left(d(x_n, Tx_n) - D_n^0\right) = \lim_{n\to\infty} \left(d(x_{n+1}, Tx_n) - D_n\right) \quad (28)$$

If (17) holds, subject to (19), with $\mu < 1$, $\left\{D_n^0\right\} \to D$ and $\{D_n\} \to D$ then $\lim_{n\to\infty} d(x_n, Tx_n) = \lim_{n\to\infty} d(x_{n+1}, Tx_n) = D$.

Proof: Since $d(A, B_n) = D_n$ then $d(Tx_n, x_{n+1}) \ge D_n$; $\forall n \in \mathbf{Z}_{0+}$. Also, if follows from (19) and (17) with $\mu = (\alpha+\beta)(1+\beta)+\beta$ that (18) holds since

$$d(x_{n+2}, x_{n+1}) \le \alpha d(x_{n+1}, x_n) + \beta(d(Tx_n, x_n) + d(x_n, x_{n+1}) - D_n) \le \mu^{n+1} d(x_0, x_1) + \beta \sum_{k=0}^{n} \mu^{n-k}\left(D_k^0 - D_k\right) \quad ; \forall n \in \mathbf{Z}_+ \quad (29)$$

for any given $x_0 \in A$. If $\mu \le 1$ and $\sum_{k=0}^{n} \mu^{n-k}\left(\left|D_k - D_k^0\right|\right) < \infty$ then the sequence $\{d(x_n, x_{n+1})\}$ is bounded. Thus, Property (i) has been proved. The relations (20) and (21) of Property (ii) follow directly from (18) of Property (i) if $\mu < 1$. On the other hand, the triangle inequality and (19) lead to (25) since

$$\left| d(x_{n+1}, Tx_n) - d(x_n, Tx_n) \right| \le d(x_n, x_{n+1})$$

The relation (23) follows from (21) and (22). To prove (24), note that if $\lim_{n\to\infty}\left(D_n - D_n^0\right) = 0$, then for any given $\varepsilon \in \mathbf{R}_+$, there is $m = m(\varepsilon) \in \mathbf{Z}_{0+}$ such that $\sup_{m\le k\le n+m+1} \left|D_k - D_k^0\right| < \varepsilon$ for any $n(\in \mathbf{Z}_+) \ge m$ and then, from (24), $\limsup_{n\to\infty} d(x_{n+1}, x_n) \le \frac{\beta\varepsilon}{1-\mu}$. Since ε is arbitrary, the limit $\lim_{n\to\infty} d(x_n, x_{n+1})$ exists and $\lim_{n\to\infty} d(x_n, x_{n+1}) = 0$. This property and (22) yield directly $\lim_{n\to\infty} \left(\left| d(x_{n+1}, Tx_n) - d(x_n, Tx_n) \right| \right) = 0$ and then Property (ii) has been fully proved. To prove Property (iii), note that (25) holds directly from $d(A, B_n) = D_n$; $\forall n \in \mathbf{Z}_{0+}$ and $\{x_n\} \subset A$. Also, (19) is identical to

$$d(x_n, Tx_n) - D_n^0 \le (1+\alpha+\beta) d(x_n, x_{n+1}); \ \forall n \in \mathbf{Z}_{0+}$$

which leads to (26) by taking into account (21). The relation (27) follows from (21), (26) and the relation:

$$d(x_{n+1}, Tx_n) \le d(x_n, Tx_n) + d(x_{n+1}, x_n); \forall n \in \mathbf{Z}_{0+}$$

Hence, Property (iii) has been proved. Property (iv) is a direct consequence of Property (iii) for the case when $d(x_n, x_{n+1}) \to 0$ and $\left|D_n - D_n^0\right| \to 0$ as $n \to \infty$ including its particular sub-case when $\{D_n^0\} \to D$ and $\{D_n\} \to D$.

Theorem 4. Under all the assumptions of Proposition 3.2 and Property (iii), Eqn. (3.10), assume also that (X, d) is complete, that A and B_n; $\forall n \in \mathbf{Z}_{0+}$, are nonempty subsets of X such that A is closed, A_0 is nonempty, the limit set $B := \lim_{n\to\infty} B_n$ exists, is closed and approximatively compact with respect to A (or the weaker condition that A_0 is closed) and $T(A_0) \subseteq B_0$. Assume also that the non-self mapping restriction $T : A|A_1 \to B$, for some nonempty subset $A_1 \subset A$, which contains A_0, is a generalized weak proximal contraction. Then, $T : A|A_1 \to B$ has a unique best proximity point if $\mu < 1$.

Proof: Since $d(x_{n+1}, x_n) \to 0$ as $n \to \infty$ $\{x_n\} \to x$ from (24). Since $\{x_n\} \subset A$ and A is closed then $x \in A$. Since $D_n \le d(x_{n+1}, Tx_n) = \{d(x, Tx_n)\}(\to D)$ since $\{D_n\} \to D$; and $\left|D_n - d(x_n, Tx_n)\right| \le d(x_n, x_{n+1})$ $\forall n \in \mathbf{Z}_{0+}$. Since B is approximatively compact with respect to A and $\{x_n\} \subset A$, then there are $y \in A_0$ and a sequence $\{\bar{x}_n\} \subset A$, such that $\{\bar{x}_n - x_n\} \to 0$, then $\{\bar{x}_n\} \to x$ since $\{x_n\} \to x$, and $\{Tx_n\} \subset B$ such that as $n \to \infty$:

$d(\bar{x}_n, T\bar{x}_n) \to d(y, B) = D$; $d(\bar{x}_n, B) \to d(y, B) = D$; $d(x_n, B) \to d(y, B) = D$

$\{d(x_n, Tx_n)\} \to D$, $\{d(x_{n+1}, Tx_n)\} \to D$ and $\{d(x, Tx_n)\} \to D$. Since B is approximatively compact with respect to A and $\{x_n\} \subset A$, then the sequence $\{Tx_n\} \subset \bigcup B_n$, such that $Tx_n \in B_n$, has a convergent subsequence $\{Tx_{n_k}\} \to z \in B$, since B is also closed and both $\{Tx_{n_k}\}$ and $\{Tx_{n_k}\}$ have the same limit $z \in B$. Also, $z \in cl\, B_0$ and $d(x, Tx_{n_k}) \to D(= d(x, z))$ as $k \to \infty$ so that x is a best proximity point of $T : A \to (\cup B_n)$. Note that since the limit set B exists, it is by construction the infinite union of intersections of the form $B = \bigcup_{n=0}^{\infty} \bigcap_{m \ge n}^{\infty} B_m$ so that $B_0 \subseteq B \subseteq \cup B_n$ so that there is a restriction $T : A|A_1 \to (\cup B_n)|B$ for some nonempty $A_1 \subseteq A$. Assume not so that $A_1 = \varnothing$. If $A_1 = \varnothing$ then A_0 is also empty which is impossible then $A_1 = \varnothing$. It is now proved by contradiction that the best proximity point is unique. Assume not so that there are two best proximity points x, y such that there are two sequences $\{x_n\} \to x$ and $\{y_n\} \to y$ contained in A. Since $T : A|A_1 \to (\cup B_n)|B$ is a geenralized weak proximal contraction, one gets from the implied logic proposition of (2) with $u = x$ and $v = y$ and $D_n = D$ that

$$\begin{aligned}(1-\alpha)d(x, y) &\le \beta(d(Tx, y) - D) \le \beta d(Tx, x) + \beta d(x, y) - \beta D \\ &= \beta d(x, y)\end{aligned}$$

which fails for $0 \le \beta < 1 - \alpha$ if $x \ne y$. Thus, $x = y$.

References

[1] S. Sadiq Basha, "Best proximity points: optimal solutions", *J. Optim. Theory Appl.*, Vol. 151, pp. 210-216, 2011.

[2] A.A. Eldred and P. Veeramani, "Existence and convergence of best proximity points", *Journal of Mathematical Analysis with Applications*, Vol. 323, pp. 1001-1006, 2006.

[3] V. Berinde, "Approximating fixed points of weak contractions using the Picard iteration", *Nonlin. Anal. Forum*, Vol. 9, pp. 43-53, 2004.

[4] M. Gabeleh, "Discrete optimizarion via various proximal non-self mappings" (in press).

[5] M. De la Sen, "Linking contractive self-mappings and cyclic Meir-Keeler contractions with Kannan self– mappings", *Fixed point Theory and Applications*, Article ID 572057, 2010, doi:10.1155/2010/572057.

[6] M. De la Sen, "About robust stability of dynamic systems with time-delays through fixed point theory", *Fixed Point Theory and Applications*, Article ID 48187, 2008, doi:10.1155/2008/48187.

[7] M. De la Sen, "Some combined relations between contractive mappings, Kannan mappings, reasonable expansnsive mappings and T- stability", *Fixed Point Theory and Applications*, Article ID 815637, 2009, doi:10.1155/2009/815637.

[8] M. De la Sen and R. P. Agarwal, "Some fixed point- results for a class of extended cyclic self-mappings with a more general contractive condition", *Fixed Point Theory and Applications*, 2011, 2011:59.

[9] M. Delasen, "Online optimization of the free parameters in discrete adaptive-control systems", *IEE Proceedings-D Control Theory and Applications*, Vol. 131, No. 4, pp. 146-157, 1984.

[10] M. Delasen, "On some structures of stabilizing control laws for linear and time-invariant systems with bounded delays and unmeasurable states", *International Journal of Control*, Vol. 59, No. 2, pp. 529-541, 1994.

A stochastic regulator inventory control model with random price dynamics and variable demand

A.Tsoularis
Business Rose Bowl, Leeds Metropolitan University, Leeds, LS1 3HB, UK
e-mail: a.tsoularis@leedsmet.ac.uk

J.Wallace
School of Management, Bradford University, Emm Lane, Bradford, BD9 4JL, UK
e-mail: jwallace1@bradford.ac.uk

Abstract— This paper considers an inventory control system meeting uncertain demand in continuous time. The demand is a function of both time and price, with the price evolving as a Wiener process with no drift. The objective of the inventory planning system is to track the demand as closely as possible whilst minimizing replenishment costs within a time frame such that the price remains within acceptable limits in it. An optimal control problem is formulated to achieve this goal and its solution is presented.

.

Keywords—Wiener process, Fokker-Planck equation, Ornstein-Uhlenbeck process, Hamilton-Jacobi-Bellman equation.

I. LITERATURE REVIEW

Thompson and Sethi [1] considered a production-inventory model which determines production rates over time to minimize an integral representing a discounted quadratic loss function and solved the model both with and without nonnegative production constraints. This model was generalized by Bensoussan at al. [2] who considered an infinite horizon stochastic production planning problem with the production rate constrained to be non-negative. Fleming et al. [3] considered a stochastic optimal production control problem with random demand. Browne and Zipkin [4] dealt with the standard re-order point/batch size inventory model with continuous stochastic demand. Gallego and van Ryzin [5] investigated the problem of dynamic pricing of inventories when demand is price sensitive and stochastic and the objective is to maximize expected revenues and derived the optimal pricing in closed form when an exponential family of demand functions. Dohi et al. [6] considered optimal production planning when the demand is a Markov diffusion process. Sethi et al. [7] concerned with the optimal production planning in a dynamic stochastic manufacturing system consisting of a single failure prone machine facing constant demand with discounted cost. Johansen and Thorstenson [8] considered an inventory system with periodic review and stochastic lead times with continuous time demand generated by a Poisson process and derived optimality conditions for the long-run average cost. Berling and Martínez-de-Albéniz [9] considered optimal inventory policies when the demand is a Poisson process and the commodity price fluctuates as a geometric Brownian motion or as Ornstein-Uhlenbeck process. This article presents an inventory model of an item facing demand which is simultaneously an exponential function of time and linear function of price. This functional form was suggested by Wee [10] and Hollier and Mak [11] who treated the price as a deterministic variable. In this work, however, the price is allowed to fluctuate around a nominal value and the problem is to determine an ordering policy that ensures the demand is tracked closely at minimum cost throughout a finite horizon chosen so as to keep the price within predetermined limits.

II. THE INVENTORY MODEL

Let $x(t)$ be the inventory and $y(t)$ the demand rate at any time *t*. If $u(t)$ is the continuous inventory replenishment rate, he inventory evolves according to the state equation:

$$dx = (u - y)dt,\ x(0) = x_0 \tag{1}$$

The demand growth rate is a linear function of the selling price, *z* but increases or decreases exponentially with time. This form was suggested by Wee [10] and earlier by Hollier and Mak [11], who assumed only exponential decline in time:

$$y(z,t) = (a - bz)e^{\alpha t} \tag{2}$$

where $a, b > 0$ and α can be either a positive or negative parameter governing exponentially the increase or decrease in the demand rate. For a positive demand rate, the price should

ISBN: 978-1-61804-230-9

be kept below the maximum level, $\frac{a}{b}$. Once z exceeds $\frac{a}{b}$, the demand rate, $y(t)$, becomes negative and stock begins to build up as is readily evident from (1). On the other hand, for $z < 0$, the demand rate becomes positive and increasing as low price drives the inventory to low levels, as again readily evident from (1).

Price dynamics

In this paper we assume that the selling price, $z(t)$, is a drift-free Wiener process with diffusion parameter, σ:

$$dz = \sigma dw\,,\quad z(0) = z_0 \tag{3}$$

The stochastic price dynamics is then

$$z(t) = z_0 + \sigma w(t) \tag{4}$$

with constant mean the initial price, z_0, variance, $\sigma^2 t$, and normal conditional density function

$$f_z(z,t \mid z_0) = \frac{1}{\sigma\sqrt{2\pi t}}\exp\left(-\frac{(z-z_0)^2}{2\sigma^2 t}\right),$$

which is the solution to the Fokker-Planck (forward) equation, $\frac{\partial f_z(z,t \mid z_0)}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 f_z(z,t \mid z_0)}{\partial z^2}$ with initial condition $f_z(z,0) = \delta(z - z_0)$.

The structure of the demand rate equation breaks down from the practical viewpoint when the price falls outside the interval $\left[0, \frac{a}{b}\right]$. Let $T(z_0)$ be the mean of the first time (that is the mean exit time) price starting from z_0, reaches either $d > z_0$ or $c < z_0$. The two extreme values are effectively acting as absorbing states. The functional form of $T(z_0)$ can be obtained from solving the differential equation [12]:

$$\frac{\sigma^2}{2}\frac{d^2 T}{dz_0^2} = -1$$

with boundary conditions, $T(c) = T(d) = 0$. The solution is

$$T(z_0) = \frac{(z_0 - c)(d - z_0)}{\sigma^2}$$

For $c = 0, d = \frac{a}{b}$,

$$T(z_0) = \frac{z_0\left(\frac{a}{b} - z_0\right)}{\sigma^2}$$

Now let

$$\pi_c(z_0) = probability(z(T) = c \mid z(0) = z_0)$$
$$\pi_d(z_0) = probability(z(T) = d \mid z(0) = z_0)$$

The differential equation [13]

$$\frac{\sigma^2}{2}\frac{d^2\pi_0}{dz_0^2} = 0$$

with the boundary conditions, $\pi_0(c) = 1,\ \pi_d(d) = 0,\quad \pi_c(z_0) + \pi_d(z_0) = 1$, has the solution

$$\pi_c(z_0) = \frac{d - z_0}{d - c}$$
$$\pi_d(z_0) = \frac{z_0 - c}{d - c}$$

For $c = 0$, $\pi_0 = 1 - \frac{z_0}{d}$ and $\pi_d = \frac{z_0}{d}$. The probability of price hitting 0 drops linearly the closer price is to its upper limit, $\frac{a}{b}$.

The probability that the price hits its maximum level, $\frac{a}{b}$, for the first time is given by

$$2\int_{\frac{a}{b}}^{\infty}\frac{1}{\sigma\sqrt{2\pi t}}\exp\left(-\frac{(z-z_0)^2}{2\sigma^2 t}\right)dz$$

There is no closed form expression for this integral. Introduce the variable transformation, $\eta = \frac{z - z_0}{\sigma\sqrt{t}}$, which gives

$$2\int_{\frac{\frac{a}{b}-z_0}{\sigma\sqrt{t}}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{\eta^2}{2}}d\eta = 2\left(1 - N\left(\frac{\frac{a}{b} - z_0}{\sigma\sqrt{t}}\right)\right)$$

By choosing a desired probability value, the time the price hits its maximum can be readily obtained from standard normal distribution tables. For instance, let us assign a probability of 0.01 for the price hitting its maximum level $\frac{a}{b} = 2$ say, and let $z_0 = 1, \sigma = 0.05$. Then the first time, T, that the price starting from $z_0 = 1$ at $t = 0$, will hit the maximum $\frac{a}{b} = 2$ is

ISBN: 978-1-61804-230-9

found from $N\left(\frac{20}{\sqrt{T}}\right) = 0.995$ to be $T \approx 60.29$. Doubling the diffusion parameter of the price to $\sigma = 0.1$ yields a time horizon, one quarter of the previous value, $T \approx 15$.

Figures 1 and 2 below show a typical price and demand evolution with positive exponential growth along with the expected values for a nominal initial price, $z_0 = 1$,diffusion $\sigma = 0.05$ and $a = 1, b = 0.5, \alpha = 0.1$.

Figure 1. Price dynamics subject to fluctuations.

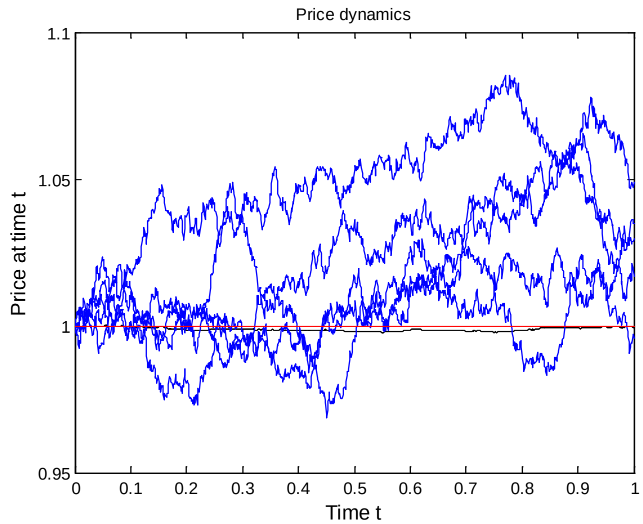
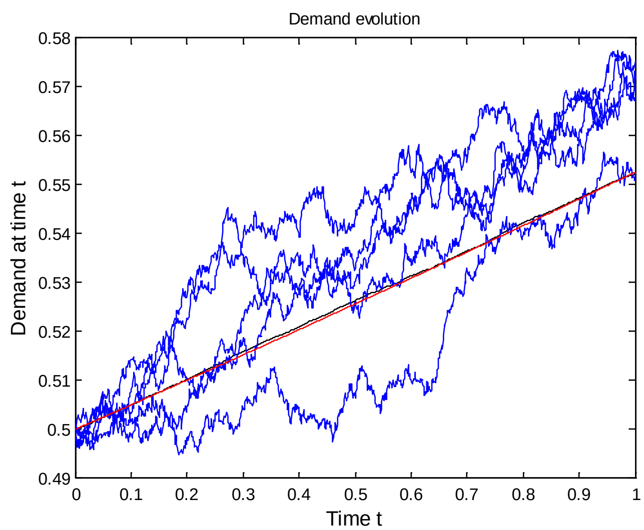


Figure 2. Demand growth subject to fluctuations in price.



The demand growth dynamics

By using Itó's formula the demand rate obeys the stochastic differential equation

$$dy = (\alpha y)dt + (-b\sigma e^{\alpha t})dw,\; y(0) = a - bz_0 \tag{5}$$

which is a time-dependent Ornstein-Uhlenbeck process.

The Lipschitz and growth conditions

$$|\alpha(y_1 - y_2)| \le K|y_1 - y_2|$$
$$\alpha^2|y|^2 + b^2\sigma^2 e^{2\alpha t} \le K^2(1 + |y|^2)$$

ensure the existence and uniqueness of the solution of (5) in $[0, T]$ provided that the growth parameter is bounded, $|\alpha| \le K$, for some arbitrary constant *K*.

Let $v(t, y) = ye^{-\alpha t}$. Then

$$dv = e^{-\alpha t}dy - \alpha v dt = -b\sigma dw$$

The demand growth follows the stochastic dynamics

$$y(t) = y_0 e^{\alpha t} - \int_0^t b\sigma e^{\alpha t} dw \tag{6}$$

with expected value, $y_0 e^{\alpha t}$. To obtain its variance, we have first

$$E[y^2(t)] = \left(y_0^2 + \frac{b^2\sigma^2}{2\alpha}\right)e^{2\alpha t} - \frac{b^2\sigma^2}{2\alpha}$$

and finally

$$Variance(y(t)) = \frac{b^2\sigma^2}{2\alpha}\left(e^{2\alpha t} - 1\right)$$

The process, $dv = -b\sigma dw$, $v_0 = a - bz_0$, has conditional probability density

$$f_v(v, t \mid v_0) = \frac{1}{b\sigma\sqrt{2\pi t}} \exp\left(-\frac{(v - v_0)^2}{2b^2\sigma^2 t}\right)$$

with initial condition $f_v(0, v) = \delta(v - v_0)$.

III. THE OPTIMAL CONTROL P|ROBLEM

The objective is to minimise a performance index incorporating the difference between the demand for an item and its inventory, $x - y$, and the replenishment costs, *u*, over a finite time horizon [0,*T*]. The overall cost to be minimised is of the form

$$J(0, x_0) = \underset{u(t)}{E}\left[\int_0^T \left(c_1(x - y)^2 + c_2 u^2\right)dt\right] \tag{7}$$

where c_1 and c_2 are positive cost parameters that may vary with time. An optimal Markovian non-anticipative replenishment policy, $u^*(x, y, t)$, is sought to maintain the

ISBN: 978-1-61804-230-9

inventory near the demand level whilst keeping the order costs as low as possible. The full linear stochastic regulator optimal control problem is thus

$$\inf_{u(t)} J(0,x_0)=V(x_0,u^*)$$

given the two state equations, for inventory and demand:

$$dx=(u-y)dt,\ \ x(0)=x_0$$
$$dy=(\alpha y)dt+(-b\sigma e^{\alpha t})dw,\ y(0)=a-bp_0$$

The Hamilton-Jacobi-Bellman equation for the value function, $V(x_0,u^*)$, is

$$0=\inf_{u(t)}\left\{\begin{array}{l}(u-y)\dfrac{\partial V^*}{\partial x}+\\ \alpha y\dfrac{\partial V^*}{\partial y}+\dfrac{b^2\sigma^2e^{2\alpha t}}{2}\dfrac{\partial^2V^*}{\partial y^2}+\\ c_1(x-y)^2+c_2u^2+\dfrac{\partial V^*}{\partial t}\end{array}\right\} \tag{8}$$

with the boundary condition

$$V^*(T,x(T))=0$$

The final time, T, must be chosen so that the price remains in the interval, $\left[0,\dfrac{a}{b}\right]$, throughout with an arbitrarily chosen probability.

The solution to the Hamilton-Jacobi-Bellman equation has the form [14]:

$$V(t,x,y)=\begin{pmatrix}x & y\end{pmatrix}Q(t)\begin{pmatrix}x\\ y\end{pmatrix}+s(t) \tag{9}$$

where $Q(t)$ is a symmetric positive semidefinite 2x2 matrix forming the certainty equivalent part and $s(t)$ is a continuously differentiable scalar function of time accounting for the stochasticity of the process [15]. Let

$$Q(t)=\begin{pmatrix}q_1 & q_2\\ q_2 & q_3\end{pmatrix}$$

so that

$$V(x,y,t)=q_1(t)x^2+2q_2(t)xy+q_3(t)y^2+s(t) \tag{10}$$

From (8) the optimal replenishment policy is given by

$$u^*=-\frac{1}{2c_2}\frac{\partial V}{\partial x}$$

The optimal replenishment policy is active when the value function declines with the inventory, that is when $\dfrac{\partial V}{\partial x}<0$, otherwise $u^*=0$, so the optimal control can in fact be captured by the law

$$u^*=max\left(-\frac{1}{2c_2}\frac{\partial V}{\partial x},0\right) \tag{11}$$

Substituting (11) into (8) and equating the coefficients we obtain the following ordinary coupled differential equations for $Q(t)$ and $s(t)$:

$$c_2\dot q_1-q_1^2+c_1c_2=0 \tag{12}$$

$$\dot q_2+q_2\left(\alpha-\frac{q_1}{c_2}\right)-q_1-c_1=0 \tag{13}$$

$$\dot q_3+2\alpha q_3-\frac{q_2^2}{c_2}-2q_2+c_1=0 \tag{14}$$

$$\dot s=\sigma^2b^2e^{2\alpha t}q_3 \tag{15}$$

with the boundary conditions $q_1(T)=q_2(T)=q_3(T)=s(T)=0$.

To avoid lengthy algebraic manipulations without loss of generality, we assign the cost parameters, the unit value, so let $c_1=c_2=1$. The solution to the uncoupled ordinary differential equation (12) is then

$$q_1(t)=\tanh(T-t) \tag{16}$$

Given (16) the integrating factor for (13) is $e^{\alpha t}\cosh(T-t)$ and the solution to (13) is

$$q_2(t)=\left\{\begin{array}{l}\dfrac{\left(e^{T-t}-e^{\alpha(T-t)}\right)\operatorname{sech}(T-t)}{\alpha-1}\ \text{ if } \alpha\neq 1\\ (t-T)e^{T-t}\operatorname{sech}(T-t)\ \text{ if }\ \alpha=1\end{array}\right\} \tag{17}$$

The derivation of closed form solutions to (14) and (15) is not possible. The Markovian optimal policy from any state (x,y) at time $t<T$ is then

$$u^*(x,y,t) = \left\{ \begin{array}{l} = max\left(\begin{array}{l} -x\tanh(T-t) - \\ y\dfrac{\left(e^{T-t} - e^{\alpha(T-t)}\right)\mathrm{sech}(T-t)}{\alpha-1}, 0 \end{array} \right) \text{ if } \alpha \neq 1 \\ = max\left(-x\tanh(T-t) - y(t-T)e^{T-t}\,\mathrm{sech}(T-t), 0\right) \text{ if } \alpha = 1 \end{array} \right\} \quad (18)$$

IV. DISCUSSION

This paper puts forward an inventory control system operating under stochastic demand and pricing subject to random fluctuations. The price is assumed to be a Wiener process with no drift. The method of solution adopted here can also be used when price is an arithmetic Brownian motion, $dz = \mu dt + \sigma dw$, and consequently the demand rate follows the process, $dy = (\alpha y - b\mu e^{\alpha t})dt + (-b\sigma e^{\alpha t})dw$. The objective of the inventory system is to meet demand by replenishing when necessary at minimum cost. From the perspective of an inventory planner the solution obtained is characterised by the following key features:

(i) The operating time of the process is chosen to ensure that the demand rate is positive with an arbitrarily chosen probability.

(ii) Ordering takes place only when necessary ($u = 0$ when appropriate).

(iii) Random demand is monitored and met as often as possible.

(iv) Inventory is driven to low levels at the end of the time horizon.

References

[1] Thompson G.L., Sethi S.P. (1980) ‘Turnpike horizons for production planning’, *Management Science*, Volume 26, No. 3, pp. 229-241.

[2] Bensoussan A., Sethi S.P., Vickson R., Derszko N. (1984) ‘Stochastic production planning with production constraints’, *SIAM Journal of Control and Optimization*, Volume 22, No. 6, pp. 920-935.

[3] Fleming W.H., Sethi P., Soner H.M. (1987) ‘An Optimal Stochastic Production Planning Problem with Randomly Fluctuating Demand’, *SIAM Journal of Control and Optimization*, Volume 25, No. 6, pp. 1494-1502.

[4] Browne S., Zipkin P. (1991) ‘Inventory Models with Continuous Stochastic Demands’, *The Annals of Applied Probability*, Volume 1, No. 3, pp. 419-435.

[5] Gallego G., van Ryzin G. (1994) ‘Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons’, *Management Science*, Volume 40, No. 8, pp. 999-1020.

[6] Dohi T., Kato N., Osaki S. (1995) ‘Optimal Production Planning under Diffusion Demand Pattern’, *Mathematical and Computer Modelling*, Volume 21, No. 11, pp.35-46.

[7] Sethi S.P., Suo W., Taksar M.I., Zhang Q. (1997) ‘Optimal production planning in a stochastic manufacturing system with long-average cost’, *Journal of Optimization Theory and Applications*, Volume 92, No. 1, pp. 161-188.

[8] Johansen S.G., Thorstenson A. (2008) ‘Optimal base-stock policy for the inventory system with periodic review, backorders and sequential lead times’, *International Journal of Inventory Research*, Volume 1, No. 1, pp. 44-52.

[9] Berling P., Martínez-de-Albéniz V. (2011) ‘Optimal Inventory Policies when Purchase Price and Demand Are Stochastic’, *Operations Research*, Volume 59, No. 1, pp. 109-124.

[10] Wee Hui-Ming, (1995) ‘Joint pricing and replenishment policy for deteriorating inventory with declining market’, *International Journal of Production Research*, Volume 40, pp.163-171.

[11] Hollier, R.H, Mak K.L. (1983) ‘Inventory replenishment policies for deteriorating items in a declining market’, *International Journal of Production Research*, Volume 21, No. 6, pp. 813-826.

[12] Gillespie, D.T (1992) *Markov Processes*. Academic Press. San Diego.

[13] Gardiner, C.W (2002) *Handbook of Stochastic Methods*. Springer Verlag, Berlin.

[14] Arnold L. (1974) *Stochastic Differential Equations*: *Theory and Applications*. J.Wiley, New York.

[15] Fleming W.H., Rischel R (1975) *Deterministic and Stochastic Optimal Control*. Springer Verlag, Berlin.

The Euler equation to characterize optimal policies of discounted Markov decision processes: applications to economic growth models

Gabriel Zacarías-Espinoza, Hugo Cruz-Suárez
and Raúl Montes-de-Oca *†‡

Abstract

In this paper, necessary and sufficient conditions to obtain optimal policies of suitable discounted Markov decision processes (MDPs) are presented. These conditions include concavity and differentiability properties required for the utility and the transition law of the MDPs considered. This permits to establish a version of the Euler equation which essentially gives the characterization of the corresponding optimal polices. The conditions assumed are standard in economic growth models, and, in particular, an economic growth model with a logarithm utility function is provided.

keywords: Markov decision process, total discounted utility, Euler equation, economic growth model.

1 Introduction

This article deals with Markov decision processes (MDPs) (see [2] and [8]) with both the state and the control spaces given by the set of nonnegative real numbers. The total discounted reward (or total discounted utility) is taken as the objective function.

In the MDPs theory it is common to characterize the optimal solution, i.e. the optimal policy and the optimal value function by means of the Bellman equation or the dynamic programming equation (see [1], [2], [8], [10], and [15]).

Besides, in specific economic applications modeled as MDPs, it is well-known that the optimal solution can be characterized by means of a functional equation which involves the optimal solution and which is called Euler equation (EE) (see [4], [11], [12], [13], [14] and [15]). This EE requires mainly differentiability conditions on the transition law and the reward function (see [3] and [5]).

On the other hand, in [5] different versions of the EE have been presented for the optimal value function, but not for the optimal policy of discounted MDPs.

Therefore, the main contribution of this paper is to provide necessary and sufficient conditions which permit to obtain a version of EE in order to characterize the optimal policies of discounted MDPs with the description as in first paragraph of this Introduction. It is important to mention that the conditions presented are of the type of concavity and differentiability on the reward and transition law, so they are satisfied for a wide class of economic models (see [4], [11], [12], [13], [14], and [15]). The theory developed is illustrated in detail by means of a stochastic growth economic model. Finally, the EE has been applied in various economic branches, for example, in asset prices (see [6] and [7]) and in economic growth (see [4], [11], [12], [13], [14], and [15]).

The paper is organized as follows. Section 2 gives the basics on discounted MDPs. In the Section 3 the main result on the EE is provided. Section 4 develops the stochastic growth economic example, and Section 5 supplies the Conclusions.

2 Discounted Markov decision processes

Let $(X, A, \{A(x)|x \in X\}, Q, r)$ be a discrete-time, stationary *Markov control model* (see [2] and [8] for notation and terminology) where $X = A = [0, \infty)$ are called the *state* and the *action* (or *control*) spaces, respectively. $\{A(x)|x \in X\}$ is a family of subsets $A(x)$ of A, $x \in X$ with $int(A(x)) \neq \emptyset$ ($int(A(x))$ denotes the interior of $A(x)$). Here $A(x)$ represents the corresponding set of admissible actions for the state x. Suppose that $A(x)$ is convex for all $x \in X$, and that the correspondence $x \to A(x)$ is convex, i.e., $\theta a + (1-\theta)a' \in A(\theta x + (1-\theta)x')$ for all $x, x' \in X$, $a \in A(x)$, $a' \in A(x')$, and $\theta \in (0, 1)$. Q is the *transition probability law* which is induced by the following difference equation:

$$x_{t+1} = L(F(x_t, a_t), \xi_t), \qquad (1)$$

$t = 0, 1, \ldots$, with the initial state $x_0 = x$, where $\{\xi_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables taking values in $S = (0, \infty)$. Let ξ be a generic element of the sequence $\{\xi_t\}$. $L : X' \times S \to X$

*Zacarías-Espinoza G. and Montes-de-Oca R. are with the Department of Mathematics, UAM-Iztapalapa, Ave. San Rafael Atlixco 186, Col. Vicentina, México 09340, México. e-mails: gabrielzaces@hotmail.com, momr@xanum.uam.mx

†Cruz-Suárez H., Facultad de Ciencias Físico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio y Río Verde, Col. San Manuel, CU, 72570 Puebla, Pue., México. e-mail: hcs@fcfm.buap.mx

‡Manuscript received...; revised ....

and $F : \mathbb{K} \to X'$ are measurable functions, with $X' \subset X$. $r : \mathbb{K} \to \mathbb{R}$ is the *reward* (or *utility*) function, where $\mathbb{K} := \{(x, a) | \, x \in X, a \in A(x)\}$.

A *control policy* π is a (measurable, possibly randomized) rule for choosing actions, and at each time $t = 0, 1, \ldots$ the control prescribed by π may depend on the current state as well as on the history of previous states and actions. We denote the set of all policies by Π. Given the initial state $x_0 = x$, any policy π defines the unique probability distribution of the state-action processes $\{(x_t, a_t)\}$. For details see, for instance, [2] or [8]. This probability distribution is denoted by P_x^π, whereas E_x^π stands for the corresponding expectation operator. Let $\mathbb{F}$ be the set of all measurable functions $f : X \to A$, such that $f(x) \in A(x)$ for every $x \in X$. A policy $\pi \in \Pi$ is *stationary* if there exists $f \in \mathbb{F}$ such that, under π, the control $f(x)$ is applied at each time $t = 0, 1, \ldots$. The set of all stationary policies is identified with $\mathbb{F}$. The focus here is on the *expected total discounted reward* presented as

$$v(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t) \right], \tag{2}$$

where the policy $\pi \in \Pi$ is used, and $x \in X$is the initial state. In (2), $\alpha \in (0, 1)$ is a given discounted factor.

A policy π^* is said to be *optimal* if $v(\pi^*, x) = V(x)$, $x \in X$, where $V(x) := \sup_{\pi \in \Pi} v(\pi, x)$, $x \in X$, is the so-called *optimal value function*.

The following Basic Assumption (BA) is essentially the dynamic programming technique. The BA is fulfilled in a wide variety of cases, see [2] or [8] and require no further comments. Throughout the paper, MDPs that satisfy BA are taken into account.

Assumption 2.1. *(BA)*

a) The optimal value function V satisfies the Optimality Equation:

$$V(x) = \sup_{a \in A(x)} \{r(x, a) + \alpha E[V(L(F(x, a), \xi))]\},$$

$x \in X$.

b) There exists $f^* \in \mathbb{F}$, such that, for each $x \in X$,

$$V(x) = r(x, f^*(x)) + \alpha E[V(L(F(x, f^*(x)), \xi))].$$

c) $v_n(x) \to V(x)$, for each $x \in X$, where

$$v_n(x) = \sup_{a \in A(x)} \{r(x, a) + \alpha E[v_{n-1}(L(F(x, a), \xi))]\},$$

for $x \in X$ and $n = 1, 2, \ldots$, with $v_0(x) = 0$.

3 The Euler equation

Let $G : \mathbb{K} \to \mathbb{R}$ given by

$$G(x, a) := r(x, a) + \alpha H(x, a), \tag{3}$$

where $H(x, a) := E\left[V\left(L(F(x, a), \xi)\right)\right]$, $(x, a) \in \mathbb{K}$.

Let X and Y be Euclidean spaces. $C^2(X, Y)$ denotes the set of functions $l : X \to Y$ with a continuous second derivative (when $X = Y$, $C^2(X, Y)$ will be denoted by $C^2(X)$). Let $\Gamma \in C^2(X \times Y, \mathbb{R})$. Γ_x and Γ_y denote the partial derivatives of Γ for x and y, respectively. The second partial derivatives of Γ are Γ_{xx}, Γ_{xy}, Γ_{yx}, and Γ_{yy}. Let f^* be the optimal policy guaranteed by BA.

Assumption 3.1. *a) r, $H \in C^2(int(\mathbb{K}); \mathbb{R})$;*

b) For each $x \in int(X)$: $r_a(x, \cdot) > 0$, $r_{aa}(x, \cdot) < 0$, $H_{aa}(x, \cdot) \leq 0$;

c) For each $s \in S$: $L(\cdot, s)$ is strictly concave in X' and $L(\cdot, s) \in C^2(int(X'); X)$ with $L'(\cdot, s) > 0$; $F \in C^2(int(\mathbb{K}); X')$, $F(\cdot)$ strictly concave in $\mathbb{K}$ and with $F_a(x, \cdot) < 0$.

d) For each $x \in X$, $f^*(x) \in int(A(x))$.

Consider W defined for $(x, a) \in \mathbb{K}$ as

$$W(x, a) := \left[r_x - r_a F_a^{-1} F_x\right](x, a), \tag{4}$$

Note that Assumption 3.1 ensures that W is well defined and is continuously differentiable. Moreover, observe that the term F_a^{-1} is well defined thanks to Assumption 3.1 c).

The proof of the next lemma can be found in [3] or [5].

Lemma 3.2. *Assumption 3.1 implies that $V \in C^2(int(X); \mathbb{R})$ and*

$$V'(x) = W(x, f(x)), \textit{for each } x \in int(X). \tag{5}$$

Now the main result of the paper will be presented.

Theorem 3.3. *Under Assumption 3.1 we have: if $f \in \mathbb{F}$ is optimal, then f satisfies the Euler equation:*

$$r_a F_a^{-1}(x, f(x)) + \alpha E[W(L(\digamma, \xi), f(L(\digamma, \xi)))L'(\digamma, \xi)] = 0, \tag{6}$$

for each $x \in int(X)$, where $\digamma := F(x, f(x))$ and W is defined in (4).

Conversely, if $f \in \mathbb{F}$ satisfies (6) for each $x \in int(X)$ and

$$\lim_{n \to \infty} \alpha^n E_x^f \left[W(x_n, f(x_n)) x_n\right] = 0, \tag{7}$$

where $\{x_n\}$ is the trajectory generated by f with $x_0 = x$, then f is optimal.

Proof. Consider a fixed $x \in X$ and let $f \in \mathbb{F}$ be optimal. Note that lemma 3.2 implies that $G \in C^2(int(\mathbb{K}); \mathbb{R})$ where G is defined in (3). Then for the first order condition

ISBN: 978-1-61804-230-9

it results that $G_a(x, f(x)) = 0$, and for the invertibility of F_a it follows that

$$\begin{aligned} & - r_a F_a^{-1}(x, f(x)) \\ & = \alpha E \left[V' \left(L \left(F(x, f(x)), \xi \right) \right) L'(F(x, f(x)), \xi) \right]. \end{aligned} \tag{8}$$

On the other hand, since V satisfies (5), it is obtained that

$$\begin{aligned} & V' \left(L \left(F(x, f(x)), \xi \right) \right) \\ & = W \left(L(F(x, f(x)), \xi), f \left(L(F(x, f(x)), \xi) \right) \right). \end{aligned} \tag{9}$$

Finally, (6) follows from substituting (9) in (8).

Conversely, take a fixed $x \in X$ and let $f \in \mathbb{F}$ such that f satisfies (6) and (7). Take $g \in \mathbb{F}$. Let $a_t = f(x_t)$, and $b_t = g(y_t)$, where $t = 0, 1, \ldots$, be the trajectories of controls generated by f and g, respectively, with $x_0 = y_0 = x$.

As r, L and F are strictly concave and belong to C^2, it results that

$$\begin{aligned} & r(x_t, a_t) - r(y_t, b_t) \\ & \geq r_x(x_t, a_t)(x_t - y_t) - r_a(x_t, a_t)(a_t - b_t). \end{aligned} \tag{10}$$

$$\begin{aligned} & x_{t+1} - y_{t+1} \\ & \geq L'(F(x_t, a_t), \xi_t)(F(x_t, a_t) - F(y_t, b_t)). \end{aligned} \tag{11}$$

$$\begin{aligned} & F(x_t, a_t) - F(y_t, b_t) \\ & \geq F_x(x_t, a_t)(x_t - y_t) + F_a(x_t, a_t)(a_t - b_t). \end{aligned} \tag{12}$$

Since $L'(\cdot, \xi) > 0$ and $F_a(x, \cdot) < 0$ and using (11) y (12), it follows that

$$\begin{aligned} & (a_t - b_t) \\ & \geq F_a^{-1}(x_t, a_t)[{L'}^{-1}(F(x_t, a_t), \xi_t)(x_{t+1} - y_{t+1}) \\ & - F_x(x_t, a_t)(x_t - y_t)]. \end{aligned} \tag{13}$$

To ease the computations, we use the following notation: for $t = 0, 1, 2 \ldots$, $rx_t := r_x(x_t, a_t)$, $ra_t := r_a(x_t, a_t)$, $Fa_t := F_a(x_t, a_t)$, $Fx_t := F_x(x_t, a_t)$, and $L_t := L'(F(x_t, a_t), \xi_t)$.

Hence, from (10) and (13) it is obtained that

$$\begin{aligned} & r(x_t, a_t) - r(y_t, b_t) \geq rx_t(x_t - y_t) \\ & + ra_t Fa_t^{-1}[L_t^{-1}(x_{t+1} - y_{t+1}) - Fx_t(x_t - y_t)]. \end{aligned}$$

Multiplying the last inequality by α^t, adding from 0 to n, and using the fact that $x_0 = y_0$, it results that

$$\begin{aligned} & \sum_{t=0}^{n} \alpha^t \left(r(x_t, a_t) - r(y_t, b_t) \right) \\ & \geq \sum_{t=0}^{n} \alpha^t rx_t(x_t - y_t) + \alpha^t ra_t Fa_t^{-1} \\ & \left(L_t^{-1}(x_{t+1} - y_{t+1}) - Fx_t(x_t - y_t) \right) \\ & = \sum_{t=1}^{n-1} \left[\alpha^t \left(ra_t Fa_t^{-1} + \alpha W(x_{t+1}, a_{t+1}) L_t \right) L_t^{-1}(x_t - y_t) \right] \\ & + \alpha^n ra_n Fa_n^{-1} L_n^{-1}(x_{n+1} - y_{n+1}) \\ & \geq \sum_{t=1}^{n-1} \left[\alpha^t \left(ra_t Fa_t^{-1} + \alpha W(x_{t+1}, a_{t+1}) L_t \right) L_t^{-1}(x_t - y_t) \right] \\ & + \alpha^n ra_n Fa_n^{-1} L_n^{-1} x_{n+1}. \end{aligned}$$

where the last inequality is a consequence of the fact that $x_{n+1} - y_{n+1} \leq x_{n+1}$ and $r_a F_a^{-1} {L'}^{-1} < 0$.

On the other hand, from (1), using the elementary properties of the conditional expectation, and as f satisfies (6), it follows that

$$\begin{aligned} & \alpha^n E_x^f \left[ra_n Fa_n^{-1} L_n^{-1} x_{n+1} \right] \\ & = -\alpha^{n+1} E_x^f \left[W(x_{n+1}, a_{n+1}) x_{n+1} \right]. \end{aligned}$$

In a similar way it is possible to conclude that

$$\begin{aligned} & E_x^f \left[\left(ra_t Fa_t^{-1} + \alpha W(x_{t+1}, a_{t+1}) L_t \right) L_t^{-1}(x_t - y_t) \right] \\ & = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} & E_x^f \left[\sum_{t=0}^{n} \alpha^t \left(r(x_t, a_t) - r(y_t, b_t) \right) \right] \\ & \geq \alpha^{n+1} E_x^f \left[W(x_{n+1}, a_{n+1}) x_{n+1} \right]. \end{aligned}$$

Now, letting n tend to ∞, and since f satisfies (7), we conclude that

$$E_x^f \left[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t) \right] \geq E_x^f \left[\sum_{t=0}^{\infty} \alpha^t r(y_t, b_t) \right].$$

Hence, f is optimal. □

4 An economic growth model with a logarithm utility (EGMLU)

Suppose that $r(x_t, a_t) = \ln(a_t)$ and that the transition probability law is given by $x_{t+1} = \xi_t (x_t^\gamma - a_t)$, with $\gamma \in (0, 1)$, $a_t \in [0, x_t^\gamma]$ and $x_0 = x > 0$, where $\{\xi_t\}$ is a sequence of i.i.d. random variables, taking values in $S := (0, 1)$. Suppose that $\mu_\gamma := E[\ln(\xi^\gamma)]$ is finite. The density of ξ is designated by $\Delta \in C^2((0, \infty))$.

ISBN: 978-1-61804-230-9

Remark 4.1. *The economic growth model presented here is a stochastic version of an economic growth model given in [3].*

The proof of the next two lemmas can be found in [4].

Lemma 4.2. *Assumption 2.1 is valid for example EGMLU*

Lemma 4.3. *For each $n = 1, 2, \ldots$, $v_n(x) = \gamma k_n \ln(x) + c_n$, and $f_n(x) = \frac{x^\gamma}{k_n}$, $x \in X$, with $k_n = \sum_{t=0}^{n-1} (\alpha\gamma)^t$, $n = 1, 2, \ldots$, where*

$$f_n(x) \in \arg \max_{a \in A(x)} \{\ln(a) + \alpha E[v_{n-1}(\xi(x^\gamma - a))]\}.$$

Note that for $x \geq 0$, $f_n(x) \to \tilde{f}(x)$, where

$$\tilde{f}(x) := x^\gamma(1 - \alpha\gamma), \tag{14}$$

and $\tilde{f}(x) \in (0, x^\gamma)$, i.e., $\tilde{f}$ is a stationary policy. Moreover, the evaluation of $\tilde{f}$ in (2) permits to obtain that $v(\tilde{f}, x) = K \ln(x) + C > -\infty$, with $K, C \in \mathbb{R}$.

Corollary 4.4. *For Example GEMLU,*

$$V(x) = \frac{\gamma}{1 - \alpha\gamma} \ln(x) + C,$$

$x > 0$, where $C = \frac{1}{1-\alpha}[\ln(1-\alpha\gamma) + \frac{\alpha\gamma}{1-\alpha\gamma}(\mu_\gamma + \ln(\alpha\gamma))]$, and $\tilde{f}$ is the optimal policy, where $\tilde{f}$ is defined in (14).

Proof. Fix $x > 0$. As $v_n(x) \to V(x)$, and $k_n \to 1/(1 - \alpha\gamma)$, $n \to \infty$, (notice that $0 < \alpha\gamma < 1$), hence $\{c_n\}$ is convergent. Let $C := \lim_{n\to\infty} c_n$. Therefore,

$$V(x) = \frac{\gamma}{1 - \alpha\gamma} \ln(x) + C.$$

On the other hand, since $U'(a) = 1/a$ and $h'(x) = \gamma x^{\gamma-1}$, then

$$\alpha E[h'(\xi(h(x) - \tilde{f}(x)))U'\left(\tilde{f}\left(\xi\left(h(x) - \tilde{f}(x)\right)\right)\right)\xi]$$
$$= \alpha E\left[\frac{\gamma(\xi(x^\gamma - x^\gamma(1 - \alpha\gamma)))^{\gamma-1}}{(\xi(x^\gamma - x^\gamma(1 - \alpha\gamma)))^\gamma (1 - \alpha\gamma)}\xi\right]$$

and a direct computation permits to obtain that $\tilde{f}$ satisfies (6). Besides,

$$\alpha^t E\left[h'(x_t)U'(\tilde{f}(x_t))x_t\right] = \frac{\alpha^t\gamma}{1 - \alpha\gamma},$$

then, letting $t \to \infty$ allows to verify that (7) holds. So Theorem 3.3 yields that $\tilde{f}$ is the optimal policy.

On the other hand, since V satisfies

$$\frac{\gamma}{1 - \alpha\gamma} \ln(x) + C$$
$$= \sup_{a\in[0,x^\gamma]} \{\ln(a) + \alpha \frac{\gamma}{1 - \alpha\gamma} E\left[\ln(\xi(x^\gamma - a))\right] + \alpha C\},$$

then, substituting $\tilde{f}$ in the last equation, yields C. □

5 Conclusion

With the theory presented in this paper it is possible to characterize the optimal policies of several economic applications of discounted MDPs. Moreover, the EE presented open the possibility to study the approximation of the optimal policies by means of numerical techniques (see [9]).

References

[1] Arrow K. J., *Note on uncertainty and discounting in models of economic growth*, J. Risk Unc. 38, 87-94, (2009).

[2] Bertsekas D. P., Dynamic Programming: Deterministic and Stochastic Models, Prentice-Hall, Belmont, 1987.

[3] Cruz-Suárez H. and Montes-de-Oca R., *An envelope theorem and some applications to discounted Markov decision processes*, Math. Methods Oper. Res. 67, 299-321, (2008).

[4] Cruz-Suárez H., Montes-de-Oca R. and Zacarías G., *A Consumption-Investment problem modelled as a discounted Markov decision process*, Kybernetika 47, 740-760, (2011).

[5] Cruz-Suárez H., Zacarías-Espinoza G. and Vázquez-Guevara V., *A Version of the Euler Equation in Discounted Markov Decision Processes*, Journal of Applied Mathematic, Volume 2012, Article ID 103698, doi:10.1155/2012/103698, 1-16, (2012).

[6] Epstein L. and Zin S., *Substitution, risk aversion, and the temporal behavior of consumption and asset returns I: Theoretical framework*, Econometrica, 57, 937-969, (1989).

[7] Gurkaynak R. S., *Econometric tests of asset price bubbles: taking stock*, J. Econom. Surveys, 22, 166186, (2008).

[8] Hernández-Lerma O. and Lasserre J. B., Discrete-Time Markov Control Processes: Basic Optimality Criteria, Springer-Verlag, New York, 1996.

[9] Kenneth L. Judd, Numerical Methods in Economics, Massachusetts Institute of technology, London, 1998.

[10] Jaśkiewics A. and Nowak A.S., *Discounted dynamic programming with unbounded returns: application to economic models*, J. Math. Anal. Appl. 378, 450-462, (2011).

[11] Kamihigashi T., *Stochastic optimal growth with bounded or unbounded utility and bounded or unbounded shocks*, J. Math. Econom. 43, 477-500, (2007).

[12] Le Van C. and Dana R.A., Dynamic Programing in Economics, Kluwer Academic Publishers, Boston, (2003).

[13] Nishimura K. and Stachurski J., *Stability of stochastic optimal growth models: a new approach*, J. Econom. Theory, 122, 100–118, (2005).

[14] Ramsey F. P., *A Mathematical theory of saving*, Econ. J. 38, 543-559, (1928).

[15] Stokey N., Lucas R. and Prescott E., Recursive Methods in Economic Dynamics, Harvard University Press, Cambridge, 1989.

Acknowledgment

This work was supported in part by CONACYT (Mexico) and ASCR (Czech Republic) under Grant No. 171396.

The Non-Physical Finite Element Method: Modelling Normal Shock Waves in Fluids

R. Darvizeh and K. Davey

***Abstract*—** This paper presents the theory and application of a newly developed finite element based shock capturing scheme. The approach is founded on the general Reynolds transport forms of the governing conservation laws replacing the more traditional differential approach and is shown to more amenable to shock physics. The new formulation makes use of transport equations involving non-physical variables, where each conservation law gives rise to a non-physical field. The non-physical variables can be shown to have limiting continuity at a shock which permits the use of the classical continuous Galerkin finite element method along with features that allow for localised approximation at a shock. An additional feature is the presence of a non-physical source term at a shock whose strength is related to the magnitude of any discontinuity present. The numerical results are found to be in near perfect correspondence with predictions from analytical models.

***Keywords*—**finite element, fluids, non-physical, shock capturing.

I. Introduction

THE present work belongs to the category of finite element based techniques in computational fluid dynamics [1]. Traditional mesh-based methods are believed to be on the whole not well suited to the treatment of material discontinuities [2]. Finite element based techniques such as the discontinuous Galerkin finite element method (DGFEM) or extended finite element method (XFEM) are capable of modelling phase change and shock problems [3,4] but they can suffer from singularity, instability, poor conditioning and significant extra computational costs (in comparison with classical FEM) for dynamic phase change and high rate problems [5,6]. They involve basis functions not typical to classical finite elements involving discontinuous or enriched shape functions and weightings and consequently their implementation in the commercial FE software is not straight forward [7,8].

The non-physical finite element method (NPFEM), is introduced as a novel approach designed to overcome these difficulties and increase the capability of classical FE based numerical algorithms and software in modelling phase change, shock and diffusion problems [9,10]. The NPFEM is formulated on the basis of the continuous classical FEM (i.e. in Sobolev space H^1) without any violation or extension (unlike DGFEM and XFEM). The central core of the NPFEM is based on three main features which are

(i) introducing a non-physical variable as a limiting-continuous mathematical object;
(ii) the use of the general Reynolds transport (integral based) conservation laws in place of standard differential forms;
(iii) relating the physical fields to their associated non-physical variables through general transport formulations;
(iv) the application of theory pertaining to a moving control volume (MCV) within another MCV (typically one MCV follows the boundary of the continuum body whilst the other tracks the material discontinuity in an element or through the whole domain).

Limiting-continuity refers to the feature that the non-physical variable is continuous over the system but has a source like behaviour at the place of a shock. This feature permits the discretisation of the non-physical variable with the continuous polynomial basis standard to the classical Galerkin FEM (CGFEM).

The NPFEM can in principle cater for any type of material discontinuity in solids and fluids [10, 11-12] but the scope of the present work is limited to showing the ability and potential of the NPFEM in capturing shock waves in fluids. The problems selected here are restricted to piston induced shock problems to showcase the methodology although the theory is presented in a generic form.

II. General Transport Equation

The integral form of the transport equation for a control volume Ω moving with velocity $\underline{\mathrm{v}}^*$ can be written as

$$\frac{D^*}{D^*t}\int_\Omega \rho\psi dV + \int_\Gamma \rho\psi\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}^*\right)\cdot \underline{n} d\Gamma = \int_\Gamma \underline{J}\cdot \underline{n} d\Gamma + \int_\Omega \rho b dV \quad (1)$$

where ψ is physical field per unit mass, ρ is density, $\underline{n}$ is an outward pointing unit normal to the CV boundary Γ, $\underline{J}\cdot\underline{n}$ is the flux, $\underline{\mathrm{v}}$ is material velocity and b represents a body force. .

Authors would like to acknowledge School of MACE, University of Manchester for the pump priming research funding which helped the authors to extend the NPFEM to different case studies. Also the authors would like to acknowledge Electricité de France (EDF) funding and support during preliminary work on the theory for a single material discontinuity.

K. Davey is in the School of MACE, University of Manchester, Manchester, UK (corresponding author to provide phone: +44(0)161-306-3834; e-mail: keith.davey@manchester.ac.uk).

R. Darvizeh is in the School of MACE, University of Manchester, Manchester, UK (e-mail: roohoolamin.darvizeh@manchester.ac.uk).

The operator D^*/D^*t is the total derivative with respect to a reference system, and is related to the physical spatial system via the identity $D^*/D^*t = \partial/\partial t|_{\boldsymbol{x}} + \underline{\mathrm{v}}^* \cdot \nabla$ where $\boldsymbol{x}$ refers to a spatial coordinate and $\nabla = \partial/\partial \boldsymbol{x}$. The derivative is also related to the material derivative through the relation $D^*/D^*t = D/Dt + \left(\underline{\mathrm{v}}^* - \underline{\mathrm{v}}\right)\cdot\nabla$ where $\underline{\mathrm{v}}^*$ is the velocity of the moving control volume. It should be noted that both the material and the control volume Ω can be arranged to contain the same spatial points despite having different velocity fields provided the condition $\underline{\mathrm{v}}^* \cdot \underline{n} = \underline{\mathrm{v}} \cdot \underline{n}$ applies on Γ.

In the absence of any discontinuity in the system, Equation (1) can be reduced to its equivalent PDE form of the conservation transport equations which is

$$\rho \frac{D^*\psi}{D^*t} + \rho\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}^*\right)\cdot\nabla\psi = -\operatorname{div}\left(\underline{J}\right) + \rho b \tag{2a}$$

which is is deduced on applying a Reynold-type Theorem and the Divergence Theorem to Equation (1) and the use of continuity equation

$$\frac{D^*\rho}{D^*t} + \rho \operatorname{div}\underline{\mathrm{v}}^* + \operatorname{div}\left(\rho\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}^*\right)\right) = 0 \tag{2b}$$

The integral form of the conservation laws, Equation (1) is called general transport equation since it can be applied to continua with discontinuity while Equation (2a) cannot be applied at a discontinuity. Applying Equation (1) to the small control volume Ω_i with a moving material discontinuity depicted in Fig. 1 results

$$\begin{aligned} \int_{\Gamma_i^\ell} \rho\psi\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}_{\mathrm{i}}\right)\cdot\underline{n}_s d\Gamma - \int_{\Gamma_i^s} \rho\psi\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}_{\mathrm{i}}\right)\cdot\underline{n}_s d\Gamma \\ = -\int_{\Gamma_i^\ell} \underline{J}\cdot\underline{n}_s d\Gamma + \int_{\Gamma_i^s} \underline{J}\cdot\underline{n}_s d\Gamma \end{aligned} \tag{3}$$

which provides the well-known Rankine-Hugoniot jump condition $\left\lfloor\left\lfloor \rho\psi\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}_{\mathrm{i}}\right)\cdot\underline{n}\right\rfloor\right\rfloor = -\left\lfloor\left[\underline{\mathrm{J}}\cdot\underline{n}\right]\right\rfloor$, where the square brackets denote evaluation at either side of the interface Γ_i, e.g. $\left\lfloor\left[\underline{\mathrm{J}}\cdot\underline{n}\right]\right\rfloor = \underline{\mathrm{J}}_\ell \cdot \underline{n}_\ell + \underline{\mathrm{J}}_s \cdot \underline{n}_s$.

III. Non-physical Transport Equation

To make Equation (1) more readily solvable using CGFEM, the physical field variable ψ should be related to a mathematically equivalent continuous variable which is defined by

$$\begin{aligned} \frac{D^*}{D^*t}\int_\Omega \widehat{\psi} dV = \frac{D^*}{D^*t}\int_\Omega \rho\psi dV + \int_\Gamma \rho\psi\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}^*\right)\cdot\underline{n} d\Gamma \\ = -\int_\Gamma \underline{J}\cdot\underline{n} d\Gamma + \int_\Omega \rho b dV \end{aligned} \tag{4}$$

where $\widehat{\psi}$ is called non-physical variable and Equation (4) is termed the non-physical or equivalent transport equation.

Applying Equation (4) to a small control volume Ω_i moving with interface Γ_i at velocity $\underline{\mathrm{v}}_i$ (as presented in Fig. 1) results in

$$\frac{D^+}{D^+t}\int_{\Gamma_i} \widehat{\psi}' d\Gamma = \int_{\Gamma_i} \left[\rho\psi\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}_{\mathrm{i}}\right)\cdot\underline{n}_s\right] d\Gamma = -\int_{\Gamma_i} \left[\underline{J}\cdot\underline{n}_s\right] d\Gamma \tag{5}$$

where $\widehat{\psi}'$ is called the non-physical source and Equation (5) is called the source equation.

The operator $D^+/D^+t\left(= \partial/\partial t|_{\mathbf{x}} + \underline{\mathrm{v}}^+ \cdot \nabla\right)$ stands for the total derivative with respect to the second CRS which moves with velocity $\underline{\mathrm{v}}^+$ with the conditions $\underline{\mathrm{v}}^+ \cdot \underline{n}_i = \underline{\mathrm{v}}_i \cdot \underline{n}_i$ on Γ_i. and $\underline{\mathrm{v}}^+ \cdot \underline{n} = \underline{\mathrm{v}} \cdot \underline{n}$ on Γ. The first condition ensures that this CRS tracks the interface Γ_i. and the second condition matches the overall movement of the newly induced CV with domain Ω since both have the same spatial points but pose different velocity fields $\underline{\mathrm{v}}$ and $\underline{\mathrm{v}}^+$, respectively.

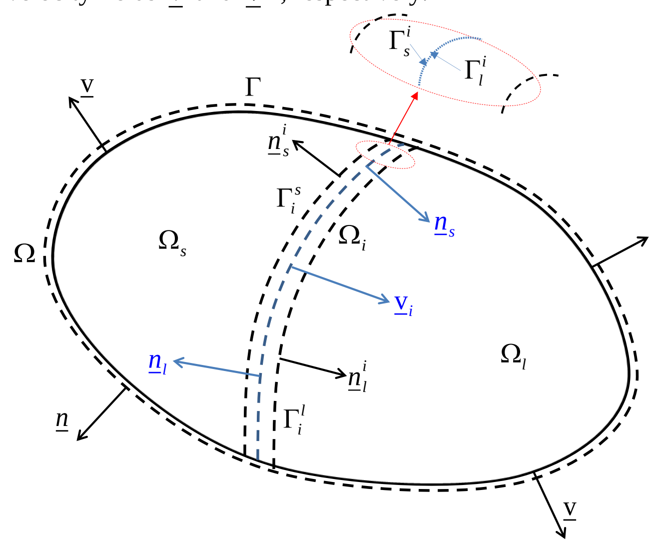


Fig. 1 moving interface in a control volume

A. *Three Possible Non-Physical Approaches*

The definition of the non-physical variable $\widehat{\psi}$ afforded by Equation (4) has the advantage of simplicity but suffers for two principal disadvantages. Ideally it would be preferable for $\widehat{\psi}$ to be some constant multiple of ψ in regions (elements) absence of any discontinuity. It is apparent from Equation (4) that this is not the case since $\widehat{\psi}$ is influenced by the advective-like term $\int_\Gamma \rho\psi\left(\underline{\mathrm{v}} - \underline{\mathrm{v}}^*\right)\cdot\underline{n} d\Gamma$. The method also has the peculiar property that $\widehat{\psi}$ changes with $\underline{\mathrm{v}}^*$ although this is more a conceptual difficulty rather than a practical one.

A principal advantage with the present definition for $\widehat{\psi}$ however is that it is permissible for $\widehat{\psi}$ to be discontinuous at element boundaries; a property arising because no boundary terms involving $\widehat{\psi}$ are present on the left hand side of Equation (4). Bringing boundary terms into the left hand side of Equation (4) is evidently a possibility. Two possible alternative formulations for the definition of $\widehat{\psi}$ are

$$\begin{aligned}&\frac{D^*}{D^*t}\int_\Omega \widehat{\psi}dV-\int_\Gamma \widehat{\psi}\underline{v}^*\cdot\underline{n}d\Gamma=\\&\frac{D^*}{D^*t}\int_\Omega \rho\psi dV+\int_\Gamma \rho\psi\left(\underline{v}-\underline{v}^*\right)\cdot\underline{n}d\Gamma=-\int_\Gamma \underline{J}\cdot\underline{n}d\Gamma+\int_\Omega \rho b dV\end{aligned}\tag{6}$$

$$\begin{aligned}&\frac{D^*}{D^*t}\int_\Omega \widehat{\psi}dV+\int_\Gamma \widehat{\psi}\left(\underline{v}-\underline{v}^*\right)\cdot\underline{n}d\Gamma=\\&\frac{D^*}{D^*t}\int_\Omega \rho\psi dV+\int_\Gamma \rho\psi\left(\underline{v}-\underline{v}^*\right)\cdot\underline{n}d\Gamma=-\int_\Gamma \underline{J}\cdot\underline{n}d\Gamma+\int_\Omega \rho b dV\end{aligned}\tag{7}$$

where in all cases continuity of $\widehat{\psi}$ is insisted upon at any shock, i.e. $\left[\left[\widehat{\psi}\underline{n}\right]\right]_{\Gamma_i}=0$ and the two associated source equations for these equations are

$$\begin{aligned}\frac{D^+}{D^+t}\int_{\Gamma_i}\widehat{\psi}'dV-\int_{\Sigma_i}\widehat{\psi}'\underline{v}^+\cdot\underline{tn}d\Gamma&=\int_{\Gamma_i}\left[\left[\rho\psi\left(\underline{v}-\underline{v}^+\right)\cdot\underline{n}\right]\right]d\Gamma\\&=-\int_{\Gamma_i}\left[\left[\underline{J}\cdot\underline{n}\right]\right]d\Gamma\end{aligned}\tag{8}$$

$$\begin{aligned}\frac{D^+}{D^+t}\int_{\Gamma_i}\widehat{\psi}'dV-\int_{\Sigma_i}\widehat{\psi}'\underline{v}^+\cdot\underline{tn}d\Gamma+\int_{\Gamma_i}\left[\left[\widehat{\psi}\left(\underline{v}-\underline{v}^+\right)\cdot\underline{n}\right]\right]d\Gamma\\=-\int_{\Gamma_i}\left[\left[\underline{J}\cdot\underline{n}\right]\right]d\Gamma\end{aligned}\tag{9}$$

where the square bracket signifies a jump and $\underline{tn}$ is a outward pointing unit normal on $\sum_i=\Gamma\cap\Gamma_i$ ($\sum_i$ is the boundary for Γ_i) that is in a tangent plane of Γ_i and in Equation (9) the integral

$$\begin{aligned}\int_{\Gamma_i}\left[\left[\widehat{\psi}\left(\underline{v}-\underline{v}^+\right)\cdot\underline{n}\right]\right]d\Gamma&=\int_{\Gamma_i}\widehat{\psi}\left[\left[\left(\underline{v}-\underline{v}^+\right)\cdot\underline{n}\right]\right]d\Gamma\\&=\int_{\Gamma_i}\widehat{\psi}\dot{m}''\left[\left[\rho^{-1}\right]\right]d\Gamma\end{aligned}\tag{10}$$

where $\dot{m}''$ is the mass flux rate.

Equation (4), which is used through this work, provides the simplest assembled system but as mentioned suffers in that it is non-local in the sense that non-physical variable $\widehat{\psi}$ is affected by the term $\int_\Gamma \rho\psi\left(\underline{v}-\underline{v}^*\right)\cdot\underline{n}d\Gamma$ apart from when $\underline{v}=\underline{v}^*=\underline{0}$. This approach can be contrasted against Equation (6) which provides a local definition for $\widehat{\psi}$ when $\underline{v}=\underline{0}$ but not otherwise. The presence of a boundary term on the left hand side of Equation (6) system makes things slightly more involved and inter-element continuity of $\widehat{\psi}$ is a requirement with this approach for $\underline{v}^*\neq\underline{0}$. Equation (7), is local to only those elements that possess a discontinuity (can set $\widehat{\psi}=\rho\psi$ otherwise) but the assembled form is more complex but typical to conduction-advection type systems. The source equation is more involved also since both source $\widehat{\psi}'$ and non-physical $\widehat{\psi}$ terms are present, so a coupled analysis is required. Inter-element continuity of $\widehat{\psi}$ is a requirement with this approach.

IV. NON-PHYSICAL FINITE ELEMENT FORMULATIONS

A. *Galerkin Weighted Equivalent Transport Equation*

The continuous Galerkin weighting of Equation (4) gives

$$\begin{aligned}&\frac{D^*}{D^*t}\int_{\Omega_e}N_i\widehat{\psi}dV=\frac{D^*}{D^*t}\int_{\Omega_e}N_i\rho\psi dV\\&-\int_{\Omega_e}\rho\psi\left(\underline{v}-\underline{v}^*\right)\cdot\nabla N_i dV+\int_{\Gamma_e}N_i\rho\psi\left(\underline{v}-\underline{v}^*\right)\cdot\underline{n}d\Gamma\\&=\int_{\Omega_e}\nabla N_i\cdot\underline{J}dV-\int_{\Gamma_e}N_i\underline{J}\cdot\underline{n}d\Gamma+\int_{\Omega_e}\rho N_i b dV\end{aligned}\tag{11}$$

Using of the continuous Galerkin shape functions N_i guarantees the continuity of the weightings across the discontinuity. Combining Equation (4) and the weighted form of the source equation, Equation (5), results

$$\begin{aligned}&\frac{D^*}{D^*t}\int_{\Omega_e/\Gamma_i}N_i\widehat{\psi}dV=\\&\int_{\Omega_e}\nabla N_i\cdot\underline{J}dV-\int_{\Gamma_e}N_i\underline{J}\cdot\underline{n}d\Gamma+\int_{\Omega_e}\rho N_i b dV\\&+\int_{\Gamma_i^e}N_i\left[\underline{J}\cdot\underline{n}_s\right]d\Gamma+\int_{\Gamma_i^e}\nabla N_i\cdot\left(\underline{v}^*-\underline{v}^\times\right)\widehat{\psi}'d\Gamma\end{aligned}\tag{12}$$

Also writing the LHS of Equation (4) and the Equation (5) for an element domain gives

$$\begin{aligned}\frac{D^*}{D^*t}\int_{\Omega_e/\Gamma_i}\widehat{\psi}dV=\frac{D^*}{D^*t}\int_{\Omega_e}\rho\psi dV+\int_{\Gamma_e}\rho\psi\left(\underline{v}-\underline{v}^*\right)\cdot\underline{n}d\Gamma\\-\int_{\Gamma_i^\times}\rho_s\left(\underline{v}_s-\underline{v}_i\right)\cdot\underline{n}_s\left[\psi\right]d\Gamma\end{aligned}\tag{13}$$

$$\begin{aligned}\frac{D^\times}{D^\times t}\int_{\Gamma_i^e}\widehat{\psi}'d\Gamma+\int_{\Sigma_i^e}\widehat{\psi}'\left(\underline{v}^*-\underline{v}^\times\right)\cdot\underline{tn}\,d\Sigma+\int_{\Gamma_i^e}\left(\underline{v}^*-\underline{v}^\times\right)\widehat{\psi}'d\Gamma\\=\int_{\Gamma_i^e}\rho_s\left(\underline{v}_s-\underline{v}_i\right)\cdot\underline{n}_s\left[\psi\right]d\Gamma=-\int_{\Gamma_i^e}\left[\underline{J}\cdot\underline{n}_s\right]d\Gamma\end{aligned}\tag{14}$$

To maintain consistency of notation, the $+$ sign is replaced by $\times$ for an element domain (e) and consequently D^+/D^+t and $\underline{v}^+$ changes to $D^\times/D^\times t$ and $\underline{v}^\times$ respectively.

Equations (12), (13) and (14) are termed the weighted equivalent governing equation, linkage equation and source equation, respectively. Because of the way of the NPFEM is constructed, the linkage Equation (13) and source Equation (14) do not need to be assembled and should be kept in a un-weighted form and solved only for the elements which experience the moving discontinuity. Table 1 summarizes the unknowns and Equations which should be solved.

Equation No.	Equation	Form	Associated Unknown
(12)	Equivalent Governing	Weighted and Assembled	Physical Feild
(13)	Linkage	Non-Weighted	Non-physical Variable
(14)	Source	Non-Weighted	Non-physical Source

Table 1 NPFEM equations which are solved using schemes (a) and (c) as depicted in Fig. 2

B. Discretisation Schemes and Solution Algorithms

The non-physical variable $\widehat{\psi}$ is approximated by $\widehat{\psi}=\sum_{j=1}^{m_e} N_j \widehat{\psi}_j$ on an element. The nodal values of $\widehat{\psi}_j$ are determined by an assembled form of Equation (12). To use Equation (12) for updating the physical field ψ_j values at each node, the non-physical variable should be related to the physical field. Due to non-uniqueness of the non-physical variable, $\widehat{\psi}$, it can defined in different ways. For example, the $\widehat{\psi}$ can be written as a function of level set variable ϕ for 2-D and 3-D problems, where a level set method [13] is used to track the interface, i.e. $\widehat{\psi}(\phi)=\widehat{C}\phi$,where $\widehat{C}$ is called the non-physical capacitance. Another strategy which is used through this work is based on defining the non-physical variable $\widehat{\psi}$ in an element as a function of physical variable values at the nodes; this approach gives

$$\widehat{\psi}_e=\widehat{C}_e\overline{\psi}_e \tag{15}$$

where $\overline{\psi}_e=\sum_i N_i\psi_e^i$, $\widehat{C}_e$ is the non-physical capacitance of the element e and ψ_e^i is the value of the physical variable ψ at i^{th} node of the element.

Also the non-physical variable $\hat{\psi}$ can be defined as

$$\widehat{\psi}_e=\sum_i N_i\widehat{C}_i\psi_e^i \tag{16}$$

where $\widehat{C}_i$ is the nodal non-physical capacitance associated with node i.

Based on Equations (15) and (16), two different discretisation schemes can be followed in solving a problem using NPFEM (See Fig. 2). The variety of schemes possible are as a consequence of the non-uniqueness associated with the non-physical variable.

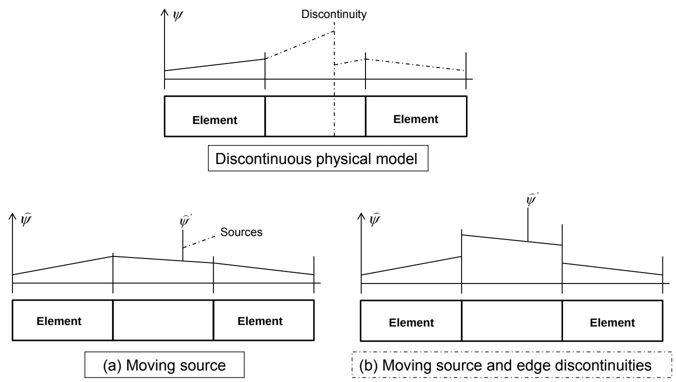


Fig. 2. 1-D representation of possible NPFEM schemes (a) and (b)

Also it can be shown that the flux term $\underline{J}$ in Equation (12) can be discretised using standard Galerkin shape functions even though $\underline{J}$ is discontinuous at elements that include a moving discontinuity. It is important to note here that there is no discontinuity on the left hand side of Equation (12). Consequently the right hand side of Equations (12) is also absent of discontinuity achieved primarily by isolating the source term. The following proposition shows that the right hand side of Equations (12) is also absent of discontinuity.

Proposition IV.B.

The time integration of $\int_{\Gamma_i^e}\widehat{\psi}_i'\left(\underline{\mathrm{v}}_N^*-(\underline{\mathrm{v}}_i^\times)_N\right)\cdot\nabla N_i d\Gamma$, (i.e. $\int_{t_n}^{t_{n+1}}\int_{\Gamma_i^e}\widehat{\psi}_i'\left(\underline{\mathrm{v}}_N^*-(\underline{\mathrm{v}}_i^\times)_N\right)\cdot\nabla N_i d\Gamma dt$) annihilates any discontinuities arising in $\int_{t_n}^{t_{n+1}}\int_{\Omega_e}\nabla N_i\cdot\underline{J}dVdt$.

Proof. Using Lebesgue integration, integration of $\int_{\Delta V_k}\nabla \mathrm{N_i}\cdot\underline{\mathrm{J}}\mathrm{dV}$ over time interval $\alpha\Delta t$, where $\alpha\in[0,1]$, results the following measure relationship.

$$\begin{aligned}\int_{\Delta t}\int_{\Delta V_k}\nabla N_i\cdot\underline{J}dVdt &=\nabla N_i\cdot\underline{J}_k\Delta V_k\Delta t\\ &+\alpha\nabla N_i\cdot\left(\underline{J}_{k+1}-\underline{J}_k\right)\Delta V_k\Delta t\end{aligned} \tag{17}$$

where the small volume ΔV_k is the swept volume by the boundary Γ_k over time Δt in the computational frame Ω. Application of the jump condition $\rho_{k+1}\left(\underline{\mathrm{v}}_{k+1}-\underline{\mathrm{v}}_{\mathrm{i}}\right)\cdot\underline{n}_k\left(\psi_k-\psi_{k+1}\right)=\left(\underline{J}_{k+1}-\underline{J}_k\right)\cdot\underline{n}_k$ and replacing ΔV_k by $A\left(\underline{\mathrm{v}}_k^\times-\underline{\mathrm{v}}^*\right)\cdot\underline{n}_k\Delta t$ in Equation (17) gives

ISBN: 978-1-61804-230-9

$$
\begin{aligned}
&\int_{\Delta t}\int_{\Delta V_k} \nabla N_i \cdot \underline{J} dVdt = \nabla N_i \cdot \underline{J}_k \Delta V_k \Delta t \\
&+\alpha \nabla N_i \cdot \rho_{k+1}\left(\underline{v}_{k+1} - \underline{v}_k^{\times}\right)\left(\psi_k - \psi_{k+1}\right)\Delta V_k \Delta t \\
&= \nabla N_i \cdot \underline{J}_k \Delta V_k \Delta t \\
&+\alpha \nabla N_i \cdot \rho_{k+1}\left(\underline{v}_{k+1} - \underline{v}_k^{\times}\right)\left(\psi_k - \psi_{k+1}\right) A\left(\underline{v}_k^{\times} - \underline{v}^{*}\right)\cdot \underline{n}_k \Delta t \Delta t
\end{aligned} \tag{18}
$$

Performing temporal Integration of the source relation, Equation (14), on the boundary Γ_k yields

$$
\hat{\psi}'_{\alpha} - \hat{\psi}'_0 = \alpha \rho_{k+1}\left(\underline{v}_k - \underline{v}_i\right)\cdot \underline{n}_k \Delta t \left(\psi_k - \psi_{k+1}\right) \tag{19}
$$

Incorporation of Equation (19) in $\int_{\Delta t}\int_{\Gamma_k} \nabla N_i \cdot \left(\underline{v}^* - \underline{v}_k^{\times}\right)\hat{\psi}' d\Gamma dt$ gives the following measure relationship.

$$
\begin{aligned}
&\int_{\Delta t}\int_{\Gamma_k} \nabla N_i \cdot \left(\underline{v}^* - \underline{v}_k^{\times}\right)\hat{\psi}' d\Gamma dt = \\
&\int_{\Delta t} \hat{\psi}'\left(\underline{v}^* - \underline{v}^{\times}\right)\cdot \nabla_{\Gamma_k} N_i dt = \hat{\psi}'_0\left(\underline{v}^* - \underline{v}_k^{\times}\right)\cdot \nabla_{\Gamma_k} N_i A \Delta t \\
&+\alpha \rho_{k+1}\left(\underline{v}_k - \underline{v}_k^{\times}\right)\cdot \underline{n}_k \left(\psi_k - \psi_{k+1}\right)\Delta t A\left(\underline{v}^* - \underline{v}_k^{\times}\right)\cdot \nabla_{\Gamma_k} N_i \Delta t
\end{aligned} \tag{20}
$$

It is evident that second terms on the RHS of Equations (18) and (20) are identical and of opposite sign and will eliminated on addition. □

Application of Proposition IV.B yields a simplified form of Equation (12), i.e.

$$
\begin{aligned}
\frac{D^*}{D^*t}\int_{\Omega_e/\{\Gamma_i : i\in k_e\}} N_i \hat{\psi} dV &= \int_{\Omega_e} \nabla N_i \cdot \underline{J} dV - \int_{\Gamma_e} N_i \underline{J}\cdot \underline{n} d\Gamma \\
&+ \int_{\Omega_e} \rho N_i \cdot b dV + \sum_{i\in k_e}\int_{\Gamma_i^e} N_i\left[\underline{J}\cdot \underline{n}_i\right] d\Gamma
\end{aligned} \tag{21}
$$

In addition the $\underline{J}$ in the element with discontinuity can be approximated using continuous shape functions as $\underline{J}_e = \sum_i N_i \underline{J}_e^i$, where N_i is shape function and $\underline{J}_e^i$ is the value of the $\underline{J}$ at i^{th} node of the element.

Therefore Equation (21) possesses no discontinuity. The possibility of removing the source term from Equation (12) and isolating it in Equation (14) which is shown through Proposition 4.1 reduces the Table 1 to Table 2.

Equation No.	Equation	Form	Associated Unknown
(12)	Equivalent Governing	Weighted and Assembled	Physical Field
(13)	Linkage	Non-Weighted	Non-physical Variable

Table 2 summary of NPFEM equations which should be solved using schemes (c) and (d) as depicted in Fig. 3

The solution schemes (a) and (b) presented in Fig. 2. are simplified to schemes (c) and (d) respectively which are depicted in Fig. 3.

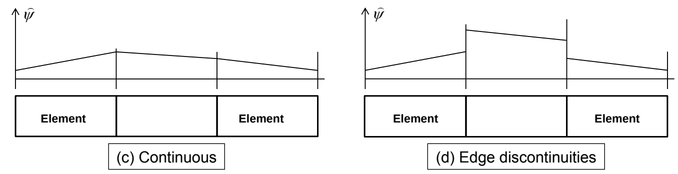


Fig. 3. 1-D representation of possible NPFEM schemes (c) and (d)

To solve the system of non-linear algebraic equations produced by the NPFEM, linear homotopy continuation method [14] is employed.

V. CASE STUDY: ATTENUATING NORMAL SHOCK

This case study considers an attenuating moving normal shock in the piston-cylinder model. The physical and discretised models are illustrated in Figs. 4 and 5. The length of the cylinder is 1 m and the piston moves with the velocity V_2 and induces a moving shock with velocity V_i. The shock wave divides the volume of the cylinder into two regions which are denoted by region 2 and region 1 (Fig. 5). The particle velocity in the compressed region, region 2, is assumed to be spatially uniform and equals to the piston velocity V_2. The particle velocity in region 1 is equal to 0 m/s. Due to the spatially uniform and linear distributions of the physical fields, the NPFEM results returned for this case study are mesh size insensitive. For the sake of simplicity and clarity a 6-element ALE mesh is considered as a discretised finite element model.

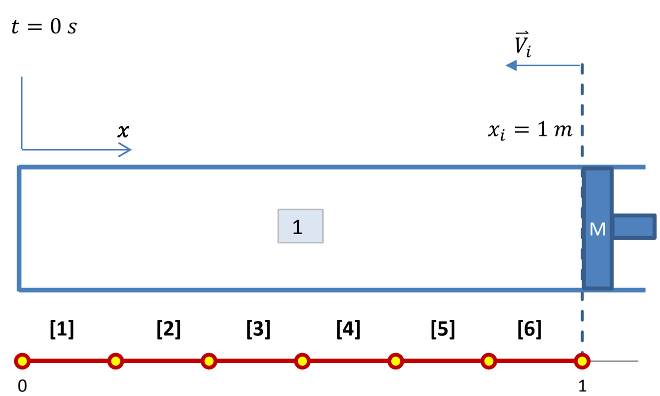


Fig. 4 piston-cylinder model at initial state

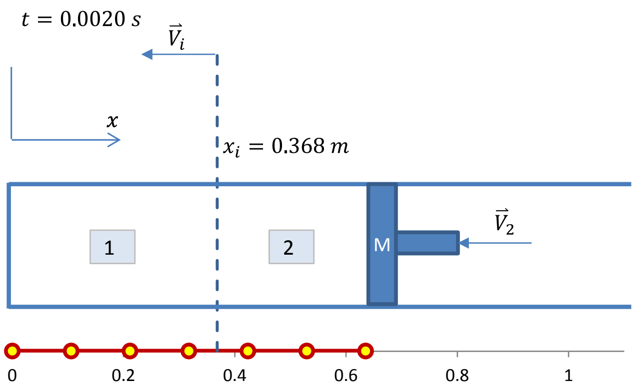


Fig. 5 intermediate state at t=0.0020 s

Theoretically, it can be shown that to have transient behaviour typical to an attenuating shock the pressure at the piston and the shock wave in region 2 must be different. A pressure boundary condition at the piston together with the use of linear shape functions induces a pressure distribution in the finite element results.

The material properties and input data for physical and non-physical variables supposed for this model are presented in Tables 3 and 4. In Table 4, $\widehat{C}_0$, $\widehat{D}_0$ and $\widehat{\mathrm{v}}_0'$ denote the initial values of the non-physical density capacitance, no-physical velocity capacitance, non-physical density and velocity sources respectively. The non-physical variables initial values are arbitrary values and it can be shown theoretically that the NPFEM results are independent of choice of the non-physical variables values [11, 12].

ρ_1 (Kg/m^3)	ρ_2 (Kg/m^3)	P_1 (KPa)	C_p $(KJ/Kg\,^oK)$
1.284	2.5	101.375	1.0049

Table 3 material properties of dry air $\gamma = 1.40$

$\widehat{C}_0$ (Kg/m^3)	$\widehat{D}_0$ (Kg/m^3)	$\widehat{\mathrm{v}}_0'$ $(Pa \cdot s)$
3	3	0

Table 4 initial values of the non-physical variables

The graphs presented in Figs. 6 to 9 are NPFEM results contrasted against analytical-model[1] results for the shock wave velocity, V_i, the piston or compressed region particle velocity, V_2 and the temporal behaviour of the pressure in the compressed region at the place of shock and piston (P_{2j} and P_2) respectively. The maximum relative error between the analytical and NPFEM results is presented in each graph. These errors are for the 6-element ALE model with a total process time $t = 0.001357\ s$ and time increment, $\Delta t = 9.05e^{-5}\ s$ which is equivalent to 15 time increments.

[1] The analytical model will be published in journal version of this paper.

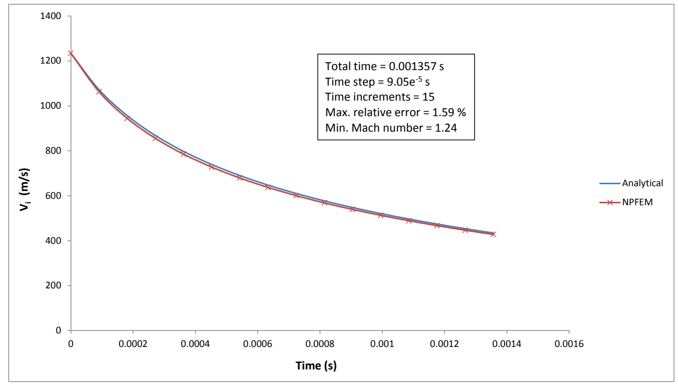


Fig. 6. Shock wave velocity versus time

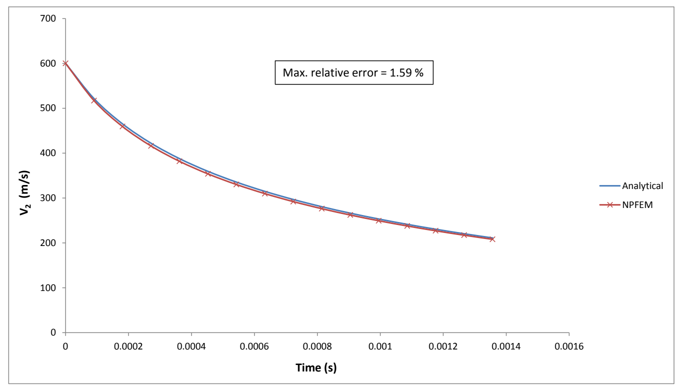


Fig. 7. Piston or particle velocity versus time

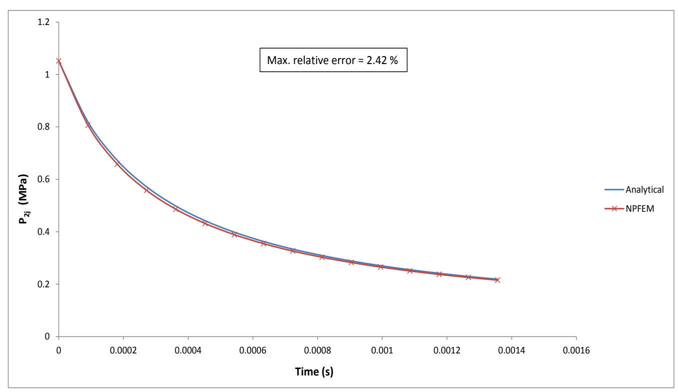


Fig. 8. Pressure at the place of shock versus time

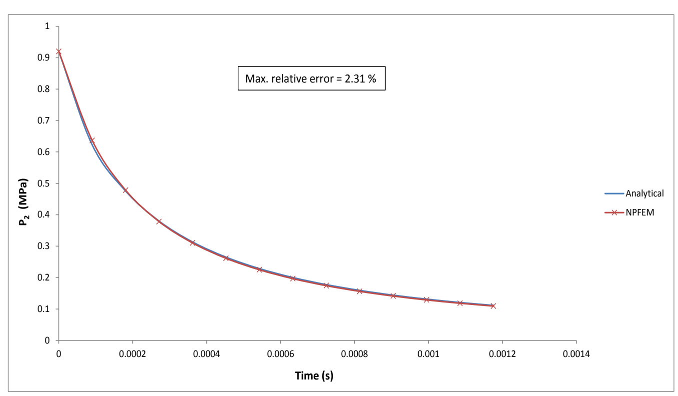


Fig. 9 pressure at the place of piston versus time

The physical results presented in Figs. 6 to 9 and their maximum relative errors show good agreement between analytical and NPFEM results for a highly coarse mesh and relatively coarse time discretisation. As mentioned earlier, because of the linear distribution of the physical fields and the way that NPFEM captures the shock, the numerical results are not mesh size sensitive. However, due to the nonlinear temporal behaviour of the physical fields, the NPFEM results are sensitive to time discretisation.

The value of the non-physical density capacitance is the same for elements at each time step and it does not make any difference that an element experiences the moving discontinuity or not. This special property comes from the nature of the conservation of mass governing equation and the way the non-physical variable is defined and linked to it, i.e.

$$\frac{D^*}{D^*t}\int_{\Omega_e/\Gamma_i}\hat{\rho}dV = 0 \tag{22}$$

In a transient model this feature does not hold for other non-physical capacitances such as non-physical velocity capacitance. The results of the case study is based on scheme (b) of the NPFEM (See Fig. 2) and initial values for the non-physical capacitances are presented in Table 4. The NPFEM results for the temporal behaviour of the non-physical density and velocity capacitances are shown in Figs. 10 and 11 respectively.

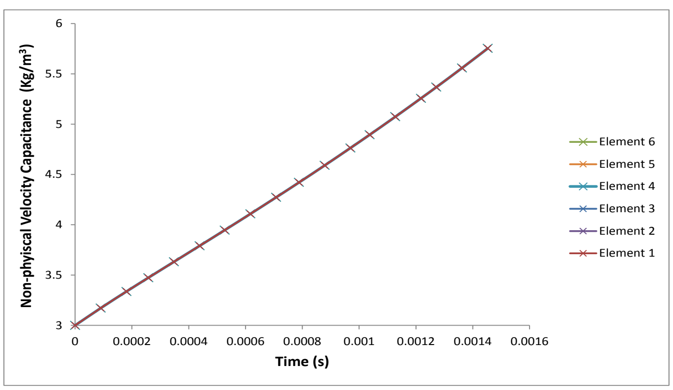


Fig. 10 elemental non-physical density capacitance

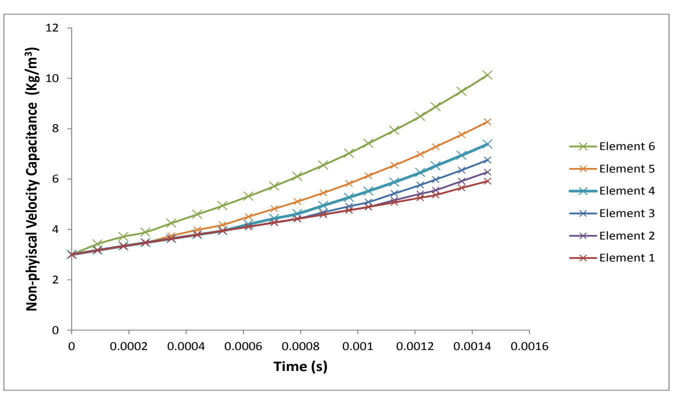


Fig. 11 elemental non-physical velocity capacitance

Fig. 11 illustrates this fact that the values of the non-physical velocities are the same for all the elements as long as they do not experience the shock wave. This feature is a direct consequence of the manner in which the non-physical variable is related to the physical fields.

Theoretically, there is no non-physical source term for density. The variation of the non-physical velocity source versus time is presented in Fig. 12. It is evident that the value of the non-physical source in each element is non-zero as long as it experiences the moving shock. Fig. 12 shows that the non-physical source for each element gets a non-zero value when the shock moves inside the element. This behavior arises through the definition of the non-physical source and its direct relation to the value of the physical field jump.

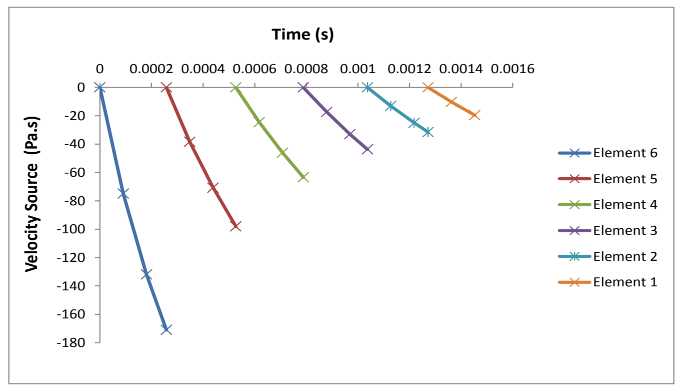


Fig. 12 non-physical velocity sources

Based on the theory of the non-physical method and the non-physical variable definition, the non-physical variable and consequently the non-physical capacitance and the non-physical source are non-unique. This property of the non-physical method is illustrated and proved in reference [9]. As a result of the non-uniqueness of the non-physical variables, the physical results of the NPFEM are insensitive to the manner in which there are defined and to their initial values.

The presented results are based on scheme (b) solution algorithm to demonstrate results for both non-physical variables and sources. It is theoretically clear that the most efficient scheme is the scheme (d). Also to show it in practical sense, the CPU time of the schemes (a-d) are compared in Fig. 13. The codes of all schemes are run in a PC with (AMD Athlon II X2, 2.90 GHz, 4.00 GB RAM) feature.

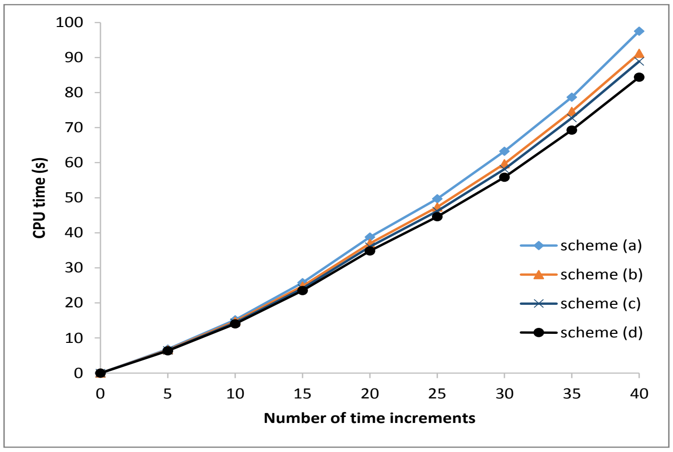


Fig. 13 comparison of CPU time for solution algorithm schemes

Fig. 13 supports the theoretical fact that the difference between computational costs of the schemes increases with decreasing the size of time discretisation. And so the distinction of scheme (d) becomes more apparent. It should be noted that the effect of spatial mesh size or the number of elements on the CPU time difference of the solution schemes is negligible for the presented case study. This is because with 1–D, 2-noded elements the difference in the number of nodes and the number of elements, and consequently their associated number of equations, is always 1.

VI. Conclusion

The integral (transport) formulation is used to represent the weak form of the conservation laws. Also the strategy of MCV in a MCV is employed to capture the moving discontinuity, i.e. moving control volume (MCV) in a control volume which the first CV surrounds the main body and the second CV tracks the shock wave.

The non-physical method and its finite element representation, NPFEM, are introduced. In the presented numerical technique, NPFEM, for each conservation law, three equations are required to be solved instead of one. This is a consequence of introducing two new variables, i.e. the non-physical variable $\hat{\psi}$ and the non-physical source $\hat{\psi}'$. It is proved mathematically that discontinuities are absent from the finite element formulation of the equivalent governing equation and therefore it can be weighted and discretised using the CGFEM. This important feature of the NPFEM formulations is due to continuous behaviour of the non-physical variable $\hat{\psi}$ at every point except at the shock position where it is shown that it has a source like behaviour $\hat{\psi}'$. This limiting continuity of the non-physical variable allows annihilation and isolation of the mathematical discontinuity in the fields in the equivalent transport equations. The verification and flexibility of NPFEM is illustrated through a 1-D transient case study and using four different solution schemes. The NPFEM results show a precise agreement with the results of an analytical model. It is shown that despite to finite difference based techniques the stationary and transient behaviour of the shock wave in 1-D shock tube can be captured in NPFEM by using coarse spatial meshes with relatively few nodes. Also in comparison with finite element based numerical methods, the NPFEM works completely in the frame work of CGFEM without any extension (e.g. CVFEM, XFEM) or violation (e.g. DGFEM). This features of the NPFEM should facilitate its straightforward employment in existing commercial FEM software without any major difficulties. Finally, the NPFEM is a novel strategy which still needs further research and development especially for 2-D and 3-D cases.

References

[1] E. R. Stein, R. de Borst, and T. J. R. Hughes, *Encyclopedia of Computational Mechanics*. John Wiley, 2004, vol. 3.

[2] A. Legay, H. Wang, and T. Belytschko, "Strong and weak arbitrary discontinuities in spectral finite elements", *Int. J. for Num.Meth. in Eng.*, vol. 4, no. 8, pp. 991-1008, 2005.

[3] T. Belytschko, R. Gracie, and G. Ventura, "A review of extended/generalized finite element methods for material modeling", *Mod. and Sim. in Mat. Sci. and Eng.*, vol. 17, no. 4, pp. 043001, 2009.

[4] F. Bassi, and S. Rebay, "High-order accurate discontinuous finite element solution of the 2D Euler equations", *J. of Comp. Phys.*, vol. 138, no. 2, pp. 251-285, 1997.

[5] B. Cockburn, and C.-W. Shu, "The Runge–Kutta discontinuous Galerkin method for conservation laws V: multidimensional systems", *J. of Comp. Phys.*, vol. 141, no. 2, pp. 199-224, 1998.

[6] I. Babuška, and U. Banerjee, "Stable generalized finite element method (SGFEM)", *Comp. meth. in app. Mech. and eng.*, vol. 201, pp. 91-111, 2012.

[7] T. Strouboulis, I. Babuška, and K. Copps, "The design and analysis of the generalized finite element method", *Comp. meth. in app. Mech. and eng.*, vol. 181, no. 1, pp. 43-69, 2000.

[8] L. Noels, and R. Radovitzky, "A general discontinuous Galerkin method for finite hyperelasticity. Formulation and numerical applications", *Int. J. for Num. Meth. in Eng.*, vol. 68, no. 1, pp. 64-97, 2006.

[9] R. Darvizeh, and K. Davey, "Strategies for incorporating material discontinuities into finite flement formulations", in *Computational Methods For Engineering Science*, B. H. V. Topping, Ed. Saxe-Coburg Publications: Stirlingshire, 2012, pp. 167-192.

[10] K. Davey, and R. Mondragon, "A non-physical enthalpy method for the numerical solution of isothermal solidification", *Int. J. for Num. Meth. in Eng.*, vol. 84, no. 2, pp. 214-252, 2010.

[11] R. Darvizeh, and K. Davey, "The annihilation of physical discontinuites in finite elements", in *Proc. 19th UK National Conf. of the Association for Computational Mechanics in Engineering*, Heriot-Watt University: Edinburgh, 2011,. pp. 273-276.

[12] R. Mondragon, and K. Davey, "Weak discontinuity annihilation in solidification modelling", *Comp. & Struct.*, vol. 89, no. 7, pp. 681-701, 2011.

[13] D.-J Wang, Y. Tang, H.-C Yu, and Z.-S Tang, "Level set methods based on distance function", App. Math. and Mech., vol. 24, no. 8, pp. 950-960, 2003..

[14] E. L. Allgower, and K. Georg, *Introduction to Numerical Continuation Methods*. Society for Industrial and Applied Mathematics, 2003.

Some Results on Harmonic Mappings Related to Janowski Alpha Spirallike Functions

Melike Aydoğan

Abstract— Let $f(z)=\mathrm{h(z)}+\overline{g(\mathrm{z})}$ be a univalent sense-preserving harmonic mapping of the open unit disc $D=\{z||z|<1\}$. If $f(z)$ satisfies the condition $|\omega(z)|=\left|\frac{g'(z)}{h'(z)}\right|<k,\ 0<k<1$ then $f(\mathrm{z})$ is called k-quasiconformal harmonic mapping in D. In the present paper we will give some properties of the class of conformal mappings related to starlike function of complex order.

Keywords— Coefficient inequality, Conformal mapping, Growth theorem, Subordination principle.

I. Introduction

LET Ω the family of functions $\phi(z)$ regular in D and satisfying the conditions $\phi(0)=0,\ |\phi(z)|<1$ for every $z\in D.$

Next, for arbitrary fixed numbers $A,\ B,\ -1<A\le 1,$ $-1\le B<A,$ denote by $P(A,B)$ if and only if

$$p(z)=\frac{1+A\phi(z)}{1+B\phi(z)} \tag{1}$$

for some $\phi(\mathrm{z})\in\Omega$ and every $z\in D.$

Moreover, let $S^*(\mathrm{A},\mathrm{B},\alpha)$ denote the family of functions $h(z)=z+c_2z^2+\dots$ regular in D and such that $h(\mathrm{z})$ is in $S^*(A,B,\alpha)$ if and only if there is a real number α for which

$$e^{i\alpha}z\frac{h'(z)}{h(z)}=p(z)cos\alpha+isin\alpha,|\alpha|<\frac{\pi}{2},p(z)\in P(A,B) \tag{2}$$

is true for every $z\in D.$ Let $s_1(z)=z+d_2z^2+\dots$ and $s_2(z)=z+e_2z^2+\dots$ be analytic functions in $D.$ If there exists a function $\phi(z)\in\Omega$ such that $s_1(z)=s_2(\phi(z))$ for all $z\in\mathcal{D},$ then we say that $s_1(z)$ is subordinate to $s_2(z)$ and we write $s_1(z)\prec s_2(z).$ Specially if $s_2(z)$ is univalent in $D,$ then $s_1(z)\prec s_2(z)$ if and only if $s_1(\mathcal{D})\subset s_2(\mathcal{D})$ implies $s_1(\mathcal{D}_r)\subset s_2(\mathcal{D}_r),$ where $\mathcal{D}_r=\{z\,||z|<r,0<r<1\}.$ ([2]-[4])

Finally, a planar harmonic mapping in the open unit disc D is a complex-valued harmonic function $f(z)$ which maps D onto the some planar domain $f(\mathrm{D}).$ Since D is a simply connected domain, the mapping f has a canonical decomposition $f(z)=h(z)+\overline{g(z)},$ where $h(z)$ and $g(z)$ are analytic in D and have the following power series expansion,

$$h(z)=\sum_{n=0}^{\infty}a_nz^n,$$

$$g(z)=\sum_{n=0}^{\infty}b_nz^n$$

where $a_n,b_n\in\mathbb{C},\ \ n=0,1,2,\dots$ as usual we call $h(z)$ the analytic part of $f(z)$ and $g(z)$ is co-analytic part of $f(z).$ An elegant and complete treatment theory of the harmonic mapping is given Duren's monograph [3].

Lewy [7] proved in 1936 that the harmonic function $f(z)$ is locally univalent in D if and only if its Jacobien

$$J_f=|h'(z)|-|g'(z)|$$

is different from zero in $D.$ In the view of this result, locally univalent harmonic mappings in the open unit disc D are either sense-preserving if $|h'(z)|>|g'(z)|$ in D or sense-reversing if $|g'(z)|>|h'(z)|$ in $D.$ Throughout this paper, we will restrict ourselves to the study of sense-preserving harmonic mappings. We also note that $f(z)=h(z)+\overline{g(z)}$ is sense-preserving in D if and only if $h'(z)$ does not vanish in D and the second dilatation $|\omega(z)|=\left|\frac{g'(z)}{h'(z)}\right|$ has the

M. Aydoğan is with the Department of Mathematics, Işık University, Meşrutiyet Köyü Şile, Istanbul, Turkey (e-mail: melike.aydogan@isikun.edu.tr).

property $\left|\omega(z)\right|<1$ for all $z\in D$. Therefore, the class of all sense-preserving harmonic mappings in the open unit disc D with $a_0=b_0=0$ and $a_1=1$ will be denoted by S_H. Thus S_H contains standard class S of univalent functions. The family of all mappings $f\in S_H$ with the additional property $g'(0)=0$, i.e., $b_1=0$ is denoted by S_H^0. Hence, it is clear that $S\subset S_H^0\subset S_H$. A univalent harmonic mapping is called k-quasiconformal ($0<k<1$) if $\left|\omega(z)\right|<k$. For the general definition of quasiconformal mapping, see[3].

The main purpose of this paper is to give some properties of the k-quasiconformal harmonic mappings

$$S_H^*(kq,A,B,\alpha)=\left\{\begin{array}{l} f(z)=h(z)+\overline{g(z)}\in S_H \mid \\ \left|\omega(z)\right|<k, 0<k<1, \\ h(z)\in S^*(A,B,\alpha)\end{array}\right\}.$$

For this investigation we will need the following theorem and lemma.

Theorem 1([6])

Let $h(z)$ be an element of $S^*(A,B,\alpha)$, then

$$rF(A,B,\alpha,-r)\le\left|h(z)\right|\le rF(A,B,\alpha,r), B\neq 0,$$
$$rF_1(A,\alpha,-r)\le\left|h(z)\right|\le rF_1(A,\alpha,r), B=0$$

where

$$F(A,B,\alpha,r)=(1-Br)^{\frac{(A-B)cos\alpha(cos\alpha-1)}{2B}}\cdot(1+Br)^{\frac{(A-B)cos\alpha(cos\alpha+1)}{2B}},$$

$$F_1(A,\alpha,r)=e^{(Acos\alpha)r}.$$

and

$$\left[(1-Ar)cos\alpha-(1-Br)sin\alpha\right]F_*(A,B,\alpha,-r)\le\left|h'(z)\right|$$
$$\le\left[(1+Ar)cos\alpha+(1+Br)sin\alpha\right]F_*(A,B,\alpha,r), B\neq 0,$$

$$(1-Ar)cos\alpha e^{-(Acos\alpha)r}\le\left|h'(z)\right|$$
$$\le(1+Ar)cos\alpha e^{(Acos\alpha)r}, B=0$$

where

$$F_*(A,B,\alpha,r)=(1-Br)^{\frac{(A-B)cos\alpha(cos\alpha-1)}{2B}}\cdot(1+Br)^{\frac{(A-B)cos\alpha(cos\alpha+1)}{2B}-1}.$$

Lemma 1 ([5])

Let $\phi(z)$ be regular in the unit disk D with $\phi(0)=0$. Then if $\left|\phi(z)\right|$ attains its maximum value on the circle $\left|z\right|=r$ at the point z_1, one has $z_1\phi'(z_1)=k\phi(z_1)$ for $k\ge 1$.

II. MAIN RESULTS

Theorem 2

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S^*(kq,A,B,\alpha)$, then

$$\frac{g(z)}{h(z)}\prec\frac{k^2(b_1-z)}{k^2-\overline{b_1}z}. \tag{3}$$

Proof. We consider the linear transformation $\frac{k^2(b_1-z)}{k^2-\overline{b_1}z}$, this transformation maps $\left|z\right|<k$ onto itself. On the other hand, we have

$$\omega(z)=\frac{g'(z)}{h'(z)}=\frac{\left(b_1z+b_2z^2+\ldots\right)'}{\left(z+a_2z^2+\ldots\right)'}$$
$$=\frac{b_1+2b_2z+\ldots}{1+2a_2z+\ldots}\Rightarrow\omega(0)=b_1.$$

Therefore, the function

$$\phi(z)=\frac{k^2\left(b_1-\omega(z)\right)}{k^2-\overline{b_1}\omega(z)}$$

satisfies the conditions of Schwarz lemma, then we have

$$\omega(z)=\frac{k^2\left(b_1-\phi(z)\right)}{k^2-\overline{b_1}\phi(z)}\Rightarrow$$
$$\omega(z)=\frac{g'(z)}{h'(z)}=\frac{k^2\left(b_1-z\right)}{k^2-\overline{b_1}z}. \tag{4}$$

On the other hand, the linear transformation $\frac{k^2(b_1-z)}{(k^2-\overline{b_1}z)}$ maps $\left|z\right|=r$ onto the disc with the centre

$$C(r)=\left(\frac{k^2\left(1-r^2\right)Reb_1}{k^2-\left|b_1\right|^2r^2},\frac{k^2\left(1-r^2\right)Imb_1}{k^2-\left|b_1\right|^2r^2}\right),$$

and the radius

$$\rho(r)=\frac{k\left(k^2-\left|b_1\right|^2\right)r^2}{k^2-\left|b_1\right|^2r^2}.$$

Then, we write

$$\omega(\mathcal{D}_r)=\left\{z \mid \left|\omega(z)-\frac{k^2(1-r^2)b_1}{k^2-|b_1|^2r^2}\right|\le\frac{k\left(k^2-|b_1|^2\right)r}{k^2-|b_1|^2r^2}\right\}. \quad (5)$$

Now we define the function $\phi(z)$ by

$$\frac{g(z)}{h(z)}=\frac{k^2(b_1-\phi(z))}{k^2-\overline{b_1}\phi(z)}. \quad (6)$$

Then, $\phi(z)$ is analytic and $\phi(0)=\frac{k^2\cdot 0}{|b_1|^2-k^2}=0.$ If we take the derivative from (6) and after the brief calculations we get

$$\omega(z)=\frac{g'(z)}{h'(z)}=\frac{k^2(b_1-\phi(z))}{k^2-\overline{b_1}\phi(z)}+\frac{k^2\left(|b_1|^2+k^2-2b_1\phi(z)\right)\phi(z)}{\left(k^2-\overline{b_1}\phi(z)\right)^2}\cdot\frac{h(z)}{h'(z)}. \quad (7)$$

On the other hand, since $h(z)\in S^*(A,B,\alpha)$, then the boundary value of $z\frac{h'(z)}{h(z)}$ is

$$\frac{1-e^{-i\alpha}B(Acos\alpha+iBsin\alpha)r^2+re^{i\theta}(A-B)cos\alpha}{1-B^2r^2}.$$

Therefore, the equality (7) can be written in the following form on $|z|=r$.

$$\omega(z)=\frac{g'(z)}{h'(z)}=\frac{k^2(b_1-\phi(z))}{k^2-\overline{b_1}\phi(z)}+\frac{k^2(|b_1|^2+k^2-2b_1\phi(z))}{(k^2-b_1\phi(z))^2}\cdot\frac{(1-B^2r^2)}{1-e^{(-i\alpha)B(Acos\alpha+iBsin\alpha)r^2+re^{i\theta}(A-B)cos\alpha)}}. \quad (8)$$

In this step, if we use Jack lemma,

$$\omega(z_1)=\frac{g'(z_1)}{h'(z_1)}=\frac{k^2(b_1-\phi(z_1))}{k^2-\overline{b_1}\phi(z_1)}+\frac{k^2\left(|b_1|^2+k^2-2b_1\phi(z_1)\right)}{\left(k^2-\overline{b_1}\phi(z_1)\right)^2}\cdot\frac{1-B^2r^2}{1-B^2r^2}\begin{bmatrix}1-e^{-i\alpha}B(Acos\alpha+iBsin\alpha)r^2\\+re^{i\theta}(A-B)cos\alpha\end{bmatrix}\notin\omega(\mathcal{D}_r).$$

But, this contradicts to (5) because $|\phi(z_1)|=1, k\ge 1$ so our assumption is wrong, i.e., $|\phi(z)|<1$ for all $z\in\mathcal{D}$. Therefore (6) shows that

$$\frac{g(z)}{h(z)}\prec\frac{k^2(b_1-z)}{k^2-\overline{b_1}z}.$$

Corollary 1

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S^*(kq,A,B,\alpha)$, then

$$\frac{k(|b_1|-kr)}{k-|b_1|r}\le|\omega(z)|\le\frac{k(|b_1|+kr)}{k+|b_1|r}$$

$$\frac{(k-k|b_1|)+(|b_1|-k^2)r}{k+|b_1|r}\le 1-|\omega(z)|\le\frac{(k-k|b_1|)-(|b_1|-k^2)r}{k-|b_1|r}$$

$$\frac{(k+k|b_1|)-(|b_1|+k^2)r}{k-|b_1|r}\le 1+|\omega(z)|\le\frac{(k+k|b_1|)+(|b_1|+k^2)r}{k+|b_1|r}$$

This corollary is a simple consequence of the inequality (5).

Corollary 2

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S^*(kq,A,B,\alpha)$, then

$$rF(A,B,\alpha,-r)\frac{k(|b_1|-kr)}{k-|b_1|r}\le|g(z)|\le rF(A,B,\alpha,r)\frac{k(|b_1|+kr)}{k+|b_1|r},$$

$$rF_1(A,\alpha,-r)\frac{k(|b_1|-kr)}{k-|b_1|r}\le|g(z)|\le rF_1(A,\alpha,r)\frac{k(|b_1|+kr)}{k+|b_1|r}.$$

where

$$F(A,B,\alpha,r)=(1-Br)^{\frac{(A-B)cos\alpha(cos\alpha-1)}{2B}}(1+Br)^{\frac{(A-B)cos\alpha(cos\alpha+1)}{2B}},$$

$$F_1(A,\alpha,r)=e^{(Acos\alpha)r}.$$

ISBN: 978-1-61804-230-9

$$\frac{k\left(\left|b_1\right|-kr\right)}{k-\left|b_1\right|r}\begin{bmatrix}\left(1-Ar\right)cos\alpha \\ -\left(1-Br\right)sin\alpha\end{bmatrix}F_*\left(A,B,\alpha,-r\right)\leq$$

$$\left|g'\left(z\right)\right|\leq\frac{k\left(\left|b_1\right|+kr\right)}{k+\left|b_1\right|r}\begin{bmatrix}\left(1+Ar\right)cos\alpha \\ +\left(1+Br\right)sin\alpha\end{bmatrix}F_*\left(A,B,\alpha,r\right),$$

$B\neq 0,$

$$\left(1-Ar\right)cos\alpha e^{-(Acos\alpha)r}\frac{k\left(\left|b_1\right|-kr\right)}{k-\left|b_1\right|r}\leq$$

$$\left|g'\left(z\right)\right|\leq\left(1+Ar\right)cos\alpha e^{(Acos\alpha)r}\frac{k\left(\left|b_1\right|+kr\right)}{k+\left|b_1\right|r},B=0$$

where

$$F_*\left(A,B,\alpha,r\right)=\left(1-Br\right)^{\frac{(A-B)cos\alpha(cos\alpha-1)}{2B}}\cdot\left(1+Br\right)^{\frac{(A-B)cos\alpha(cos\alpha+1)}{2B}-1}.$$

Proof. Using Theorem 2 and inequality (5) we can write

$$\frac{k\left(\left|b_1\right|-kr\right)}{k-\left|b_1\right|r}\left|h\left(z\right)\right|\leq\left|g\left(z\right)\right|\leq\frac{k\left(\left|b_1\right|+kr\right)}{k+\left|b_1\right|r}\left|h\left(z\right)\right|, \quad (9)$$

$$\frac{k\left(\left|b_1\right|-kr\right)}{k-\left|b_1\right|r}\left|h'\left(z\right)\right|\leq\left|g'\left(z\right)\right|\leq\frac{k\left(\left|b_1\right|+kr\right)}{k+\left|b_1\right|r}\left|h'\left(z\right)\right|. \quad (10)$$

If we use Theorem 1 in the inequalities (8) and (9) we obtain the desired result.

Corollary 3

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S^*\left(kq,A,B,\alpha\right)$, then

$$[(1-Ar)cos\alpha-(1-Br)sin\alpha]^2[F_*(A,B,\alpha,-r)]^2\cdot G(k,|b_1|,r)\leq J_f\leq[(1+Ar)cos\alpha+(1+Br)sin\alpha]^2\cdot[F_*(A,B,\alpha,r)]^2G(k,|b_1|,-r),B\neq 0,$$

$$G\left(k,\left|b_1\right|,-r\right)\left[H_*\left(A,\alpha,-r\right)\right]^2\leq J_f\leq G\left(k,\left|b_1\right|,r\right)\left[H_*\left(A,\alpha,r\right)\right]^2,B=0.$$

Proof. Since

$$J_f=\left|g'\left(z\right)\right|^2-\left|h'\left(z\right)\right|^2=\left|g'\left(z\right)\right|^2-\left|h'\left(z\right)\omega\left(z\right)\right|^2=\left|h'\left(z\right)\right|^2\left(1-\left|\omega\left(z\right)\right|^2\right),$$

then if we use Theorem 1 and the inequality (5), in this step we obtain the desired result.

Corollary 4

$$\int\frac{k\left(1-\left|b_1\right|\right)+\left(\left|b_1\right|-k^2\right)r}{k+\left|b_1\right|r}\begin{bmatrix}\left(1-Ar\right)cos\alpha- \\ \left(1-Br\right)sin\alpha\end{bmatrix}\cdot F_*\left(A,B,\alpha,-r\right)dr\leq\left|f\right|\leq\int\frac{k\left(1+\left|b_1\right|\right)+\left(\left|b_1\right|+k^2\right)r}{k+\left|b_1\right|r}\cdot\begin{bmatrix}\left(1+Ar\right)cos\alpha \\ -\left(1+Br\right)sin\alpha\end{bmatrix}F_*\left(A,B,\alpha,r\right)dr,$$

$$\int\frac{k\left(1-\left|b_1\right|\right)+\left(\left|b_1\right|-k^2\right)r}{k+\left|b_1\right|r}\left(1-Ar\right)e^{-(Acos\alpha)r}cos\alpha dr\leq$$

$$\left|f\right|\leq\int\frac{k\left(1+\left|b_1\right|\right)+\left(\left|b_1\right|+k^2\right)r}{k+\left|b_1\right|r}\left(1+Ar\right)e^{(Acos\alpha)r}cos\alpha dr.$$

Theorem 3

Let $f(z)=h(z)+\overline{g(z)}$ be an element of $S^*\left(kq,A,B,\alpha\right)$, then

$$\sum_{m=2}^{n}k^4\left|b_m-b_1a_m\right|^2\leq(\left|b_1\right|^2-k^2)^2+\sum_{m=2}^{n}\left|\overline{b_1}b_m-k^2a_m\right|^2 \quad (11)$$

Proof. Using Theorem 2, then we can write

$$\frac{g\left(z\right)}{h\left(z\right)}=\frac{k^2\left(b_1-\phi\left(z\right)\right)}{k^2-\overline{b_1}\phi\left(z\right)}\Leftrightarrow$$

$$k^2g\left(z\right)-k^2b_1h\left(z\right)=\left(\overline{b_1}g\left(z\right)-k^2h\left(z\right)\right)\phi\left(z\right).$$

Therefore, we have

$$\sum_{m=2}^{n}(k^2b_m-k^2b_1a_m)z^m+\sum_{m=n+1}^{\infty}d_mz^m=\left[\left(\left|b_1\right|^2-k^2\right)z+\sum_{m=2}^{n}\left(\overline{b_1}b_m-k^2a_m\right)\right]\phi\left(z\right) \quad (12)$$

where the coefficients d_m have been chosen suitably and the equality (12) can be written in the form

$$F\left(z\right)=G\left(z\right)\phi\left(z\right),\qquad\left|\phi\left(z\right)\right|<1$$

then we have

$$\left|F\left(z\right)\right|^2=\left|G\left(z\right)\phi\left(z\right)\right|^2=\left|G\left(z\right)\right|^2\left|\phi\left(z\right)\right|^2<\left|G\left(z\right)\right|^2$$

$$\Rightarrow\frac{1}{2\pi}\int_0^{2\pi}\left|F\left(re^{i\theta}\right)\right|^2d\theta\leq\frac{1}{2\pi}\int_0^{2\pi}\left|G\left(re^{i\theta}\right)\right|^2d\theta$$

$$\Rightarrow \sum_{m=2}^{n} k^4 \left|b_m - b_1 a_m\right|^2 r^{2k} + \sum_{m=n+1}^{\infty} \left|d_m\right|^2 r^{2k}$$
$$\leq \left[\left(\left|b_1\right|^2 - k^2 \right)^2 r^{2k} + \sum_{m=2}^{n} \left|\overline{b_1} b_m + k^2 a_m\right|^2 r^{2k} \right]$$

passing to the limit as $r \to 1$ we obtain (11). The method of this proof was based on the Clunie method [1].

REFERENCES

[1] Clunie, J. , *On Meromorphic Schlicht Functions,* J. London Math. Soc. 34, (1959), 215-216.

[2] Duren, P. , *Univalent Functions,* Springer Verlag, (1983).

[3] Duren, P. , *Harmonic Mappings in the Plane,* Vol. 156 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge UK, (2004).

[4] Goodman A. W. , *Univalent Functions,* Volume I, Mariner publishing Company INC, Tampa Florida, (1983).

[5] Jack, I.S. , *Functions starlike and convex of order alpha,* J. London Math. Soc. 3, (1971), 369-374.

[6] Lewys, H. ,*On the non-vanishing of the Jacobian in certain one-to-one mappings,* Bull. Amer. Math. Soc. 42, (1936), 689-692.

The 4-ordered property of some chordal ring networks

Shin-Shin Kao*, Shih-Chin Wey and Hsiu-Chunj Pan

Abstract—A graph *G* is *k-ordered* if for any sequence of *k* distinct vertices of *G*, there exists a cycle in *G* containing these *k* vertices in the specified order. Obviously, any cycle in a graph is *1*-ordered, *2*-ordered and *3*-ordered. Thus the study of any graph being *k*-ordered always starts with $k = 4$. In this paper, we study the *4*-orderedness of certain chordal rings, denoted by $CR(n;1,q)$ for *n* being an even integer with $n \geq 6$ and *q* an odd integer with $3 \leq q \leq n/2$. More specifically, we prove that $CR(n;1,5)$ is *4*-ordered for $n \geq 14$, and $CR(n;1,7)$ is *4*-ordered for $n \geq 18$. The proof is based on computer experimental results by Mr. M. Tsai, which can be found in [8], and mathematical induction.

Keywords**—**4-ordered, chordal ring, cycle embedding, hamiltonian, cycles,.

I. INTRODUCTION

We consider finite, undirected and simple graphs only. Let $G = (V, E)$ be a graph, where *V* is the set of vertices of *G* and $E \subseteq \{(u,v) \mid u, v \in V\}$ is the set of edges of *G*, respectively. Let *u, v* be two vertices of *G*. If $e = (u,v) \in E$, then we say that the vertices *u* and *v* are *adjacent* in *G*. The *degree* of any vertex *u* is the number of distinct vertices adjacent to *u*. We use $N(u)$ to denote the set of vertices which are adjacent to *u*. A path *P* between two vertices v_0 and v_k is represented by $P=\langle v_0, v_1, \ldots\ldots, v_k\rangle$ where each pair of consecutive vertices are connected by an edge. We use P^{-1} to denote the path $\langle v_k, v_{k-1}, v_{k-2}, \ldots, v_0\rangle$. We also write the path $P=\langle v_0, v_1, \ldots\ldots, v_k\rangle$ as $\langle v_0, v_1, \ldots\ldots, v_i, Q, v_j, v_{j+1}, \ldots\ldots, v_k\rangle$, where *Q* denotes the path $\langle v_i, v_{i+1}, \ldots, v_j\rangle$. A *hamiltonian path* between *u* and *v*, where *u* and *v* are two distinct vertices of *G*, is a path joining *u* to *v* that visits every vertex of *G* exactly once. A *cycle* is a path of at least three vertices such that the first vertex is the same as the last vertex. A *hamiltonian cycle* of *G* is a cycle that traverses every vertex of *G* exactly once. A *hamiltonian graph* is a graph with a hamiltonian cycle. A graph *G* is *k-ordered* (or *k-ordered hamiltonian*, resp.) if for any sequence of *k* distinct vertices of *G*, there exists a cycle (or a hamiltonian cycle, resp.) in *G* containing these *k* vertices in the specified order. Obviously, any cycle in a graph is 1-ordered, 2-ordered and 3-ordered. Thus the study of *k*-orderedness (or *k*-ordered hamiltonicity) of any graph always starts with $k = 4$.

This work was supported in part by National Science Council of R.O.C. under Contract NSC102-2115-M-033 -004 -.

Shin-Shin Kao is a professor in Department of Applied Mathematics, Chung Yuan Christian University, Chung Pei Rd, Chung Li, Taiwan 32023, R.O.C. (Corresponding author. TEL: 886-3-2653116; FAX: 886-3-2653199; e-mail: shin2kao@gmail.com.).

Shih-Chin Wey received his M.S. from Chung Yuan Christian University in 2011. (e-mail: a4576942@yahoo.com.tw).

Hsiu-Chunj Pan is a graduate student in Chung Yuan Christian University now. (e-mail: g10201101@cycu.edu.tw).

The concept of *k*-orderedness and *k*-ordered hamiltonicity has attracted various studies since it was first introduced by Ng and Schultz [7] in 1997. Readers can refer to [2-6] for more recent research results.

In this paper, we are interested in the 4-orderedness of certain types of chordal rings. This work was initiated by Dr. T.Y. Ho and his partial results. The chordal ring family has been adopted as the underlying topology of certain interconnection networks [1] and is studied for the real architecture for parallel and distributed systems due to the advantage of a built-in hamiltonian cycle, symmetry, easy routing and robustness. See [9] and its references. The chordal rings $CR(n;1,q)$, where *n* is an even integer with $n \geq 6$ and *q* an odd integer with $3 \leq q \leq n/2$, is defined as follows. Let $G(V,E)=CR(n;1,q)$, where $V=\{a_1, a_2, \ldots, a_n\}$ and $E=\{(a_i, a_{(i+1) mod(n)}: 1 \leq i \leq n\} \cup \{(a_i, a_{(i+q) mod(n)}: i \text{ is odd and } 1 \leq i \leq n\}$. See Figure 1 for an illustration.

The following two lemmas are proved by computer experiments. See [8].

Lemma 1.1 $CR(n;1,5)$ is 4-ordered for any even integer *n* with $14 \leq n \leq 50$.

Lemma 1.2 $CR(n;1,7)$ is 4-ordered for any even integer *n* with $18 \leq n \leq 50$.

II. THE 4-ORDEREDNESS OF CR(N;1,5)

Theorem 2.1. $CR(20 + 6k;1,5)$ is 4-ordered for $k \geq 0$.

Proof. By Lemma 1.1, $CR(20;1,5)$ is a 4-ordered graph. It is interesting to see whether or not $CR(20+6k;1,5)$ is 4-ordered for $k \geq 1$. We can embed $CR(20+6k;1,5)$ into $CR(26+6k;1,5)$ as follows. Let *R* be a subset of $V(CR(20+6k;1,5)) \cup E(CR(20+6k;1,5))$. We define a function *f*, which maps *R* from $CR(20+6k;1,5)$ into $CR(26+6k;1,5)$ in the following way: (1) If $a_i \in R \cap V(CR(20+6k;1,5))$, where $1 \leq i \leq 20+6k$, then $f(a_i)=b_i$.(2) If $(a_i, a_j) \in R \cap E(CR(20+6k;1,5))$, where $1 \leq i, j \leq 20+6k$, then

$$f((a_i, a_j))=\begin{cases}(b_i, b_{i+1}) & \text{for } 1 \leq i \leq 19+6k \text{ and } j = i+1;\\ (b_i, b_{i+5}) & \text{for } i = odd \text{ with } 1 \leq i \leq 15+6k \text{ and } j = i+5;\\ \text{undefined} & \text{otherwise.}\end{cases}$$

Therefore, $CR(26+6k;1,5) - f(CR(20+6k;1,5))$ consists of the vertex set $\{b_{21+6k}, b_{22+6k}, b_{23+6k}, b_{24+6k}, b_{25+6k}, b_{26+6k}\}$ and the edge

ISBN: 978-1-61804-230-9

set $\{(b_{20+6k},b_{21+6k}), (b_{21+6k}, b_{22+6k}), (b_{22+6k}, b_{23+6k}), (b_{23+6k},b_{24+6k}), (b_{24+6k}, b_{25+6k}), (b_{25+6k},b_{26+6k}), (b_{26+6k},b_1),(b_{17+6k},b_{22+6k}), (b_{19+6k},b_{24+6k}), (b_{21+6k},b_{26+6k}), (b_{23+6k},b_2), (b_{25+6k},b_4)\}$. Figure 2.1 gives an illustration, in which f maps R from $CR(20;1,5)$ into $CR(26;1,5)$. We can see that (1) $f(a_i) = b_i$ for $1 \le i \le 20$, denoted by black vertices on both graphs. (2) $f((a_i, a_{i+1})) = (b_i,b_{i+1})$ for $1\le i \le 19$, denoted by green edges on both graphs. (3) $f((a_i, a_{i+5})) = (b_i,b_{i+5})$ for i is odd with $1\le i \le 15$, denoted by blue edges on both graphs. (4) $f((a_{20}, a_1)) = \phi$, $f((a_{17}, a_2)) = \phi$ and $f((a_{19}, a_4)) = \phi$, denoted by dashed edges on $CR(20;1,5)$. (5) $CR(26;1,5) - f(CR(20;1,5))$ consists of the vertex set $\{b_{21}, b_{22}, b_{23}, b_{24}, b_{25}, b_{26}\}$ and edge set $\{(b_{20}, b_{21}), (b_{21},b_{22}), (b_{22}, b_{23}), (b_{23}, b_{24}), (b_{24}, b_{25}), (b_{25}, b_{26}), (b_{26}, b_1), (b_{17}, b_{22}), (b_{19}, b_{24}), (b_{21},b_{26}), (b_{23}, b_2), (b_{25}, b_4)\}$.

We first present the construction of the required cycle in $CR(26;1,5)$ using the known cycle of $CR(20;1,5)$, denoted by C', as an illustration. There are 20 vertices $a_1, a_2,..., a_{20}$ in $CR(20;1,5)$, and 26 vertices $b_1, b_2,..., b_{26}$ in $CR(26;1,5)$. To prove the theorem, we do case studies by considering different situations. Take $G = CR(26;1,5)$. Let x_1, x_2, x_3 and x_4 be four arbitrary vertices of G. We want to construct a cycle C in G that visits x_i's in the given order. Note that we can always find at least one set of six consecutive vertices, denoted by $S = \{b_i,b_{i+1},b_{i+2},...,b_{i+5}\}$, such that $S \cap \{x_1, x_2, x_3, x_4\} = \phi$. Without loss of generality, let $x_1 = b_1$ and $S = \{b_{21},b_{22},...,b_{26}\}$. Removing the vertices of S and all edges adjacent to S in G, we obtain a subgraph of $CR(20;1,5)$. Obviously, $S \cap f(CR(20;1,5)) = \phi$. Note that $CR(20;1,5)$ is 4-ordered and hence contains a cycle that visits x_i's in the given order, denoted by C'. We will obtain C by embedding $CR(20;1,5)$ into $CR(26;1,5)$ and rerouting the cycle C'. There are two cases.

Case 1 $(a_{17}, a_2) \in C'$.

Case 1.1 $|\{a_{19}, a_{20}\} \cap C'| = 0$. It means that only the edge $(a_{17}, a_2) \in C'$. Let $C' = \langle a_1,Q_1, a_{17}, a_2,Q_2, a_1 \rangle$, where Q_1 is a path between a_1 and a_{17}, Q_2 is a path between a_2 and a_1, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), f(Q_1), f(a_{17}), b_{22},b_{23}, f(a_2), f(Q_2), f(a_1)\rangle = \langle b_1, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_1\rangle$. See Figure 2.2 and Figure 2.3 for an illustration.

Case 1.2 $|\{ a_{19}, a_{20}\} \cap C'| = 1$.

Case 1.2.1 $\langle a_{18}, a_{19}, a_4\rangle \in C'$. It means that $(a_{17}, a_2) \in C'$ and $\langle a_{18}, a_{19}, a_4 \rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{17}, a_2, Q_2, a_{18}, a_{19}, a_4, Q_3, a_1\rangle$, where Q_1 is a path between a_1 and a_{17}, Q_2 is a path between a_2 and a_{18}, Q_3 is a path between a_4 and a_1, and $Q_i \cap Q_j = \phi$ for each $i \ne j$ and $\{i, j\} \subseteq \{1,2,3\}$. We construct $C = \langle f(a_1), f(Q_1), f(a_{17}), b_{22}, b_{23}, f(a_2), f(Q_2), f(a_{18}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_3), f(a_1) \rangle = \langle b_1, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_{18}, b_{19}, b_{24}, b_{25}, b_4, f(Q_3), b_1\rangle$. See Figure 2.4 and Figure 2.5 for an illustration.

Case 1.2.2 $\langle a_{15},a_{20},a_1 \rangle \in C'$. It means that $(a_{17}, a_2) \in C'$ and $\langle a_{15}, a_{20}, a_1\rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{17}, a_2, Q_2, a_{15}, a_{20}, a_1 \rangle$, where Q_1 is a path between a_1 and a_{17}, Q_2 is a path between a_2 and a_{15}, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), f(Q_1), f(a_{17}), b_{22}, b_{23}, f(a_2), f(Q_2), f(a_{15}), f(a_{20}), b_{21}, b_{26}, f(a_1) \rangle = \langle b_1, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_{15}, b_{20}, b_{21}, b_{26}, b_1\rangle$. See Figure 2.6 and Figure 2.7 for an illustration.

Case 1.3 $\{a_{19}, a_{20}\} \cap C'| = 2$.

Case 1.3.1 $\langle a_{18}, a_{19}, a_{20}, a_{15}\rangle \in C'$. It means that $(a_{17}, a_2) \in C'$ and $\langle a_{18}, a_{19}, a_{20}, a_{15}\rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{17}, a_2, Q_2, a_{18}, a_{19}, a_{20}, a_{15}, Q_3, a_1\rangle$, where Q_1 is a path between a_1 and a_{17}, Q_2 is a path between a_2 and a_{18}, Q_3 is a path between a_{15} and a_1, and $Q_i \cap Q_j = \phi$ for each $i \ne j$ and $\{i, j\} \subseteq \{1, 2, 3\}$. We construct $C = \langle f(a_1), f(Q_1), f(a_{17}), b_{22}, b_{23}, f(a_2), f(Q_2), f(a_{18}), f(a_{19}), f(a_{20}), f(a_{15}), f(Q_3), f(a_1) \rangle = \langle b_1, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_{18}, b_{19}, b_{20}, b_{15}, f(Q_3), b_1\rangle$.

Case 1.3.2 $\langle a_1, a_{20}, a_{19}, a_4\rangle \in C'$. It means that $\langle a_1, a_{20}, a_{19}, a_4\rangle \in C'$ and $(a_{17}, a_2) \in C'$. Let $C' = \langle a_1, a_{20}, a_{19}, a_4, Q_1, a_{17}, a_2, Q_2, a_1\rangle$, where Q_1 is a path between a_4 and a_{17}, Q_2 is a path between a_2 and a_1, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), b_{26}, b_{21}, f(a_{20}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_1), f(a_{17}), b_{22}, b_{23}, f(a_2), f(Q_2), f(a_1) \rangle = \langle b_1, b_{26}, b_{21}, b_{20}, b_{19}, b_{24}, b_{25}, b_4, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_1\rangle$.

Case 1.3.3 $\langle a_{15}, a_{20}, a_{19}, a_4\rangle \in C'$. It means that $(a_{17}, a_2) \in C'$ and $\langle a_{15}, a_{20}, a_{19}, a_4\rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{17}, a_2, Q_2, a_{15}, a_{20}, a_{19}, a_4, Q_3, a_1\rangle$, where Q_1 is a path between a_1 and a_{17}, Q_2 is a path between a_2 and a_{15}, Q_3 is a path between a_4 and a_1, and $Q_i \cap Q_j = \phi$ for each $i \ne j$ and $\{i, j\} \subseteq \{1,2,3\}$. We construct $C = \langle f(a_1), f(Q_1), f(a_{17}), b_{22}, b_{23}, f(a_2), f(Q_2), f(a_{15}), f(a_{20}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_3), f(a_1) \rangle = \langle b_1, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_{15}, b_{20}, b_{19}, b_{24}, b_{25}, b_4, f(Q_3), b_1\rangle$.

Case 1.3.4 $\langle a_{18}, a_{19}, a_{20}, a_1\rangle \in C'$. It means that $(a_{17}, a_2) \in C'$ and $\langle a_{18}, a_{19}, a_{20}, a_1\rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{17}, a_2, Q_2, a_{18}, a_{19}, a_{20}, a_1\rangle$, where Q_1 is a path between a_1 and a_{17}, Q_2 is a path between a_2 and a_{18}, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), f(Q_1), f(a_{17}), b_{22}, b_{23}, f(a_2), f(Q_2), f(a_{18}), f(a_{19}), f(a_{20}), b_{21}, b_{26}, f(a_1) \rangle = \langle b_1, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_{18}, b_{19}, b_{20}, b_{21}, b_{26}, b_1\rangle$.

Case 1.3.5 $\langle a_{18}, a_{19}, a_4\rangle \in C'$, $\langle a_{15}, a_{20}, a_1\rangle \in C'$. It means that $(a_{17}, a_2) \in C'$, $\langle a_{18}, a_{19}, a_4\rangle \in C'$ and $\langle a_{15}, a_{20}, a_1\rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{17}, a_2, Q_2, a_{18}, a_{19}, a_4, Q_3, a_{15}, a_{20}, a_1\rangle$, where Q_1 is a path between a_1 and a_{17}, Q_2 is a path between a_2 and a_{18}, Q_3 is a path between a_4 and a_{15}, and $Q_i \cap Q_j = \phi$ for each $i \ne j$ and $\{i, j\} \subseteq \{1,2,3\}$. We construct $C = \langle f(a_1), f(Q_1), f(a_{17}), b_{22}, b_{23}, f(a_2), f(Q_2), f(a_{18}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_3), f(a_{15}), f(a_{20}), b_{21}, b_{26}, f(a_1) \rangle = \langle b_1, f(Q_1), b_{17}, b_{22}, b_{23}, b_2, f(Q_2), b_{18}, b_{19}, b_{24}, b_{25}, b_4, f(Q_3), b_{15}, b_{20}, b_{21}, b_{26}, b_1\rangle$.

Case 2 $(a_{17}, a_2) \notin C'$.

Case 2.1 $|\{ a_{19}, a_{20}\} \cap C'| = 0$. It means that $C' = \langle a_1, Q_1, a_1\rangle$, but the edges (a_{17}, a_2), (a_{15}, a_{20}), (a_{19}, a_4), (a_{18}, a_{19}), (a_{19}, a_{20}), $(a_{20}, a_1) \notin C'$.

Case 2.2 $|\{ a_{19}, a_{20}\} \cap C'| = 1$.

Case 2.2.1 $\langle a_{18}, a_{19}, a_4\rangle \in C'$. It means that $\langle a_{18}, a_{19}, a_4\rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{18}, a_{19}, a_4, Q_2, a_1\rangle$, where Q_1 is a path between

ISBN: 978-1-61804-230-9

a_1 and a_{18}, Q_2 is a path between a_4 and a_1, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), f(Q_1), f(a_{18}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_2), f(a_1) \rangle = \langle b_1, f(Q_1), b_{18}, b_{19}, b_{24}, b_{25}, b_4, f(Q_2), b_1 \rangle$.

Case 2.2.2 $\langle a_{15}, a_{20}, a_1 \rangle \in C'$. It means that $\langle a_{15}, a_{20}, a_1 \rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{15}, a_{20}, a_1 \rangle$, where Q_1 is a path between a_1 and a_{15}. We construct $C = \langle f(a_1), f(Q_1), f(a_{15}), f(a_{20}), b_{21}, b_{26}, f(a_1) \rangle = \langle b_1, f(Q_1), b_{15}, b_{20}, b_{21}, b_{26}, b_1 \rangle$.

Case 2.3 $|\{ a_{19}, a_{20}\} \cap C'| = 2$.

Case 2.3.1 $\langle a_{18}, a_{19}, a_{20}, a_{15} \rangle \in C'$. It means that $\langle a_{18}, a_{19}, a_{20}, a_{15} \rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{18}, a_{19}, a_{20}, a_{15}, Q_2, a_1 \rangle$, where Q_1 is a path between a_1 and a_{18}, Q_2 is a path between a_{15} and a_1, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), f(Q_1), f(a_{18}), f(a_{19}), f(a_{20}), f(a_{15}), f(Q_2), f(a_1) \rangle = \langle b_1, f(Q_1), b_{18}, b_{19}, b_{20}, b_{15}, f(Q_2), b_1 \rangle$.

Case 2.3.2 $\langle a_1, a_{20}, a_{19}, a_4 \rangle \in C'$. It means that $\langle a_1, a_{20}, a_{19}, a_4 \rangle \in C'$. Let $C' = \langle a_1, a_{20}, a_{19}, a_4, Q_1, a_1 \rangle$, where Q_1 is a path between a_4 and a_1. Let $C = \langle f(a_1), b_{26}, b_{21}, f(a_{20}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_1), f(a_1) \rangle = \langle b_1, b_{26}, b_{21}, b_{20}, b_{19}, b_{24}, b_{25}, b_4, f(Q_1), b_1 \rangle$.

Case 2.3.3 $\langle a_{15}, a_{20}, a_{19}, a_4 \rangle \in C'$. It means that $\langle a_{15}, a_{20}, a_{19}, a_4 \rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{15}, a_{20}, a_{19}, a_4, Q_2, a_1 \rangle$, where Q_1 is a path between a_1 and a_{15}, Q_2 is a path between a_4 and a_1, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), f(Q_1), f(a_{15}), f(a_{20}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_2), f(a_1) \rangle = \langle b_1, f(Q_1), b_{15}, b_{20}, b_{19}, b_{24}, b_{25}, b_4, f(Q_2), b_1 \rangle$.

Case 2.3.4 $\langle a_{18}, a_{19}, a_{20}, a_1 \rangle \in C'$. It means that $\langle a_{18}, a_{19}, a_{20}, a_1 \rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{18}, a_{19}, a_{20}, a_1 \rangle$, where Q_1 is a path between a_1 and a_{18}. We construct $C = \langle f(a_1), f(Q_1), f(a_{18}), f(a_{19}), f(a_{20}), b_{21}, b_{26}, f(a_1) \rangle = \langle b_1, f(Q_1), b_{18}, b_{19}, b_{20}, b_{21}, b_{26}, b_1 \rangle$.

Case 2.3.5 $\langle a_{18}, a_{19}, a_4 \rangle \in C'$, $\langle a_{15}, a_{20}, a_1 \rangle \in C'$. It means that $\langle a_{18}, a_{19}, a_4 \rangle \in C'$ and $\langle a_{15}, a_{20}, a_1 \rangle \in C'$. Let $C' = \langle a_1, Q_1, a_{18}, a_{19}, a_4, Q_2, a_{15}, a_{20}, a_1 \rangle$, where Q_1 is a path between a_1 and a_{18}, Q_2 is a path between a_4 and a_{15}, and $Q_1 \cap Q_2 = \phi$. We construct $C = \langle f(a_1), f(Q_1), f(a_{18}), f(a_{19}), b_{24}, b_{25}, f(a_4), f(Q_2), f(a_{15}), f(a_{20}), b_{21}, b_{26}, f(a_1) \rangle = \langle b_1, f(Q_1), b_{18}, b_{19}, b_4, f(Q_2), b_{15}, b_{20}, b_{21}, b_{26}, b_1 \rangle$.

Given four arbitrary vertices $\{x_i | 1 \le i \le 4\}$ in $CR(26;1,5)$, we have presented a constructive skill for finding a cycle C in $CR(26;1,5)$ from the known cycle C' in $CR(20;1,5)$ that visits x_i's in the right order. The same technique is applied to derive a cycle $\bar{C}$ in $CR(26+6k;1,5)$ from a cycle $\bar{C}'$ in $CR(20+6k;1,5)$ that passes four arbitrary vertices in the required order. More specifically, using the induction hypothesis, we assume that the statement holds for $CR(20+6k^*;1,5)$ for some integer $k^* \ge 1$. Replacing the vertex label a_i (or b_i, resp.) with a_{i+6k^*} (or b_{i+6k^*} resp.) in the above derivation, we can show that the statement in the theorem holds for $CR(26+6k^*;1,5)$. Hence the theorem is proved by mathematical induction. □

With Lemma 1.1, it is known that $CR(22;1,5)$ and $CR(24;1,5)$ are 4-ordered. It is easy to see that our technique in Theorem 2.1 can be utilized to obtain the following two theorems.

Theorem 2.2. $CR(22+6k;1,5)$ is 4-ordered for $k \ge 0$.

Theorem 2.3. $CR(24+6k;1,5)$ is 4-ordered for $k \ge 0$.

Combining Lemma 1.1 and Theorems 2.1–2.3, we have the following theorem.

Theorem 2.4. $CR(n;1,5)$ is 4-ordered for any even integer n with $n \ge 14$.

III. The 4-orderedness of $CR(n;1,7)$

With the similar derivations as in Section II, the authors proved the following theorems. Readers can refer to [8] for details.

Theorem 3.1. $CR(26+8k;1,7)$ is 4-ordered for $k \ge 0$.
With Lemma 1.2, it is known that $CR(28;1,7)$, $CR(30;1,7)$ and $CR(32;1,7)$ are 4-ordered. It is easy to see that our technique in Theorem 3.1 can be utilized to obtain the following three theorems.

Theorem 3.2. $CR(28+8k;1,7)$ is 4-ordered for $k \ge 0$.

Theorem 3.3. $CR(30+8k;1,7)$ is 4-ordered for $k \ge 0$.

Theorem 3.4. $CR(32+8k;1,7)$ is 4-ordered for $k \ge 0$.

Combining Lemma 1.2 and Theorem 3.1- 3.4, we have the following theorem.

Theorem 3.5. $CR(n;1,7)$ is 4-ordered for any even integer n with $n \ge 18$.

IV. 4. Conclusion

Let $n \ge 6$ be an even integer. In this paper, we show the 4-orderedness of $CR(n;1,5)$ for $n \ge 14$, and that of $CR(n;1,7)$ for $n \ge 18$, by combining computer experimental results for small n and mathematical induction for general n's. An natural question to be explored is the 4-ordered hamiltonicity of the chordal rings. In particular, the 4-ordered hamiltonicity for the graphs in $CR(n;1,5)$ and $CR(n;1,7)$. Currently, computer experiments already shows that the 4-ordered hamiltonicity only exists on $CR(n;1,5)$, or $CR(n;1,7)$, and some other chordal rings for specific n's. In addition, our rerouting technique based on embedding the known cycle in a smaller chordal ring, denoted by G', into the larger one, denoted by G, does not include all the vertices in G obtained from enlarging G'. Thus the 4-ordered hamiltonicity of the chordal ring family remains an open problem.

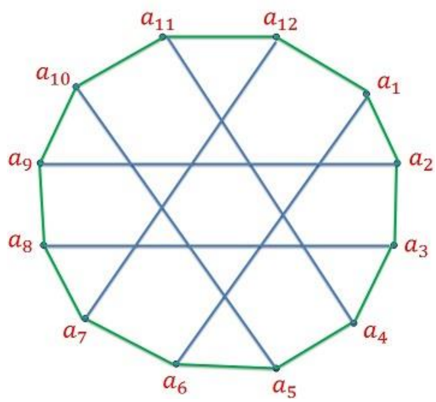


Figure 1: CR(12; 1, 5)

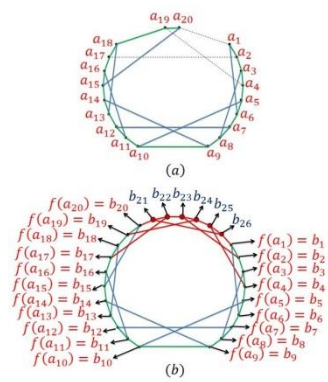


Figure 2.1: (a) CR(20; 1, 5); (b) CR(26; 1, 5) and the function *f*.

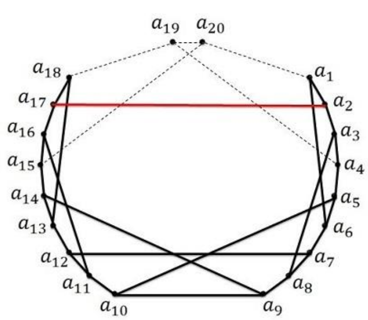


Figure 2.2: Case 1.1 in Theorem 2.1., where $(a_{17}, a_2) \in C'$.

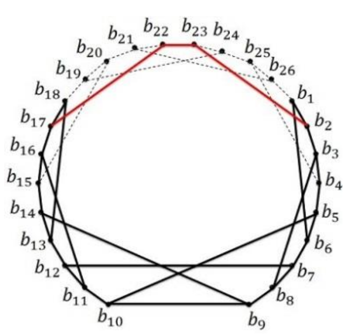


Figure 2.3: The cycle C constructed in Case 1.1 in Theorem 2.1.

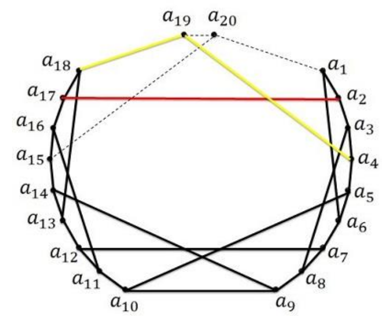


Figure 2.4: Case 1.2.1 in Theorem 2.1., where $(a_{17}, a_2) \in C'$ and $(a_{18}, a_{19}, a_4) \in C'$.

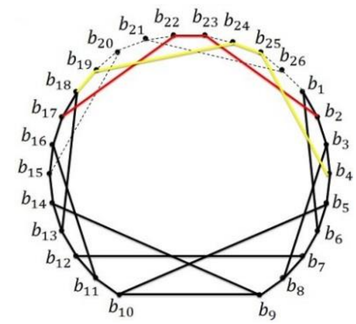


Figure 2.5: The cycle C constructed in Case 1.2.1 in Theorem 2.1

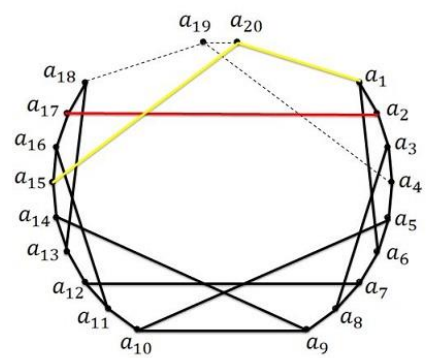


Figure 2.6: Case 1.2.2 in Theorem 2.1., where $(a_{17}, a_2) \in C'$ and $(a_{15}, a_{20}, a_1) \in C'$.

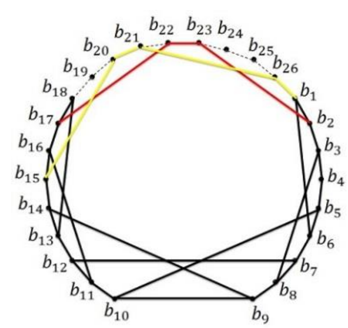


Figure 2.7: The cycle C constructed in Case 1.2.2 in Theorem 2.1.

ISBN: 978-1-61804-230-9

ACKNOWLEDGMENT

This paper adapted the thesis of S.-C. Wey [8]. The authors would like to express their gratitude to Dr. T.Y. Ho and Ming Tsai for their contribution in this work.

REFERENCES

[1] R.N. Farah, M.Othman, and M.H. Selamat, "Combinatorial properties of modified chordal rings degree four networks", *Journal of Computer Science* 6, 3, 2010, pp. 279–284.
[2] Ralph J. Faudree, "Survey of results on k-ordered graphs", *Discrete Mathematics*, 229, 2001, pp. 73–87.
[3] Chun-Nan Hung, David Lu, Randy Jia, Cheng-Kuan Lin, Laszlo Liptak, Eddie Cheng, Jimmy J.M. Tan, and Lih-Hsing Hsu, "4-ordered-Hamiltonian problems of the generalized Petersen graph GP(n,4)", *Mathematical and Computer Modelling* 57, 2013, pp. 595–601.
[4] Ruijuan Li, "A Fan-type result on k-ordered graphs", *Information Processing Letters*, 110, 2010, pp. 651– 654.
[5] Ruijuan Li, Shengjia Li, and Yubao Guo, "Degree conditions on distance 2 vertices that imply k-ordered hamiltonian", *Discrete Applied Mathematics*, 158, 2010, pp. 331–339.
[6] Karola Meszaros, "On 3-regular 4-ordered graphs", *Discrete Mathematics*, 308, 2008, pp. 2149–2155.
[7] Lenhard Ng and Michelle Schultz, "k-ordered hamiltonian graphs", *Journal of Graph Theory*, 24, 1997, No. 1, pp. 45–57.
[8] S.-C. Wey, "The 4-ordered property of some chordal ring networks", Master thesis, Department of Applied Mathematics, Chung-Yuan Christian University, Taiwan, R.O.C., 2011.
[9] Xianbing Wang and Yong Meng Teo, "Global data computation in chordal rings", *J. Parallel and Distributed Computing*, 69, 2009, pp. 725–736.

Ease of Doing Business: an efficiency comparison of the G20 economies

Antonio Pavone, Paola Pianura

Abstract— This work has a twofold objective: first, to apply the production possibility frontier concept to compare the "ease of doing business" among economies; second, to distinguish the performance of countries with respect to the operating environment in which the adjustments are included. The paper focuses on the G20 economies, split into Mature and Rapid Growth Markets. The evaluation is limited to starting a business and trading across borders regulation areas. Data cover the period 2009–2013 and were acquired by the World Bank source. In the second stage, we analyze the dependency of "ease of doing business" based on ICT environment.

Keywords— Administrative business regulation, international comparisons, inter efficiency decomposition.

I. Introduction

The challenge for government is to deliver effective and efficient regulation: that is effective in addressing an identified problem, and efficient in terms of minimizing compliance costs on the citizens and firms whereas maximizing the benefits to society.

In recent years, the interest in making international comparisons has increased. Supranational institutions have produced an international set of indicators that evaluate the performance of policy.

Regarding business regulations, perhaps, the most well-known survey, produced at the international level, is the Doing Business by the World Bank. It offers an annual ranking of "ease of doing business". In the comparison, Doing Business assumes a common market and regulatory system. In reality, homogeneity is seldom present. The lack of homogeneity is one of the pitfalls mentioned by Dyson et al. (2001).

The external control of the operating environment in which production occurs is important when one suspects heterogeneity between groups, otherwise, direct comparisons may be misleading.

To overcome this aspect, we decompose the inter-country efficiency into intra efficiency and inter catch-up components. The inter catch-up component is relevant for policy purposes because reflects the environments in which countries operate.

For our application, we focus the analysis on G20 countries which cover more than 85% of the world economy, and we distinguish them between Mature Markets (MMs) and Rapid-Growth Markets (RGMs).

This paper unfolds as follows: data are described in Section 2; section 3 introduces the decomposition of efficiency based on convex and non-convex production models; section 4 presents the results of the application; some concluding remarks are presented in section 5; appendix reports statistical data at a country level.

II. The Data

Doing Business covers 11 areas and 189 economies, relating to regulation of the life cycle of domestic firms. Each of them consists of a set of indicators constructed by laws and regulations in accordance with: procedures to be undertaken, time needed, costs that businesses have to support and other composite indicators based on multiple parameters.

For our analysis, we focus on "starting a business" and "trading across borders" (divided into import and export trade).

Selected countries cover the period between 2009 and 2013. To take advantage of the amount of information available, the sample data are pooled. Table I[1] summarizes data collected across countries and over time and distinguished by market groups.

In general, it appears that the burden of compliance in MMs is lower than the G20 average even if for "starting a business" RGMs appear less expensive.

F. A. Author is with the Italian National Institute of Statistics, Italy, (corresponding author to provide phone: 0039 06 4673 6680; e-mail: pavone@istat.it).

S. A. Author is with the Italian National Institute of Statistics, Italy, (corresponding author to provide phone: 0039 06 4673 6681; e-mail: pianura@istat.it).

[1] Original value is expressed as percentage of economy's income per capita. We convert it in economic value using a Purchasing Power Parity (PPP) conversion factor. Moreover, for the 2012, we impute GNI data.

ISBN: 978-1-61804-230-9

Table I G20 data average, 2009- 2013

Area	Indicators	MMs	RGMs	G20
Starting a business	Procedures (number)	6	10	8
	Time (days)	9	34	22
	Cost	1,776	1,396	1,576
	Minimum capital	354	2,021	1,231
Trading across borders export	Documents (number)	3	6	5
	Time (days)	10	16	13
	Cost (per container)	1,099	1,306	1,207
Trading across borders import	Documents (days)	4	8	6
	Time (days)	9	20	15
	Cost (per container)	1,159	1,483	1,330

Source: our elaboration on World Bank data

Univariate analysis is just a first descriptive stage of evaluation, before obtaining an overall picture of "ease of doing business". A more powerful technique refers to the efficiency measure, where multiple inputs and outputs are involved in the production process.

Debreu (1951) and Farrell (1957) introduced a measure of technical efficiency. With an input requirement set, their measure is defined as (one minus) the maximum equiproportionate (i.e., radial) reduction in all inputs that is feasible with given technology and outputs[2].

We apply this technique to measure the degree of "ease of doing business" decomposing it between intra- and inter-market differences.

III. Material and Methods

The estimation of an efficient frontier requires a specification of the input and output factors of the production process.

Homogeneity is one of the most important assumptions to make reliable comparisons in the context of productivity analysis.

Dyson et al. (2001) suggest four key criteria to consider when evaluating performance:

(i) the factors cover the full range of resources used;

(ii) the factors capture all activity levels and performance measures;

(iii) the factors are common to all productive units;

(iv) environmental variation has been assessed and captured if necessary.

In our application we presume that, fixed regulation goals, the "ease of doing business" depends not only on national responsibility to minimize the burden for firms in the compliance of regulation, but also on the characteristics of the environment in which countries operate.

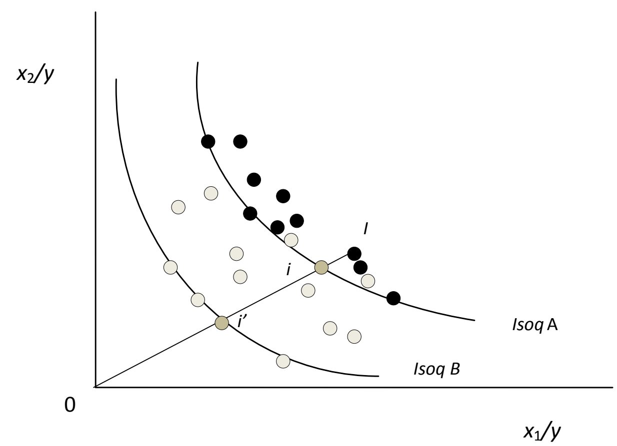
If we suppose that the homogeneity assumption is not satisfied, due to regulatory and market differences, it is required to separate the two effects on productive efficiency in intra-market and inter-market components.

An illustrative example will clarify the procedures followed. The figure shows a set of producers which uses two inputs (x_1 and x_2) to produce a single output.

The producers operate under two different environment technological groups; those labelled with the black dot operate under the group A, whereas, those labelled with the white dot operate under the group B. The technological frontier of each group is represented by a unit-isoquant and is constructed using the most efficient observations within that group.

Let us consider a producer *I* and its frontier group A denoted by the surface Isoq A. The level of intra-unit efficiency is defined as the ratio (0*i*/0*I*) where *i* is the projection point of *I* onto the Isoq A.

Fig.1 The inter unit efficiency decomposition



To appraise the component of efficiency which can be attributed to the difference between Isoq A and Isoq B, internal group inefficiency needs to be eliminated. Thus, the original input levels have been replaced with the corresponding adjusted values *i*. So, for the producer *I*, the ratio of (0*i*'/0*i*) measures the inter-unit catch-up efficiency. The inter unit efficiency is the product of inter- catch-up and intra efficiency components.

Thus, for the producer *I* we have:

$0i'/0I = (0i/0I) \times (0i'/0i)$

The basic model to estimate the degree of efficiency is:

[2] From the efficiency analysis, a productive unit is efficient in an input-oriented framework, if it is on the boundary of the input requirement set, or, for the output oriented case, on the boundary of the output correspondence set.

ISBN: 978-1-61804-230-9

$$\min_{\lambda,z} \lambda$$

$$\text{s.t} \quad \sum_{i=1}^{I} z_i\left(x_{ik}/y_i\right) \leq \lambda\left(x_{ik}/y_i\right), \quad k = 1,...,q$$

$$z_i \geq 0 \qquad ; \qquad \text{convex constrain}$$

$$z_i \in \{0,1\} \qquad ; \text{ non convex constrain}$$

$$\sum_{i=1}^{I} z_i = 1, \quad \forall i \in \{1,...,I\}$$

where:

λ is the input technical efficiency measure;

x_{ik} is the quantity of k to obtain a unit of output y_i, of producer i;

z_i is the intensity variable for producer i.

The scalar value of lambda calculated by the model is a radial measure providing the minimum proportional reduction of all inputs.

The model is run ones for each producer in a dataset.

Non convex model relies on the same idea as convex model, except that the frontier envelops the observations more closely than convex model does (see Tulkens, 1993).

IV. Empirical analysis

Our analysis provides results (for the Mann-Whitney rank order statistic and the independent sample test[3]) to determine whether the efficiency differences between MMs and RGMs are significant.

Table II gives the efficiency average scores for starting a business regulation for MMs and RGMs groups. The summaries are also distinct with respect to the reference envelopes used (convex and non-convex model).

Irrespective of which model is assumed, the hypothesis test indicates that there is a significant difference in the efficiency spreads within groups. On average, this reveals a worse performance of MMs respect to RGMs.

However, considering the inter- catch-up component, we find that the efficiency of the RGMs is much lower than the MMs. This comparison suggests a heterogeneous environmental context between the markets.

Looking at trading across borders, tables III and IV, the statistical test indicates the main difference between MMs and RGMs regarding the inter- catch-up component.

To explore the environmental context, in which business regulation is made and administered, we examined the ICT diffusion between the markets.

Namely, we analyze the data collected by INSEAD and WIPO and classified-in four dimensions.

- ACCESS which captures ICT readiness and includes five infrastructure and access indicators (fixed-telephone subscriptions, mobile cellular telephone subscriptions, international internet bandwidth per internet user, percentage of households with a computer, and percentage of households with internet access).

Table II Starting a business - decomposition of productive efficiency

Efficiency decomposition	Group	Test for indipendent means				Mann-Whitney rank test			
		Non convex model		Convex model		Non convex model		Convex model	
		Average*	t	Average*	t	Average	Z	Average	Z
Intra	MMs	0.56	-2.39	0.49	-2.45	42.64	-1.82	44.17	-1.30
	RGMs	0.68		0.62		52.82		51.45	
Inter catch-up	MMs	1.00	12.55	1.00	16.46	71.27	-8.14	72.13	-8.27
	RGMs	0.44		0.41		27.06		26.28	
Inter	MMs	0.56	5.75	0.49	5.64	62.94	-5.04	64.27	-5.46
	RGMs	0.30		0.25		34.55		33.36	

[3] With respect to the multiplicative nature of the decomposition, geometric averages are used. Moreover, to evaluate the equivalent of geometric means between MMs and RGMs groups, we apply an independent test on arithmetic means of the logarithms of the efficiency values.

ISBN: 978-1-61804-230-9

Table III Trading across borders, Export - decomposition of productive efficiency

Efficiency decomposition	Group	Test for indipendent means				Mann-Whitney rank test			
		Non convex model		Convex model		Non convex model		Convex model	
		Average*	t	Average*	t	Average	Z	Average	Z
Intra	MMs	0.97	3.55	0.90	1.76	56.20	-3.16	48.39	-0.13
	RGMs	0.88		0.84		40.62		47.65	
Inter catch-up	MMs	1.00	10.95	1.00	12.08	68.00	-7.47	70.69	-7.96
	RGMs	0.74		0.73		30.00		27.58	
Inter	MMs	0.97	8.18	0.90	7.45	66.96	-6.68	65.16	-5.77
	RGMs	0.65		0.62		30.94		32.56	

*Geometric average

Table IV Trading across borders, Import - decomposition of productive efficiency

Efficiency decomposition	Group	Test for indipendent means				Mann-Whitney rank test			
		Non convex model		Convex model		Non convex model		Convex model	
		Average*	t	Average*	t	Average	Z	Average	Z
Intra	MMs	0.95	2.45	0.90	2.60	55.13	-2.70	54.18	-2.10
	RGMs	0.89		0.82		41.58		42.44	
Inter catch-up	MMs	1.00	10.70	1.00	13.43	68.00	-7.47	71.17	-8.03
	RGMs	0.66		0.68		30.00		27.15	
Inter	MMs	0.95	8.56	0.89	8.64	67.13	-6.69	67.58	-6.60
	RGMs	0.58		0.56		30.78		30.38	

*Geometric average

USE which captures ICT intensity and includes three ICT intensity and usage indicators (percentage of internet users, fixed (wired)-broadband subscriptions, and active mobile broadband subscriptions).

E-GOVERNMENT SERVICE, by United Nations, is a composite indicator that gives a measure of ‘how much’ governments are putting online.

E-PARTECIPATION which focuses on the following components: use of the internet to facilitate provision of information by governments to citizens (“e-information sharing”), interaction with stakeholders (“e-consultation”), and engagement in decision-making processes (“e-decision making”).

These dimensions are presented in table V.

Looking at the first ICT dimension, we observe a noteworthy difference between MMs and RGMs, except for the mobile cellular subscriptions. MMs have the highest percentages of households with a computer and also of households with internet access, whereas ICT is not so widespread in RGMs.

Table V ICT summary statistics, 2012

Dimension	Indicators	MMs	RGMs	G20
ICT access	Fixed-telephone subscriptions per 100 inhabitants	51.79	17.59	33.79
	Mobile-cellular subscriptions per 100 inhabitants	113.29	121.35	117.53
	International Internet bandwidth Bit/s per Internet user	79’67	21’78	49’20
	Percentage of households with computer	81.71	40.71	60.13
	Percentage of households with Internet access	82.16	36.28	58.01
	Internet secure servers per 1.000.000 inhabitants	1,046.03	286.94	606.56
ICT use	Percentage of individuals using internet	80.59	40.77	59.63
	Fixed (wired)-broadband subscriptions per 100 inhabitants	31.04	8.03	18.93
	Active mobile broadband subscriptions per 100 inhabitants	73.83	26.23	48.78
Government’s on line service		86.58	58.75	71.93
On line e-participation		73.68	34.75	53.19

Our elaboration on INSEAD and WIPO data.

In addition, we consider also Internet secure servers per 1.000.000 inhabitants by the World Bank that reflect the level of ICT security. In RGMs the percentage is much lower than in MMs which could be indirectly evidence of few people accessing the internet for business transactions.

If we regard the second dimension, ICT use, the gap between MMs and RGMs is relevant. There is a low percentage use of internet in RGMs with respect to MMs. However, RGMs use more mobile than fixed services reflecting the greatest diffusion of smartphone and other mobile equipments.

For the third and fourth dimensions, we consider two composite indicators: online service component and e-participation. The first measures the willingness and capacity of national administrations to use information and communication technology to deliver public services. Whereas the second reflects how useful these features are and how effectively they have been deployed by the government compared to all other countries. On average, we note that there is a considerable difference between the markets.

These data provide statistic evidence of barriers to develop a high level of ICT adoption as far as the core services of public administration. In other words, not all RGMs are equally prone to get involved into e-government.

V. CONCLUSION

This paper analyzes the performance of groups of units, rather than individual performance.

Focusing on G20 countries, distinguished between MMs and RGMs and two regulation areas for businesses (“starting a business” and “trading across borders” divided into import and export trade), we estimate intra- and inter-country measures of technical efficiency. Intra-country technical efficiency measures the use of the country’s own production potential and represents an absolute measure of technical efficiency. Inter-country efficiency is a relative efficiency measure revealing the performance of a particular country respect to the “best practice” available in the market. The inter-firm catch-up component reflects the environmental differences in which countries operate.

The inter evaluation shows that MMs are more efficient than RGMs. The main difference is captured by the inter catch-up component. This result suggests a broad conclusion: MMs are more active in keeping up with new technologies rather than exploiting their existing production potential. We suppose that the role of ICT appears to contribute significantly to reduce the cost and complexity of the procedures imposed on business. In fact, in the comparison, we observe that, where ICT infrastructure is well developed, also efficiency in the business regulation increases.

Appendix

Starting a business: average efficiency*, 2009- 2013

Country Code	Convex model Intra	Convex model Inter catch-up	Convex model Inter	Non convex model Intra	Non convex model Inter catch-up	Non convex model Inter
ARG	0.45	0.32	0.15	0.60	0.33	0.20
AUS	1.00	1.00	1.00	1.00	1.00	1.00
BRA	0.38	0.46	0.17	0.38	0.61	0.23
CAN	1.00	1.00	1.00	1.00	1.00	1.00
CHN	0.49	0.69	0.34	0.53	0.90	0.48
DEU	0.24	1.00	0.24	0.32	1.00	0.32
FRA	0.61	1.00	0.61	0.77	1.00	0.77
GBR	0.62	0.96	0.60	0.66	1.00	0.66
IDN	0.48	0.27	0.13	0.48	0.30	0.14
IND	0.46	0.31	0.14	0.57	0.31	0.18
ITA	0.44	1.00	0.44	0.50	1.00	0.50
JPN	0.20	1.00	0.20	0.23	1.00	0.23
KOR	0.40	1.00	0.40	0.49	1.00	0.49
MEX	0.94	0.45	0.42	1.00	0.47	0.47
RUS	0.73	0.43	0.31	0.76	0.47	0.36
SAU	0.70	0.28	0.20	0.91	0.26	0.24
TUR	1.00	0.46	0.46	1.00	0.50	0.50
USA	0.53	1.00	0.53	0.55	1.00	0.55
ZAF	0.93	0.61	0.56	0.94	0.60	0.57

*Geometric average

Export: average efficiency*, 2009- 2013

Country Code	Convex model Intra	Convex model Inter catch-up	Convex model Inter	Non convex model Intra	Non convex model Inter catch-up	Non convex model Inter
ARG	0.88	0.62	0.55	0.93	0.66	0.62
AUS	0.78	1.00	0.78	0.89	1.00	0.89
BRA	0.87	0.59	0.52	0.87	0.62	0.55
CAN	0.90	1.00	0.90	1.00	1.00	1.00
CHN	0.97	0.99	0.95	1.00	1.00	1.00
DEU	0.86	1.00	0.86	0.93	1.00	0.93
FRA	1.00	1.00	1.00	1.00	1.00	1.00
GBR	0.84	1.00	0.84	0.96	1.00	0.96
IDN	1.00	0.97	0.97	1.00	1.00	1.00
IND	0.78	0.76	0.60	0.80	0.77	0.62
ITA	0.81	1.00	0.81	1.00	1.00	1.00
JPN	0.92	1.00	0.92	1.00	1.00	1.00
KOR	0.98	1.00	0.98	1.00	1.00	1.00
MEX	1.00	0.67	0.67	1.00	0.75	0.75
RUS	0.50	0.65	0.32	0.55	0.62	0.34
SAU	0.98	0.78	0.76	1.00	0.79	0.79
TUR	0.96	0.69	0.66	1.00	0.67	0.67
USA	1.00	1.00	1.00	1.00	1.00	1.00
ZAF	0.66	0.71	0.47	0.75	0.66	0.49

*Geometric average

ISBN: 978-1-61804-230-9

Import: average efficiency*, 2009- 2013

	Convex model			Non convex model		
Country Code	Intra	Inter catch-up	Inter	Intra	Inter catch-up	Inter
ARG	0.69	0.60	0.42	0.88	0.49	0.44
AUS	0.79	1.00	0.79	0.88	1.00	0.88
BRA	0.73	0.57	0.42	0.83	0.52	0.43
CAN	0.86	1.00	0.86	1.00	1.00	1.00
CHN	1.00	0.98	0.97	1.00	1.00	1.00
DEU	0.92	1.00	0.92	1.00	1.00	1.00
FRA	1.00	1.00	1.00	1.00	1.00	1.00
GBR	0.96	1.00	0.96	0.97	1.00	0.97
IDN	0.91	0.94	0.86	0.93	1.00	0.93
IND	0.82	0.70	0.57	0.86	0.70	0.61
ITA	0.85	1.00	0.85	1.00	1.00	1.00
JPN	0.72	0.99	0.71	0.72	1.00	0.72
KOR	1.00	1.00	1.00	1.00	1.00	1.00
MEX	1.00	0.66	0.66	1.00	0.74	0.74
RUS	0.55	0.55	0.31	0.65	0.48	0.31
SAU	0.96	0.67	0.65	1.00	0.69	0.69
TUR	0.98	0.61	0.60	1.00	0.62	0.62
USA	1.00	1.00	1.00	1.00	1.00	1.00
ZAF	0.67	0.64	0.43	0.79	0.58	0.46

*Geometric average

ICT access dimension - 2012

	Fixed-telephone subscriptions per 100 inhabitants	Mobile-cellular subscriptions per 100 inhabitants	International Internet bandwidth Bit/s per Internet user	Percentage of households with computer	Percentage of households with Internet access	Internet secure servers per 1.000.000 inhabitants
ARG	24.30	142.50	21.97	56.00	47.50	41.79
AUS	45.70	106.20	69.46	85.20	81.40	1724.46
BRA	22.30	125.20	25.08	49.90	45.40	54.28
CAN	51.90	75.70	100.98	86.60	83.00	1233.44
CHN	20.60	81.30	4.17	40.90	37.40	3.14
DEU	61.80	131.30	75.53	87.00	85.00	1090.92
FRA	61.90	98.10	84.55	81.00	80.00	408.96
GBR	52.60	130.80	188.88	87.00	88.60	1478.28
IDN	15.50	115.20	17.21	15.10	6.50	3.95
IND	2.50	68.70	5.19	10.90	9.50	3.57
ITA	35.50	159.50	76.25	67.00	63.00	208.07
JPN	50.80	109.40	33.04	80.00	86.00	750.05
KOR	61.90	110.40	26.04	82.30	97.40	2751.56
MEX	17.40	86.80	16.30	32.20	26.00	28.07
RUS	30.10	183.50	32.95	60.60	51.20	38.95
SAU	16.70	184.70	35.87	67.70	66.60	30.54
TUR	18.60	90.80	40.35	50.20	47.20	116.76
USA	44.00	98.20	62.27	79.30	75.00	1474.08
ZAF	7.90	134.80	18.70	23.60	25.50	83.75

ICT use dimension - 2012

Country Code	Percentage of individuals using internet	Fixed (wired)-broadband subscriptions per 100 inhabitants	Active mobile broadband subscriptions per 100 inhabitants
ARG	55.80	10.90	20.80
AUS	82.30	25.10	102.70
BRA	49.80	9.20	37.30
CAN	86.80	32.90	50.00
CHN	42.30	13.00	17.20
DEU	84.00	34.00	41.10
FRA	83.00	37.80	52.20
GBR	87.00	34.00	72.00
IDN	15.40	1.20	31.90
IND	12.60	1.10	4.90
ITA	58.00	22.10	52.10
JPN	79.10	27.90	113.10
KOR	84.10	37.60	106.00
MEX	38.40	10.90	10.20
RUS	53.30	14.50	53.00
SAU	54.00	6.80	44.70
TUR	45.10	10.50	16.30
USA	81.00	28.00	75.30
ZAF	41.00	2.20	26.00

E-Government dimensions - 2012

Country code	Online Service	Online e-partecipation
ARG	52.90	28.90
AUS	86.30	76.30
BRA	67.30	50.00
CAN	88.90	68.40
CHN	52.90	21.10
DEU	75.20	76.30
FRA	87.60	57.90
GBR	97.40	92.10
IDN	49.70	21.10
IND	53.60	18.40
ITA	57.50	26.30
JPN	86.30	73.70
KOR	100.00	100.00
MEX	73.20	57.90
RUS	66.00	65.80
SAU	79.70	63.20
TUR	46.40	5.30
USA	100.00	92.10
ZAF	45.80	15.80

ISBN: 978-1-61804-230-9

References

[1] [1] Brockett, P. L., & Golany B., 1996. Using Rank Statistics for Determining Programmatic Efficiency Differences in Data Envelopment Analysis. Management Science 42 (3): 466 – 472.W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] Charnes, A., Cooper, W.W., Rhodes, E., 1981. Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through, Management Science, 27, 668-697.

[3] Charnes, A., W.W. Cooper, and E. Rhodes, 1978, Measuring the efficiency of decision making units, European Journal of Operational Research 2, 429-444.

[4] Debreu, G., 1951. The coefficient of resource utilization. Econometrica, 19: 273-292.

[5] Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico,C.S., Shale, E.A., 2001. Pitfalls and protocols in DEA. European Journal of Operational Research 132, 245-259.

[6] Deprins, D., Simar, L., and Tulkens, H. 1984. Measuring Labor Inefficiency in Post Offices, in Marchand, M., Pestieau, P. and Tulkens, H. (eds.) The Performance of Public Enterprizes: Concepts and Measurements, 243-267.

[7] Farrell, M. J. I. 1957. The Measurement of Productivity Efficiency, Journal of the Royal Statistical Society, Ser. A 120, 253-281.

[8] Tulkens, H. (1993). On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit, Journal of Productivity Analysis, 4(1-2), 183-210.

On the Stability of a Discrete Time Ramsey Growth Model with Stochastic Labor

Gabriel Zacarías-Espinoza, Hugo Cruz-Suárez
and Enrique Lemus-Rodríguez [*†‡]

March 27, 2014

Abstract

In this paper we study a version of Ramsey's discrete time Growth Model where the evolution of Labor through time is stochastic. Taking advantage of recent theoretical results in the field of Markov Decision Processes, a first set of conditions on the model are established that guarantee a long-term stable behaviour of the underlying Markov chain.

keywords: Discounted Markov Decision Processes, Stochastic Euler Equation, Stability of Markov Chains, ergodic convergence, Ramsey Growth Model.

1 Introduction

Ramsey's seminal work on economic growth has been extended in many ways, but, to the best of our knowledge, the study of a random discrete time version is still in its youth (see [18]) and provides an opportunity for a fruitful interaction between economists and mathematicians that should lead to better simulations and consequently, to a better understanding of the effects of the random deviations in the growth of an economy and its impact on the population. In this paper a first random model is proposed where the population, i.e., the Labor (force) grows in a stochastic manner.

The structure of the paper is simple: the model is spelled out, as well as a set underlying Assumptions, and its long-term stability is established. Due to space constraints, for the technical proofs the reader is referred to a theoretical paper ([20]) where a general framework is built.

*Gabriel Zacarías-Espinoza, Department of Mathematics, UAM-Iztapalapa, Ave. San Rafael Atlixco 186, Col. Vicentina, México 09340, México. e-mails: gabrielzaces@hotmail.com

†Hugo Cruz-Suárez, Facultad de Ciencias Físico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio y Río Verde, Col. San Manuel, CU, 72570 Puebla, Pue., México. e-mail: hcs@fcfm.buap.mx

‡Enrique Lemus-Rodríguez, Escuela de Actuaría, Universidad Anáhuac México-Norte, Av. Universidad Anáhuac 46, Col. Lomas Anáhuac, 52786 Huixquilucan, Edo. de México. México. e-mail: elemus@anahuac.mx

2 Ramsey Growth Model driven by Stochastic Labor

Our model considers a random number of consumers, the population or Labor L_t, whose growth is Markovian:

$$\begin{aligned} L_{t+1} &= L_t\eta_t, \\ & L_0 \text{ is known} \end{aligned} \tag{1}$$

where L_t denotes the number of consumers at time t, $t = 0, 1, ...$, and $\{\eta_t\}$ is a sequence of independent and identically distributed (iid) random variables, such that $P(\eta_t > 0) = 1$.

Remark 2.1. *In the literature of economic growth models is usual to assume that the number of consumers grow very slowly in time (see [12] and [18]). Observe that the model presented in this paper is a first step in an effort to weaken that constraint of the model.*

The production function for the economy is given by

$$\begin{aligned} Y_t &= F(K_t, L_t), \\ & K_0 \text{ is known,} \end{aligned}$$

i.e. the production Y_t is a function of capital, K_t, and labor, L_t, where the production function, F, is a homogeneous function of degree one. The output must be split between consumptions $C_t = c_tL_t$ and the gross investment I_t, i.e.

$$C_t + I_t = Y_t. \tag{2}$$

Let $\delta \in (0, 1)$ be the depreciation rate of capital. Then the evolution equation for capital is given by:

$$K_{t+1} = (1 - \delta)K_t + I_t. \tag{3}$$

Substituting (3) in (2), it is obtained,

$$C_t - (1 - \delta)K_t + K_{t+1} = Y_t. \tag{4}$$

In the usual way, all variables can normalized into *per capita* terms, namely, $y_t := Y_t/L_t$ and $x_t := K_t/L_t$. Then (4) can be expressed in the following way:

$$c_t - (1 - \delta)x_t + K_{t+1}/L_t = y_t = F(x_t, 1).$$

ISBN: 978-1-61804-230-9

Now, using (1) in the previous relation, it is obtained

$$x_{t+1} = \xi_t(F(x_t, 1) + (1 - \delta)x_t - c_t),$$

$t = 0, 1, 2, ...,$ where $\xi_t := (\eta_t)^{-1}$. Define $h(x) := F(x, 1) + (1 - \delta)x$, $x \in X := [0, \infty)$, h henceforth to be identified as the production function.

The transition law of the system is given by

$$x_{t+1} = \xi_t(h(x_t) - c_t),$$

where $c_t \in [0, h(x_t)]$ and $\{\xi_t\}$ is a sequence of iid random variables, with $x_0 = x$ known.

Let ξ a generic element of $\{\xi_t\}$. Suppose that ξ has a density Δ, which is strictly positive and continuous second derivative, i.e, $\Delta \in C^2((0, \infty))$. Furthermore, suppose that $E[\xi^p]$ and $E[\xi^{-1}]$ exist and both are finite, where $p > 1$ and E denotes the expectation operator. Let A be defined by

$$A = \cup_{x \in X} [0, h(x)].$$

A is called control set.

Definition 2.2. *A plan or consumption sequence is a sequence $\pi = \{\pi_t\}_{t=0}^{\infty}$ of stochastic kernel π_t on the control set A given the history*

$$h_t = (x_1, c_1, ..., x_{t-1}, c_{t-1}, x_t), \tag{5}$$

for each $t = 0, 1, ...$ and satisfying the constraints $\pi_t(A(x_t)|h_t) = 1$, $t = 0, 1, \cdots$. The set of all plans will be denoted by Π.

Let $\mathbb{F}$ be the set of all measurable functions $f : X \to A$, such that $f(x) \in A(x)$ for every $x \in X$. A plan $\pi \in \Pi$ is *stationary* if there exists $f \in \mathbb{F}$ such that, under π, the control $f(x)$ is applied at each time $t = 0, 1, \ldots$. In this case, a stationary plan is denoted by f.

Given an initial capital $x = x_0 \in X$ and a plan $\pi \in \Pi$ then $\mathbb{P}_x^\pi$ denotes the probability measure on the canonical space $(\Omega, \Im)$, where $\Omega := (X \times A)^\infty$ and $\Im$ is their corresponding σ-algebra of Borel on Ω, where the performance index used to evaluate the quality of the plan π is determined by

$$v(\pi, x) = \mathbb{E}_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t U(c_t) \right]$$

where $U : [0, \infty) \to \mathbb{R}$ is a measurable function known as utility function and $\alpha \in [0, 1]$ is the discount factor.

The goal of the controller is to maximize utility of consumption on all plans $\pi \in \Pi$, that is:

$$V(x) := \sup_{\pi \in \Pi} \mathbb{E}_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t U(c_t) \right],$$

$x \in X$.

For ulterior reference, this model be called Ramsey Growth Model under Stochastic Labor, or **RSL**.

The RSL model is a Markov Decision Process (MDPs) (see [8] and [9]). In fact, the Markov Control Model could be identified in the following way: $X = A = [0, \infty)$, $A(x) = [0, h(x)]$, $x \in X$, the transition law Q is given by

$$Q(B|x, c) = \int I_B((h(x) - c)s) \Delta(s)ds,$$

$(x, c) \in \mathbb{K} := \{(x, c)|x \in X, c \in A(x)\}$. Futhermore, the reward function is given by an utility function U.

In this context, a plan is called a policy.

The following assumptions are well known in the context of economic growth models (see [2], [5], [11], [12] [13], [15], [16], [17], [19] and [18]), guaranteeing the existence of an optimal plan, the validity of the dynamic programming algorithm and a characterization of the optimal plan via a version of an stochastic Euler equation in the context of MDPs (see [5] and [6]). Through of this work, Assumptions 2.3 and 2.4 below, are supposed to hold.

Assumption 2.3. *The production function h, satisfies:*
$h \in C^2((0, \infty))$, is a concave function on X,$h' > 0$ and $h(0) = 0$ and,
Let $h'(0) := \lim_{x \downarrow 0} h'(x)$. Suppose that $h'(0) \geq 1$ and

$$\alpha h'(0) > E[\xi^{-1}]. \tag{6}$$

Assumption 2.4. *The utility function U satisfies:*

a) $U \in C^2((0, \infty), \mathbb{R})$, with $U' > 0$ and $U'' < 0$, $U'(0) = \infty$ and $U'(\infty) = 0$,

b) There is a function ϑ defined on S with $E[\vartheta(\xi)] < \infty$, and this satifies that

$$|U'(h(s(h(x) - c)))h'(s(h(x) - a))s\Delta(s)| \leq \vartheta(s), \tag{7}$$

$s \in S$, $c \in (0, h(x))$.

3 Basic Properties of the RSL

Assumptions 2.3 and 2.4 guarantee that two powerful tools, Dynamic Programming and Euler equation, can be used in the study of our RSL.

In paper [20], using results from [8], [9] and [10] it is shown that the standard Dynamic Programming techniques hold for the RSL, in particular, in [20] it is shown how the result in [3] ensure uniqueness of the optimal policy. Moreover, in [5] it is shown that for such optimal policy the optimal action is an interior point in $[0, h(x)]$ for each $x \in X$.

In short, under the above assumptions, the value function of the corresponding MDP satisfies Bellman's optimality equation, the optimal comsumption plan is unique and value iteration works. For details see [20].

Similarly, in [20], the results in [5] an [6] allow us to establish a MDP Euler equation in the context of the classical Value Iteration Algorithm, i.e.:

The optimal plan f satisfies

$$U'(f(x)) = \alpha E[h'(\xi(h(x) - f(x)))U'(f(\xi(h(x) - f(x))))\xi], \tag{8}$$

ISBN: 978-1-61804-230-9

$x > 0$. Reciprocally, if $f \in \mathbb{F}$ satifies (8) and

$$\lim_{t\to\infty} \alpha^t E_x^f \left[h'(x_t)U'(f(x_t))x_t\right] = 0, \tag{9}$$

then f is an optimal plan.

For further results and details, see [20].

4 Stability of the RSL

Inspired by Nishimura and Stachurski in [16], it is possible to use Euler's equation ([20], [5] and [6]) to establish ergodic convergence towards a invariant probability measue of the optimal processse using density functions of the driving noise defined on $(0, \infty)$ the optimal process also converges in L^p if $p \geq 1$.

Let $f \in \mathbb{F}$ be the stationary optimal plan for the RSL model, the stochastic optimal process is given by

$$x_{t+1} = \xi_t(h(x_t) - f(x_t)),$$

$t = 0, 1, 2, \ldots, x_0 = x \in X = [0, \infty)$, known.

In order to avoid trivialities, we assume $x_0 > 0$. We know that:

$$Q(B|\, x, f(x)) = \int\limits_{\{s|s(h(x)-f(x))\in B\}} \Delta(s)ds, \, x > 0,$$

due to the fact that $h - f$ is strictly positive and that the density Δ is continuous and also strictly positive, we get that kernel $Q(\cdot|\, x, f(x))$ is determined by

$$q_1(y|\, x, f(x)) := \Delta\left(\frac{y}{h(x) - f(x)}\right)\frac{1}{h(x) - f(x)}, \, x > 0.$$

Given $x > 0$, inductively, for each time $t \geq 1$, x_t distribution has density q_t determined by

$$q_t(y) = \int q_1(y|\, x, f(x))q_{t-1}(x)dx.$$

For each $A \in B(X)$ we define measure:

$$\Xi(A) := \int\limits_B \Delta(s)ds.$$

Lemma 4.1. *The optimal process $\{x_n\}$ is Ξ-irreducible and strongly aperiodic.*

Proof. See [20] □

Let W be defined for$x \in (0, \infty)$ as

$$W(x) := [U'(f(x))h'(x)]^{1/2} + x^p + 1, \tag{10}$$

donde $p > 1$.

Lemma 4.2. *W is a Lyapunov function.*

Proof. See [20] □

From the previous results, in a dense technical deduction involving classical Markov Decision Processes techniques (see [20]) it is possible to prove the fundamental theorem of this paper, that due to space limitations, we will simply state:

Theorem 4.3. *There exists a unique invariant probability measure P for the optimal process of the RSL. In fact, this process is W-geometrically ergodic.*

Proof. See [20] □

And a very interesting side result is the following:

Theorem 4.4. *The corresponding optimal process of the RSL converges in mean.*

Proof. See [20]

□

5 Concluding remarks.

This first RSL has noteworthy stability properties: a unique invariant measure and geometric convergence. The Markov Decision Process Euler Equation approach is a powerful tool for the analysis of economic models like the RSL.

This paper is a first step in a much wider research programme where a further study of this and related models is pursued. In particular, actual simulations and actual calculation of the invariant probability measure should provide us with a better picture of the behaviour of this stochastic model. Also, the study of the robustness of the optimal policies, the trajectories and the invariant probability under perturbations of the model is of great importance.

References

[1] Arrow K. J., *A note on uncertainty and discounting in models of economic growth*, J. Risk Unc. 38, 87-94, (2009).

[2] Brock W. and Mirman L., *Optimal economic growth and uncertainty: the discounted case*, J. Econ. Th. 4, 479-513, (1972).

[3] Cruz-Suárez D., Montes-de-Oca R. and Salem-Silva F., *Conditions for the uniqueness of optimal policies of discounted Markov decision processes*, Math. Methods Oper. Res. 60, 415-436, (2004).

[4] Cruz-Suárez H. and Montes-de-Oca R., *An envelope theorem and some applications to discounted Markov decision processes*, Math. Methods Oper. Res. 67, 299-321, (2008).

[5] Cruz-Suárez H., Montes-de-Oca R. and Zacarías G., *A Consumption-Investment problem modelled as a discounted Markov decision process*, Kybernetika 47, 740-760, (2011).

[6] Cruz-Suárez H., Zacarías-Espinoza G. and Vázquez-Guevara V., *A Version of the Euler Equation in Discounted Markov Decision Processes*, Journal of Applied Mathematic, Volume 2012, Article ID 103698, doi:10.1155/2012/103698, 1-16, (2012).

[7] Gut A., Probability: A Graduate Course, Springer, New York, 2005.

[8] Hernández-Lerma O. and Lasserre J. B., Discrete-Time Markov Control Processes: Basic Optimality Criteria, Springer-Verlag, New York, (1996).

[9] Hernández-Lerma O. and Lasserre J. B., Further Topics on Discrete-Time Markov Control Processes, Springer-Verlag, (1999).

[10] Jaśkiewics A. and Nowak A.S., *Discounted dynamic programming with unbounded returns: application to economic models*, J. Math. Anal. Appl. 378, 450-462, (2011).

[11] Kamihigashi T., *Stochastic optimal growth with bounded or unbounded utility and bounded or unbounded shocks*, J. Math. Econom. 43, 477-500, (2007).

[12] Le Van C. and Dana RA., Dynamic Programing in Economics, Kluwer Academic Publishers, Boston, (2003).

[13] Levhari D. and Srinivasan T. N., *Optimal savings under uncertainty*, Rev. Econ. Stud. 36, 153-163, (1969).

[14] Meyn S. P. and Tweedie R. L., Markov Chains and Stochastic Stability, Springer-Verlag, London, 1993.

[15] Mirman L. and Zilcha I., *On optimal growth under uncertainty*, J. Econ. Th. 2, 329-339, (1975).

[16] Nishimura K. and Stachurski J., *Stability of stochastic optimal growth models: a new approach*, J. Econom. Theory, 122, 100–118, (2005).

[17] Ramsey F. P., *A Mathematical theory of saving*, Econ. J. 38, 543-559, (1928).

[18] Sladký K., *Some Remarks on Stochastic Versions of the Ramsey Growth Model*, Bulletin of the Czech Econometric Society, Vol 19, 139-152, (2012).

[19] Stachurski J., *Stochastic growth: asymptotic distributions*, Econom. Theory. 21, 913-919. (2003).

[20] Zacarías-Espinoza G., Cruz-Suárez H., and Lemus-Rodríguez E., *Stability of the Optimal Process of Stochastic Ramsey Growth Models*, Submitted.

Acknowledgment

This work was supported in part by CONACYT (Mexico) and ASCR (Czech Republic) under Grant No. 171396.

Computational investigation of Möbius strip in tensioned fabric structure

H.M. Yee[1]* and M.A. Samsudin[2]

Abstract— Form-finding of tensioned fabric surface bordered by Möbius strip is investigated. Möbius strip has the mathematical property of being non-orientable and with only one side and only one boundary component. In this paper, the possibility of adopting the form of Möbius strip as surface shape for tensioned fabric structure has been studied. The combination of shape and internal forces for the purpose of stiffness and strength is an important feature of tensioned fabric surface. For this purpose, form-finding needs to be carried out. Nonlinear analysis method is used for form-finding analysis in this paper. Pattern of pre-stress in the resulting tensioned fabric surface is also studied. Form-finding has been found to converge for Möbius strip with midcircle radius R over half-width W, R/W= 0.7, 1.2, 1.7, 2.2 and 2.7. It provides an alternative choice for structural designer to consider the Möbius strip, R/W= 0.7, 1.2, 1.7, 2.2 and 2.7 applied in tensioned fabric structure.

Keywords— Form-Finding, Möbius strip, Nonlinear analysis method and Tensioned fabric surface.

I. Introduction

Tensioned Fabric Structure (TFS) include a wide variety of systems that are distinguished by their reliance upon tensile only members to support load. TFS have been employed throughout recorded history as in rope bridges and tents. However, large permanent tension structures were generally a 19th century development in bridges and a 20th century development in buildings. The design of large TFS has been fully dependent upon the use of computers. Many of the developments in TFS have occurred in the last 30 years, precisely, because of the accessibility of powerful computers.

One of the most exciting shapes that captured the focus of the mathematics is Möbius Strip surface. Möbius Strip surface is a surface with only one side, one boundary component and can be made by twisting a strip by 360° and joining both ends of the strip. It was discovered by two German mathematicians name August Ferdinand Möbius and Johann Benedict Listing in 1858.

[1]H. M. Yee is with the Faculty of Civil Engineering, Universiti Teknologi MARA, 13500 Permatang Pauh, Pulau Pinang, Malaysia (e-mail: yhooimin@yahoo.com, * Corresponding Author).

[2]M. A. Samsudin is with the Faculty of Civil Engineering, Universiti Teknologi MARA, 13500 Permatang Pauh, Pulau Pinang, Malaysia (e-mail: madzimsamsudin@gmail.com).

Möbius Strip has the mathematical property that is non-orientable surface. [1] have mentioned that Möbius strip surface has great potential be as an architecture form. [2] have carried out the possibility implement Möbius strip surface in a shell structure. [3] and [4] have proposed nonlinear analysis method for form-finding of tensioned fabric structures in the form of Möbius strip. Applicability of the computational strategies proposed has been verified by form-finding carried out models of tensioned fabric structures in the one of the form of Möbius strip (with R/W > 1.3699) have shown good agreement with mathematically defined surfaces. Form-finding of Möbius strip TFS models by assuming an initial assumed shape with the opening at the center is not able to yield converged result for value R/W≤ 1.3699. The shape obtained after form-finding are found to be different from the form of typical Möbius strip surface with opening at the center. The opening at the center of Möbius strip is found to be non-existent. Shape of Möbius strip (without opening) has been verified through soap film model with R/W = 1. Form-finding has been found to converge for Möbius strip TFS model (with R/W≤ 1.3699) with initial assumed shape specified to follow the topology without opening which has been observed in experiment.

In this paper, only form-finding using nonlinear analysis method of Möbius strip TFS with R/W = 0.7, 1.2, 1.7, 2.2 and 2.7 has been carried out. Form-finding is the step to determine the initial equilibrium shape that satisfies the prescribed prestress system and boundary condition.

II. Generation of Möbius Strip Surface in TFS

For this paper, the software [5] has been used for the purpose of model generation. Aspect of modeling of surface of Möbius strip and form as well pre-stress pattern of the resulting TFS through form-finding using nonlinear analysis method are studied. Möbius Strip as shown in Fig. 1 can be represented parametrically by the following set of equations [6]:

$$X = \left(R + S\cos\frac{\tau}{2} \right)\cos\tau$$

$$Y = \left(R + S\cos\frac{\tau}{2} \right)\sin\tau$$

$$Z = S\sin\left(\frac{\tau}{2}\right)$$

$$\text{for } \{S{:}-W,W\} \text{ and } \{\tau : 0, 2\pi\} \qquad (1)$$

The problem with this generation model is single continuous surface involve only one curve along the band in a Cartesian plane. The problem is impossible to join the band border and to

ISBN: 978-1-61804-230-9

overcome this problem, [3] has proposed by applying discontinuous surface in modeling with small gap across the width of the strip.

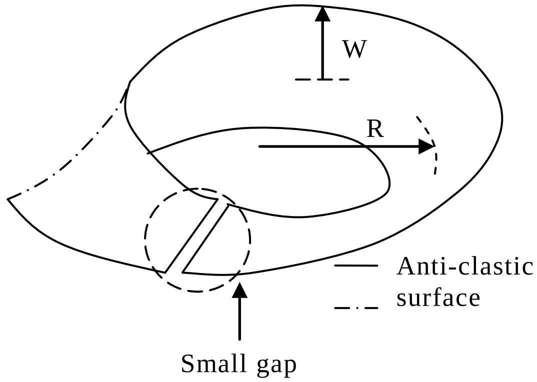


Fig. 1: A Möbius strip with midcircle radius *R* and half-width *W*

III. Form-Finding Analysis Using Nonlinear Analysis Method

A nonlinear finite element analysis program by [3] for the analysis of tensioned fabric structures has been used in this study. The procedure adopted is based on the work by [3]. 3-node plane stress element has been used as element to model the surface of TFS. All x, y and z translation of nodes lying along the boundary edge of the Möbius strip have been restrained. Similarly, all nodes lying on either side of the gap introduced were also restrained from translating in all directions. The member pretension in warp and fill direction, is 2000N/m, respectively. The shear stress is zero.

Two stages of analysis were involved in the procedures of form-finding analysis in one cycle proposed by [3]. First stage (denoted as SF1) is analysis which starts with an initial assumed shape in order to obtain an updated shape for initial equilibrium surface. The initial assumed shape can be obtained from any pre-processing software and reference [5] is chosen for this paper. This is then followed by the second stage of analysis (SS1) aiming at checking the convergence of updated shape obtained at the end of stage (SF1). During stage (SF1), artificial tensioned fabric properties, E with very small values are used. Both warp and fill tensioned fabric stresses are kept constant. In the second stage of (SS1), the actual values of tensioned fabric properties are used. Resulting warp and fill tensioned fabric stresses are checked at the end of the analysis against prescribed tensioned fabric stresses. Then, iterative calculation has to be carried out in order to achieve convergence where the criteria adopted is that the average of warp and fill stress deviation should be < 0.01. The resultant shape at the end of iterative step n (SSn) is considered to be in the state of initial equilibrium under the prescribed warp and fill stresses and boundary condition if difference between the obtained and the prescribed membrane stresses relative to the prescribed stress is negligibly small. Such checking of difference in the obtained and prescribed stresses has been presented in the form of total stress deviation in warp and fill direction versus analysis step. As a first shape for the start of form-finding procedure adopted in this paper, initial assumed shape is needed. For the generation of such initial assumed shape, knowledge of the requirement of anti-clastic nature of TFS is used. The incorporation of anti-clastic feature into the model will help to produce a better initial assumed shape.

IV. Computational Analysis

Different combination parameters radius, R and haft-width, W for Möbius strip have been studied. The determination of R and W has been carried out as follow: W = 10 and R = 7, 12, 17, 22 and 27. Möbius Strip R/W = 0.7, 1.2, 1.7, 2.2 and 2.7 has been studied. The opening of the opening at the center becomes larger when R has been increased.

Two different variables R/W = 0.7 and R/W = 1.2 which is R/W ≤ 1.3699 for Möbius strip have been chosen for the first discussion. Total number of nodes and elements used in the model are 343 and 576, respectively. Fig. 2 shows the initial assumed shape for Möbius strip R/W = 0.7. The surface models R/W = 0.7 has been found intersect with each other after form-finding as shown in Fig. 3. The convergence of R/W = 0.7 not able to achieve due to intersection of surface.

Another set of Möbius strip R/W = 0.7 has been carried out by assuming the initial assumed shape similar to Fig. 3. Total number of nodes and elements used in the model are 154 and 270, respectively. Fig. 4 shows the convergence of the Möbius strip R/W = 0.7. The convergence curve in Fig. 5 shows the total warp and fills deviation < 0.01. Similar case obtained also for the Möbius strip R/W = 1.2.

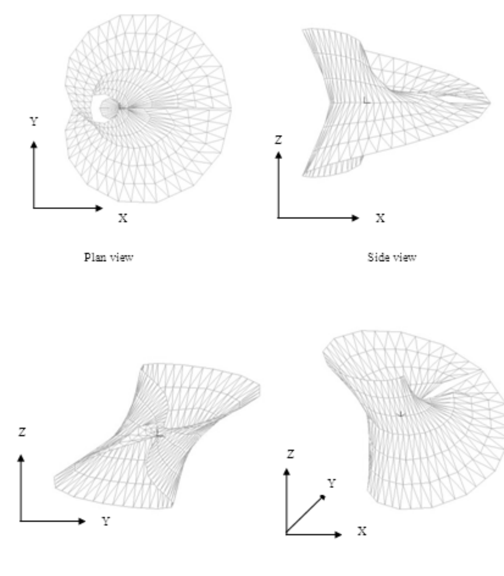


Fig. 2: Initial assumed shape for Möbius strip (R/W = 0.7)

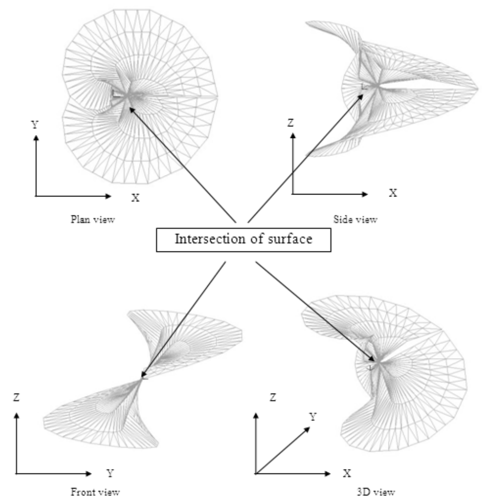


Fig. 3: Models of TFS intersect with each other in the form of Möbius strip after form-finding (R/W = 0.7)

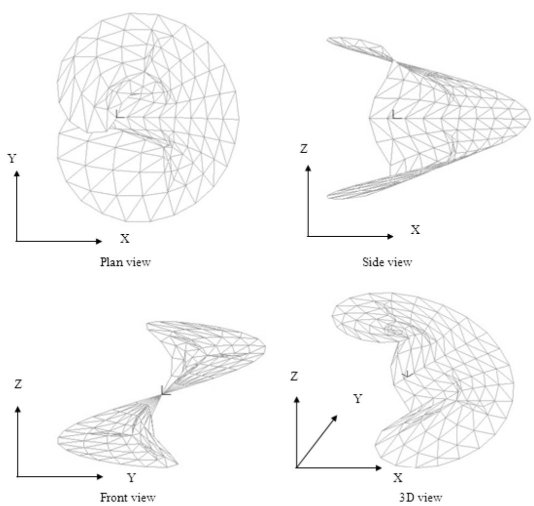


Fig. 4: TFS model in the form Möbius strip (without opening) after form-finding (R/W = 0.7)

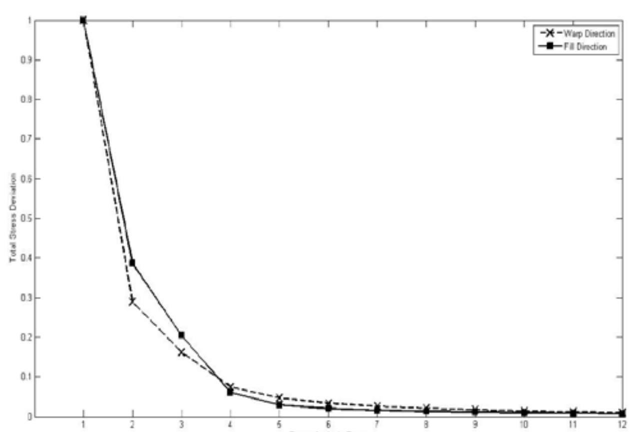


Fig. 5: Variation of total stress deviation in warp and fill direction verses stress analysis stage for Möbius strip TFS Model (R/W = 0.7)

Möbius strip R/W = 1.7, 2.2 and 2.7 have been chosen for the following discussion. The surfaces of a Möbius strip also generated from (1). Total number of nodes and elements used in the model are 343 and 576, respectively. Fig. 6 shows the initial assumed shape for Möbius strip (R/W = 1.7). Fig. 7 shows the convergence of the Möbius strip R/W = 1.7. The convergence curve in Fig. 8 shows the total warp and fills deviation < 0.01. The similarity convergence from the mathematical shape can be clearly seen in Fig. 7. Similar case obtained also for the Möbius strip R/W = 2.2 and 2.7.

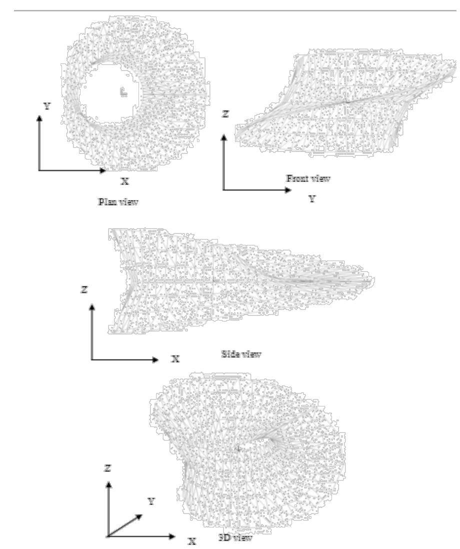


Fig. 6: Initial assumed shape for Möbius strip (R/W = 1.7)

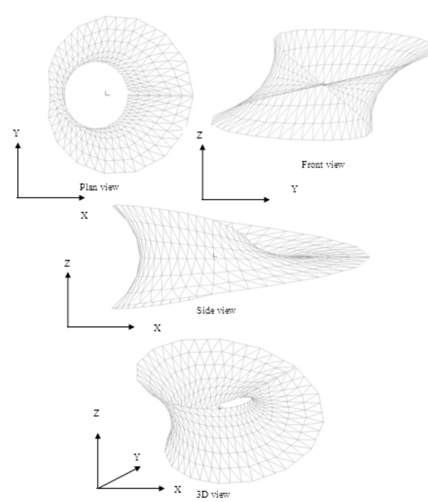


Fig. 7: TFS model in the form Möbius strip after form-finding (R/W = 1.7)

ISBN: 978-1-61804-230-9

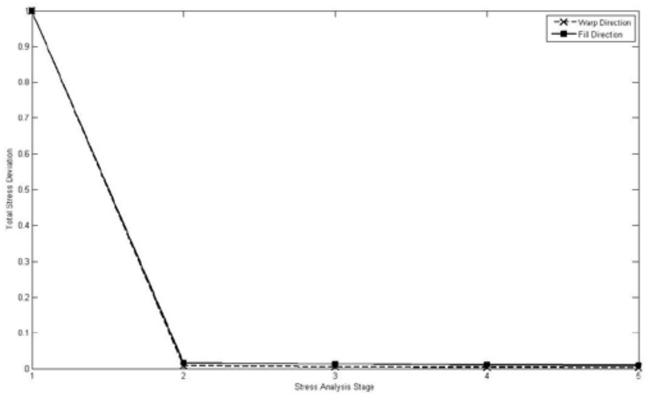


Fig. 8: Variation of total stress deviation in warp and fill direction verses stress analysis stage for Möbius Strip TFS Model (R/W = 1.7)

V.CONCLUSION

Form-finding on TFS with surface in the form of Möbius strip R/W= 0.7, 1.2, 1.7, 2.2 and 2.7 has been carried out successfully using the procedure adopted which is based on nonlinear analysis method. The results from this computational study show that TFS in the form of Möbius strip with proper selection of surface parameter is a structurally viable surface form to be considered.

REFERENCES

[1] V. Petresin and L. Robert, "The double Möbius strip studied", *Nexus Network Journal*, vol. 4, pp. 4, 2002.

[2] K.K. Choong and M.S. Kuan, "Structural behaviour of shell surface in the form of Möbius strip Published Conference Proceedings style)," in *Proc. 6th Sixth International Conference on Computation of Shell and Spatial Structures*, New York, 2008, pp. 1–4.

[3] H.M. Yee, "A computational strategy for form-finding of tensioned fabric structure using nonlinear analysis method," Ph.D. dissertation, School of Civil Engineering, Universiti Sains Malaysia, Pulau Pinang, Malaysia, 2011.

[4] H.M. Yee, K.K. Choong and J.Y. Kim, "Form-finding analysis of tensioned fabric structures using nonlinear analysis method," *Advanced Materials Research*, vol. 243-249, pp. 1429–1434, May 2011.

[5] *ADINA*, System 8.1. R&D Inc., 2003.

[6] Weisstein and W. Eric. (2013, May 5). Networks [Online]. Available: http://mathworld.wolfram.com/ MöbiusStrip.html

Hooi Min YEE borned in Sungai Petani, Kedah, Malaysia, graduated with a doctorate degree structural engineering in 2011, Master of Science Engineering in 2007 and Bachelor of Engineering (Honours) Civil Engineering in 2006 from the Universiti Sains Malaysia, Pulau Pinang, Malaysia.

During her PhD study, she worked as fellow at Universiti Sains Malayisa. After graduating, she worked as ASSISTANT PROFESSOR at Faculty of Civil Engineering, Universiti Teknologi MARA, Pulau Pinang, Malaysia. Her research interest is Computational Mechanics, Computational Analysis of Shell and Spatial Structures, Nonlinear Analysis and Architectural Engineering.

Ir. Dr. Yee is a Professional Engineer of Board of Engineers Malaysia, Institution of Engineers Malaysia, Construction Industry Development Board Malaysia and Concrete Society of Malaysia. Ir. Dr. Yee has obtained 6 awards in Invention, Innovation and Research Design Platform and has published 36 up-to-date publications.

ESTIMATES OF THE FIRST EIGENVALUE OF THE LAPLACIAN WHICH ACTS ON SYMMETRIC TENSORS

Mikeš J., Tsyganok I.I., Stepanova E.S.

Abstract— In the present paper we show spectral properties of a little-known second-order elliptic differential operator acting on symmetric tensors.

Keywords—Riemannian manifold, second order elliptic differential operator on symmetric tensors, eigenvalues and eigentensors.

I. INTRODUCTION

On a Riemannian manifold (*M*, *g*), the differential operators given information on the geometry of (*M*, *g*) must be tied to its metric *g*. In the present paper we will be interested in the special differential operator $\Box$ of second order, which acts on symmetric tensors. This operator is an analogue of the well-known Hodge-de Rham Laplacian Δ which acts on exterior differential forms. Moreover, the operator $\Box$ is a self-adjoint Laplacian operator and its kernel is a finite-dimensional vector space on the compact Riemannian manifold (*M*, *g*). The Laplacian $\Box$ admits the Weitzenböck decomposition and hence to study it, we use the Bochner technique. The main purpose of this paper is to estimate the first eigenvalue of the operator $\Box$.

II. DEFINITIONS AND NOTATIONS

2.1. Let (*M*, *g*) be a compact oriented C^∞-Riemannian manifold of dimension *n* and S^pM be the symmetric tensor product of order *p* of the tangent bundle *TM* of *M*. On arbitrary tensor spaces on *M* we have the canonical point-wise scalar product $g(\,\cdot\,,\cdot\,)$ and their C^∞-sections the *global scalar product* $\langle\,\cdot\,,\cdot\,\rangle$. In particular, for any $\varphi,\varphi'\in C^\infty S^pM$ we have

$$\langle\varphi,\psi\rangle=\int_M\frac{1}{p!}g(\varphi\,,\varphi'\,)\,dv\,, \qquad (2.1)$$

The paper was supported by Grant P201/11/0356 of the Czech Science Foundation and by the Project POST-UP CZ 1.07/2.3.00/30.0004.

Mikeš J., Stepanova E.S. are with Department of Algebra and Geometry, Palacky University, 77146 Olomouc, Listopadu 12, Czech Republic,
e-mail: josef.mikes@upol.cz, stephelena@list.ru
Tsyganok I.I. is with Department of Mathematics, Finance University under the Government of Russian Federation, 125468 Moscow, Leningradsky Prospect, 49-55, Russian Federation, e-mail: i.i.tsyganok@mail.ru

where *dv* is the volume element of (*M*, *g*). If g_{ij}, $\varphi_{i_1..i_p}$ and $\varphi'_{i_1..i_p}$ denote the components of the metric *g* and tensor fields $\varphi,\varphi'\in C^\infty S^pM$ with respect to a local coordinate system $x^1,\dots,x^n$ on (*M*, *g*) then

$$\langle\varphi,\psi\rangle=\int_M\frac{1}{p!}\left(g^{i_1j_1}\dots g^{i_pj_p}\varphi_{i_1..i_p}\varphi'_{j_1..j_p}\right)dv$$

for $g^{ik}g_{jk}=\delta^i_j$.

Next, if *D* is a differential operator between some C^∞-sections of arbitrary tensor bundles over *M*, then its formal adjoin D^* is uniquely defined by the formula $\langle D\,\cdot,\cdot\rangle=\langle\,\cdot,D^*\cdot\rangle$ (see [1, p. 460]). For example, the covariant derivative $\nabla:C^\infty S^pM\to C^\infty\left(T^*M\otimes S^pM\right)$, whose formal adjoint ∇^* will be also denoted by (see [1, p. 54])

$$\delta=\nabla^*:C^\infty\left(T^*M\otimes S^pM\right)\to C^\infty S^pM$$

In local coordinates, this operator is defined by equalities $(\delta\,\xi)_{i_1..i_p}=-g^{ij}\nabla_j\,\xi_{i\,i_1..i_p}$ for any $\xi\in C^\infty\left(T^*M\otimes S^pM\right)$.

We define the operator

$$\delta^*:C^\infty S^pM\to C^\infty S^{p+1}M$$

of degree 1 by the formula $\delta^*\varphi=\mathrm{Sym}(\nabla\varphi)$ for an arbitrary $\varphi\in C^\infty S^pM$ (see [1, p. 35]). In local coordinates we have

$$\left(\delta^*\varphi\right)_{i_0i_1..i_{p-1}i_p}=\nabla_{i_0}\varphi_{i_1..i_{p-1}i_p}+\ldots+\nabla_{i_p}\varphi_{i_0i_1..i_{p-1}}\,.$$

Then $\delta:C^\infty S^{p+1}M\to C^\infty S^pM$ is the adjoint operator of δ^* with respect to the global product (2.1).

2.2. In [2] was defined the operator

$$\Box=\,\delta\,\delta^*-\delta^*\delta:C^\infty S^pM\to C^\infty S^pM$$

of degree 2 which acts on C^∞-sections of the bundle S^pM. This operator is an analogue of the well-known Hodge-de Rham Laplacian $\Delta:C^\infty\Lambda^pM\to C^\infty\Lambda^pM$ which acts on C^∞-sections of the bundle Λ^pM of covariant skew-symmetric *p*-tensors; in other words, exterior differential *p*-forms (see [1, p. 34]; [3, p. 204]). The operator $\Box$ is a self-adjoint Laplacian operator and its kernel is a finite-dimensional vector space on the compact Riemannian manifold (*M*, *g*) (see [4]). In this case, by virtue of the *Fredholm alternative* (see [1, p. 464]; [3, p. 205]) the vector spaces Ker $\Box$ and Im $\Box$ are orthogonal complements of each other with respect to the global scalar product (2.1), i.e. $C^\infty S^p M = \mathrm{Ker}\ \square \oplus \mathrm{Im}\ \square$.

Compare the operator $\square$ with the *rough Laplacian* $\nabla^*\nabla$ (see [1, p. 54]). First, it is easy to see that these two operators coincide if (M, g) is a locally Euclidean space. Second, the operator $\square - \nabla^*\nabla$ has order zero and can be defined by symmetric endomorphism B_p of the bundle S^pM, where B_p can be algebraically (even linearly) expressed through the curvature and Ricci tensors of (M, g) (see [1, p. 53]; [2]). This can be expressed by the Weitzenböck formula $\square = \nabla^*\nabla + B_p$ and means that $\square$ is a *Lichnerowicz Laplacian*. In local coordinates we have

$$B_p(\varphi)_{i_1..i_p} = -\sum_k g^{js} r_{i_k j}\, \varphi_{i_1..i_{k-1} s i_{k+1}..i_p} + \\ + \sum_{k\neq l} g^{js} g^{mt} R_{i_k j i_l m}\, \varphi_{i_1..i_{k-1} s i_{k+1}..i_{l-1} t i_{l+1}..i_p},$$

where r_{ij} and R_{ijkl} are components of the Ricci tensor *Ric* and curvature tensor R with respect to a local coordinate system $x^1,\ldots,x^n$.

Remark. The original Lichnerowicz Laplacian (see [1, p. 53-54]) has the form $\Delta_L T = \nabla^*\nabla T + \Gamma_p(T)$ where

$$\Gamma_p(\varphi)_{i_1..i_p} = \sum_k g^{js} r_{i_k j}\, T_{i_1..i_{k-1} s i_{k+1}..i_p} - \\ - \sum_{k\neq l} g^{js} g^{mt} R_{i_k j i_l m}\, T_{i_1..i_{k-1} s i_{k+1}..i_{l-1} t i_{l+1}..i_p}$$

for local components $T_{i_1..i_p}$ of $T \in C^\infty\left(\otimes^p T^*M\right)$.

In particular, if (M, g) is a locally Euclidean manifold then the equation $\square\ \varphi = 0$ becomes $\sum_k \frac{\partial^2 \varphi_{i_1\ldots i_p}}{\left(\partial x^k\right)^2} = 0$ with respect to a local Cartesian coordinate system $x^1,\ldots,x^n$. This means that all components of a symmetric p-tensor φ on a locally Euclidean manifold (M, g) are harmonic functions belonging to the kernel of the Laplacian $\square$. This property is also typical for harmonic exterior differential p-forms (see [3, p. 205-212]). Therefore, the symmetric tensor $\varphi \in \ker \square$ was named in [2] as a *harmonic symmetric tensor* on a Riemannian manifold (M, g).

Remark. An application of the theory of harmonic symmetric tensors can be find in our paper [5].

2.3. For $p \geq 2$ the explicit expression for $\square$ is complicated but for $p = 1$ it has the form $\square = \nabla^*\nabla + B_1$ where $B_1 = -Ric$. It is well known that $\Delta = \nabla^*\nabla + Ric$ then we obtain $\square = \Delta - 2\,Ric$, where, in local coordinates, $(\square\,\varphi)_i = (\Delta\varphi)_i - 2\,r_{ij}\,\varphi^j$ (see [4]). This form of the operator $\square$ was used by K. Yano (see [6, p. 40]) for the investigation of *local isometries* of (M, g) preserving the metric g. Therefore, in paper [4], we called it as the *Yano operator* or, in other words, the *Yano Laplacian.*

On the other hand, a vector field X on (M, g) is called an *infinitesimal harmonic transformation* if the one-parameter group of local transformations of (M, g) generated by X consists of local harmonic diffeomorphisms (see [8]). At the same time, the equality $\square\,\varphi = 0$ for $\varphi = g(X,\cdot)$ is necessary and sufficient condition for a vector field X to be an infinitesimal harmonic transformation on (M, g) (see [4]).

Infinitesimal isometric transformations of Riemannian manifolds (see [6, p. 44]), infinitesimal conformal transformations of two-dimensional Riemannian manifolds (see [6, p. 47]), affine Killing vector fields (see [6, p. 45]) and geodesic vector fields (see [7]) of Riemannian manifolds, a vector field that transforms a Riemannian metric into a *Ricci soliton* metric are examples of infinitesimal harmonic transformations (see [4]).

III. MAIN RESULTS

3.1. The curvature tensor R defines a linear endomorphism $\overset{\circ}{R} : S^2(M) \to S^2(M)$ by the formula $\overset{\circ}{R}(\varphi)_{il} = R_{ijkl}\,\varphi^{jk}$ where R_{ijkl} are local components of R and $\varphi^{ij} = g^{ik} g^{jl} \varphi_{kl}$ for local components φ_{ij} of an arbitrary $\varphi \in S^2(M)$. The symmetries of R imply that $\overset{\circ}{R}$ is a selfadjoint operator, with respect to the pointwise inner product on $S^2(M)$. Hence the eigenvalues of $\overset{\circ}{R}$ are all real numbers at each point $x \in M$. Thus, we say $\overset{\circ}{R}$ is positive (resp. negative), or simply $\overset{\circ}{R} > 0$ (resp. $\overset{\circ}{R} < 0$), if all the eigenvalue of $\overset{\circ}{R}$ are positive (resp. negative). This endomorphism $\overset{\circ}{R} : S^2(M) \to S^2(M)$ was named the *operator of curvature of the second kind* (see [10]).

Remark. The definition, properties and applications of $\overset{\circ}{R}$ can be found in the following monograph and papers [1]; [9]; [10]; [11]; [12] and etc.

3.2. A real number λ^r for which there is a tensor $\varphi \in C^\infty S^p M$ which is not identically zero such that $\square\ \varphi = \lambda^p\,\varphi$ is called an *eigenvalue* of $\square$ and the corresponding $\varphi \in C^\infty S^p M$ is called an *eigentensor* of $\square$ corresponding to λ^p. Then the following theorem is true.

Theorem. *Let* (M, g) *be an n-dimensional compact oriented Riemannian manifold. Suppose the curvature operator of the second kind* $\overset{\circ}{R} : S^2(M) \to S^2(M)$ *is negative and bounded above by some negative number* $-\varepsilon$ *at each point* $x \in M$. *Then an arbitrary eigenvalue* λ^p *of the Yano Laplacian* $\square : C^\infty S^p M \to C^\infty S^p M$ *satisfies the inequality*

$$\lambda^p > p(n+p-2)\varepsilon\,.$$

In the case $p = 1$ we find the corollary from the above theorem.

Corollary. *Let* (M, g) *be an n-dimensional compact orient-*

ISBN: 978-1-61804-230-9

ed Riemannian manifold. Suppose the Ricci tensor satisfies the inequality $Ric \le -(n-1)\varepsilon\, g$ *for some positive number* ε*, then an arbitrary eigenvalue* λ^1 *of the Yano Laplacian* $\Box : C^\infty S^1 M \to C^\infty S^1 M$ *satisfies the inequality*

$$\lambda^1 > (n-1)\varepsilon.$$

IV. PROOFS OF STATEMENTS

4.1. Let (M, g) be a compact oriented C^∞-Riemannian manifold of dimension n and $\Box : C^\infty S^p M \to C^\infty S^p M$ be the Yano Laplacian. Then the following proposition is true.

Lemma 1. *Let* (M, g) *be a compact oriented* C^∞*-Riemannian manifold of dimension* $n \ge 2$ *and* $\Box : C^\infty S^p M \to C^\infty S^p M$ *be the Yano Laplacian then*

$$\langle \Box\,\varphi, \varphi \rangle = -\langle B_p(\varphi), \varphi\rangle + \langle \nabla\varphi, \nabla\varphi\rangle = 0 \quad (4.1)$$

where $\langle B_p(\varphi), \varphi\rangle = \int_M \frac{1}{(p-1)!} g(B_p(\varphi), \varphi)\, dv$.

Proof. We introduce in our discussion a vector field X with local components $\varphi^j{}_{i_2..i_p} \nabla_j \varphi^{i i_2..i_p}$ where

$$\varphi^{j_1..j_q}{}_{i_{q+1}..i_p} = g^{i_1 j_1} \ldots g^{i_q j_q} \varphi_{i_1..i_q i_{q+1}..i_p}$$

for the Kronecker delta δ^i_j. Then

$$\begin{aligned}\operatorname{div} X &= \nabla_i \left(\varphi^j{}_{i_2..i_p} \nabla_j \varphi^{i i_2..i_p}\right) = \\ &= \nabla_i \varphi_{j i_2..i_p} \nabla^j \varphi^{i i_2..i_p} + \varphi^j{}_{i_2..i_p} \nabla_i \nabla_j \varphi^{i i_2..i_p} = \\ &= \nabla_i \varphi_{j i_2..i_p} \nabla^j \varphi^{i i_2..i_p} + r_{ij}\, \varphi^{i i_2..i_p} \varphi^j{}_{i_2..i_p} - \\ &- (p-1)\, R_{ijkl} \varphi^{i k i_3..i_p} \varphi^{jl}{}_{i_3..i_p}\end{aligned},$$

where $r_{ij} = R^k{}_{ikj}$ are local components of the Ricci tensor Ric. In turn, for the vector field Y with local components $\varphi^i{}_{i_2..i_p} \nabla_j\, \varphi^{j i_2..i_p}$, we have the equalities

$$\begin{aligned}\operatorname{div} Y &= \nabla_i \left(\varphi^i{}_{i_2..i_p} \nabla_j\, \varphi^{j i_2..i_p}\right) = \\ &= \nabla_i\, \varphi^i{}_{i_2..i_p} \nabla_j\, \varphi^{j i_2..i_p} + \varphi^i{}_{i_2..i_p} \nabla_i \nabla_j\, \varphi^{j i_2..i_p}\end{aligned}.$$

Then it follows from the Green theorem $\int_M \operatorname{div} Z\, dv = 0$ for $Z = X - Y$ that

$$\begin{aligned}\int_M \Big(& r_{ij}\, \varphi^{i i_2..i_p} \varphi^j{}_{i_2..i_p} - (p-1) R_{ijkl} \varphi^{i k i_3..i_p} \varphi^{jl}{}_{i_3..i_p} + \\ &+ \nabla_{i_0} \varphi_{i_1 i_2 \ldots i_p} \nabla^{i_1} \varphi^{i_0 i_2..i_p} - \nabla_i\, \varphi^i{}_{i_2..i_p} \nabla_j\, \varphi^{j i_2..i_p} \Big) dv = 0\end{aligned}. \quad (4.2)$$

Since

$$\begin{aligned}&\nabla_{i_0} \varphi_{i_1 i_2 \ldots i_p} \nabla^{i_1} \varphi^{i_0 i_2..i_p} = \\ &= \frac{1}{p(p+1)} \left(\nabla_{i_0} \varphi_{i_1 i_2 \ldots i_p} + \ldots + \nabla_{i_p} \varphi_{i_0 i_1 \ldots i_{p-1}}\right) \cdot \\ &\cdot \left(\nabla^{i_0} \varphi^{i_1 i_2 \ldots i_p} \ldots + \nabla^{i_p} \varphi^{i_0 i_1 \ldots i_{p-1}}\right) - \\ &- \frac{1}{p} \left(\nabla_{i_0} \varphi_{i_1 i_2 \ldots i_p}\right) \left(\nabla^{i_0} \varphi^{i_1 i_2 \ldots i_p}\right)\end{aligned}$$

the integral formula (4.2) can be given by the form

$$\langle B_p(\varphi), \varphi\rangle + \langle \delta^*\varphi, \delta^*\varphi\rangle - \langle \delta\varphi, \delta\varphi\rangle - \langle \nabla\varphi, \nabla\varphi\rangle = 0 \quad (4.3)$$

where

$$\langle B_p(\varphi), \varphi\rangle = \int_M \frac{1}{(p-1)!} g(B_p(\varphi), \varphi) dv = \int_M \frac{1}{(p-1)!} \cdot \left(r_{ij}\, \varphi^{i i_2..i_p} \varphi^j{}_{i_2..i_p} - (p-1) R_{ijkl} \varphi^{i k i_3..i_p} \varphi^{jl}{}_{i_3..i_p}\right) dv;$$

$$\langle \delta^*\varphi, \delta^*\varphi\rangle = \int_M \frac{1}{(p+1)!} \left(\nabla_{i_0} \varphi_{i_1 i_2 \ldots i_p} + \ldots + \nabla_{i_p} \varphi_{i_0 i_1 \ldots i_{p-1}}\right) \cdot \left(\nabla^{i_0} \varphi^{i_1 i_2 \ldots i_p} \ldots + \nabla^{i_p} \varphi^{i_0 i_1 \ldots i_{p-1}}\right) dv;$$

$$\langle \delta\varphi, \delta\varphi\rangle = \int_M \frac{1}{(p-1)!} \left(\nabla_i\, \varphi^i{}_{i_2..i_p}\right)\left(\nabla_j\, \varphi^{j i_2..i_p}\right) dv;$$

$$\langle \nabla\varphi, \nabla\varphi\rangle = \int_M \frac{1}{p!} \left(\nabla_{i_0} \varphi_{i_1 i_2 \ldots i_p}\right)\left(\nabla^{i_0} \varphi^{i_1 i_2 \ldots i_p}\right) dv.$$

The operator $\Box$ satisfies the property

$$\langle \Box\,\varphi, \varphi\rangle = \langle \delta^*\varphi, \delta^*\varphi\rangle - \langle \delta\,\varphi, \delta\,\varphi\rangle,$$

which follows immediately from its definition. Therefore, the integral formula (4.3) can be rewritten in the form (4.1).

4.2. Let (M, g) be an n-dimensional Riemannian manifold. If there exists a positive constant ε such that

$$R_{ijkl}\, \varphi^{il} \varphi^{jk} \le -\varepsilon\, \varphi_{ij} \varphi^{ij} \quad (4.4)$$

holds for any $\varphi \in C^\infty S^2 M$, then (M, g) is said to be a *Riemannian manifold of negative restricted curvature operator of the second kind.*

Lemma 2. *Let* (M, g) *be an n-dimensional compact oriented Riemannian manifold and* $\Box$ *be the Yano Laplacian which acts on symmetric tensors. Suppose the curvature operator of the second kind* $\overset{\circ}{R} : S^2(M) \to S^2(M)$ *is negative and bounded above by some negative number* $-\ \varepsilon$ *at each point* $x \in M$*, then*

$$\langle \Box\,\varphi, \varphi\rangle \ \ge\ p(n+p-2)\varepsilon \langle \varphi, \varphi\rangle + \langle \nabla\varphi, \nabla\varphi\rangle \quad (4.5)$$

Proof. From (4.4) we obtain the following inequalities

$$\begin{aligned}&-(p-1) R_{ijkl}\, \varphi^{i k i_3..i_p}\, \varphi^{jl}{}_{i_3..i_p} = \\ &= (p-1) R_{ijlk}\, \varphi^{i k i_3..i_p}\, \varphi^{jl}{}_{i_3..i_p} \le -(p-1)\varepsilon\ \varphi_{i_1..i_p}\, \varphi^{i_1..i_p}\end{aligned};$$

$$r_{ij}\,\varphi^{i\,i_2\ldots i_p}\varphi^{j}{}_{i_2\ldots i_p} \le -(n-1)\varepsilon\,\varphi_{i_1\ldots i_p}\,\varphi^{i_1\ldots i_p},$$

which show that $\langle B_p(\varphi),\varphi\rangle \le -p\,(n+p-2)\langle \varphi,\varphi\rangle$. If we take account of (4.1), we obtain (4.5).

Then for an eigentensor φ corresponding to an eigenvalue λ^p (4.5) becomes the inequalities

$$\lambda^p\langle \varphi,\varphi\rangle \ge p(n+p-2)\varepsilon\langle \varphi,\varphi\rangle + \langle\nabla\varphi,\nabla\varphi\rangle \ge$$
$$\ge p(n+p-2)\varepsilon\langle \varphi,\varphi\rangle \quad (4.6)$$

which prove that

$$\lambda^p \ge p(n+p-2)\varepsilon > 0. \quad (4.7)$$

If the equality is valid in (4.7), then from (4.6) we obtain $\nabla\varphi = 0$ and $B_p(\varphi)=0$. In this case we have the identity $\varphi = 0$, because $B_p(\varphi)$ is strictly negative for an arbitrary nonzero $\varphi\in C^\infty S^pM$. This completes the proof.

4.3. Let (*M*, *g*) be an *n*-dimensional compact and oriented Riemannian manifold, and define the Ricci tensor *Ric* of (*M, g*) by its components r_{ij} with respect to a local coordinate system $x^1,\ldots,x^n$. If there exists a positive constant ε such that

$$r_{ij}\,\varphi^i\,\varphi^j \le -(n-1)\varepsilon\;\varphi_i\,\varphi^i \quad (4.8)$$

holds for any $\varphi\in C^\infty S^1M$, then (*M, g*) is said to be a *Riemannian manifold of negative restricted Ricci curvature.*

Next, suppose the Ricci is negative and bounded above by some negative number – $(n-1)\varepsilon$ at each point $x\in M$, i.e. $Ric \le -(n-1)\varepsilon\, g$, then from $\square = \nabla^*\nabla - Ric$ we obtain the inequality

$$\langle \square\,\varphi,\varphi\rangle \ge (n-1)\varepsilon\langle \varphi,\varphi\rangle + \langle\nabla\varphi,\nabla\varphi\rangle \quad (4.9)$$

for an arbitrary $\varphi\in C^\infty S^1M$. Then for an eigenform φ corresponding to an eigenvalue λ^1, (4.9) becomes the inequalities

$$\lambda^1\langle \varphi,\varphi\rangle \ge (n-1)\varepsilon\langle \varphi,\varphi\rangle + \langle\nabla\varphi,\nabla\varphi\rangle \ge$$
$$\ge (n-1)\varepsilon\langle \varphi,\varphi\rangle \quad (4.10)$$

which prove that

$$\lambda^p \ge (n-1)\varepsilon > 0. \quad (4.11)$$

If the equality is valid in (4.11), then from (4.10) we obtain $\nabla\varphi = 0$ and $r_{ij}\,\varphi^i\varphi^j = 0$. In this case we have the identity $\varphi = 0$, because *Ric* is strictly negative.

References

[1] Besse A., *Einstein manifolds*, Springer-Verlag, Berlin – Heidelberg, 1987.
[2] Samson J.H., On a theorem of Chern, *Transactions of the American Mathematical Society* **177** (1973), 141-153.
[3] Petersen P., *Riemannian Geometry*, Springer Science, New York, 2006.
[4] Mikeš J., Stepanov S.E., Tsyganok I.I., From infinitesimal harmonic transformations to Ricci solitons, *Mathematica Bohemica* **138**:1 (2013), 25-36.
[5] Mikeš J., Stepanova E., A five-dimensional Riemannian manifold with an irreducible *SO*(3)-structure as a model of abstract statistical manifold, *Annals of Global Analysis and Geometry* (published in July 2013 with open access at Springerlink.com, http://link.springer.com/article/10.1007%2Fs10455-013-9390-0/fulltext.html).
[6] Yano K., *Integral formulas in Riemannian geometry*, Marcel Dekker, New York, 1970.
[7] Yano K., Nagano T., On geodesic vector fields in a compact orientable Riemannian space, *Comment. Math. Helv.* **35** (1961), 55-64.
[8] Nouhaud O. Transformations infinitesimales harmoniques, *C. R. Acad. Paris, Ser*. A **274** (1972), 573-576.
[9] Bouguignon J.-P., Karcher H., Curvature operators: pinching estimates and geometric examples, *Ann. Sc. Éc. Norm. Sup.* **11**(1978), 71-92.
[10] Tachibana S., Ogiue K., Les variétés riemanniennes dont l'opérateur de coubure restreint est positif sont des sphères d'homologie réelle, *C. R. Acad. Sc. Paris* **289** (1979), 29-30.
[11] Nishikawa S., On deformation of Riemannian metrics and manifolds with positive curvature operator, *Lecture Notes in Mathematics* **1201** (1986), 202-211.
[12] T. Kashiwada, On the curvature operator of the second kind, *Natural Science Report, Ochanomizu University* **44**:2 (1993), 69-73.

Critical Success Factors for Sharing Economy among SMEs

Soon Goo Hong, Hyun Jong Kim, Hyung Rim Choi, Kangbae Lee, Min-Je Cho

***Abstract*—**Small- and medium-sized enterprises (SMEs) face a range of management problems for which the application of sharing economy may be considered a potential method of resolution. The introduction of sharing economy may well save costs and resources as well as improve negotiation power and reliability. This study attempts to prioritize success factors that should be considered when introducing sharing economy among SMEs and then derive suggested points to realize sharing economy among them. To meet these objectives, this study conducted a Delphi survey three times on experts from companies and groups with experience collaboration and partnerships similar to that of sharing economy and organizations related to SMEs. This study is academically meaningful in that it is the first study on the success factors for sharing economy among businesses and it may be utilized as reference data for the establishment of strategies when introducing sharing economy by SMEs and of relevant policies by the government.

***Keywords*—**Sharing Economy, Small- and medium-sized enterprises (SMEs), Delphi survey.

I. Introduction

SMEs account for a large percentage of the national economy. The number of SMEs in the top seven countries with the highest GDP made up 99% of those represented and the number of people engaged in SMEs also formed a high percentage, at 70% of the total (Mun Jin-uk et al., 2011). However, SMEs confront a number of difficulties in technological competitiveness, capital, and human resources. To resolve these challenges, in addition to their own efforts, different methods have been sought for such things as government aid, the formation of coexisting relationships with large businesses, and collaboration among SMEs. Nonetheless, most SMEs are not yet equipped competitively (Han Jeong-wha, 2006).

Sharing economy may be considered as one of the measures to solve the various problems that SMEs face. Sharing economy essentially means that a company lends idle goods and services, individually owned by it, to other businesses or companies jointly investing in goods and services that are expected to become idle when a single company owns them. This may save costs and resources as well as enhance negotiation power and reliability for the participants.

At present, sharing economy practices, in the form of sharing idle resources primarily among individuals, are spreading throughout the world. Hitherto, although there has been some research on sharing economy among individuals, there has been almost no research on sharing economy among businesses. In reality, when there is no research on cases of sharing economy among businesses, it is expected that it will be difficult for SMEs to introduce and realize sharing economy. Therefore, research directed toward the successful introduction and realization of sharing economy by SMEs is necessary.

Accordingly, this paper will derive success factors to introduce sharing economy among SMEs. To achieve the goals of this paper, a Delphi survey was conducted three times on experts from companies and groups with experience in collaboration and partnerships similar to that of sharing economy and organizations related to SMEs.

II. Theoretical Background

A. The Concept of Sharing Economy

Sharing economy can be defined as collaborative activities under which a company or an individual lends idle goods and services individually owned to other businesses or individuals, or individuals or companies jointly invest in goods and services that are expected to become idle when they own them individually (Hong Sun-gu et al., 2013).

The term “sharing economy” was coined by Professor Lawrence Lessing at Harvard University, in 2008; it emerged as a new paradigm of consumption that prevents excessive consumption based on collaborative consumption and shares, exchanges, and lends goods for resource and cost saving, not owning them (Botsman, 2010). Currently, the practice of sharing houses, offices, home appliances, and vehicles is expanding among individuals (The Korea Transport Institute, 2012). The most representative example of sharing economy among individuals is Airbnb, which provides accommodation services to travelers. Airbnb, which started in the U.S., is expanding its service globally. Zicar also introduced the method of sharing economy and lends vehicles in hour units, instead of day units.

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A5A2A03034683).

S. G. Hong is Professor of Dong-A University, Korea (e-mail: shong@dau.ac.kr).

H. J. Kim is with Dong-A University, Korea (e-mail: hatdol@gmail.com).

H. R. Choi is Professor of Dong-A University, Korea (e-mail: hrchoi@dau.ac.kr).

K. Lee is Professor of Dong-A University, Korea (e-mail: kanglee@dau.ac.kr).

M. J. Cho is Professor of Dong-A University, Korea (e-mail: mhcho@dau.ac.kr).

Such a concept of sharing economy is applicable to businesses. Sharing economy among companies can be defined as collaborative activities under which a company lends idle goods and services individually owned by it to other businesses, or companies invest jointly in and use goods and services that are expected to become idle when they own them individually.

Sharing economy among companies may be divided largely into two types. Type 1 is to share idle resources owned by individual companies, and Type 2 is to jointly own resources, through joint investment, that are expected to become idle when owned individually. Specifically, Type 1 is to lend tangible and intangible goods, such as facilities, equipment, and intellectual property rights, owned by individual companies to other companies. Type 2, which is to jointly own goods through joint investment, may be classified into detailed types according to their types and subjects, as shown in Table 1. Other existing concepts that are similar to sharing economy among businesses include collaboration, strategic partnership, outsourcing, and joint business. Companies have made efforts to save costs and strengthen competitiveness through such methods. These kinds of efforts are conducted through agreements and consultations among a small number of businesses, or by cooperative associations with a certain purpose and toward a certain subject on a short-term one-time basis. Sharing economy has a broader purpose and subject in terms of resource sharing and the number of companies that participate in it is not limited. Therefore, the concept of sharing economy encompasses all of the above.

Table 1. Type of Sharing Economy

Item	Type	Subject	Content
Goods	Joint facilities	Facilities, equipment, and land	Jointly investing in (owning) and sharing facilities that are expected to become idle
	Joint purchase	Purchases raw and subsidiary materials and human resources	Procuring cheaply and stably through large purchases and dividing idle purchases resulting from large purchases.
	Joint development	Research and development resources and results	Joint development of brands and products
	Joint marketing	Marketing resources	Jointly using marketing resources such as advertisement fees, marketing human resources, spaces, advertisement channels, and distribution and selling networks
Services	Joint education	Educational resources	Sharing resources necessary for education of human resources among businesses
	Joint service	Airline seats, follow-up service	Jointly utilizing service provision routes aimed at increasing profits for service improvement and in consideration of service product characteristics.

B. Delphi Method

A Delphi method is used to predict the future through experts. Since executing a study using this method is easy and simple, it may be utilized when estimating conflicting relationships among interest groups, as a tool to receive opinions from the many, and for predicting the future. Although the Delphi method uses an expert group to produce group estimates and it may involve uncertainty and ambiguity, the technique is used to find exact estimates using a statistical method under the assumption that these group estimates are highly likely to include the range of correct answers.

In an early study of sharing economy, because the respondents were not well aware of the concept, it was difficult to derive success factors by conducting a general survey. Therefore, the Delphi method that examines success factors by obtaining opinions from experts in the relevant area is appropriate.

C. Figures

All tables and figures will be processed as images. **However, EUROPMENT cannot extract the tables and figures embedded in your document.** The figures and tables you insert in your document are only to help you gauge the size of your paper, for the convenience of the referees, and to make it easy for you to distribute preprints.

D. Document Modification

If you modify this document for use with other **EUROPMENT**, journals or conferences, you should save it as type “Word - RTF (*.doc)” so that it can be opened by any version of *Word*.

III. Study Method and Content

A. Study Design

A Delphi method was used to derive success factors aimed at introducing and realizing sharing economy. Since the Delphi method depends on knowledge and collaboration of the participating experts, the selection of the experts and the determination of the number of surveys are important. When the number of experts is around 15, the difference in the median value is not large (Lee Seong-wung, 1987). Under the Delphi method, a reasonable number of surveys should be determined for deriving an agreement among the participants. Changes in the experts’ answers to the surveys are the largest in the first and second rounds of the surveys; from the fifth round of the survey on, there is no change in their answers (Roh Seung-yun, 2006). In existing research, the survey was conducted three to four times; therefore, this study also conducted three rounds of surveys.

For the Delphi survey, eighteen subjects were selected from experts in companies and groups that have proceeded with collaboration and partnerships similar to that of sharing economy and institutions related to SMEs. The subjects

ISBN: 978-1-61804-230-9

consisted of 10 experts who were affiliated with SME cooperatives and eight experts in institutions related to SMEs. All of the experts had at least five years of experience.

B. Delphi Survey

1) The first round of Delphi surveys

The survey used an open questionnaire, in which the experts could write their opinions freely regarding the success factors of sharing economy. Similar answers among the collected diverse answers, about success factors, were categorized and their frequency was analyzed. As shown in Table 2, five factors in Type 1 and seven factors in Type 2 were derived.

Table 2. First Surveys

Item	Success	Frequency	Percentage (%)
Type 1	Economic interests of businesses	14	77.8
	Trust between the providing company and the receiving company	13	72.2
	Sharing of information about idle resources between companies	7	38.9
	Proximity between companies (location, distance)	7	38.9
	Usefulness of provided infrastructure	6	33.3
	Participating companies' will and efforts	3	16.7
Type 2	Economic interests of participating companies	18	100
	Hosting company or groups will and efforts	17	94.4
	Support by	18	100
	Hosting company or groups trust	11	61.1
	Trust among participating companies	9	50
	Usefulness of provided infrastructure	7	38.9
	Participating companies' will and efforts	6	33.3

2) The second round of Delphi surveys

The importance level of the five factors in Type 1 and the seven factors in Type 2, which were derived from the results of the first survey round, were examined using a Likert 7-point scale (7 = very high and 1 = very low). As displayed in Table 3, the result was that a majority of the questions obtained 4 (moderate) or higher points. In Type 1, economic interests, trust between the providing company and the receiving company, and sharing information about idle resources between companies obtained 6 points (high) or higher. In Type 2, the participating companies' economic interests and support by the central and local governments obtained 6 points or higher. The lowest scores obtained were for the utility of the provided infrastructure and the participating companies' will and efforts to participate.

The validity of the method was verified through the degree of agreement among the experts' opinions. To this end, the agreement, convergence, and variance rates of each question were examined. The smaller the variance value was, the closer the convergence rate was to 0, and the closer the agreement rate was to 1, opinions converged were agreed. Among the total 11 questions, the convergence rates of four questions were larger than 0.5 and the agreement rates of four questions were relatively small; however, the remaining questions were agreed among the respondents.

Table 3 Second Surveys

Item	Success	Average	Variance	Median value	Convergence rate	Agreement rate	Coefficient of variation
Type 1	Economic interests	6.44	0.27	6	0.5	0.83	0.08
	Trust between the providing company and the receiving company	6.28	0.42	6	0.5	0.83	0.11
	Sharing of information about idle resources between companies	6.11	0.49	6	0.75	0.75	0.17
	Proximity between companies (location, distance)	5.22	0.56	5	0.5	0.80	0.14
	Usefulness of provided infrastructure	5.00	0.49	5	0.5	0.80	0.14
Type 2	Economic interests of a participating business	6.44	0.47	7	0.5	0.85	0.10
	Support by	6.33	0.45	6	0.5	0.83	0.11
	Hosting company or groups will and efforts	5.83	0.69	6	0.5	0.83	0.14
	Hosting company or groups trust	5.61	0.65	6	0.5	0.83	0.15
	Trust among participating companies	5.11	0.6	5	0.75	0.70	0.15
	Usefulness of provided infrastructure	5.11	0.69	5	1	0.60	0.18
	Participating companies' will and efforts to participate	4.94	1.09	5	1	0.60	0.19

3) The third round of Delphi surveys

In the third round of Delphi surveys, the median values of the results from the second round of surveys were presented to the experts and the importance of the items was measured again using the Likert 7-point scale. This survey shows how the opinions of the experts have been revised and their agreement has been reached. The survey results were compared with the average, the variance, and the results of the second Delphi round; to verify how much agreement has been reached, the convergence rate, agreement rate, and the coefficient of variation were compared. According to the survey results shown in Table 4, the average value moved close to the median value presented in the results of the second Delphi round and the variance also decreased. Compared to the results of the second round of Delphi surveys, the convergence rate became closer to 0, the agreement rate became closer to 1, and the coefficient of variance decreased, showing changes in the direction of agreement.

Table 4. Third Surveys

Item	Success	Average	Variance	Median value	Convergence rate	Agreement rate	Coefficient of variation
Type 1	Economic interests	6.39	0.25	6	0.5	0.83	0.08
	Trust between the providing company and the receiving company	6.28	0.21	6	0.5	0.83	0.07
	Sharing of information about idle resources between companies	6.00	0.35	6	0	1	0.10
	Proximity between companies (location, distance)	5.22	0.30	5	0.5	0.80	0.11
	Usefulness of provided infrastructure	5.11	0.34	5	0.25	0.90	0.11
Type 2	Economic interests of a participating business	6.67	0.24	7	0.5	0.86	0.07
	Support by	6.28	0.21	6	0.5	0.83	0.07
	Hosting company or groups will and efforts	6.00	0.35	6	0	1	0.10
	Hosting company or groups trust	5.83	0.38	6	0.5	0.83	0.11
	Participating companies' will and efforts to participate	5.11	0.46	5	0.5	0.80	0.13
	Utility of provided infrastructure	5.06	0.41	5	0.25	0.90	0.13
	Trust among participating companies	4.89	0.46	5	0.5	0.80	0.14

C. *Suggestive Points*

The following suggestive points have been presented through success factors of Type 1, which shares idle resources among individual SMEs and Type 2, which shares resources expected to become idle among SMEs.

First, economic interests were found to be most important as a success factor to realize sharing economy among SMEs in Types 1 and 2. This appears to be natural, because interest should be obtained through sharing economy. Companies largely show lukewarm responses to activities that include sharing economy in cases where there is no tangible economic interest to them. In particular, inducing participation by SMEs under economically difficult conditions will be hard if there is no tangible interest to them. To realize sharing economy among SMEs, it is crucial to induce participation by companies; to this end, values and interests that may be obtained through sharing economy should be clearly presented. Participation by companies will only be induced when the industries and types of business sharing to introduce, and the business models of sharing economy by each shared subject, are established and the benefits to the SMEs are clearly presented.

Second, trust and sharing information among companies were found to be important in Type 1. This means that obtaining information for companies to directly share idle resources and trust among companies, shared subjects, are vital. In general, there are many cases in which reliability among companies including SMEs is not high. Trust among companies is lacking because they worry about technology and information leakage, do not trust their competitors, and have experienced failure in existing partnership business (Kim Seung-il, 2007). In addition to worry about technology and information leakage, in the process of sharing economy where they lend or borrow idle resources, they may have a sense of anxiety about the quality of shared resources, damages to them when they use them, and payment of utilization fees. Measures to remove such worries and heighten reliability among participating companies are clearly needed. To this end, past use history and reputation of companies that provide and receive idle resources should be able to be checked and safety devices to hold companies responsible for provision and use should be made.

For smooth sharing economy among individual companies, information regarding idle resources should be easily and quickly available. Therefore, an effective system for sharing economy should be developed to provide services. An intermediate system should be operated under which the kinds and usable time of idle resources and where they exist can be examined and reservations to use them may be made with ease. This is well demonstrated by Airbnb, which lends idle residential facilities. With Airbnb, users who need accommodation may obtain information such as available time slots and fees through search and evaluate facilities using past users' reviews. Furthermore, since settlements are made by credit card, problems arising from payment of user fees may be preempted. Such an infrastructure is critical to the smooth sharing of idle resources among companies.

ISBN: 978-1-61804-230-9

Third, in Type 2, support by the central and local governments, as well as the role of the hosting companies and groups were all found to be crucial. This shows that an initial capital investment is needed for companies to share resources. Furthermore, the importance of hosting companies that push ahead with and operate such an investment is high. In Type 2, since the burden of the initial capital investment to SMEs is heavy, support by the central and local governments are essential. Support of the initial investment capital for the sharing economy by the central and local governments is needed to strengthen the competitiveness of SMEs.

There is a hosting company, or group, in cases where business is performed jointly among companies. Because a hosting company, or group, plays the most important role in preparing and operating business, its will, its efforts, and its reliability are significant. In the preparatory stage of sharing economy, a hosting company, or group, should seek methods to heighten the likelihood of business success through thorough advance preparation and by building trust so that SMEs may take part in sharing economy. Efforts to make it transparent and beneficial to participating companies should be made in operations as well. For example, through thorough preparation and active persuasion, Busan Marine Equipment Association induced participation by shipbuilding equipment companies, large shipbuilding companies, and the government, and was thus able to obtain capital to establish its joint logistic center. There were 120 member companies when the association initiated operation of the center; this number increased to 450 within three years. Participating SMEs saved 30% on their logistic cost, as compared to the past (Korea Federation of SMEs, 2011).

IV. Conclusion

This study conducted Delphi surveys to derive the needed success factors to realize sharing economy among SMEs. The study obtained opinions from experts at the Korea Federation of SMEs and SME-related institutions, derived success factors and priorities, and presented suggestive points to realize sharing economy based on the results.

Success factors to realize a sharing economy among SMEs were surveyed: in Type 1, these factors included participating companies' economic interests, sharing information on idle resources, and their trust among each other. In Type 2, they included economic interests, support by the central and local governments, and hosting companies' will and efforts. To introduce sharing economy among companies, tangible interests should be clearly shown to participating companies. Information system and services to share information aimed at showing such interests is also needed and support by the central and local governments regarding initial investment costs to push forward sharing economy is a priority.

This study is academically meaningful in that sharing economy among individuals was applied to companies, thereby deriving success factors to realize sharing economy among SMEs and inducing follow-up research in a reality where research on sharing economy among companies is lacking. Fundamentally, this study may be conducive to establishing strategies necessary to introduce sharing economy among SMEs and be used as reference data in the government and relevant institutions establishing policies to activate sharing economy among companies.

The limitations of this study are that the case research and the survey subjects were confined to Korean SMEs and it is therefore difficult to generalize the results. Furthermore, because such reliability may differ among companies that provide idle resources and those that receive them, additional research on trust among companies in Type 1 should be conducted. In addition, differing viewpoints and opinions among hosting groups and participating companies in Type 2 should also be examined.

References

[1] J. Han, “Korean economy, what's the problem? The SME market and the government, what's the problem?,” National Strategy Institute, 2005.

[2] K. Kim, “SME status in 2011,” Korea federation of small and medium business

[3] Korea federation of small and medium business. “Cooperative Association joint-venture successful casebook,” 2008.

[4] lee, Sung Woong. Study on the usefulness of Delphi method in technological forecasting. Chonbuk National University, a master's degree thesis.

[5] Nho, Seung-Yong. "The Delphi technique: with professional insight to predict the future. easy to understand Research Methodology, 2006

[6] J. Mun, I. Go, Y. Lee, C. Lee, E. Mun, J. Im, C. Chang (2011), Status of SMEs trade between Member States compared to $ 1 trillion, KOTRA, Global Issue Report, 11-010

[7] Lessig, Lawrence, “Remix: Making Art and Commerce Thrive in Hybrid Economy,” Penguin Books, . 2008.

[8] S.Hong, H. Kim, “The Critical Success Factors on a Sharing Economy for SMEs,” 30th Pan Pacific Conference, 2013.

[9] R. Botsman, R. Rogers, “What Mine is Yours,”Harpercollins books, 2010.

[10] S. Kim, "A Study on the Status and Policy of Cooperation among SME," Small Business InstituteReality, 2007.

The prediction of mortality by causes of death in Critical Illness

Valeria D'Amato[1], Vincenzo Passannante[2] , Marilena Sibillo[3]

***Abstract*—** The paper proposes a new contractual structure built within the critical illness policy model. The new product is represented by an accelerated critical illness with a special accelerated benefit in case of death for a specific cause. The inclusion of the benefit in case of a specific cause death does not involve additional cost to the life office beyond the critical illness benefit. On the contrary the new design ensures less expensive conditions in comparison with the standard policy and it is appealing from the market point of view, looking for more and more personalized clauses.

We propose a novel form for modelling framework of the product under consideration. In order to highlight the profiles of commercial attractiveness of the product, we measure the premium price of the specific accelerated critical illness, which will be illustrated for different ages and compared to a standard accelerated product. Furthermore, actuarial valuations are performed for indicating how much money is needed to fulfil the obligations of the insurance contracts defining premiums to be received and benefits to be paid..

***Keywords*—** Causes of death, Lee-Carter, Critical Illness, Insurance, forecasting.

I. Introduction

THE study of cause-specific mortality time series is one of the main sources of information for public health monitoring. Models for trends in mortality rates for different ages and sexes as well as for different countries are often based on the assumption that past trends in historical data will continue in the future. Mortality trends and related fluctuations determine changes in the causes of deaths. These causes have different age patterns and have shown different trends over recent years. At the same time, systematic changes in causes of death have been common across the industrialized economies.

Recent literature has addressed the issue of cause-specific mortality analysis. In particular, Maccheroni et al. (2007) examine how the Lee-Carter model is not suitable for the analysis by causes of death. Sherris et. al. (2010) discuss the factors driving mortality changes based on causes of death. Tuljapurkar et al. (2000) show how mortality declines have had common trends in the G7 countries, although there is evidence of variability in those trends. Booth et al. (2006) also demonstrate the difficulties related to the projections obtained by the decomposition of the population according to causes of death. Wilmoth (1995) shows how taking into account causes of death can influence projected trends and effectively highlights how cause of death influence is hidden in aggregated data.

The World Health Organization (World Health Organization, 2009) has revised the international classification of diseases (ICD) approximately every 10 years since 1900. The purpose of revision is to stay abreast of advances in medical sciences, changes in medical terminology and to ensure the international comparability of health statistics. However, the ICD revision often causes major discontinuities in trends of mortality and morbidity statistics because of changes in classification rules for selecting underlying causes of death. The ranking of leading causes of death is also affected by this revision. These discontinuities lead not only to a misinterpretation of trends in mortality, but also to misinformation about the changes in life expectancy (Kochanek et al., 1994). Furthermore, without properly correcting these discontinuities, trends in age-specific death-rates may become distorted; this distortion may lead to unreliable forecasts of life expectancy. The problem induced by the study of cause-specific mortality, related to the jumps caused by the reclassification ICD, can be mitigated by the model of Haberman et al. (2013).

The Critical Illness Insurance market, in particular in UK, growth up to 1999, was followed by a plateau in 2000 and 2001. Sales peaked in 2002, when over one million accelerated critical illness policies were sold (CMI, 2010). These policies pay an assurance benefit on the occurrence of a serious event, such as the diagnosis of an illness. Most are sold as 'accelerated benefits' riders with life insurance policies. In this context reliable projections of survival probabilities are crucial to correctly determine insurance premiums, technical provisions and other actuarial valuations.

The contribution of this work is to propose an original insurance policy, in the market of the Critical Illness, designing a contract which includes a benefit in case of specific death-cause, beyond the benefit in case of specified illness for a policyholder more and more demanding.

We introduce a modelling framework of the product under consideration and price the accelerated benfit by forecasting survival probabilities throughout tables for specific death-causes.

1 Valeria D'Amato is with the Department of Economics and Statistics, University of Salerno, Italy (e-mail: vdamato@unisa.it).

2 Vincenzo Passannante is with the Department of Economics and Statistics, University of Salerno, Italy (e-mail: vpassannante@unisa.it).

3 Marilena Sibillo is with the Department of Economics and Statistics, University of Salerno, Italy (e-mail: msibillo@unisa.it).

The layout of the paper is the following: in section 2, we introduce the novel critical illness coverage and a mark-to-market valuation framework. Section 3 describes a specific death-cause mortality model we use for projecting specific life tables. Numerical applications are illustrated in section 4. Concluding remarks are offered in section 5.

II. Critical Illness Cover with a specific accelerated benefit

The basic critical illness policy is a very simple product. Normally, you have a lump sum cash payment that's paid upon the occurrence or diagnosis of one of a number of specified diseases or conditions. There are some conditions for setting the coverage. A benefit is paid if the assured suffers or dies for one the following critical conditions (Heart Attack, Coronary Artery Bypass Surgery, Stroke, Cancer, Major Head Trauma, Severe Burns

There are two main contractual options: the Standalone Cover and the Accelerated Cover.

The acceleration is the coverage in case of two events, one relating to the risk of falling ill and the other relating to the risk of dying.

Actuarial models for Dread Disease (DD) insurance can be built up starting from a multistate structure. In Fig. 5 the following states are considered:

$\alpha =$ healthy;
$i =$ ill, dread disease suffer;
$d(D) =$ dead being due to dread disease;
$d(O) =$ deaths being due to other causes
$\lambda =$ the portion payable on DD diagnosis
$1-\lambda =$ the portion payable on death

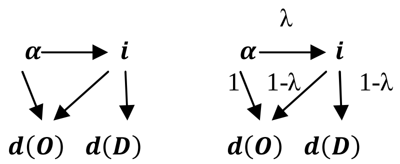


Fig. 1 Multistate schemes for DD insurance.

The following intensities define the time-continues probabilistic structure:

$\mu_y^{ai} =$ intensity of transition from a to i;

$\mu_y^{ad(O)} =$ intensity of transition from a to d(O);

$\mu_{y,r}^{id(O)} =$ intensity of transition from i to d(O);

$\mu_{y,r}^{id(D)} =$ intensity of transition from i to d(D);

The mortality of DD suffers is concerned, the calculation of actuarial values only requires the use of the total force of mortality $\mu^{id(O)} + \mu^{id(D)}$, since benefits are usually independent on the cause of death. The probabilities of interest for actuarial calculations are:

$$_tp_y^{aa} = exp\left[-\int_0^t (\mu_{y+u}^{ai} + \mu_{y+u}^{ad(O)})du\right] \tag{1}$$

$$_\tau p_{y+u,0}^{ii} = exp\left[-\int_0^t (\mu_{y+u+r,r}^{id(O)} + \mu_{y+u+r,r}^{id(D)})dr\right] \tag{2}$$

The attained age is denoted by y and r denotes the time elapsed since DD inception.

Let us consider a temporary assurance with a DD acceleration benefit. We assume that, for a given sum S, the amount λS (where $0 < \lambda \leq 1$, the acceleration parameter) is payable on DD diagnosis, while the remaining $(1-\lambda)S$ is payable on a specific cause of death if this occurs within the policy term *n*. In our case we let $_hq_{x,c}^{aa}$ denote the probability of death at time h for c cause and for simplicity we restrict our attention to an acceleration benefit with $\lambda = 1$.

If we calculate the single premium meeting the death benefit as well as the DD acceleration benefit, we have:

$$A_{x,n}^{D+DD;1} = \sum_{h=0}^{n-1} {}_hp_x^{aa}(q_{x+h,c}^{aa} + w_{x+h})v^{h+1/2} \tag{3}$$

where:

- w_x is the probability of becoming a DD sufferer within one year
- $_hp_x^{aa}$ is the probability of being healthy at age x+1
- $v^{h+1/2}$ is a discounted factor
-

The relevant annual premium is given by:

$$P_{x:n|} = \frac{A_{x,n}^{D+DD;1}}{\ddot{a}_{x:n|}^{aa}} \tag{4}$$

In particular the value of the reserve is given by:

$$V_{x+t,n-t}^{a} = V_{x+t,n-t}^{(D+DD;1)} - P_{x:n|}\ddot{a}_{x+t:n-t|}^{aa} \tag{5}$$

Let us indicate by k_x the healthy curtate future lifetime of the insured aged x at issue. The cash flow scheme related to the policy at time s is the following, in the case of anticipated annual premiums:

$$X_s = \begin{cases} {}_{/m}P_{x,s+1} & k_x \geq s \quad 0 \leq s \leq m-1 \\ 0 & k_x \geq s \quad s \geq m \\ B_s = F & s-1 \leq k_x \leq s \quad 1 \leq s \leq n \end{cases}$$

in which $s = 1,2,\dots,n$, $P_{n+1} = 0$, $X_0 = P_1$ and where $_{/m}P_{x,s+1}$ is the premium amount payable up to time m and B_s is the critical illness benefit equal to the face value F. In light

ISBN: 978-1-61804-230-9

of a fair valuation of the policy (Coppola et al. 2009), the stochastic flow of the portfolio f_s at times, $s > t$ by a trading strategy can be expressed as follows:

$$f_0 = -c_{/m}P_{x,1} \quad \text{if } s = 0$$

$$f_s - {}_{/m}P_{x,s+1}y_s + B_s y_s^u + B_s(Y_s - Y_{s-1}) \text{ if } s = 1,2,\dots,n$$

where y_s is the number of healthy assured among the survivors (briefly healthy survivors) at time s;
Y_s^u is the number of unhealthy assured among the survivors (briefly unhealthy survivors) at time s, in particular having $Y_s^u = Y_s - Y_{s-1}$, with Y_s the number of the healthy and unhealthy survivors at time s.
We formulate the stochastic provision at time t in its fair value form, replicating the stochastic flow F_s at time s as in the following equations:

$$V_t = E\{L_t/I_t\} = E\left\{\sum_{s=t+2}^{n}\left[-{}_{/m}P_{x,s+1}y_s + B_s y_s^u + B_s(Y_s - Y_{s-1})\right] v(t,s)/ It \right. \tag{6}$$

where L_t is the stochastic loss in t of the portfolio of c contracts in-force and I_t is represented by the filtration $\{I_t\} \subset I$ containing the information flow at time t . On the basis of the conditional expectation calculus, we can write:

$$V_t = \left\{\sum_{s=t+2}^{n}\left[-{}_{/m}P_{x,s+1}c \;\; {}_tp_{xs-t}p_{x+t} + B_s c q_x^H\right] v(t,s)/ ItEv(t,s)/It \right. \tag{7}$$

where ${}_tP_x$is the survival probability of assured aged x up to time t, q_x^h is the rate of the $h-th$ death or diagnosis of a critical illness, referred to an assured aged x, whichever occurs first and v(t,s) the stochastic present value at time t of one monetary unit at time s.

III. Lee-Carter model with cause of death coding adjustments

Assume that the number of deaths are independent Poisson responses $D_{xt} \sim Poisson(e_{xt}\mu_{xt})$. Let $S = \{s_1, s_2 \dots, s_h\}$ be the times at which coding changes occur. In order to account for the coding changes, we assume as in Haberman et al., 2013, that the force of mortality is given by:

$$log_{\mu_{xt}} = \alpha_x + \beta_x k_t + \sum_{i=1}^{h} \delta_x^{(i)} f^{(i)}(t) \tag{8}$$

where:

- μ_{xt} is the age-specific death rate for the *x* interval and the year *t*.
- α_x is the average age-specific mortality.
- k_t is the mortality index in the year t.
- β_x is a deviation in mortality due to changes in the k_t index.
- $f^{(i)}(t) = I_{s_{i-1}} \leq t < s_i$ is an indicator function.
- $\delta_x^{(i)}$ measures the magnitude of coding change at age x.

This model assumes that there are different age-patterns $\alpha_x + \delta_x^{(i)}$for each period $[s_{i-1}, s_i)$ where different causes of death coding system is used. The parameters in (8) can be estimated by maximizing the Poisson log-likelihood of the model. The model is over parameterized since the structure is invariant under either the parameter transformations, that is for any constants $b_1, b_2 \neq 0$:

$$\{\widetilde{\alpha_x}, \widetilde{k_t}\} = \{\alpha_x + b_1\beta_x, k_t - b_1\} \tag{9}$$

$$\{\widetilde{\beta_x}, \widetilde{k_t}\} = \left\{\frac{1}{b_2}\beta_x, b_2 k_t\right\} \tag{10}$$

$$\left\{\widetilde{\delta_x^{(i)}}, k_t\right\} = \left\{\delta_x^{(i)} + a_i\beta_x, k_t - a_i f^{(i)}(t)\right\}, \quad i = 1,\dots,h \tag{11}$$

Transformation (9) and (10) are the original ones from the Lee-Carter model, whilst the family of transformation defined by (11) are induced by the new parameters $\delta_x^{(i)}$ (Haberman et al., 2013).
In order to ensure the complete characterization of the model, the following constraints need to be imposed:

$$k_{t_n} = 0 \tag{12}$$

$$\sum_x \beta_x = 1 \tag{13}$$

In the model the underlying mortality trend is captured only by k_t whilst parameters $\delta_x^{(i)}$ capture the discontinuities in mortality trend induced by the changes in the coding system of the causes of death. In order to accomplish this, we use the family of transformation defined in formula (11).
Inspired by the procedure introduced by Ray et al. (2011), we set the constants $a_i, i = 1,\dots,h$, by fitting the model

$$k_t = g(t) + \sum_{i=1}^{h} \delta_x^{(i)} f^{(i)}(t) + \epsilon_t \tag{14}$$

where $g(t)$ is a continuous function fitted by a thin plate penalized regression spline and ϵ_t is an error term. Model (14) decomposes the time trend k_t into:

- a smooth function $g(t)$ representing the underlying mortality trend;
- the jumps in mortality $\sum_{i=1}^{h} a_i f^{(i)}(t)$ induced by data production changes
- the noise ϵ_t around the underlying trend.

The smoothes parameter of $g(t)$ is derived using generalized cross-validation.
Given constants $a_i, i = 1,\dots,h$ from model (14), respectively for k_t and $\delta_x^{(i)}$, are given by:

$$k_t \longrightarrow k_t - \sum_{i=1}^{h} a_i f^{(i)}(t) \tag{15}$$

$$\delta_x^{(i)} \longrightarrow \delta_x^{(i)} f^{(i)} + a_i\beta_x \qquad i = 1,\dots,h \tag{16}$$

which transfers the jumps in mortality due to data production changes to the $\delta_x^{(i)}$ parameters while k_t represent the underlying mortality trend plus the fluctuations around this trend.

IV. Numerical Application

In order to highlight the profiles of commercial attractiveness of the product, we measure the premium price of the specific accelerated critical illness, which will be illustrated for different ages and compared to a standard accelerated product. First we calculate mortality rates as the number of persons for each age (x), sex (s), and country (c) who die in a particular year (t) of a specified cause (d), divided by the number of persons of that age and sex in the country alive at the beginning of the year.

$$\mu_{xt} = \frac{D_{x,s,c,d,t}}{E_{x,s,c,d,t}} \tag{17}$$

Data were obtained from the Mortality Database administered by the World Health Organization [2009] (WHO) which contains demographic information, including the number of deaths according to the underlying cause of death, for Italy over the last 50 years from 30 to 89 age.

Causes of death are defined by the International Classification of Diseases (ICD), which ensures consistencies between countries. In this study, only the primary causes of death are consider to be used for the construction of the new policy that we intend to propose.The ICD changed three times between 1950 and 2006, from ICD-7 to ICD-10, in order to take into account changes in science and technology and to refine the classification. In Italy there were two reclassifications (ICD7-8 in 1968, ICD8-9 in 1979) The six main causes of death are diseases of the circulatory system, cancer, diseases of the respiratory system, external causes, infectious and parasitic diseases and other.

In Figure 2 it is possible to observe the fitted parameters and in Figure 3 the fitting trend of the mortality index k_t respect to that observed (that has two jumps in the years in which the ICD changed); both the graphs are referred to the cause of death related to the circulatory system.. Figures 4 and 5 are referred to the cause of death referred to the respiratory system.

With these graphs we can see how the new transformation transfers the jumps in mortality due to data production changes to the $\delta_x^{(i)}$ parameters and leaves k_t representing the underlying mortality trend plus the fluctuations around this trend.

In the plots there are represented the evaluation of the parameters of the model (8). The model eliminates the discontinuities in the mortality rates.

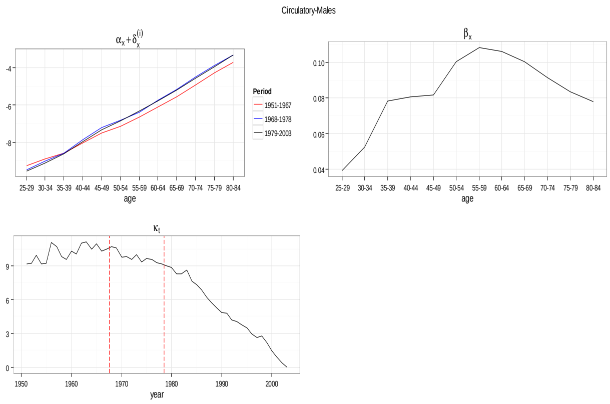


Fig. 2 Multistate schemes for DD insurance.

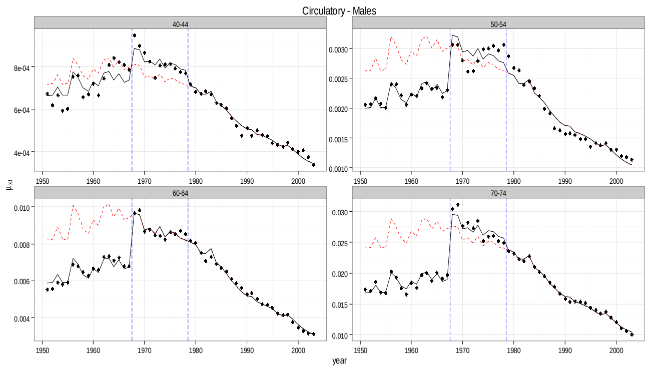


Fig. 3:$k_t$with the coding changes for different group age

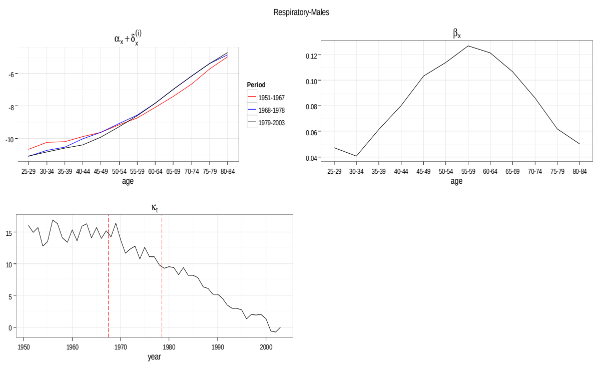


Fig. 4: Fitting parameters Respiratory System

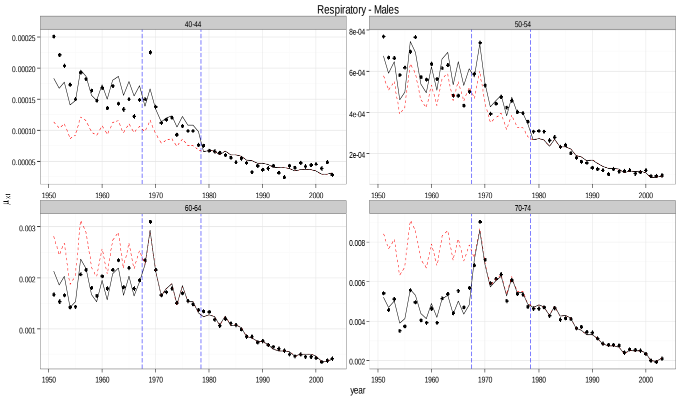


Fig. 5: $k_t$with the coding changes for different group age

In order to project the age-specific death rates, the model assumes the constancy of α_x, β_x and $\delta_x^{(i)}$. The only parameter to be projected through a procedure Box - Jenkis that serves to determine an appropriate ARIMA is k_t.
According to the model, the mortality rate is a linear trend on the basis of an ARIMA (0,1,0), which is well adapted to the representation of the evolution of the index over time. It therefore refers to the following model (Lee and Carter, 1992):

$$k_t = k_{t-1} - c - e_t \quad (18)$$

In the next plots we will see the performance of the insurance premium for different Critical Illness cover, a Stand-Alone, a Stand-alone Accelereted (with an accelerated benefit for death) and a Stand Alone with a benefit for specific cause of death.

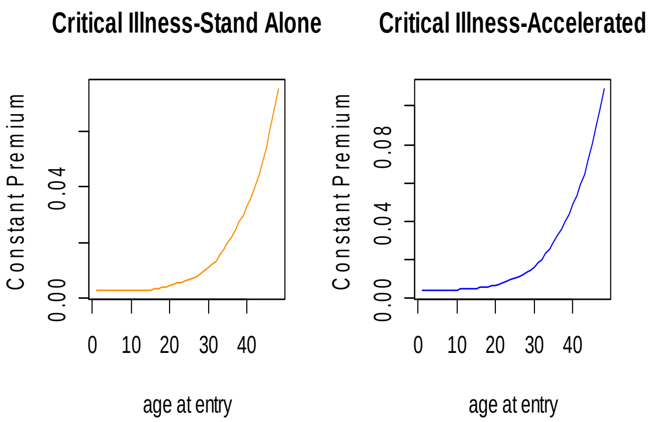


Fig. 6: Performance of the periodic premium for a TCM (Standalone and Accelerated Benefit), term 20 years

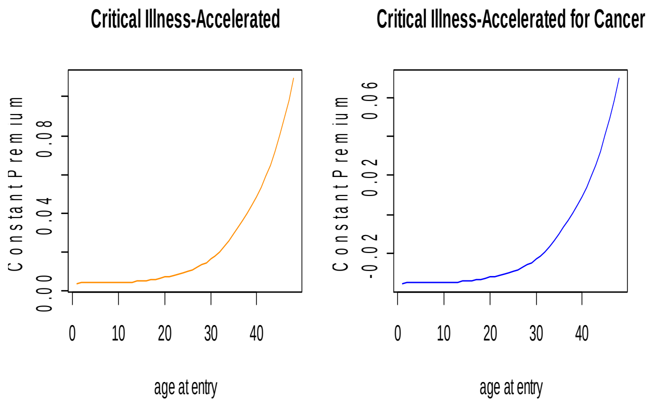


Fig. 7: Performance of the periodic premium for a TCM (Accelerated Benefit and Accelerated Benefit for Cancer), term 20 years

As you can see from the graphs, the premium for the policy that we propose in this work grows slower than the other. This means that it is less expensive than the Standard Stand Alone.
It could be more palatable and accessible to families with middle and low incomes.

V. Concluding Remarks

Actuarial services are offered in several forms, more or less structured and complex, according to models in continuous transformation to the aim of matching as much as possible the need of coverage of each potential insured. The insurance companies are strongly interested in designing new products in which the individual can recognize his own characteristics and as a consequence contracts drafted according to generic profiles reveal to be unfit and poorly attractive. This tendency is going to strengthen the use of personalized models in actuarial valuations and in particular of the death probabilities disaggregated by cause of death. The paper considers this topic in the critical illness framework, matching the advantage of a specialized mortality description with a contract strictly connected with the longevity phenomenon, thanks to which health insurance products are going to be increasingly issued. In particular, in the paper we propose a new contractual model in which an accelerated critical illness is offered with a special accelerated benefit in case of death for a specific cause. Referring to specific cause of death involves smaller expenses and allows for covering the death risk too. The approach combining disaggregated death probabilities and contracts growing in longevity improvement conditions, opens in our opinion new horizons in the actuarial models. Advantages for both the counterparties can be got and marketing competitiveness can be improved.

References

[1] Booth H, Hyndman R, Tickle L., (2006), Lee-Carter Mortality Forecasting: A Multi-Country Comparison of Variants and Extensions, Demographic Research, vol. 15, no. 9, pp. 289-310.

[2] CMI WP 49 (2010) CMI Mortality Projection Model, Institute and Faculty of Actuaries.
[3] CMI WP 50, (2011), Continuos Mortality Investigation Committee Working Paper-CMI Critical illness diagnosis rates for accelerated buisiness, 2003-2006, Institute and Faculty of Actuaries.
[4] Coppola M., D'Amato V., Sibillo M., 2009, Fair value and demographic aspects of the insured loan, Banks and Bank System, vol. 1, 19-29
[5] Haberman S., Villegas. A, (2013), Modelling mortality by cause of death and socio-economic stratification: an analysis of mortality differentials in England, The 17th International Congress on Insurance: Mathematics and Economics
[6] Haberman S., Pitacco E., (1998), Actuarial Models for Disability Insurance, Chapman & Hall
[7] Kochanek K. D., Maurer J. D., Rosenberg H. M., (1994), Why did life expectancy decline from 1984 to 1989 in the United States? American Journal of Public Health, 84, 938-944.
[8] Lee R.D, Carter L. (1992), Modeling and Forecasting the Time Series of U.S. Mortality, Journal of the American Statistical Association 87 (September): 659–671.
[9] Maccheroni C., Barugola T., (2007), Sensitivity Analysis of the Lee-Carter Model Fitting Mortality by Causes of Death. (Italy, 1982-2002), SIS, Rischio e Previsione, Atti della riunione intermedia, pages 481-482, CLEUP, Padova.
[10] Pitacco E., Olivieri A., (1997), Introduzione alla teoria attuariale delle assicurazioni di pensione, Quaderni dell'Unione Matematica Italiana, Pitagora Editrice, Bologna 1997.
[11] Rey G, Aouba A., Pavillon G., Hoffmann R., Plug I., Westerling R., Jougla E., Mackenbach J., (2011), Cause-specific mortality time series analysis: a general method to detect and correct forabrupt data production change, Population Health Metrics 2011, 9:52.
[12] Sherris M., Gaille-Arnold S. (2010), Improving Longevity and Mortality Risk Models with Common Stochastic Long-Run Trends, UNSW Australian School of Business Research Paper No. 2010ACTL13
[13] Tuljapurkar S., Li N., Boe C., (2000). A universal pattern of mortality decline in the G7 countries. Nature 405: 789-792.
[14] Wilmoth J.R., (1995). Are mortality projections always more pessimistic when disaggregated by cause of death? Mathematical Population Studies 5(4): 293-319.
[15] World Health Organization. Who mortality database, January (2009), http://www.who.int/whosis/mort/download/en/index.html

ISBN: 978-1-61804-230-9

Forecasting Trend of Traffic Fatalities in the United Arab Emirates

Ibrahim, M. Abdalla Alfaki

Abstract— This paper utilizes time series methods to fit and forecast rates of traffic fatalities in the United Arab Emirates (UAE). Such data-driven techniques are increasingly used to assess transport safety and the performance of engineering and legislative intervention measures through the use of traffic safety targets. Despite the continuous rise of traffic accidents in the UAE, the country, however, hasn't yet managed to create a mechanism to setting road safety targets in order to help monitoring the effect of potential safety countermeasures and possible reduction in traffic fatalities. Several competing models are fitted and compared in terms of goodness of fit and forecasting power. The main implications of results obtained from this work are intended to aid the UAE to explore different mechanisms to setting safety targets with the aim of reducing traffic fatality numbers and rates and the creation of sustainable prevention and safety management strategy.

Keywords— Traffic fatality, Safety target, Intervention measures, Fatality trend.

I. INTRODUCTION

ECONOMIC growth, population increase and demographic and cultural changes in the United Arab Emirates (UAE) have contributed in a dramatic increase in the number of vehicles in the last few years. The increase in the number of vehicles on the road, along with emergence of high risk road users have converged in not only a dramatic number of road accidents, but a dramatic number of fatal accidents and fatal injuries, Abdalla [1]-[2]. Compared to western countries and some other countries in the gulf region, the number of road accidents in the UAE is high, in particular, accidents that are caused by wrong driver's behavior. According to the UAE National Bureau of Statistics report of 2008, 18 percent of the accidents were caused by careless driving, and 43 percent were due to other human mistakes.

Monitoring progress in transport and road safety is essential for effective use of resources and reduction of direct and indirect costs of traffic accidents. Among the many measures that are employed by policy makers and practitioners to measure safety of transport systems is the number of people killed in traffic accidents.

The continuous rise of traffic accidents and death toll in the United Arab Emirates constitute major concerns for public health and economic and social planning (Bener and Crundall [5]; El-Sadig et al. [6]). The country, however, hasn't yet managed to create a mechanism to setting road safety targets in order to help monitoring the effect of potential safety countermeasures and possible reduction in traffic fatalities.

Three approaches to setting safety targets are discussed in the literature; namely, aspirational targets, model-based targets and extrapolation and evidence-led judgment, Marsden and Bonsall [11]. The aspiration target setting approach is based on setting an arbitrary number that represents the intended achievement in specified period of time, e.g. reaching zero fatalities in specified number of years starting from a specified base year, often without giving any justification on how the number is derived. The model-based approach, on the other hand, calls for setting reduction targets based on available data and a clear understanding of the relationship between traffic safety and other influencing factors. The success of the approach depends on the ability of the model to reflect reality, accuracy of assumptions and the extent of real implementation of assumed policy interventions. In reality it is difficult to model all influencing indicators with sufficient accuracy. Extrapolation and professional judgment can therefore be utilized to resolve this issue, using an evidence-led judgment approach. The success of the evidence-led judgment approach in setting the targets depends on the availability of sufficiently long time series data on accidents and traffic fatalities, the ability to identify the different components of the time series and on availability of substantial evidence on the impact of different policy intervention measures on accident and fatality reductions, Kweon [10].

The main objective of this study is to describe trend and forecast developments in the number and rates of traffic fatalities in the UAE. The main implications of results obtained from this work are intended to aid the UAE as a country to explore different mechanisms to setting safety targets aiming at the reduction of traffic fatality numbers and rates and the creation of sustainable safety management strategy**.**

II. AN OVERVIEW OF FATALITY PREDICTION MODELS

Statistical models using time series approaches to fitting and forecasting number and rates of traffic casualties are widely discussed in the literature at both aggregate and disaggregate prediction levels (see for example Broughton [16]; Yannis and Antoniou [15]). In several studies allowances

Ibrahim, M. Abdalla Alfaki is an Associate Professor with the College of Business and Economics, United Arab Emirates University, Al-Ain, UAE, P. O. Box 15551 (Tel. +9713-7136253; e-mail: i.abdalla@uaeu.ac.ae).

ISBN: 978-1-61804-230-9

are made to cater for intervention variables and special events, however, the majority are univariate in nature, incorporating limited number of explanatory variables (Raeside and White [17]-[18]).

Earlier in 1949, R. J. Smeed [13] published his famous law, $F = 0.0003\,(V.P)^{(1/3)}$, for predicting traffic fatalities based on the number of registered vehicles, V, and population size, P. The formula postulates a decrease in traffic fatalities per vehicles as a result of the increase in vehicle ownership, $F/V = 0.0003(V/P)^{(-2/3)}$, it also, however, suggests an increase in the number of fatalities and fatalities per population due to the increase in vehicle ownership, $F/P = 0.0003(V/P)^{(1/3)}$. Several attempts are made to validate Smeed's law, occasionally with different estimates of the parameters, using data coming from both developed and developing countries (see for example Gharaybeh [7]; Koren and Borsos [9]. The law successfully predicts the increase in traffic fatalities up to certain years, but then the validity of the formula is disputed in several occasions as it failed to match the noticeable decrease in road deaths in some of the developed countries, producing considerable deviations between the expected and the actual number of traffic fatalities (Adams [3]; Koren and Borsos [9]; Bener et al. [4]; Safe Speed [12]).

It is worthy to note that the original relationship used to derive Smeed's law in the late 60's (Smeed [14]) is formulated as

$$F\ /\ V = \alpha(V\ /\ P)^{\beta} + \varepsilon, \qquad (1)$$

where α and β are the model parameters and ε are the disturbances which are assumed to be independent and identically normally distributed. Given the nature of traffic fatality data as time series realizations, utilization of Smeed's model (1) in its current form to fit such data might not account for possible serial correlation of the model's disturbances, ε. Attempting to improve model fit and to allow for possible dependence or autocorrelation of disturbances Yannis and Antoniou [15] utilized Smeed's model (1) as a benchmark to further develop a log-transformed version of model (1) and two further autoregressive non-linear models, defined respectively in (2), (3) and (4) below, based on using vehicle ownership, $V\ /\ P$, as a macroscopic predictor of traffic fatalities. Models (3) and (4) are the autoregressive versions of models (1) and (2) respectively.

$$log\left({}^{F_t}/_{V_t}\right) = \alpha + \beta.\,log\left({}^{V_t}/_{P_t}\right) + \varepsilon_t, \qquad (2)$$

$$(F\ /\ V)_t = \varphi.\,(F\ /\ V)_{t-1} + \alpha(V\ /\ P)_t^{\beta} - \varphi.\alpha(V\ /\ P)_{t-1}^{\beta} + \varepsilon_t, \qquad (3)$$

$$log(F\ /\ V)_t = \varphi.log(F\ /\ V)_{t-1} + (1-\varphi).\,\alpha + \beta.\,log(V\ /\ P)_t - \varphi.\,\beta.\,log(V\ /\ P)_{t-1} + \varepsilon_t \qquad (4)$$

The subscript t denotes the discrete time $(t = 1, \ldots, T)$ at which the dependent and the predictor variables are measured. Yannis and Antoniou [15] fitted the four models (1 to 4) using European traffic data spanning the period from 1970 to 2002 using non-linear regression technique. They report an overall superior performance, in terms of goodness of fit and forecasting power, of the autoregressive models compared to the non-linear base model (1) and its similar log-transformed version, respectively. They further indicate that the log-transformed model gives better prediction compared to the non-linear base model (1), but suffered from correlated residuals.

Raeside and White [17]-[18] employed a variation of Broughton [16] model to fit a negative exponential model of the natural logarithm of the number of traffic fatalities, serious injuries and slight casualties in Great Britain covering the period from 1970 to 2010. Using fatal casualties, they present a time series model with an autoregressive and linear trend terms as depicted in equation (5). This model is a deviation from Broughton's [16] model who used fatality rates (traffic fatalities per vehicle-kilometers) as a dependent variable and an intervention term, seat-belts use as a predictor.

$$log(F_t) = \alpha + \beta.\,t + \varphi.\,log(F_{t-1}) \qquad (5)$$

Several authors adopted the state space methodology or structural time series methods to fit and forecast traffic fatalities. Commandeur and Koopman [19] present a broad overview of the application of this methodology on traffic fatality data. They discuss several scenarios where produced models can adequately describe the data to situations where model explanatory power is increased by incorporating explanatory or intervention terms. They used the technique to adequately fit the annual numbers of traffic fatalities in Finland, 1970 through 2003, using a deterministic level and a stochastic slope model, and to forecast fatalities for the years 2004-2008 with 90% confidence limits.

Hermans et. al. [8] use the unobserved components or structural time series methodology employing the state-space form to discover the long-term time series trend of killed and seriously injured traffic casualties in Belgium 1974 to 1999. The models are designed to allow for the impact of weather conditions, economic factors and several laws and intervention variables. Developed models are used to predict the dependent variable for 2000 and 2001 and to compare results with the actual observations. They later compared the performance of the state-space models with a regression model with ARMA errors, concluding a lot of similarity and the same direction and a comparable magnitude. However, their analysis did not include a measure of exposure to risk of accidents.

Bener et al. [4] employed regression techniques to estimate traffic fatalities using time series data on vehicles and populations of three Mediterranean countries, including the United Arab Emirates (UAE). They reported lower average absolute errors for estimation using regression analysis compared to estimation based on Smeed's equation. However, their analysis failed to investigate possible serial correlations in the data and to ascertain whether or not regression residuals are auto-correlated. Presence of positive or negative first order residual autocorrelation, for instance, is a violation of model assumptions and will lead, respectively, to serious under or over-estimation of error variance, resulting in optimistic or pessimistic conclusions and over or under-estimation of the significance of predictors employed in the model.

III. Developing Traffic Fatality Model using UAE Data

Annual traffic fatality data together with population size and number of registered vehicles estimates spanning the period from 1977 to 2008 are considered to describe trend and to predict future numbers and rates of traffic fatalities in the UAE. The data are obtained from official sources in the country; namely, the UAE Bureau of Statistics and the Ministry of Interior Annual Statistical Abstracts together with other relevant sources. Prediction models utilized in this development are macroscopic in nature, depending mainly on annually aggregated data. This kind of models are useful when the overriding objective is the provision of appropriate models that describe trend of traffic fatalities and forecast future values. Indeed, road safety trends and future forecasts can be investigated by studying the effects of various macroscopic variables including motorization level, road expenditure and intervention measures such as implementation of seat-belt use, on traffic fatalities. However, the use of several explanatory variables is usually constrained by the quality and availability of data, particularly in developing countries.

It is worthy to note that accident prediction models are usually constructed with an exposure variable that controls for the total road traffic movements within the road network. In many studies, vehicles-Kilometers travelled is used as an exposure variable. This is particularly true in studies covering mostly developed countries where good traffic and transportation reporting systems exist. In most developing countries the situation might be different and reliance is mainly on other proxy measures of exposure such as the number of registered vehicles and population size. The UAE is no exception, the country lacks a reliable reporting system that accounts for traffic movements in the road network. Therefore, the number of registered vehicles and population size are used to control for exposure in this study.

The time series data employed in this study are partitioned into two datasets. A training dataset, 1977 to 2003, which is used to describe trend and to develop traffic fatality prediction models, and a validation dataset, 2004 to 2008, which is used to validate the developed models. Goodness of fit of developed models and deviation of fitted values from actually observed values (for both within sample: 1977-2003 and out of sample 2004-2008) are measured by estimating the root mean square error ($RMSE$) as given in equation (6). The lower the value of the $RMSE$, the better the performance and forecast of the model compared to its counterparts.

$$RMSE = \sqrt{1/T \sum (F_t - \hat{F}_t)^2}, \qquad (6)$$

where F_t is the observed traffic fatality at time t, and $\hat{F}_t$ is the fitted annual traffic fatalities obtained from the constructed model.

IV. Exploring the UAE Traffic Fatality Situation

Fig. 1 presents annual UAE road traffic fatalities, 1977 to 2008. It is interesting to note the steady increase in traffic fatalities in the UAE since 1985 with seemingly linear trend. However, the overall pattern between 1977 and 2008 is far from linear. The pre-1985 period witnessed considerable fluctuations in the number of traffic fatalities. The drop in the number, starting 1984 up to 1987, is arguably linked with the collapse in oil prices and the tankers war in the Arabian Gulf erupted between Iran and Iraq, 1986-1987, which had an adverse effect on trade transportation in the whole gulf region.

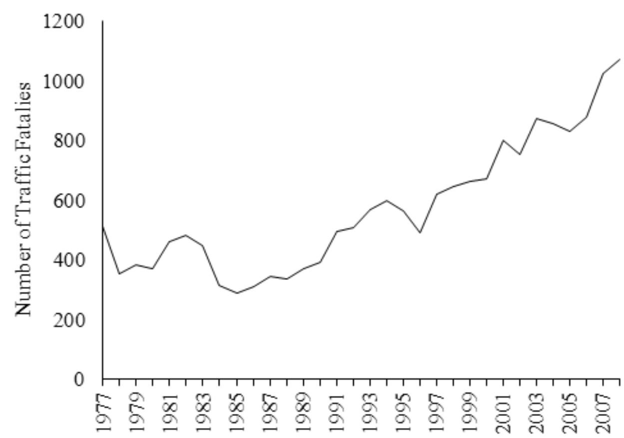


Fig. 1: UAE Traffic Fatalities, 1977 – 2008

Evidence shown in Fig. 2 which presents the sample Autocorrelation Function (ACF) of the annual traffic fatalities in the UAE, confirms the presence of serial correlation in the data. Significant autocorrelation coefficients (falling outside the 95% confidence limits) are reported at different lags, $p - value = 0.0$, based on 16 lags Box-Ljung test with $\chi^2 = 97.5$. As a result any modeling efforts that does not account for the dependence in the data might render final conclusions useless.

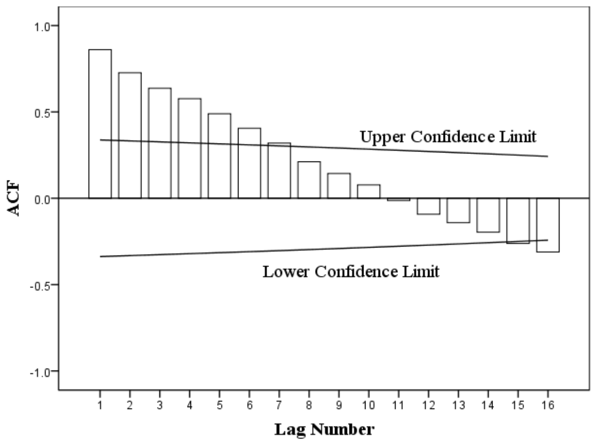


Fig. 2: Sample ACF, UAE Traffic Fatalities, 1977 – 2008

The actual variables that are used in developing the UAE traffic fatality prediction models in this study included fatalities/vehicle, F / V, as a dependent variable and car ownership or vehicles/population, V / P, as a predictor variable. While the number of traffic fatalities, F, in the UAE are showing an increasing trend as displayed in Figure 1, fatality rates per registered vehicles, F / V, have sharply decreased in the period from 1977 to 1980, Fig. 3, possibly as a result of several intervention and road safety measures including improvements in infrastructure, enforcement of speed limits and penalties and enactment of a secondary seat-belt law. After 1980 a consistently slight decreasing trend is observed, Fig. 3. Vehicles ownership (V / P, the dotted line in Fig. 3), on the other hand, is depicting an opposite increasing trend.

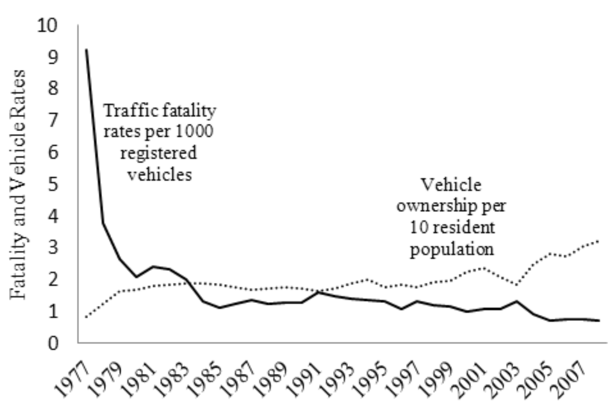


Fig. 3: UAE Traffic Fatality Rates per 1000 Registered Vehicles and Vehicles Ownership per 10 Resident Population

Given the time series nature of the data and the decreasing trend of the dependent variable, several models discussed earlier in Section II are thought to provide adequate fit to the UAE traffic fatality data. In what follows parameters estimation and model fit as well as comparisons of models' performances are discussed.

V. Fitting a Negative Exponential Model to the UAE Traffic Fatality Data

The analysis in this section is based on Brought [16] model derived according to Raeside [18] as depicted in equation (5). This is an ARMA like model with an autoregressive component (AR) of order 1, and a trend term, t, as a predictor variable. The model introduced in this study deviates from Raeside's model in (5) by imposing the number of registered vehicles, V, as an exposure variable. The dependent variable, F / V, was log-transformed to eliminate non-stationarity of the data. Therefore, the final model developed using the training UAE traffic fatality data, 1977-2003, is defined as

$$log((F / V)_t) = \alpha + \beta.t + \varphi.log((F / V)_{t-1}) \qquad (7)$$

The same training dataset is used to fit models (1), (2), (3) and (4). Table 1 depicts parameter estimates for all developed models. It is clear from Table 2 that the log-transformed AR model and the negative exponential (AR with linear trend) model achieved parity in terms of goodness of fit and prediction accuracy for both within sample and out of sample datasets. However, the log-transformed AR model, provided a slightly better performance, lower within sample $RMSE$, compared to the negative exponential model. In fact it outperformed all other fitted models. However, it is important to note that the negative exponential model, in addition to the autoregressive process, includes only a linear trend term. In this sense it provides a simple more parsimonious option. Another point in favor of the negative exponential model is that forecasting future traffic fatality rates using the competing log-transformed AR model requires future forecasts of vehicle ownership, V / P, the predictor variable.

Table 1: Fatality Prediction Models, UAE Data, 1977-2003

Model	Coefficient	α	φ	β
Non-linear (Smeed Model equation(1))	Estimate	5.8662	-	-2.2914
	Standard Error	0.1968	-	0.0867
	p-value	0.0000		0.0000
Log-transformed (equation (2))	Estimate	1.6833	-	-2.1678
	Standard Error	0.1437	-	0.2404
	p-value	0.0000	-	0.0000
Non-linear AR (equation (3))	Estimate	1.8853	0.3398	-0.4818
	Standard Error	0.7486	0.0396	0.6366
	p-value	0.0192	0.0000	0.4569
Log-transformed AR (equation (4))	Estimate	0.7425	0.6664	-0.7675
	Standard Error	0.2706	0.0787	0.4147
	p-value	0.0155	0.0000	0.0771
Negative Exponential (equation (7))	Estimate	1.6900	0.8920	-0.0660
	Standard Error	0.5200	0.1030	0.0320
	p-value	0.0030	0.0000	0.0460

The most important observation is that the non-linear base model and the log-transformed model did not pass the

ISBN: 978-1-61804-230-9

residuals serial correlation test, $p - value < 0.01$, based on Box-Ljung criteria as displayed in Table 2, consequently violating the assumption of independent residuals. It remains to acknowledge that the log-transformed model is an improvement over the non-linear base model, providing lower within sample and out of sample $RMSE$, Table 2. Inclusion of an autoregressive component in both models resulted in eliminating serial correlation and in improving model prediction, Table 2. Results discussed here are similar to conclusions reported by Yannis and Antoniou [15] using European data.

Table 2: Test Model Residuals Autocorrelation (Box-Ljung test,16 lags) and Model Goodness of fit and Accuracy ($RMSE$), UAE Fatality Data

	Box-Ljung		RMSE	
Model	χ^2	$p - value$	Within sample 1977-**2003**	Out of sample 2004-2008
Non-linear (Smeed Model, equation (1))	43.4	0.0002	0.38	0.23
Log-transformed (equation (2))	42.7	0.0003	0.23	0.21
Non-linear AR (equation (3))	16.5	0.4159	0.27	0.26
Log-transformed AR (equation (4))	17.3	0.3652	0.12	0.18
Negative Exponential (equation (7))	4.7	0.9998	0.18	0.18

VI. FORECASTING RATES OF UAE TRAFFIC FATALITIES

It was evident from the analysis of the UAE fatality data that all proposed models in this study showed a downward trend in the UAE traffic fatality rates. Forecasts of future fatality rates, 5 years ahead, with lower and upper 95% confidence limits, are generated using the negative exponential model, model (7) above. Forecasts of traffic fatalities for 2013 is 0.4847 traffic fatalities per 1000 registered vehicles, ranging over a 95% prediction interval from 0.2043 to 0.9909 fatalities per 1000 registered vehicles. This is about 30% reduction of the 2008 rate. To convert forecasted fatality rates into fatality numbers, the total number of registered vehicles in that particular year is needed. Several scenarios to forecasting the number of registered vehicles are discussed including the development of a forecasting model for the number of vehicles or alternatively use official vehicles forecasts produced by official agents (Broughton [16]).

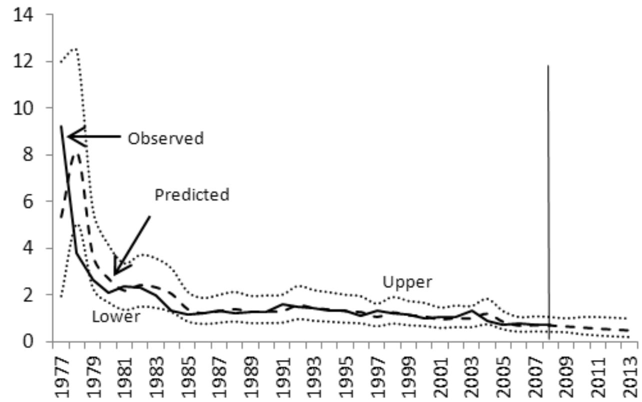


Fig. 4: Observed and Predicted Traffic Fatalities with 95% Lower and Upper Confidence Limits

VII. CONCLUDING REMARKS

Many countries have established safety plans and traffic safety targets. Targets provide a clear focus for the work of those involved in transport planning and traffic safety management. There is an overriding consensus that traffic safety targets should be specific, measurable, relevant and time-bound. The UAE is still lagging behind when it comes to the creation of safety plans based on appropriate targets for the reduction of traffic fatalities and injuries. The country is suffering from increased levels of traffic fatalities, however, accounting for the notable increase in the number of registered vehicles in the country together with the positive growth in population size, traffic fatality rates are showing a slightly declining trend. This decline might be attributable to changes in driving behaviors in addition to the introduction of several legislative and engineering intervention measures.

This paper intended to explore and describe the traffic fatality trend in the UAE, and further produce appropriate predictive models capable of forecasting future fatality trend. This is an important step needed for safety target setting. Results presented in the paper identified appropriate models that provided good fit and forecast of the UAE fatality data.

REFERENCES

[1] I. M. Abdalla Alfaki, "Fatality risk assessment and modeling of divers responsibility for causing traffic accidents in Dubai," *Journal of Safety Research*, vol. 33, 2002, pp. 483-496.

[2] [1] I. M. Abdalla Alfaki, "Effectiveness of safety belts and Hierarchical Bayesian analysis of their relative use," *Safety Science*, vol. 43, pp. 91-103.

[3] J.G.U. Adams. (1987). Smeed's law: Some further thoughts, *Traffic Engineering and Control*, pp. 70-73 [online, accessed December 2013]. Available: http://john-adams.co.uk/wp-content/uploads/2006/smeed's%20law.pdf

[4] A. Bener, D. J. Hussain, M. A. Al-Maliki, M. M. Shotar, M. F. Al-Said, and K. S. Jadaan, " Road traffic fatalities in Qatar, Jordan and the UAE: Estimates using regression analysis and the relationship with economic growth, " *Eastern Mediterranean Health Journal*, vol. 16, no. 3, 2010, pp. 318-323.

[5] A. Bener and D. Crundall, "Road Traffic Accidents in the United Arab Emirates Compared to Western Countries," *Advances in Transportation Studies an International Journal, Section A6*, 2005, pp. 5-11.

[6] M. El-Sadig, N. M. Norman, O. Lloyd, P. Romilly and A. Bener, "Road Traffic Accidents in the United Arab Emirates: Trends of Morbidity and Mortality during 1977-1998," *Accidents Analysis and Prevention*, vol. 34, 2002, pp. 465-476.

[7] F. A. Gharaybeh, (1994) "Application of Smeed's formula to assess development of traffic safety in Jordan," *Accident Analysis and Prevention*, vol. 26, no. 1, 1994, pp.113–120.

[8] E. Hermans, G. Wets and F. Van den Bossche. (2005). Frequency and Severity of Belgian Road Traffic Accidents Studied by State-Space Methods, Association for European Transport and contributors, [online] http://abstracts.aetransport.org/paper/index/id/2260/confid/11.

[9] C. Koren and A. Borsos, "Is Smeed's Law Still Valid? A World-Wide Analysis of the Trends in Fatality Rates," *Journal of Society for Transportation and Traffic Studies*, vol. , no. 11, 2010, pp. 64:74.

[10] Y. Kweon, "Data-driven reduction targets for a highway safety plan," *Transport Policy*," vol. 17, 2010, pp. 230-239.

[11] G. Marsden and P. Bonsall, "Performance Targets in Transportation Policy," *Transport Policy*, vol. 13, no. 3, 2006, pp.191-203.

[12] Safe Speed (SS) .(2013). Smeed and beyond: predicting road deaths, updated: http://www.safespeed.org.uk/smeed.html, [online, accessed December. 4, 2013].

[13] J. R. Smeed, "Some Statistical Aspects of Road Safety Research," *Journal of the Royal Statistical Society, Series A*, vol. 1, 1949, pp. 1-34.

[14] J. R. Smeed, "Variation in the Pattern of Accident Rates in Different Countries and their Causes," *Traffic Engineering and Control*, 1968, pp. 364-371.

[15] G. Yannis and C. Antoniou, "Autoregressive nonlinear time-series modeling of traffic fatalities in Europe," *European Transport Research Review*, vol. 3, 2011, pp. 113-127.

[16] J. Broughton, "Forecasting Road Accidents Casualties in Great Britain," *Accident Analysis and Prevention*, vol. 23, no. 5, 1991, pp. 353-362.

[17] R. Raeside and D. White. 2004. "Predicting Casualty Numbers in Great Britain," Paper No. 04-3684. *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, January 11–15, 2004, Washington, DC.

[18] R. Raeside, "Predicting and Monitoring Casualty Numbers in Great Britain," *Journal of Transportation and Statistics*, vol. 7, no.1, 2004, [online] http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/journal_of_transportation_and_statistics/volume_07_number_01/html/paper_04/index.html

[19] J. J. F. Commandeur and S. J. Koopman, An Introduction to State Space Time Series Analysis, 2007, Oxford University Press, Oxford, UK.

ISBN: 978-1-61804-230-9

Solving third order boundary value problem with fifth order block method

A. S. Abdullah, Z. A. Majid, and N. Senu

***Abstract*—** We develop a fifth order two point block method for the numerical solution of nonlinear boundary value problems (BVPs) directly. Most of the existence research involving BVPs will reduce the problem to a system of first order Ordinary Differential Equations (ODEs). However, the proposed method will solve the third order BVPs directly without reducing to first order ODEs. These methods will solve the nonlinear third order BVPs by shooting technique using constant step size. Numerical example is presented to illustrate the applicability of the propose method. The results clearly show that the propose method is able to solve boundary value problems (BVPs).

***Keywords*—** Boundary value problem, shooting technique, two point block method.

I. Introduction

BOUNDARY value problems (BVPs) are used in many branches of science. Some of them are in the field of optimization theory, engineering and technology. Since the boundary value problem has wide application in science research, therefore faster and accurate numerical solution of boundary value problem are very importance. Therefore, it has many boundary value problems (BVPs) solution technique is proposed. In literature contains several methods has been proposed to solve BVPs.Loghmani and Ahmadinia (2006) use a third degree B-spline function to construct an approximate solution for third order linear and nonlinear boundary value problems coupled with the least square method. Quartic nonpolynomial spline method was proposed by El-Danaf(2008) for the numerical solution of third order two point boundary value problems. El-Salam *et al*.(2010) are presented second and fourth order convergent methods based on Quartic nonpolynomial spline function for the numerical solution of a third order two-point boundary value problem. While Phang *et al*. (2011) had solved second order boundary value problem using two step direct method by shooting technique. The fifth order two point block method also use shooting technique to solve the boundary value problem directly.

The author gratefully acknowledged the financial support of Graduate Research Fund (GRF) from Universiti Putra Malaysia and MyMaster from the Ministry of Higher Education.

A. S. Abdullah is with the Institute for Mathematical Research, University Putra Malaysia, 43400 Serdang, Malaysia. (e-mail: a_shah@lyahoo. com).

Z. A. Majid is with the Institute for Mathematical Research, University Putra Malaysia, 43400 Serdang, Malaysia. (phone: 603-8946-6874; fax: 603-8936- 7958; am_zana@upm.edu.my).

N. Senu is with the Department of Mathematics, Faculty of Science, University Putra Malaysia, 43400 Serdang, Malaysia. (e-mail: norazak@upm.edu.my).

In this paper, we propose a fifth order block method for solving boundary value problems of the form as follows

$$y''' = f(x, y, y', y''),\ a \le x \le b \tag{1}$$

with boundary conditions

$$y(a) = \gamma,\ y'(a) = \alpha,\ y'(b) = \beta \tag{2}$$

where a,b,α,β,γ are the given constant. The guessing values estimated by implement the Newton method. The advantage of these methods is to solve BVPs without reduce it to the system of first order ordinary differential equations (ODEs).

II. Formulation of the Method

In this research, the direct method of multistep method is developed for the numerical solution of nonlinear boundary value problems (BVPs) directly.

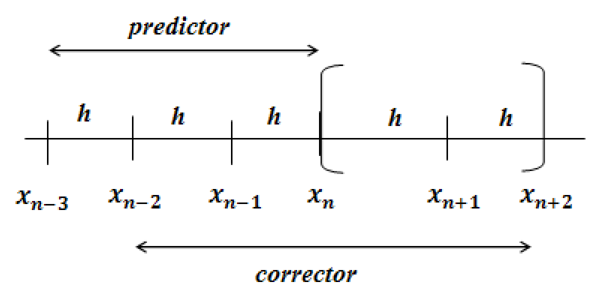


Fig 1: Two Point Direct Block Method

The interval $[a,b]$ is divided into a series of blocks with each block containing two points as shown in Fig. 1. Two value points will be found simultaneously using the same back value i.e. y_{n+1} and y_{n+2} .The point y_{n+1} at x_{n+1} can be obtained by integrating Eq. (1) over the interval $[x_n, x_{n+1}]$ once, twice and thrice that shown in Eq. 3-5:

Integrate once:

$$\int_{x_n}^{x_{n+1}} y'''(x)dx = \int_{x_n}^{x_{n+1}} f(x, y, y', y'')dx \tag{3}$$

Integrate twice:

$$\int_{x_n}^{x_{n+1}}\int_{x_n}^{x} y'''(x)dxdx = \int_{x_n}^{x_{n+1}}\int_{x_n}^{x} f(x, y, y', y'')dxdx \quad (4)$$

Integrate thrice:

$$\int_{x_n}^{x_{n+1}}\int_{x_n}^{x}\int_{x_n}^{x} y'''(x)dxdxdx = \int_{x_n}^{x_{n+1}}\int_{x_n}^{x}\int_{x_n}^{x} f(x, y, y', y'')dxdxdx. \quad (5)$$

The same process will be applied to find the second point y_{n+2}. Eq. (1) will be integrated over the interval $[x_n, x_{n+2}]$ once, twice and thrice gives,

Integrate once:

$$\int_{x_n}^{x_{n+2}} y'''(x)dx = \int_{x_n}^{x_{n+2}} f(x, y, y', y'')dx \quad (6)$$

Integrate twice:

$$\int_{x_n}^{x_{n+2}}\int_{x_n}^{x} y'''(x)dxdx = \int_{x_n}^{x_{n+2}}\int_{x_n}^{x} f(x, y, y', y'')dxdx \quad (7)$$

Integrate thrice:

$$\int_{x_n}^{x_{n+2}}\int_{x_n}^{x}\int_{x_n}^{x} y'''(x)dxdxdx = \int_{x_n}^{x_{n+2}}\int_{x_n}^{x}\int_{x_n}^{x} f(x, y, y', y'')dxdxdx \quad (8)$$

The function $f(x, y, y', y'')$ in Eq. 3-8 will be approximated using Lagrange interpolating polynomial, P. The interpolation points involved are (x_{n-3}, f_{n-3}), (x_{n-2}, f_{n-2}), (x_{n-1}, f_{n-1}), (x_n, f_n) and (x_{n+1}, f_{n+1}) we will obtain the Lagrange interpolating polynomial:

$$\begin{aligned}
P_4 = &\frac{(x-x_{n-2})(x-x_{n-1})(x-x_n)(x-x_{n+1})}{(x_{n+2}-x_{n-2})(x_{n+2}-x_{n-1})(x_{n+2}-x_n)(x_{n+2}-x_{n+1})} f_{n+2} \\
&+\frac{(x-x_{n-2})(x-x_{n-1})(x-x_n)(x-x_{n+2})}{(x_{n+1}-x_{n-2})(x_{n+1}-x_{n-1})(x_{n-1}-x_n)(x_{n+1}-x_{n+2})} f_{n+1} \\
&+\frac{(x-x_{n-2})(x-x_{n-1})(x-x_{n+1})(x-x_{n+2})}{(x_n-x_{n-2})(x_n-x_{n-1})(x_n-x_{n+1})(x_n-x_{n+2})} f_n \\
&+\frac{(x-x_{n-2})(x-x_n)(x-x_{n+1})(x-x_{n+2})}{(x_{n-1}-x_{n-2})(x_{n-1}-x_n)(x_{n-1}-x_{n+1})(x_{n-1}-x_{n+2})} f_{n-1} \\
&+\frac{(x-x_{n-1})(x-x_n)(x-x_{n+1})(x-x_{n+2})}{(x_{n-2}-x_{n-1})(x_{n-2}-x_n)(x_{n-2}-x_{n+1})(x_{n-2}-x_{n+2})} f_{n-2}
\end{aligned} \quad (9)$$

Taking $s = \dfrac{x - x_{n+2}}{h}$ and replacing $dx = hds$, changing the limit of integration from -2 to -1 for Eq. 3-5 can be written as:

$$y''_{n+1} - y''_n = \int_{-2}^{-1} P_4 hds \quad (10)$$

$$y'_{n+1} - y'_n - hy''_n = \int_{-2}^{-1} -(s+1)P_4 h^2 ds \quad (11)$$

$$y_{n+1} - y_n - hy'_n - \frac{h^2}{2!} y'' = \int_{-2}^{-1} (-s-1)^2 P_4 h^3 ds \quad (12)$$

and from -2 to 0 for Eq. 6-8 can be written as:

$$y''_{n+2} - y''_n = \int_{-2}^{0} P_4 hds \quad (13)$$

$$y'_{n+1} - y'_n - 2hy''_n = \int_{-2}^{0} -sP_4 h^2 ds \quad (14)$$

$$y_{n+2} - y_n - 2hy'_n - \frac{h^2}{2!} y''_n = \int_{-2}^{0} (-s)^2 P_4 h^3 ds \quad (15)$$

Evaluate these integral using MAPLE and the corrector formulae can be obtained. The method is the combination of predictor one order less than the corrector. The same process is applied to find the predictor formulae.

Fifth Order Block Method:

Predictor:

$$\begin{aligned}
y''_{n+1} - y''_n &= \frac{h}{12}(-9 f_{n-3} + 37 f_{n-2} - 59 f_{n-1} + 55 f_n) \\
y'_{n+1} - y'_n - hy''_n &= \frac{h^2}{360}(-38 f_{n-3} + 159 f_{n-2} \\
&\qquad - 264 f_{n-1} + 323 f_n) \\
y_{n+1} - y_n - hy'_n - \frac{h^2}{2} y''_n &= \frac{h^3}{720}(-17 f_{n-3} + 72 f_{n-2} \\
&\qquad - 123 f_{n-1} + 88 f_n)
\end{aligned} \quad (16)$$

$$\begin{aligned}
y''_{n+2} - y''_n &= \frac{h}{3}(-8 f_{n-3} + 31 f_{n-2} - 44 f_{n-1} + 27 f_n) \\
y'_{n+2} - y'_n - 2hy''_n &= \frac{h^2}{90}(-62 f_{n-3} + 246 f_{n-2} \\
&\qquad - 366 f_{n-1} + 272 f_n) \\
y_{n+2} - y_n - hy'_n - \frac{(2h)^2}{2} y''_n &= \frac{h^3}{630}(-26 f_{n-3} \\
&\qquad + 105 f_{n-2} - 162 f_{n-1} + 143 f_n)
\end{aligned} \quad (17)$$

ISBN: 978-1-61804-230-9

Corrector:

$$
\begin{aligned}
y''_{n+1} - y''_n &= \frac{h}{720}(11 f_{n-2} - 74 f_{n-1} \\
&\quad + 456 f_n + 346 f_{n+1} - 19 f_{n+2}) \\
y'_{n+1} - y'_n - hy''_n &= \frac{h^2}{1440}(11 f_{n-2} - 76 f_{n-1} \\
&\quad + 582 f_n + 220 f_{n+1} - 17 f_{n+2}) \\
y_{n+1} - y_n - hy'_n - \frac{h^2}{2} y''_n &= \frac{h^3}{10080}(23 f_{n-2} \\
&\quad - 162 f_{n-1} + 1482 f_n + 370 f_{n+1} - 33 f_{n+2})
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
y''_{n+2} - y''_n &= \frac{h}{90}(- f_{n-2} - 4 f_{n-1} \\
&\quad + 24 f_n + 124 f_{n+1} + 29 f_{n+2}) \\
y'_{n+2} - y'_n - 2hy''_n &= \frac{h^2}{90}(f_{n-2} \\
&\quad - 8 f_{n-1} + 78 f_n + 104 f_{n+1} + 5 f_{n+2}) \\
y_{n+2} - y_n - hy'_n - \frac{(2h)^2}{2} y''_n &= \frac{h^3}{630}(9 f_{n-2} \\
&\quad - 64 f_{n-1} + 516 f_n + 384 f_{n+1} - 5 f_{n+2})
\end{aligned}
\tag{19}
$$

For the beginning, the direct Adams Bashford method will be used to calculate the starting initial points. Then, the initial points we will be used for starting the predictor and corrector direct method. The predictor and corrector direct method can be applied until the end of interval. The shooting technique is used for solving the boundary value problems. In order to get better approximation for the initial points, the value of h will be reduced to $\frac{h}{128}$. However, the predictor and corrector direct method will remain using the choosing step size h.

III. IMPLEMENTATION OF THE METHOD

Shooting technique was applied in the propose method and it is an analogy to procedure of firing objects at a stationary target. We start with the initial guess, t_0 that determines the solution of the derivative $y''(a)$ as the following:

$$y''' = f(x, y, y', y''), \ a \le x \le b \tag{20}$$

$$y(a) = \gamma\ , \ y'(a) = \alpha\ , \ y'(b) = \beta\ , \ y''(a,t) = t_0 \tag{21}$$

Differentiate Eq. (20) with respect to t, and it is simplified as follows:

$$
\begin{aligned}
\frac{\partial y'''}{\partial t}(x,t) &= \frac{\partial}{\partial t} f(x, y(x,t), y'(x,t), y''(x,t)) \\
&= \frac{\partial}{\partial x}(x, y(x,t), y'(x,t), y''(x,t)) \frac{\partial x}{\partial t} \\
&\quad + \frac{\partial t}{\partial y}(x, y(x,t), y'(x,t), y''(x,t)) \frac{\partial y}{\partial t}(x,t) \\
&\quad + \frac{\partial f}{\partial y'}(x, y(x,t), y'(x,t), y''(x,t)) \frac{\partial y'}{\partial t}(x,t) \\
&\quad + \frac{\partial f}{\partial y''}(x, y(x,t), y'(x,t), y''(x,t)) \frac{\partial y''}{\partial t}(x,t) \\
&= \frac{\partial t}{\partial y}(x, y(x,t), y'(x,t), y''(x,t)) \frac{\partial y}{\partial t}(x,t) \\
&\quad + \frac{\partial f}{\partial y'}(x, y(x,t), y'(x,t), y''(x,t)) \frac{\partial y'}{\partial t}(x,t) \\
&\quad + \frac{\partial f}{\partial y''}(x, y(x,t), y'(x,t), y''(x,t)) \frac{\partial y''}{\partial t}(x,t) .
\end{aligned}
$$

Using $z(x,t)$ to denote $(\partial y / \partial t)(x,t)$, we have the initial-value problem

$$
\begin{aligned}
z''' &= \frac{\partial f}{\partial y}(x, y, y', y'')z + \frac{\partial f}{\partial y'}(x, y, y', y'')z' \\
&\quad + \frac{\partial f}{\partial y''}(x, y, y', y'')z'', \ a \le x \le b
\end{aligned}
\tag{22}
$$

$$z(a) = 0, \ \ z'(a) = 0, \ \ z''(a) = 1. \tag{23}$$

For the first initial guessing, t_0 we considered

$$t_0 = \frac{\beta - \alpha}{b - a} \tag{24}$$

See Faires and Burden (1998).

The solution of y' from Eq. (19) is determined when,

$$\varphi(t) = y'(b,t) - \beta = 0 \tag{25}$$

Newton method will be used to get a very rapidly converging iteration. We compute the $\{t_k\}$ defined as:

$$t_{k+1} = t_k - \frac{\varphi(t)}{\varphi'(t)} . \tag{26}$$

From Eq. (35), we know $\frac{\partial y}{\partial t} = z(x,t)$, so $\frac{\partial y'}{\partial t} = z'(x,t)$, and we find the solution for $y'(b,t_k)$ from Eq. (19). The solutions were applied in Newton's method to find the next guess, t_{k+1}.

$$t_{k+1} = t_k - \left(\frac{y'(b,t_k) - \beta}{z'(b,t_k)} \right). \tag{27}$$

Both Eq. (20) and Eq. (22) will be solved simultaneously using the direct method. The process will stop until the error $|\beta - y'(b,t_k)| \le tolerance$, where $tolerance = 10^{-5}$. The algorithm of the proposed method was developed in C language.

IV. RESULT AND DISCUSSION

We now consider three numerical example illustrating the comparative performance of the propose method FOBM over other existing methods. All calculations are implemented by Microsoft Visual C++ 6.0.

Notation:

h Step size

FOBM Fifth order block method

Problem 1:

$$y''' = xy + (x^3 - 2x^2 - 5x - 3)e^x, \quad 0 \le x \le 1,$$
$$y(0) = 0,\ y'(0) = 1,\ y'(1) = -e$$

Exact solution: $y(x) = x(1-x)e^x$

Source: El-Salam *et al*. (2010).

Problem 2:

$$y''' = -y + (x-4)\sin x + (1-x)\cos x, \ \ 0 \le x \le 1,$$
$$y(0) = 0,\ y'(0) = -1,\ y'(1) = \sin 1$$

Exact solution: $y(x) = x(x-1)\sin x$

Source: El-Salam *et al*. (2010).

Problem 3:

$$y''' = -y + (7 - x^2)\cos x + (x^2 - 6x - 1)\cos x, \ 0 \le x \le 1,$$
$$y(0) = 0,\ y'(0) = -1,\ y'(1) = 2\sin 1$$

Exact solution: $y(x) = (x^2 - 1)\sin x$

Source: El-Danaf (2008).

***Table 1*:** The observed maximum errors for Problem 1.

h	FFOBM	(Al-Said and Noor, 2007)
$\frac{1}{16}$	1.5836 x10^{-5}	8.1224 x10^{-4}
$\frac{1}{32}$	2.0237 x10^{-6}	2.1812 x10^{-4}
$\frac{1}{64}$	2.5611 x10^{-7}	5.5859 x10^{-5}
$\frac{1}{128}$	3.2231 x10^{-8}	1.4091 x10^{-5}

***Table 2*:**The observed maximum errors for Problem 2.

h	FOBM	(Al-Said and Noor, 2007)
$\frac{1}{16}$	8.9890 x10^{-6}	4.5978 x10^{-5}
$\frac{1}{32}$	1.2055 x10^{-6}	1.2530 x10^{-5}
$\frac{1}{64}$	1.5538 x10^{-7}	3.2356 x10^{-6}
$\frac{1}{128}$	1.9701 x10^{-8}	8.1999 x10^{-7}

***Table 3*:**The observed maximum errors for Problem 3.

h	FOBM	(El-Danaf, 2008)
2^{-3}	8.5594 x10^{-5}	1.6501 x10^{-4}
2^{-4}	6.6892 x10^{-6}	9.8380 x10^{-6}
2^{-5}	4.6012 x10^{-7}	5.8773 x10^{-7}
2^{-6}	3.0056 x10^{-8}	3.5687 x10^{-8}
2^{-7}	1.9184 x10^{-9}	2.1968 x10^{-9}

In problem 1 and 2, the maximum errors will be obtained when the step size, $h = \frac{1}{16}, \frac{1}{32}, \frac{1}{64}$ and $\frac{1}{128}$. The maximum errors were compared with Al-Said and Noor (2007). For problem 3, the results were compared with El-Danaf (2008). Table 1 and 2 show the maximum errors for FOBM are better than the results in Al-Said and Noor (2007). In Table 3, the maximum errors for both methods are comparable. The results are more precise when the number of *h* is reduced.

V. CONCLUSION

In this research, we conclude that fifth order block method with shooting technique using constant step size is suitable to solve third order nonlinear boundary value problems directly.

ACKNOWLEDGMENT

The author gratefully acknowledged the financial support of Graduate Research Fund (GRF) from Universiti Putra Malaysia and MyMaster from the Ministry of Higher Education.

REFERENCES

[1] G. B. Loghmani and M. Ahmadinia, "Numerical Solution of Third-order Boundary Value Problems," *Iranian Journal of Science & Technology, Tran. A*, Volume 30, Number A3, 2006, pp. 291-295.

[2] S. Talaat El-Danaf, "Quartic Nonpolynomial Spline Solutions for Third Order Two-Point Boundary Value Problem," *World Academy of Science, Engineering and Technology,* 45, 2008, pp. 453-456.

[3] F.A. Abd El-Salam, A.A. El-Sabbagh, and Z.A. Zaki, "The Numerical Solution of Linear Third Order Boundary Value Problems using Nonpolynomial Spline Technique," *Journal of American Science*, 6(12),2010,pp. 303-309.

[4] E.A. Al-Said and M.A Noor, "Numerical solutions of third-order system of boundary value problems," *Applied mathematics and computation,* 190, 2007, pp.332- 338.

[5] P. S. Phang, Z. A. Majid and M. Suleiman, "Solving Nonlinear Two Point Boundary Value Problem using Two Step Direct Method," *Journal of Quality Measurement and Analysis,* 7(1), 2011, pp. 129-140.

[6] Z. A. Majid N. Z. Mokhtar and M. Suleiman, "Direct Two-Point Block One-Step Method for Solving General Second-Order Ordinary Differential Equations," *Mathematical Problems in Engineering,* 2012, pp. 1-16.

[7] S. O. Fatunla, "Block methods for second order ODEs," *International Journal of Computer Mathematics,* vol. 41, 1991, pp. 55-63.

[8] D. Faires and R.L. Burden, *Numerical Methods.*2nd Ed. Pacific Grove: International Thomson Publishing Inc, 1998.

ISBN: 978-1-61804-230-9

Behavior of small and medium-sized enterprises in terms of their strategic management and their goals

L. Rolínek, D. Holátová, M. Březinová, L. Kantnerová

***Abstract*—**The strategic management represents current direction of management. It is applied in all types of enterprises, including small and medium sized ones. Using principles of process management within SMEs has its limitations caused mainly by size of the enterprise and the related focus on operational management. Other specificities are function accumulation, informal leadership, preference for oral communication before written, etc. Nevertheless, implementation of strategic management can increase competitiveness, reduce costs, improve decision-making, facilitate implementation of the employee motivation system, shorten delivery times, higher quality of customer satisfaction, etc. The aim of this article is to find how much SME include principals of the strategic management. In the research project was used secondary data (financial statements of SMEs who took part in the research) and primary data which were obtained primarily through quantitative methods questionnaire supplemented by qualitative method of in-depth interviews. Majority of examined small and medium enterprises had formulated strategy. They are aimed mainly at stability, further at quality, development and profit.

***Keywords*—**management, enterprise, SME, strategy, stability, development, profit

I. Introduction

SMEs are indispensable in all economies, can be described as a driving force of business, growth, innovation, competitiveness, and are also very important employer. In the Czech Republic performed on 31th December 2011 some business activity 1,066,787 legal and natural persons who are placed in the category of small and medium-sized enterprises. The total number of active small and medium-sized enterprises participated in 2011 is 99.84%. The share of employees in small and medium-sized enterprises amounted to 60.85% in relation to the employees of the Czech economy.

According to [7] the small and medium-sized businesses will create and offer new and quality jobs. According to [10] management of small and medium-sized enterprise has many specifics. In small companies due to the small number of employees and managers many functions are accumulated within the competence of only a few workers predominates, informal leadership is more common, oral communication is preferred than written, etc. According to [11] however, a high degree of flexibility to adapt rapidly to changing SME factors, it determines the face and reduce the growing globalization tendencies associated with the onset of multinational corporations and chains. But the principle of important role of strategic management comes here in the force too [8]. Strategic management, based on long-term forecasts, helps the company to anticipate future challenges and opportunities [3]. According [9] strategic management gives clear goals and directions for the future of the company and its workers a sense of security. Strategic management helps to increase the quality of management, leads managers to improve their deciding. Reference [3] further states that strategic management helps improve communication in the company, project coordination, staff motivation and improve resource allocation. The strategy is the basic tool to reach advantages in the market [11]. By [5] the strategic management should be flexible to changed conditions in the market. By [2] the management should have the strategic plan and strategy.

The management of enterprises of different size and specialization are today under the press of advantages, challenges and problems connected with the function of worldwide markets. The different meaning of the term of management can be related as follows: people who wish to have a career as a manager must study the discipline of management as a means toward practicing the process of management. The process consists of certain basic management functions [13]. Global competition carries with it an ever increasing set of demands by the customers. Better quality, more features, better delivery, performance and reduced costs – all become part of the expectation for our customers [14]. Strategic management is that set of managerial decisions and actions that determines the long run performance of a corporation [15]. Strategy can be viewed as building defenses against the competitive forces or as finding a position in an industry where the forces are weaker. Changes in the strength of the forces signal changes in the competitive landscape critical to ongoing strategy formulation. In exploring the implications of the five forces framework [6].

Reference [1] gives two definitions of strategy, traditional and modern. Traditional definition approaches to understand the strategy of the company as a document in which there are long-term objectives of the company, determined the process of operations and allocation the resources needed to meet these goals. The modern concept sees strategy as the company prepared for the future. The strategy includes the long-term objectives of the company, the process of strategic operations and deployment of enterprise resources needed to meet these goals so the strategies would be based on business needs, take into account the changes in its resources and capabilities, and at the same time adequately respond to changes in the environment of the company. The central concept of strategic management is a strategy that is closely linked to the objectives it tracks. According to [4] strategies express the basic idea of which way the company goals will be achieved. There is no one best way to manage say [16].

The storied British banker and financier Nathan Rothschild noted that great fortunes are made when cannonballs fall in the harbor, not when violins play in the ballroom. Rothschild understood that the more unpredictable the environment, the greater the opportunity if you have the leadership skills to capitalize on it say.

The article deals with the mapping of the fundamental characteristics of SMEs in terms of their strategic management.

II. METHODS

This article is provided as one of the outputs of the research project "Process management and the possibility of its implementation in small and medium-sized enterprises" of Grant Agency of the University of South Bohemia GAJU 068/2010/S and serves as one of the sources for grant GAJU 039/2013/S.

In the research project were used secondary data (financial statements of SMEs who took part in the research) and primary data which were obtained primarily through quantitative methods questionnaire supplemented by qualitative method of in-depth interviews.

For the classification of small and medium-sized enterprises has been used a new definition of the European Union (European Commission: A new definition of SMEs 2006) in accordance with the Law No. 47/2002 Coll. as amended. This system has been adapted for the needs of the research, according to the following table number 1, where the group of small businesses was divided into two groups according to the number of employees in order to achieve detailed breakdown.

Table. 1 Classification of SMEs

Enterprise size	Number of employees Number of employees (Annual Work Unit)	Annual turnover	Annual balance sheet
Medium enterprise	50 – 249	< 50 mil. €	< 50 mil. €
Small enterprise	25 – 49	< 10 mil. €	< 10 mil. €
	10 – 24	< 10 mil. €	< 10 mil. €
Micro enterprise	0 – 9	< 2 mil. €	< 2 mil. €

In the South Bohemian region was registered in 2011 68,826 economic entities from the selected category. Enterprises with up to 9 employees (micro-enterprises) represent 18% of the total number. Small enterprises represent less than 4% of all the small and medium-sized enterprises in the region. There are registered 78% of medium-sized enterprises within the region (Statistical bulletin- South Bohemian region, 2011). According to the legal form of business in the basic aggregate there dominate private entrepreneurs (77%), after that companies (9%), associations (5%), corporations of proprietors and cooperatives (3%), foreign persons (2.5%) and others (European Commission). Key figures on European business with a special feature on SMEs 2011.The prevailing business activities according to the classification of economic activities CZ-NACE in the examined aggregate constitutes wholesale and retail (22%). Building industry and industry are represented within the research identically with 13%, within the region there are about 21,000 or 19,000 as the case may be and so they occupy the second and third position from the point of view of CZ-NACE. Business activities according to CZ-NACE such as section A - agriculture, forestry and fishing, section B - mining and extraction and section I – accommodation, catering and restaurants, were eliminated from the research because of their different perception and management of company processes – these are completely specific groups. There was chosen a research sample from the base aggregate with the method of improbability of random choice by reason of difficult conditions of data collection.

The research sample consists of 187 small and medium-sized enterprises of South Bohemia region. Composition of the research sample according to size (number of employees) and activity is indicated in table 2 and 3.

Table. 2: Research Sample of Small and Medium-Sized Enterprises According to Number of Employees

Number of employees	Number of enterprises	
0 - 9	20	11%
10 - 24	70	37%
25 - 49	36	19%
50 - 249	61	33%

Table. 3: Research Sample of Small and Medium-sized Enterprises according to Business Activity

Activity	Number of enterprises	
Building industry	51	27%
Engineering	57	30%
Wood production	22	12%
Food production	11	6%
Trade	22	12%
Services and Transport	24	13%

III. RESULTS

From the point of view of categories due to number of employees is evident, that the majority (always more than 60 %) of enterprises has formulated strategy. Differences concerning strategies and their formulation at individual categories of examined small and medium-sized enterprises are indicated in table 4.

At the medium-sized enterprises (with more than 50 employees) the frequency is about 87 %. It is surprising that 72 % of micro-enterprises (up to 9 employees) also have formulated strategy. From the point of view of written formulation of the examined strategy the number of enterprises is lower, due to the fact that the number logically grows with the enterprise size. Micro-enterprises have then written formulated strategy only in 9 % of cases, on the contrary medium-sized enterprises in 63 % of cases, at the same time where enterprises with more than 150 employees have written formulated strategy in 83 % of cases. It is possible to compare the given results in the figure number 1, where curves of frequencies of examined enterprises having the formulated strategy, and of enterprises having the strategy formulated in the written form, are obvious.

Table. 4: Frequencies of Formulated Strategies of Examined Small and Medium-sized Enterprises according to the Number of Employees

Number of employees	Number of enterprises	Formulated strategy		Written formulated strategy	
		yes	no	yes	no
0 – 9	20	72%	28%	9%	81%
10 - 24	70	61%	39%	30%	70%
25 - 49	36	69%	31%	61%	39%
50 - 249	61	87%	13%	63%	37%

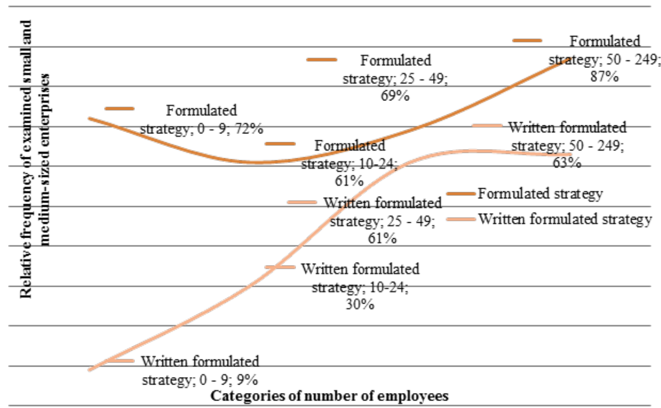


Fig. 1: Curves of Frequencies of Examined Small and Medium-sized Enterprises Having Formulated Strategy According to the Number of Employees

The formulation of strategy from the point of view of categories due to business activity is shown in the table 5. It is similarly for all enterprises, that always minimally 60 % of them have formulated strategy. Most of the enterprises with formulated strategy are in the service sector (81 %), and also with the written strategy (59 %). The enterprises in the sphere of commerce show a little bit lesser frequency, when 76 % of them have formulated strategy and 44 % strategy in written form. Generally enterprises in the field of production show the frequency about 74 % in relation to the formulated strategy and 40 % in relation to the written formulated strategy.

Strategies are most often created by enterprises which are concentrated on engineering production (81%, 49% written strategies). On the other side only 57% of enterprises which are focused on wooden production create strategies and only 14% of them have got a written version. Also the construction enterprises focus on strategies less - only 63% of them prepare their strategies and only 39% out of the examined enterprises have written form. The data can be seen in the figure number 2. The curves are basically copying the same line just with different values.

Table.5: Frequency of Examined SMEs with a Strategy according to Business Area

Business area	Number of companies	Prepared strategy		Written version	
		yes	no	yes	no
Construction	51	63%	37%	39%	61%
Engineering production	57	81%	19%	49%	51%
Wooden production	22	57%	43%	14%	86%
Food production	11	75%	25%	37%	63%
Trade	22	76%	24%	44%	56%
Services and Transport	24	81%	19%	59%	41%

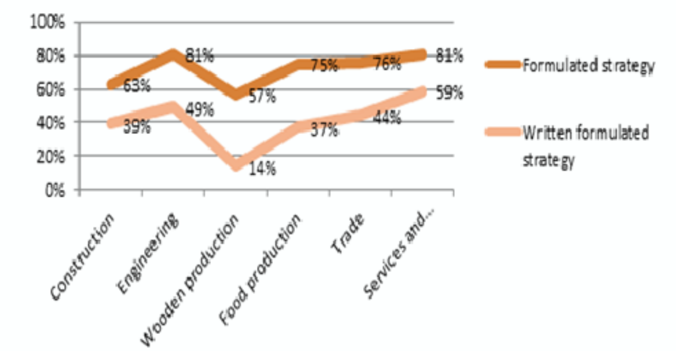


Fig. 2: Curves of Frequency of Examined SMEs with a Strategy according to Business Area

Table. 6: Frequency of Followed Strategies in examined SMEs according to the Number of Employees

Number of employees	Number of companies	Not specified	Quality	Stabilization	Development	Profit
0 - 9	20	36%	36%	14%	9%	5%
10 - 24	70	40%	34%	14%	9%	3%
25 - 49	36	36%	28%	19%	11%	6%
50 - 249	61	16%	58%	16%	5%	5%

Out of 187 examined enterprises 31% of them didn't indicate their strategy type. The rest of the examined SMEs are focused on strategy of survival. Four categories were determined:

1) "Quality" (41% of examined enterprises) - enterprises focus on quality in all aspects of business, on business ethics, they aim to have a good name, they are oriented on customers and make all the effort to satisfy them, these enterprises are working towards higher flexibility and reliability in everyday praxis and also try to make their employees satisfied
2) "Stability" (16% of examined enterprises) - enterprises focus on stability within the market but also stabilization of its internal environment, for example keeping the same number of employees
3) "Development" (8% of examined enterprises) - enterprises focus on development of the enterprise in many areas - extension of production or provided services, market share and innovations
4) "Profit" (4% of examined enterprises) - enterprises focus on a long-standing growth of profit, this includes the growth of revenues and also decrease of costs.

Table 6 and figure 3 shows the differences between strategies of SMEs from different categories. It can be seen that the strategies structure of examined enterprises is similar, similarity can also be seen in terms of order of relative frequencies. Categories differ in terms of followed strategies. Dominant strategy is "Quality" followed by 30% of small enterprises (less than 50 employees), on the other side more than half (58%) of enterprises with more than 50 employees follow this strategy. Second most common strategy is "Stability" - around 20% of enterprises choose this strategy to reach their goals. "Development" strategy is followed by 10% of enterprises, but enterprises with more than 150 employees don't implement this strategy at all. Last strategy "Profit" is followed by 5% of enterprises.

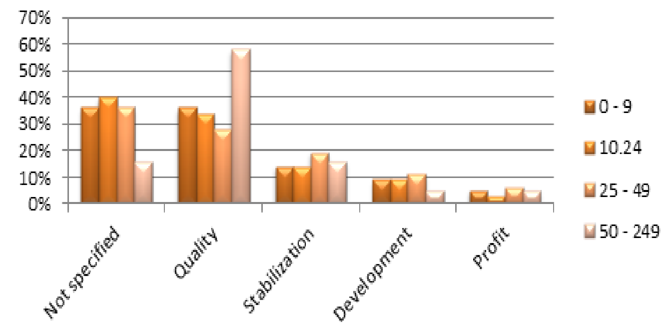


Fig. 3 : Frequency of followed Strategies in examined SMEs according to the Number of Employees.

Table 7 and figure 4 highlight a similar analysis of SMEs presenting the results according to a business activity. The structure of implemented strategies is similar as above where the categories were created according to the number of employees, but some differences can be seen in order of relative frequencies. Enterprises working in construction usually follow "Quality" and "Stability" strategies, 9% of enterprises follow "Profit" strategy and only 7% follow "Development". “Quality“ was the most followed strategy by manufacturing enterprises, even more dominantly as it was chosen by 40% businesses as the device to reach their goals. 17% manufacturing enterprises in total follow “Stabilization“ strategy, roughly half less of them follow strategy focused on “development“, and just 2% of companies follow strategy of “Profit“.

Engineering production companies follow dominantly strategy of “Quality“, wood and food production companies follow strategy of “Stabilization“ similarly. Most trading companies (Trade, 52%) follow strategy of “Quality“, 20% of companies are focused on strategy of “Stabilization“, just 4% strive for fulfillment of “Development“. No trading company set profit strategy as its goal. Service companies follow mostly quality strategy (57%), the second most frequent strategy was “Development”. Just few service companies follow “Profit“(5%) or “Stabilization“(3%) strategy.

ISBN: 978-1-61804-230-9

Table. 7: Frequency of SME observed Strategies according to the Business Activity

Business activity	Frequency	Not specified	Quality	Stabilization	Development	Profit
Construction	51	38%	24%	23%	7%	8%
Engineering	57	36%	41%	10%	11%	2%
Wood production	22	19%	38%	33%	5%	5%
Food production	11	37%	38%	25%	0%	0%
Trade	22	24%	52%	20%	4%	0%

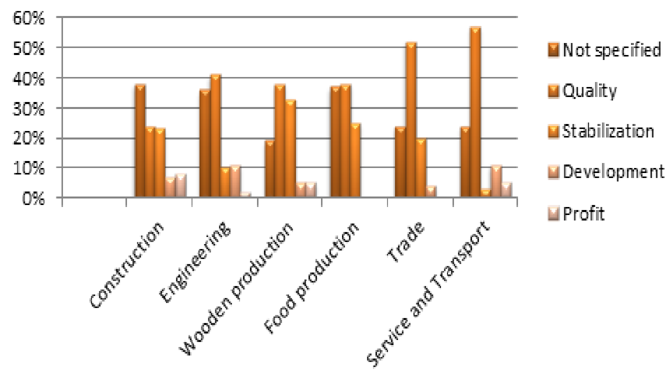


Fig. 4: Frequency of SME observed Strategies according to the Business Activity

Causal effects of strategy choice among observed small and medium enterprises according to the number of employee's category and the type of business activity were tested using statistical method of Kruskal-Wallis test. First, relation between strategy choice and number of employees was tested. Based on analyzed data we were able to confirm with 95% probability the null hypothesis that says there is no relation between tested variables (H = 6,675946; df = 4; p-value = 0,1540).

Further, relation between strategy choice and category of business activity was tested. Based on analyzed data we were able to reject with 95% probability the null hypothesis in favor of the alternative hypothesis that says there is relation between examined variables (H = 7,849167; df = 3; p-value = 0,0492). This relation was tested using correlation coefficient r. Interdependence of strategy choice and particular business activity is evident in table 8 and graphically in figure 5. The strongest correlation (80%) was detected by strategy choice of product engineering and trading companies. Strong correlation (74%) was also found by determining strategy of building companies and wood processing companies. 68% correlation is visible by strategy choice of wood processing and service (transport included) companies. Last significant correlation, higher than 50%, were discovered by wood processing and trading companies; building and service companies (transport included). On the other hand no correlation was detected by determining strategy of production engineering and trading companies, food industry and trading companies.

Table. 8: Correlation Coefficient of MSE Strategy according to the Business Activity

	Building Industry	Production Engineering	Wood Processing Industry	Food Industry	Commerce	Service and Transport
Building Industry		0,438736	0,741967	-0,102815	0,808746	0,518120
Production Engineering	0,438736		0,294688	0,022168	0,000000	0,319173
Wood Processing Industry	0,741967	0,294688		-0,146748	0,543214	0,683588
Food Industry	-0,102815	0,022168	-0,146748		0,000000	0,345930
Commerce	0,808746	0,000000	0,543214	0,000000		0,147087
Service and Transport	0,518120	0,319173	0,683588	0,345930	0,147087	

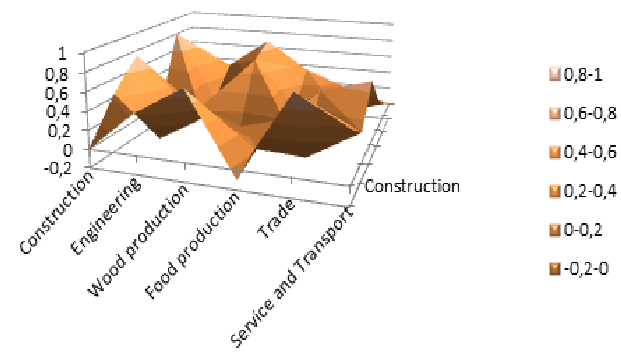


Fig. 5: Correlation Coefficient of MSE Strategy according to the Business Activity

Dedouchová [1] presented that majority of the small and medium sized companies has formally /as written document/ or not formally determined strategy. Presented research brought short term goals and brought also the similar results. Strategy has the goal to push forward the company towards to achieve short term goals and finally to achieve its basic purpose. The goals of the small and medium sized companies are often based on personal goals and preferences of the owners. Determined goals in such a way are logically very subjective, they are often not chosen correctly in comparison with the milieu where the companies create activities. The goal of the company should always correspond with the chosen mission and should put into respect strong and weak aspects of undertaking, opportunities and distress, which the company can use or must face them. In general the goal of the company should answer the basic questions of undertaking and present requested result. According to Minárová and Malá [12] target of any company consists from different aims, which management wants to reach and which should be guaranteed by performance of different activities. These targets should be ambitious, easy to accomplish, quantifiable.

In the research sample of the small and medium sized companies in the frame of research there were identified 6 categories of the company goals. The category "Quality" was dominant. The companies presented for their goal both satisfaction of the customers (also employees) increasing the quality, good name of the company and flexibility. Following most often presented goal was "Development", the development of the whole company and separate development of the partial parts too. Innovation was also presented as the main goal in this category. Category "Stability" has also very strong representation, the companies presented the whole stability at the market and also stability of prices. In this category the companies also mentioned keeping the number of employees. Another goal is connected with this matter and it is covering the market and Survival, which was followed in minimum cases. Concerning studying small and medium sized companies the goal Profit was, of course, dominant one.

In the table 9 and figure 6 numerousness of searching of the goals of some small and medium sized companies according to the number of employees is presented. Following characteristics can be seen from these results. Concerning micro companies the goals are concentrated on "Quality" (64 percent), further "Development" (50), and "Stability" (41) on the contrary, no company follows the goal "Survival". The small companies to 24 employees follow the goal Stability in 57 percent cases. In the small companies with more than 25 employees the main goals were "Quality", "Development" and "Stability" (55 percent). These companies (the same as micro companies) do not follow the goal "Survival", in comparison with the small companies to 24 employees, that follow this goal, but only in 4 percent. The medium sized companies follow in a dominant way the goal "Profit" (55 percent), the second most often followed category is "Development" (46). 43 percent medium sized companies follow the goal "Stability" and 34 percent the goal "Quality". Minimum medium sized companies follow the goal "Survival" and "Covering" the market.

Table. 9: Numerousness of the studying Goals of searched MSP according to the Number of Employees

Number of employees	Number of the comp.	Goal					
		Quality	Covering of market	Survival	development	Stability	Profit
0 – 9	20	64%	18%	0%	50%	41%	32%
10 – 24	69	43%	4%	6%	32%	57%	40%
25 – 49	35	55%	8%	0%	55%	55%	33%
50 – 249	60	34%	8%	8%	46%	43%	51%

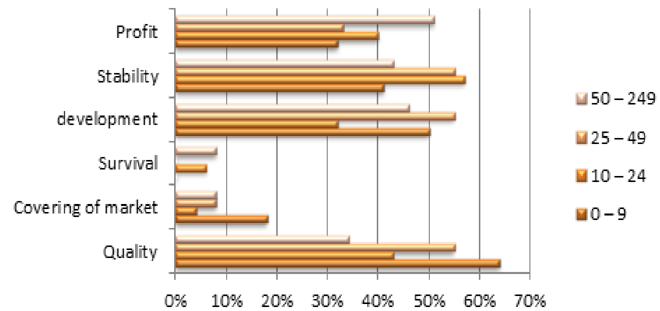


Fig. 6: Numerousness of the studying Goals searching MSP according to the Number Employees**.**

Table 10 and graph 7 show numerousness concerning diversity according to the activity of the studying small and medium sized companies.6 categories of goals were analysed here. Category Quality is dominant in the business companies (72 percent).Also less than 50 percent companies follow this goal in engineering production, on the contrary, food companies follow this goal minimum /18/,this group follows mainly category Stability in 73 percent cases. Companies in construction and wood working ones follow the goal very strongly .Business companies follow this goal very rarely Following the goal in the category of Covering the market was minimum in all studying activities of research of small and medium sized companies. Companies working in services including transport follow this goal in 15 percent and companies in engineering production follow this goal least of all in 4 percent. The goal Survival is minimum followed, food companies and business companies do not follow this goal at all. The rest of the companies shows only in 5 -7 percent. In total 60 percent small and medium sized companies in engineering production follow the goal Development, 48 percent in services including transport and 38 percent wood working companies. Construction companies follow this goal minimum (27).The goal Profit is strongly dominant for no category of the companies. This goal is followed by 55 percent construction companies, 43 percent woodworking companies, 42 percent companies in services. Food companies follow this goal only in 18 percent.

Table 10: Numerousness of the Goals examined MPS according to the activities.

Activity	Number of companies	Goals					
		Quality	Covering of market	Survival	Development	Stability	Profit
Construction	50	40%	7%	7%	27%	62%	55%
Engineering	57	49%	4%	7%	60%	38%	36%
Woodworking	22	29%	5%	5%	38%	71%	43%
Food	11	18%	9%	0%	36%	73%	18%
Trade	21	72%	12%	0%	36%	28%	36%
Service and Transport	23	36%	15%	3%	48%	42%	42%

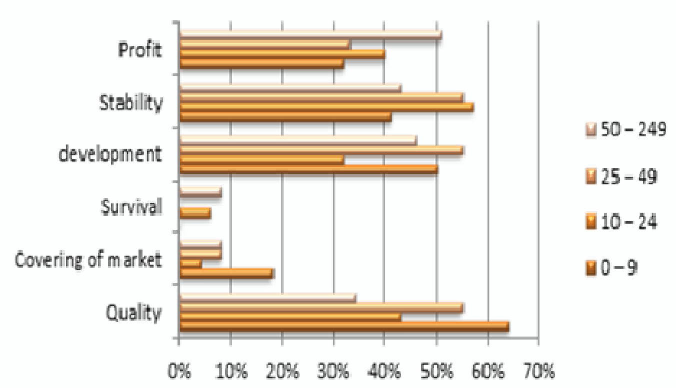


Fig. 7: Numerousness of studying Goals examined MPS according to the activities

As it was mentioned above, the goal of the company should set off the basic mission of the company with the respect to the weak and strong side. These matters were also analyzed in the examined small and medium sized companies and are presented in the figures 8 and 9. It can be seen that the companies consider very strong sides resp. Competitive advantage "Range" (38 percent), and "Quality of work" (36). 22 percent companies consider very strong page "Flexibility" and 16 percent numerousness "Know-how". The remained strong pages that were presented by the small and medium sized companies were not more than 10 percent. There was, for example, personal approach, good name of the company, quality of employees, reliability etc.

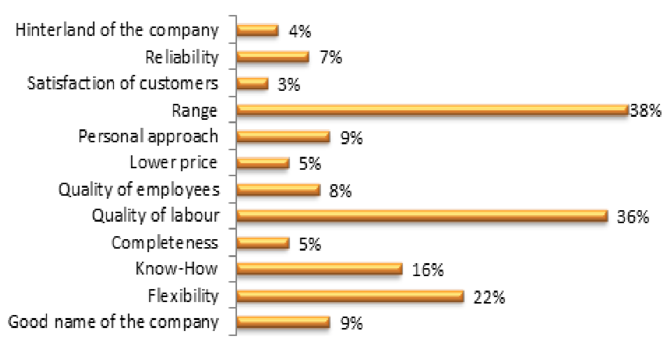


Fig. 8: Numerousness of strong Sides (competitive Advantages) examined MSP

Most often presented weak side resp. threat was insolvency of the customer-40 per cent examined small and medium sized companies. In total 25 percent companies presented decline of sale, less than policy of the state, growth of costs, and 18 percent companies presented competition. About 10 percent companies presented decline of demand, lack of qualified workers, economic crisis, but also low innovation or poor knowledge of languages of employees.

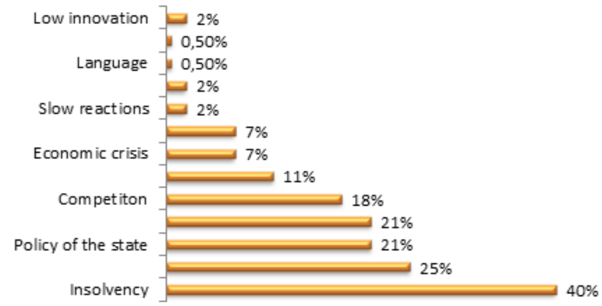


Fig. 9: Numerousness of weak Side (threats) of examined MSP

IV. CONCLUSION

Majority of examined small and medium enterprises had formulated strategy, however, mostly just enterprises with more than 25 employees has also a written version. From business activity point of view is distinctly lower frequency of formulated strategy by wood processing companies, strategy is formulated and put down in words most often by trading and service (transport included) enterprises. Based on tested data the most frequent followed strategy is quality and stabilization, regardless of employees number category or business activity.

In the case, the company has determined goal, it is aimed mainly at stability, further at quality, development and profit. Small companies follow more quality and development, but medium sized companies follow mainly development and profit. Examined small and medium sized companies consider their strong pages mainly range and quality of work and on the contrary as threat insolvency.

REFERENCES

[1] Dedouchová, M. (2001). *Strategie podniku*. Praha: C.H. Beck.

[2] Charvát, J. (2006). *Firemní strategie v praxi*. Praha : C. H. Beck.

[3] Keřkovský, M. ,Vykypěl O. (2006). *Strategické řízení: teorie pro praxi*. Praha: H. Beck.

[4] Kotler, P.,Keller K.L. (2012). *Marketing management*. New Jersey: Prentice Hall.

[5] Mallya, T.: *Základy strategického řízení a rozhodování*. Praha: Grada.

[6] Porter M. E. (2008). *The Five Competitive Forces That Shape Strategy*. Harward Business Revue.

[7] Řehoř, P. (2007). Analysis of the labour market due to the achieved level of entrepreneurial activities in the regions of the Czech Republic. Conference: 5th Scientific International Conference on Hradec Economical Days 2007, Location: Hradec Kralove, the Czech Republic. Date: FEB 06-07.

[8] Savov, R., Lančarič, D., Paška, D. (2008). Strategický manadžment kvality v podmienkach agropodnikatelských subjektov na Slovensku, Nitra: SPU.

[9] Sedláčková, H., Buchta, K. (2006). *Strategická analýza*. Praha: C. H. Beck.

[10] Srpová, J., Řehoř, P. (2010). *Základy podnikání : teoretické poznatky, příklady a zkušenosti českých podnikatelů.* Praha: Grada.

[11] Kolektív autorov (2012). *Manadžment malých a stredných podnikov*. Nitra: Gramond.

[12] Minárová, M., Malá, D. (2011). *Strategický manadžment*. Banská Bystrica : Univerzita Mateja Bela, Ekonomická fakulta v Banskej Bystrici.

[13] Ivancevich J.M., Donelly J. H., Gibon J.L. (1989). *Management – Priniciples and Function*. Library of Congres Cataloging Publication Data.

[14] Vollmann, T.E., Berry W.L., Whybark D.C. (1988). *Manufactoring Planning and Control Systems* Illinois: Irwin.
[15] Wheelen T.L., Hunger J.D. (2008). *Strategic Management and Business Policy.* New Jersey: Upper Saddle River.
[16] Fisher III, William W., Oberholzer-Gee, Felix (2013). Strategic management of intellectual property. Available at: http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&sid=089eb7c9-fcea-4f4c-8dc9-2f3ac8b9e8e7%40sessionmgr114&hid=120. Accessed 25 October 2013.
[17] Paul J. H., Krupp, S., Howland, S.: Strategic Leadership: Essential Skills. Available at: http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=5&sid=089eb7c9-fcea-4f4c-8dc9-2f3ac8b9e8e7%40sessionmgr114&hid=120. Accessed 25 October 2013.

ISBN: 978-1-61804-230-9

ARMA models for blackouts forecasting and Markov method for interruption modelling in electrical power systems

Iberraken Fairouz, Medjoudj Rabah, Aissani Djamil and Klaus Dieter Haim

Abstract— **Customers' requirements in terms of availability of supply and quality of energy are increasingly growing and the electricity distribution utilities are asked to meet them and to provide a high quality of service. They are required to activate in a multidisciplinary area including the mastery of data acquisition and processing, secured information transfer and an easy communication mode with customers. This paper aims to provide knowledge and to think creatively to assist decision-makers to understand the impacts of information and communication technologies (ICT) integration in the conventional power network giving birth to a smart energy grid. We have also investigated the attributes of combining smart systems integration with renewable energy resources insertion to mitigate the occurrence of cascading events such as blackouts. The basic tool of our work is the mastery of time series analysis and the forecasting models. This work is an excellent opportunity to highlight the dominance of ARMA model in forecasts and that of the Markov model in interruptions modelling in the field of reliability evaluation of electrical networks.**

Index Terms— **ARMA models, blackouts, interruption modeling, renewable energy, smart grid.**

I. Introduction

Traditionally, the research and the development of reliability analysis methods have focused on the generation and the transmission part of the electric power systems. Due to economic factors, they are commonly run near their operational limits. Majors cascading disturbances of blackouts of the transmission systems have various consequences. From reliability point of view, these events demonstrate that operating practices, regulatory policies, and technological tools for dealing the changes are not yet in place to assure an acceptable level of reliability. This paper investigates the contribution of the integration of the technologies of information and communication combined to the insertion of renewable energy resources giving birth to a smart energy grid. These smart systems allow mitigating these cascading events, from one hand and accelerating the network reconstruction in the case where those events were not prevented, from the other hand. Very far in the past, the issue of blackouts was introduced in biomedical engineering, where researchers have defined a blackout as loss of vision [1]. Effectively, the definition regarding the electrical engineering studies is not so far as it means the loss of light. A typical large blackout has an initial disturbance or trigger events followed by a sequence of cascading events [2]. The progression of blackout can be divided into several steps, such as: precondition, initiating events, cascade events, final state and restoration. Among these five steps, cascade events can be further divided into three phases in the process of some blackouts [3]-[4]: steady-state progression, triggering events and high-speed cascade. Karamitsos and Orfanidis [5] enumerate different causes of blackouts occurrence using several structures of networks. They have used power flow techniques to model the behavior of the systems and have stated that the frequency and the size of the blackouts depend weakly on the topology of the network and their distribution is a weak function of the latter. For some authors, suggesting some approaches, it is possible to understand and to control blackouts [6]. The authors have analyzed a 15 years, of North American electric power system blackouts for evidence of self-organized criticality. They have proposed three measures tools, such as: the energy unserved in MWh, the amount of power losses in MW and the number of customers affected. They have given the proportion of the contribution of weather in blackouts occurrence, estimated to 50% of the total number. This phenomenon takes into account the intensive operation of both cooling and heating systems. To reduce the risk of major blackouts trough improved power system visualization, Overbye and Wiegmann [7] describe several visualization techniques that can be used to provide information and control at time using automation system. They have stated that it is possible to mitigate blackouts through corrective and extreme control; such as: quickly load shedding at right location and opening tie-lines. Many researchers were interested in the problem of blackouts, and more works were conducted on the deterministic side. Nevertheless, the stochastic aspect has been addressed with the introduction of several probabilistic models highlighting the risk of blackouts occurrence. To our knowledge, the forecast aspect has not been addressed, and it is in this context that we make a modest contribution; it is to show the tendency of the evolution of these events dreaded by both energy distributors and consumers. The work reported in this paper was made by a research team specialized in electrical engineering reliability (FSE2) of the research unit LaMOS from the University of Bejaia, Algeria. After reviewing about 200 works on the blackouts, and based on the synthesis work conducted by the IEEE PES

ISBN: 978-1-61804-230-9

Task Force [8] and while having ideas about what can make smart systems, we have introduced forecast models to show the evolution of the latter in the future and to predict the behavior of the conventional network if ever information systems are integrated and in the case it becomes better communicating.

Using box and Jenkins models for North American blackouts forecasting, where the trend is quite constant, we have demonstrated that the ICT integration performs both power system operation and some reliability indices. With smart meters, a smart grid is able to collect real-time information about grid operations, through a reliable communication networks deployed in parallel to the power transmission and distribution systems.

The occurrence of a blackout can be modeled using the Markov method. A competing failure process is proposed gathering the degradation process (due to the normal operation of the system subject to ageing and to the wear of equipment) and the shock process (due to the loss a generator or a main line). Based on the three experiences of the Algerian, Italian and American blackouts occurred in year 2003; a lot of efforts may be made to reduce restoration times. Regarding the delays of turbines starting, it is demonstrated that renewable sources can play an important role in this issue, thanks to the instantaneous delivery assured using batteries storage. A particular interest is given to the contribution of both integration of smart systems and insertion of renewable sources to better manage the peaks of demands, to reduce power losses and to increase the availability of electricity. The achievements of these goals depend on the engagement of managers and the comprehension of customers to adopt the decisions. This issue can be investigated using economic criteria inspired from game theory as developed recently by Medjoudj et al [9]. Several countries have already taken a step ahead in the field of smart grids integration in the contexts of sustainable development and the insurance of the security of energy supply. The rest of the paper is organized as follows: section 2 is devoted to time series modelling, using Box and Jenkins method. In section 3 a particular attention is given to cascade degradation processes modeling where we have introduced competing failure processes including both degradation and a shock processes. In the case of a blackout occurrence, a particular interest should be granted to the restoration process. This issue is developed in section 4 with applications for three special cases of blackouts occurred in year 2003. In section 5, we have judged interesting to think creatively to produce technical measures to develop some societal indicators in conjunction with smart systems to mitigate blackouts, as a future work. Section 6, is devoted to conclusion and discussions.

II. TIME SERIES MODELLING

There are two types of models to account for time series. Initial works consider that the data is a function of time $y = f(t)$. This category model can be adjusted by the least squares method, or other iterative methods. The model analysis by Fourier transform is a sophisticated version of this type of model. A second class of models seeks to determine the value of each series based values which proceed $y_t = f(y_{i-1}, y_{i-2}, \ldots)$. This is the case of ARIMA models (Auto-Regressive - Integrated -. Moving Average), this class of models has been popularized and formalized by Box and Jenkins [10].

A. Main time series models building

A time series is a chronological sequence of observations on a particular variable. Usually the observations are taken at regular intervals (days, months, years). A time series analysis consists of two steps: (1) building a model that represents a time series, and (2) using the model to predict (forecast) future values. For our studies, we introduce the following models:

1) ARMA models: The general Auto-Regressive Moving Average (ARMA (p,q)) model for a univariate stationary time series can be presented analytically as:

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + u_t \\ &\quad + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q} \end{aligned} \tag{1}$$

Where: ϕ and θ are polynomial functions of degrees p and q respectively. It is a combination of two processes such as: Auto-Regressive and Moving Average. The autoregressive components represent the memory of the process for the preceding observations. This model can be presented as follows:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots \phi_p Y_{t-p} + u_t \tag{2}$$

The moving average components represent the memory of the process for preceding random shocks. It can be formulated as:

$$Y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q} \tag{3}$$

2) ARIMA models: A no stationary series is provided by the ARIMA (p,d,q) processes, it has the general form:

$$\Delta^d Y_t (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = (1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_p L^q) u_t \tag{4}$$

3) SARIMA models: We use the Seasonal Autoregressive Integrated Moving Average processes when we have to deal

ISBN: 978-1-61804-230-9

with time series with trends, seasonal pattern and short time correlations.

If we denote:

- Y_t, The number of events passing through the observed link during the time interval $[(t-1)\Delta\,;\,t\Delta]$ of duration $\Delta > 0$ for t=0,1,2,…

- B, the backshift operator which affects the time series y_t given by $(B^d y)_t = y_{t-d}$ for al integers d.

The $SARIMA(p,d,q)\times(P,D,Q)_s$ process y_t is given by the following equation:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t \qquad (5)$$
$$=\theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

Where: ε_t is a white noise sequence, Φ and Θ are polynomial functions of degrees P and Q respectively.

B. *Forecasting time series*

In general Box and Jenkins have popularized a three-stage method aimed at selecting an appropriate (parsimonious) ARIMA model for the purpose of estimating and forecasting a univariate time series. The three stages are: (a) identification, (b) estimation, and (c) diagnostic checking.

1) *Model Identification:* A comparison between Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) samples with those of various theoretical ARIMA processes may suggest several plausible models.

If the series is non-stationary the ACF of the series will not die down or show signs of decay at all. If data are non-stationary, transformations are required to achieve stationarity. The time series differentiation is involved to remove the trend and the seasonal trends. It is often considered the following operations:

The first difference: $z_t = (y_t - y_{t-1})$

The second difference: $z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$

From the other hand, we can consider a Log transformation to handle the exponential growth of the series and to stabilize their variability.

Once we have reached stationarity, the next step is the identification of the p and q orders of the ARIMA model. Model Estimation: *In this second stage, the estimated models are compared using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).* Model Validation: *In this stage we examine the goodness of fit of the model. We must be careful here to avoid over fitting (the procedure of adding another coefficient is appropriate). The special statistic tests we use here are the Box-Piece statistic (BP) and the Ljung-Box (LB) Q-statistic, ones which serve to test for autocorrelations of the residuals.*

C. *Real cases applications*

In this paper we have made two applications using Box and Jenkins analysis and forecasting and where the dominance of ARMA model is demonstrated.

1) *Forecasting the number of blackouts of northern USA*

The phenomenon which binds several countries in the field of power systems is the blackout. Just look at the spectacular events of 2003, where Algeria has shared the same adventure with the Italians and Americans. America is very striking this phenomenon because it is repetitive and a database has been created in this way, and where several research works have exploited it. To this end, we wanted to use it to establish forecasting models. We have reported the events as an original time series plot in figure 1.

We can see from this plot that there seems to be seasonal variation. So we need to introduce the differentiation of the series. Both of ACF and PACF show a single spike at the first lag. An ARMA (1, 1) model is indicated. The analysis of the autocorrelation coefficients and the partial autocorrelation of the transformed series bring up a significant peak delay 1 for the autocorrelation peak and a significant delay 2 for the partial autocorrelation. Based on these results, we propose as a possible model SARIMA (1,0,1) $(1,1,2)_4$. The estimated model is given as follows.

$$Y_t = 0.9560\,Y_{t+1} - 0.8694\,Y_{t-6}$$
$$-\varepsilon_t + 0.7638\,\varepsilon_{t-1} + 0.9999\,\varepsilon_{t-4}$$

The forecasting values are shown in the plot of the figure 2, where the behavior of the future events seems to be constant, thanks to the integration of smart systems in the American power network. This country has a great advance in the development of ICT, and it is oriented towards the reliability issue and to mitigate these dreaded events.

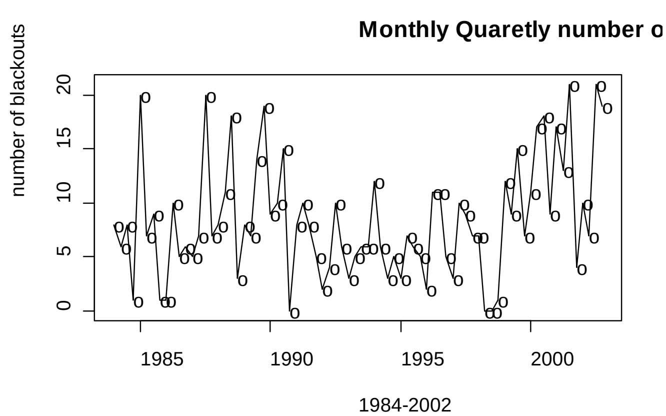


Fig. 1. Plot of the Original series

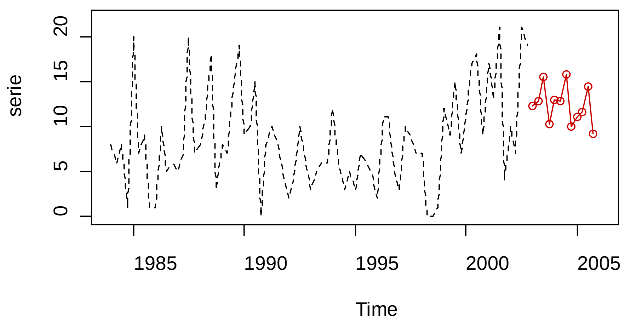


Fig. 2. Result of forecasting

2) *Forecasting failures of components of an electrical distribution network*

As stated in several recent publications, 90% of the total failures are recorded in the distribution part. It is essential to recall that actually power systems are modelled as both multi-components and multi-degraded systems. It is useful to distinguish and to classify failures based on their origins and by the components it concern. To this end, using Box and Jenkins method, we formulate forecasting models for such equipment of the Bejaia city, (Algeria) distribution system, as follows:

- overhead lines forecasting failures model

$$Y_t = 0.34851 Y_{t-1} + 0.77956 Y_{t-2} + 0.83544 \varepsilon_{t-1} + \varepsilon_t$$

- underground cables forecasting failure models

$$Y_t = 0.29148 Y_{t-1} + 0.83538 Y_{t-2} + 0.85063 \varepsilon_{t-1} + \varepsilon_t$$

- joint nodes forecasting failures model

$$Y_t = 1.01066 Y_{t-1} + 0.06623 Y_{t-2} + 0.33257 \varepsilon_{t-1} + \varepsilon_t$$

- MV/LV sub-station forecasting failures models

$$Y_t = 0.04987 Y_{t-1} + 0.75822 Y_{t-2} + \varepsilon_t$$

- Mean voltage feeder forecasting failures model

$$Y_t = 0.30520 Y_{t-1} + 0.21176 Y_{t-2} + 0.30099 Y_{t-3} + \varepsilon_t$$

The time series analysis was done on annual intervals and the data were collected on a period of ten years. The models are generally ARMA (p, q) and the forecasts show that failure evolve in increasing manner with the exception of the last two years due to the renewal of some ageing sections of cables during the recent years.

III. CASCADE DEGRADATION PROCESSES MODELING

Power transmission networks are heterogeneous systems with a large number of components that interact with each other through various ways. When the limits are exceeded for a component, it triggers its protective device. Therefore, it becomes faulty in the sense that it becomes unavailable to transmit electrical energy. The component can also fail in the direction of misoperation or damage due to ageing, or low maintenance. In any case, the power will be redistributed to other network components, according to the laws of the mesh nodes and electrical circuits, or by manual or automatic redistribution. This power will be added to the already existing power carried by these components. Therefore, their overload is inevitable if they are at their operating limits. So, this scenario leads to the propagation of failures through the network. This propagation can be local or it may be general, if the overload caused by the first degradation is very important. Any future deterioration comes to instantly change the configuration and operational parameters of the network. It makes the system unstable and the seat of transients very violating the majority of cases, such as the collapse of voltage and frequency and the loss of synchronism. Usually, the system can be pulled back to normal condition by its protection and control system. But, sometimes, the system cannot return to normal condition in good time and some new events can trigger the cascade incidents, which may interact and rapidly worsen the situation. Finally, blackout can happen. Then, every disturbance triggers a next one, and so on; the system will pass from state i to state $i-1$ due to gradual degradations or to state F due to random shock as given in figure 3.

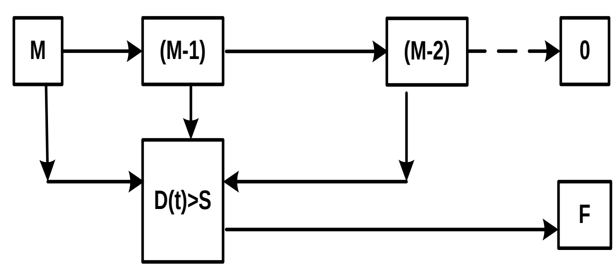


Fig. 3. System states diagram subject to two failures processes where M, (M-1), are degradation states,O is a degradation failure state and F is a catastrophic failure state.

Let us consider a repairable system connected to a load, where the system available capacity (SAC) and hourly system load (HSL) are shown in figure 4 [11].

The behavior of SAC curve shows that the generating system follows several states. These states can involve partial or total failure of a simple unit or of several units. The appearance of dips in the same curve reflects units' breakdowns and the resumption to the initial level of capacity indicates that repairs were made. The shaded area under the curve indicates the energy not supplied (ENS) and their corresponding time intervals denote durations where

the consumption exceeded the production. They learn about the times when the expected production is actually not available in its entirety. If the study period is 24 hours, we will talk about the unavailability of the system for about 3 hours. Each decreasing in the SAC curve behavior

ISBN: 978-1-61804-230-9

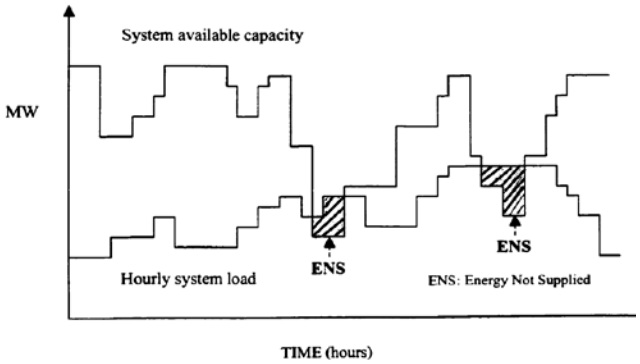


Fig. 4. Superimposition of the system available capacity and the load model

corresponds to a degrading state, subsequently to a decreasing level of system reliability. Load points are linked to supply resources by electric lines that have their own physic characteristics. The increasing of demand implies: the increasing of the current transit, the decreasing of voltage level, the increasing of reactive power consumption and etc. After the activation of system defense (compensators), regarding physic characteristics of lines, when thresholds values are reached, the protections operate and isolate the line. If the studied system is in looped configuration (supposed more reliable and more flexible in faults conditions), there is a load transfer to another line which at its turn becomes loaded and the line opening's scenario is repeated leading to cascade degradation in the lines of the system. The final results become the loss of load. Another scenario is probable; it is the luck of coordination between the items in the system defense. The problem stay in the physics category and the system loading can deal to voltage and frequency collapses. The persistence of the phenomena during a lap's time causes generating units stopping or stall. These are scenarios of several blackouts. When a failure occurs in generating units, the standby units are connected and activated. If their contribution is insufficient, the supply becomes lower than the consumption. When the polar angle of the engine reaches a certain threshold, they drop and consequently we have loss of load independently on the reliability of the connection between the supply system and the load point.

IV. Restoration processes modeling

A. *Case of a conventional system*

The network reconstruction after the blackout is modeled using Markov chain method as given in figure 5 where the degradation states are inspired from figure 2. Each state corresponds to a special structure of the network and to particular values of voltage and frequency.

As it is not easy to master the occurrence of the blackout in the current state of the network, we have attempted to model the network restoration while suggesting the contribution of smart systems in the acceleration of the service restoration process. To this end, we have studied two cases occurred in year 2003, namely the Algerian and the Italian blackouts. A particular interest was paid to the following parameters, such as: the restoration rate, the partial power restored at each stage and finaly the cumulative power restored at each stage. The tables 1 and 2 were filled following the stages of the figure 5, where the R_i are the restoration states with i=6 for the Algerian case and i=5 for the Italian one. We difined the used parameters as: $\rho_i, i = 1, n$ are degradation transition rates, and

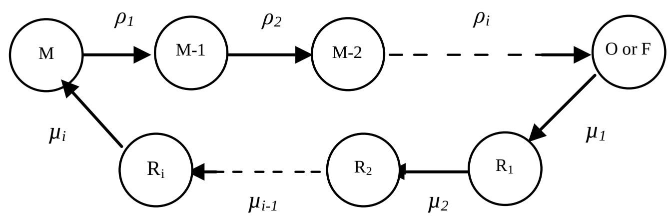


Fig. 5. Transition state diagramm: degradation tripping vs. restoration

$\mu_i, i = 1, m$ are restoration rates.

In a general manner, we have observed an inertia in the restoration processes of the networks but it is more apparent in the Algerian case. The first information about the results show a lack of effective measures for a fast restoration. To highlight the attributes of smart systems integration regarding the three countries affected by the blackouts of 2003, we have developed a comparison presented in table 3. We deduce that the order of effectiveness is decreasing according to the classification as follows: USA, Italy and Algeria because the USA had firstly integrated the smart system in their conventional network to increase the reliability and the security of supply. The Italian network behavior is acceptable, however Algeria, can take this opportunity to learn from the two experiences.

The main objective of the restoration service is to minimize the number of customers faced with the

TABLE I
STATE RESTORATION IN CASE OF 2003 ALGERIAN BLACKOUT

States	0	1	2	3	4	5
$\mu(s^{-1})$	0.15e-4	5.0e-4	4.2e-4	2e-4	2.2e-4	2.7e-4
Restored power (%)	15	15	30	13.6	23.3	3.4
Total restored power (%)	15	30	60	73.3	96.6	100

TABLE II
STATE RESTORATION IN CASE OF 2003 ITALIAN BLACKOUT

States	0	1	2	3	4
$\mu(s^{-1})$	4.0e-5	7.9e-5	6.9e-5	5.0e-5	6.0e-5
Restored power (%)	5	40	25	20	10
Total restored power (%)	5	45	70	90	100

TABLE III
PARAMETERS COMPARISON BETWEEN ALGERIAN, ITALIAN AND USA BLACKOUTS

	Algeria	Italy	USA
Power restored(GW)	5.003	27	61.8
Total restoration time(s)	15960	43200	104400

interruption of power delivery by transferring them to support feeders via network reconfiguration, with respect to components operational constraints. The reaction time is a pertinent factor to take into account where disconnected areas should be restored as quickly as possible. This scenario could be considered in the case of the integration of smart grid.

B. Case of a smarter system

It will be better to simulate events which will occur in the case where smart systems are integrated in the power grid. Three scenarios are discussed in the following.

1) Scenario 01: the initiating event is the peak demand

The smart grid concept uses smart metering which is designed to manage consumption used at peak times by encouraging more off peak power by households and small businesses, therefore shifting the load. Most outage management system and distribution management system (OMS/DMS) analyze and optimize network performance and reactively determine outage locations. Smart grid algorithms that incorporate spatial analysis will be part of a decision support system that can help determining risk and potential customer impact and recommend preventive measures by integrating real-time weather monitoring system (WMS). Note that this scenario is similar to the 2003, USA blackout.

2) Scenario 02: the event has already happened due to a loss of a component

This scenario is similar to the Algerian blackout (2003). The loss of generator coupled to a period of peak demand lead to a cascading event and finishes to a blackout. Smart grid via sensors and intelligent devices could avoid this undesired event by:

-Integrating more renewable energy power sources to the power transmission and distribution systems. They will relieve stress by adjusting automatically their operation. Then, smart grid could switch to solar and wind mode (or energy storage) to mitigate peaks demand.

-Deciding to shed appropriate system load, by temporarily switching off distribution of energy to different geographical area proportional to the severity of power system disturbance.

3) Scenario 03: the system is in a state O or F of figure 3.

Smart grid coordinates units, loads, transmission system, and their associated characteristics to a fast restoration of power to consumer by establishing priorities. By using location intelligence capabilities, it quickly diagnoses outages and determine the location of a fault caused by physical damage of the transmission and distribution facilities due to weather by measuring the optical distance along the fiber.

V. CREATIVE THINKING FOR FUTURE WORK

Creativity is essentially the act of an individual. It can come solely from within or it can be generated by a team working together in such a way that a team consciousness arises to which the individual members constitute [12]-[13].

A. Presentation of problem

In times of peak demand, in the case of insufficient of power availability in a region or in a locality, in the case of fear that automation does not manage the power flow and to be in a blackout situation, the electricity distribution company shall do shedding of a part of the load. The relevant question is which entity will be relieved?

Firstly, in such situations, contrary to popular belief where the priority of the availability of electrical power supplied by a grid was given to hospitals, industrial production units and authorities, we present an alternative management vision by reviewing in depth the assignment order of these priorities.

B) Proposition

It escapes to no one that earlier cited entities are equipped with backup sources during the design of the projects , therefore from moral point of view , we see that the high priorities are given to those who are not in need. From an economic point of view, it is necessary to help those who do not have a secure delivery. The question is how to manage load shedding with the emergency source?

- It is imperative that the energy distributor has an updated file of customers equipped with emergency sources.
- The network should be more communicative, and in real time.

ISBN: 978-1-61804-230-9

- Load shedding and network automation must be functional. The harmonics are often seen as a problem in the provided power quality. However, nowadays, it is a boon in the field of communication and data transfer. This is to address the notion of power lines, which are already operating in the field of television and internet in home networks. Regarding smart management of electrical distribution networks, we proposed the concept managing the schedules of openings and closures of protections at the upstream of the customer installation by the energy dispatchers from one hand and enabling connection of decentralized renewable energy resources to the transmission or to the production electricity networks. This solution can help to mitigate blackouts by injecting power in time when needed instantaneously.

VI. Conclusion and discussions

In a grid, even if the transportation and distribution parts are highly reliable, if the production units fail, the whole system collapses. When the load demand exceeds the production capacity, there is loss of load. To determine its proportion of time, we use loss of load probability model. A smarter power grid could be the solution to these woes. This new technologies highlight the following features: Ability to perform forecast peak demand and to ensure its management, anticipation of the start of the emergency; risk assessment of equipment failures; management of shedding their workforce at the appropriate times and select the consumer prior to relieve via power lines concept. This new technologies and concepts can significantly reduce barriers to the integration of renewable resources. It aims to build smart renewable-energy generation using micro-grids to enable houses, buildings, and villages to be energy self-sufficient. In the literature, it gave us to see the lack of research works in stochastic modeling of the occurrence of blackouts. To highlight our competence in the forecast, we introduced other opportunities of modeling using the concept of Box and Jenkins. We were able to prove the dominance of ARMA models in the field of energy distribution. Also in this context, the integration of smart system into the conventional one can help to exploit the weather forecasts to predict changes in consumer demand and anticipate on the main technical and organization measures. As it is necessary to consider the decentralization of the production units and encourage maintenance actions on the clean energy production units considered obsolete. Another aspect that is relevant is the provision of emergency sources of energy by the distributor and not by the customer. It was demonstrated that this decision contribute to avoid moral prejudice often felt by the customer and the financial one often sustained and supported par the distributor of the energy. Finally, the studies done on individual cases can be generalized to all entities throughout the national territory.

VII. References

[1] N.C.Miller and J.F.Green, “A Model Describing Acceleration-Induced Blackout”; Annals of biomedical engineering 2, 58-68 (1974)

[2] F. Iberraken, R. Medjoudj and D. Aissani: “Power system reliability modeling and decision making for quality of service improvement under smart system integration and renewable resources insertion”. In the proceedings of the international workshop; performance evaluation and quality of service (EPQoS-2013), Bejaia, May, 05, 06, 2013.

[3] D. L. Ha, S. Ploix, E. Zamai, and M. Jacomino ‘’Prévention des blackouts grâce au système de gestion d’énergie intelligent du bâtiment,’’, INP Grenoble, UJF, CNRS, 2005.

[4] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, “Complex Systems Analysis of Series of Blackouts: Cascading Failure, Critical Points, and Self-organization,’’ CHAOS, 17(2), June 2007.

[5] I.Karamitsos and K. Orfanidis; “An Analysis of Blackouts for Electric Power Transmission Systems”, PWASET, volume 12, March 2006, ISSN 1307-6884.

[6] B. A. Carreras, D. E. Newman, I. Dobson and A. B. Poole, “Evidence for Self-Organized Criticality in Electric Power System Blackouts”, Hawaii International Conference on System Sciences, January 2001, Maui, Hawaii.

[7] T. J. Overbye and D. A. Wiegmann; “Reducing the risk of major blackouts through improved power system visualization”; 15th PSCC, Liege, 22-26 August 2005.

[8] IEEE PES CAMS Task Force on Understanding, Prediction, Mitigation and Restoration of Cascading Failures; IEEE power engineering society general meeting, Pittsburgh PA, USA, July 2008.

[9] R. Medjoudj, D. Aissani and K.D.Haim, Power customer satisfaction and profitability analysis using multi-criteria decision making methods. International Journal of Electrical Power and Energy Systems (Elsevier), 45 (1) (2013) pp:331-339

[10] G.E.P Box and G.M.Jenkins, “Time series analysis: forecasting and control, revised edition. Hlden Day, 1976.

[11] J. D. Kuech, B. J. Kirby, P. N. Overholt and L.C. Markel. Measurement practices for reliability and power quality. U.S Department of energy, ORNL/TM-2004/91.

[12] T. L. Saaty and K.Peniwati. ‘’Group decision making: Drowing out and reconciling differences’’. Edited by RWS Publications, Pittsburgh, USA. 2008

[13] F. De Felice, A. Petrillo, ‘’Multicriteria approach for process modelling in strategic environmental management planning”. International Journal of Simulation and Process Modelling, Volume 8, Issue 1, 2013, Pages 6-16.

BIOGRAPHIE

Fairouz Iberraken,is with electrical engineering department, university of Bejaia, LaMOS research Unit, targua ouzemour, 06000, Algeria, email: fairouz.ib87@gmail.com, phone: 00213559054379.

Rabah Medjoudj is with electrical engineering department, university of Bejaia, targua ouzemour, 06000, Algeria, email: r.medjoudj66@gmail.com..

Djamil Aissani is with operational research department, University of Bejaia, targua ouzemour, 06000, Algeria. Email: lamos_bejaia@hotmail.com..

Klaus Dieter Haim is with Applied Sciences of Technology, university of Zittau, Theodor korner, Allee 16, Germany. Email: kd.haim@hszg.de

A Mathematical approach for the reference contingent problem

D. Barilla - G. Caristi - D. Marino - A.- Puglisi and E. Saitta
University of Messina
Department S.E.A.M
Via dei Verdi, 75, 98122- Messina
Italy
dbarilla@unime.it
gcaristi@unime.it
prof.domenicomarino@gmail.com
puglisia@unime.it
esaitta@unime.it

Abstract - Behavioural finance has made it possible to discover and, partly, to explain certain types of behaviour by investors which does not conform to the principles of rationality. The most important ones are the disposition effect (Shefrin and Statman, 1985) and the equity premium puzzle (Mehra and Prescott, 1985). By following this approach, consumer psychology has allowed us to study some classical phenomena of preference inconsistency (attraction and compromise) and the effect of different methods of evaluation (for example, joint and separate evaluation; Hsee, Loewenstein, Blount and Bazerman, 2006). In this paper, the idea is to deal with such situations through analysis of multi objective maximization consisting of reference point approaches.

Keywords: Utility Theory, Optimization Theory, Reference Point, Pareto minimum point

I. Introduction

The standard consumer theory, which considers supply as how the amount of a good X changes with varying of the price P_X preserving constant income R and the prices of other good, is a function expressed in terms of wealth and prices, but does not depend on the endowment composition. Thaler, in his work of 1980, first coined the term "endowment effect" to describe the experimental discovery that individuals prefer one object to another if this is part of their endowment.

Thus the "endowment effect" can be defined as the tendency of people to give a different value to the objects that they own from those that they don't own, or from the objects that they want to buy. From an economic point of view, it seems to be a quite rational strategy to ask for more money to sell an object than to buy it. Therefore, sellers should try to get as much profit as possible, while sellers try to get the best price. However, research has shown that this type of behavior is not really a conscious strategy followed by buyers and sellers but, rather, an automatic psychological reaction caused by owning a certain object. It is surprising that this phenomenon can be also be induced by objects that are presented to the participants in an experiment just a few moments before asking them at what price they would be willing to sell them. This experience of "minimum possession" also has an effect that induces those who have received an object to ask a higher price than what those who have not received it are willing to pay. This does not depend on affective issues or on memories of particular experiences using that object, since it was given to the participants only a few moments earlier. Behavioural finance has made it possible to discover and, partly, to explain certain types of behaviour by investors which does not conform to the principles of rationality. The most important ones are the disposition effect (Shefrin and Statman, 1985) and the equity premium puzzle (Mehra and Prescott, 1985). By following this approach, consumer psychology has allowed us to study some classical phenomena of preference inconsistency (attraction and compromise) and the effect of different methods of evaluation (for example, joint and separate evaluation; Hsee, Loewenstein, Blount and Bazerman, 2006). In this perspective, the idea is to deal with such situations through analysis of multi-objective maximization consisting of "reference point" approaches.

II. Reference Point

The reference point, which technically corresponds to the initial endowment of objects, can be influenced by other factors, such as "their aspirations, expectations, the behaviour of other agents and social norms that regulate community life".

The fundamental intuition of some authors is that our perceptual apparatus has become accustomed to the evaluation of variations rather than to the evaluation of absolute magnitudes.

When we respond to attributes such as luminosity, noise, and temperature past and present experiences define the adaptation level or reference point, and motivations are perceived in relation to this reference.

Since Kahneman and Tversky (1979) it has been well-known that people make choices between pairs of lotteries and that these choices are reversed when the equivalent pairs are framed in the domain of losses rather than gains. Faced with the choice between a certain gain and a risky gain, people are risk adverse and choose the certain gain.

At the moment of choosing between a certain loss and a risky loss, however, they become risk-keepers and they take a risky bet to avoid loss.

Thus, the prospects are not evaluated on the basis of their impact on the decision in reference to the total wealth of the player (as they would be according to the now outdated theory of expected utility), but on the basis of a certain reference point, often a "status quo" which is created at the moment in which the decision is made.

Bardsley (2008) extends the approach to a "dictatorship" type game, in such a way as to allow donors to offer money to beneficiaries.

This manipulation should not affect those individuals who have donated money in conventional dictator-games and therefore the percentage of donors should not change. In experiments, however, a significantly lower number of donors give money if keeping everything is no longer the most selfish thing to do.

Framing effects have been identified in social dilemma games and literature also contains studies on the effects of framing in anti-social preferences, which we refer to in the sense of pleasure derived from the reduction of someone else's welfare.

In order to better understand the analysis of the context in which we intend to operate, it is necessary to introduce the fundamental concepts already found in literature, including the very important "theory of welfare".

The theory of welfare economicsS (1930-1950) studies the operation of an economy of production and exchange, asking the question what should be the optimal configuration of an economic system, in which there are multiple heterogeneous individuals, with different systems of preferences and different initial factor endowments, such as the ability to work and capital, and of commodities.

Its purpose is to define a social optium regardless of the description of the institutional context (decentralized economy or planned). Even in this case a model is constructed in which the essential aspect is the process of social welfare maximization. Welfare economics is based on certain widely accepted general assumptions which in fact constitute value judgments (also including ethical evaluations which cannot therefore be attributed merely to the concept of rationality):

1. individualistic philosophical view: individuals are rational and the best judges of themselves ;
2. non-organicist vision of society: the State exists but is not an independent source of value. Its will depends on the aggregation of the wills of the individuals belonging to it;
3. the principle of efficiency is accepted, according to which the production, exchange and distribution of commodities conforms to the criterion on the basis of which a reallocation of resources that improves the welfare of an individual without causing harm to others is an improvement in the welfare of society (Pareto principle).

The fundamental problem is that of defining a social optimum, which means maximizing community welfare, and to do this we need a social welfare function.

The parameters on which social welfare depends must be established: in general, social welfare is defined in terms of individual utilities.

What does defining a situation of social optimum mean? Mixed response; optimum for whom? It is necessary to introduce the concept of a social welfare function.

$$W = W(U_1, U_2)$$

It measures individuals' welfare only on the basis of their utility without any interdependence (there is no idea of a state commodity in itself).

Its definition is needed in order to establish the priority of economic policy options.

Individual utility is measured by the amount of goods and services which individuals have at their disposition and which they can enjoy.

$$U_i = U_i(X_{i1}, X_{i2}) \qquad i = 1, 2$$

There is no interdependence in the welfare of two individuals (selfish point of view).

If such a function exists: given certain initial endowments of factors and commodities, given certain production techniques, how should the production and distribution of goods between different individuals in society be organized in order to achieve maximum social welfare?

First important conclusion:

Without the need to characterize the institutional context, it is possible to define certain conditions that must be met in order to achieve maximum collective welfare. We conside

- two individuals: 1 and 2
- two goods: X_1 and X_2
- two factors: L and K

The problem is to maximize social welfare, taking into account the constraints consisting of individual preferences, production techniques and endowments and entitlements on inputs:

$$MaxW = W(U_1, U_2)$$

st.

$$\left.\begin{array}{l} U_1 = U_1\left(X_{11}, X_{12}\right) \\ U_2 = U_2\left(X_{21}, X_{22}\right) \end{array}\right\} \text{contraint preferences}$$

st.

$$\left.\begin{array}{l} X_1 = X_1\left(K_1, L_1\right) \\ X_2 = X_2\left(K_2, L_2\right) \end{array}\right\} \text{contraint preferences}$$

st.

$$\left.\begin{array}{l} K_1 + K_2 = K \\ L_1 + L_2 = L \end{array}\right\} \text{contraint preferences}$$

and

$$\begin{array}{l} X_{11} + X_{21} = X_1 \\ X_{12} + X_{22} = X_2 \end{array}$$

For this maximisation it is possible to identify two sets

A) the conditions for the social welfare function (distribution)

B) the conditions for Pareto efficiency in an optimum situation:

1. the marginal rate of substitution between goods X1 and X2 of individual 1 will be equal to the marginal rate of substitution for individual 2; exchange efficiency
2. the marginal rates of technical transformation of inputs K and L are the same in the production of both goods; production efficiency
3. for the system as a whole, the marginal rate of transformation calculated on the production frontier is equal to the substitution marginal rate for the two individuals. Overall efficiency – in product composition.

When these conditions are fulfilled we are in the presence of a Pareto efficient situation.

Finally, in order to explore the influence of context, we start with a simple manipulation of reformulation tasks, the transition from the gains domain to the losses domain. This seems a good starting point, because, in the first place, it is a prominent factor known to be influential in other areas, and, secondly, it is an effect of pure presentation. No difference in the material, such as the effort put into earning endowment or a different game history, can plausibly affect behaviour.

Applying this idea to the analysis of individual economic behaviour, these authors propose a theory in which the preferences with respect to consumer goods are defined in terms of changes, rather than their final levels, and satisfy two fundamental behavioural assumptions:

a) aversion to losses: loss, or a variation in commodities that produces a distancing from the initial or reference endowment, produces a larger variation in welfare, in absolute terms, than that generated by a gain of equal dimension;

b) decreasing sensitivity: the marginal value of both gains and losses decreases as their distance from the initial endowment increases.

Reference point approaches aim at the definition of an ideal point. The ideal point is defined as a vector whose elements are the maximum values of the individual criteria functions. The best alternative is the one by which there is greater proximity between the criteria values of the alternative itself and the values of the ideal point, i.e. the compromise solution is defined as the alternative, in the set of efficient solutions, by which the distance from the ideal solution is minimal.

An efficient solution (Pareto solution) is a solution for which the value of a criterion (or objective) cannot be improved without reducing the value of a contrasting criterion (or objective).

A reference dependent utility problem can, therefore, occur in general in the form of an inconsistent linear system.

III. Preliminaries

In 1973 ([2]), I. A. Marusciac defined a class of extremal approximate solution of a linear inconsistent system that contains as particular cases the least square solution and the Tschebychev's best approximation solution of the system, the two main methods used to obtain an approximate solution of an inconsistent system. The least squares method was applied by M. Fekete and J.M Walsh in 1951 ([1]) in order to obtain an approximate solution of an inconsistent system, whereas the Tschebychev's best approximation method was used for the same reason by R.L. Remez in 1969 ([3]). First of all we will start with some definitions and known results.
Let consider the following system of m equations and n unknowns:

$$f_k(z) = \sum_{j=1}^{n} a_{kj} z - b_j = 0, \quad k \in M = \{1, 2, \ldots, m\} \qquad (1)$$

We denote by $S_0(A, b)$, $A_0(A, b)$ and, respectively $A_0^*(A, b)$ the subsets of infrasolutions, Pareto minimum solutions, respectively weak Pareto minimum solutions z of the system (1)

Lemma 0.1 *The system (1) is consistent if and only if every solution z of the system is also an infrasolution, **i.e.** $S(A, b)$ coincides with the set of all solutions of (1).*

Definition 0.1 *$z \in^n$ is a Pareto minimum solution or Pareto minimum point of the system (1) if there is no $u \in^n$ such that:*

- $|f_k(u)| \leq |f_k(z)|$ *for all $k \in M$.*
- *There is a $k_0 \in M$ so that $|f_{k_0}(u)| < |f_{k_0}(z)|$*

Definition 0.2 *$z \in^n$ is called a weak Pareto minimum solution of the system (1) if there is no $u \in^n$ such that*

$$|f_k(u)| < |f_k(z)| \qquad \text{for all } k \in M.$$

IV. Main Results

Considering the following minimization problem:

$$(*) \quad (u_1, u_2, \ldots, u_m) \to \min$$

under the conditions:

$$|f_k(z)| \leq u_k, \quad k \in M, \quad z \in^n, \quad u \in_+^m$$

Using this problem, we have the following characterization of the Pareto minimum solutions of system (1):

Theorem 0.1 *$z_0 \in A(A, b)$ if and only if (z_0, u_0) is a Pareto minimum solution to the problem defined above, where $u_0 = (|f_1(z_0)|, |f_2(z_0)|, \ldots, |f_m(z_0)|) \in_+^m$.*

Proof Let $z_0 \in A(A, b)$. Assume that (z_0, u_0) (with u_0 defined above) is not a Pareto minimum solution for the defined problem. Then, there exists $(z, u) \in^n \times_+^m$ so that $|f_k(z)| \leq u_k$ for all $k \in M$ and $u < u_0$ (i.e. $u \neq u_0$). It follows tat there is a $k_0 \in M$ such that $u_{k_0} < u_{0k_0}$ and, therefore:

$$|f_k(z)| \leq u_k \leq u_{0k} = |f_k(z_0)| \quad for \quad all \quad k \in M$$

$$|f_{k_0}(z)| \leq u_{k_0} < u_{0k_0}$$

which is a contradiction. Hence (z_0, u_0) is a Pareto minimum solution of the problem $(*)$.
Conversely, assume that $(z^*, u^*) \in^n \times_+^m$ is a Pareto minimum solution of the problem $(*)$. If $z_0 \notin A(A, b)$, then we can find $z_0 \in^n$ such that for all $k \in M$:

$$|f_k(z_0)| \leq |f_k(z^*)|$$

and there exists $k_0 \in M$ so that:

$$|f_{k_0}(z_0) < |f_{k_0}(z^*)|$$

Let now $u_0 = (|f_1(z_0)|, f_2(z_0)|, \ldots, f_m(z_0)|) \in_+^m$ Then, we have for all $k \in M$

$$|f_k(z_0)| = u_{0k} \leq u_k^*$$

and there is a $k_0 \in M$ so that:

$$|f_{k_0}(z_0)| \leq u_{0k_0} < u_{k_0}^*$$

Therefore, for all $k \in M$ we have $|f_k(z_0)| \leq u_{0k}$ and $u_0 \leq u^*$, $u_0 \neq u^*$. The contradiction shows that $z^* \in A(A, b)$

Remark 0.1 *It is obvious that in the real case the corresponding problem is linear. The problem of finding a Pareto minimum solution for the nonlinear vector-minimization problem $(*)$ is equivalent, by Theorem 4.1, with the problem of finding a Pareto minimum solution of a complex linear inconsistent system.*

Theorem 0.2 *Let (1) be an inconsistent system for which the matrix A has the "H-property". An approximate solution $z_0 \in^n$ with $f_k(z_0 \neq 0,\ k \in M$ of this system is a Pareto minimal solution of (1) if and only if there exists a subsystem of (1) with p equations $(n+1 \leq p \leq 2n+1)$ (we assume that $f_k(z) = 0,\ k = 1, 2, \ldots, p$) and there exist also p positive numbers d_k such that:*

$$z_{0k} = \frac{\sum d_{k_1} d_{k_2} \ldots d_{k_n} \bar{D}(k_1, k_2, \ldots, k_n; A) D(k_1, k_2, \ldots, k_n; A; b)}{\sum d_{k_1} d_{k_2} \ldots d_{k_n} |D(k_1, k_2, \ldots, k_n; A)|^2} \qquad (2)$$

where:

$$D(k_1, k_2, \ldots, k_n; A) := \begin{vmatrix} a_{k_1 1} & a_{k_1 2} & \ldots & a_{k_1 n} \\ a_{k_2 1} & a_{k_2 2} & \ldots & a_{k_2 n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k_n 1} & a_{k_n 2} & \ldots & a_{k_n n} \end{vmatrix}$$

and $D(k_1, k_2, \ldots, k_n; A; b)$ is the determinant obtained from $D(k_1, k_2, \ldots, k_n; A)$ by replacing the column $(a_{k_1 j} a_{k_2 j} \ldots a_{k_n j})^t$ with $(b_{k_1} b_{k_2} \ldots b_{k_n})^t$. $\bar{D}(k_1, k_2, \ldots, k_n; A)$ is the complex conjugate of $D(k_1, k_2, \ldots, k_n; A)$ and the sum $\sum$ is taken for all the values of $k_1, k_2, \ldots, k_n$ in the set $\{1, 2, \ldots, p\}$.
*If the matrix A is real (i.e. the system is real) and has the "H-property", by **Corolarry 4.1** we have that $p = n + 1$ and, hence, relations (2) becomes (for $j = 1, 2, \ldots, n$):*

$$z_{0j} := x_{0j} = \frac{\sum_{k=1}^{n+1} d_1 d_2 \ldots d_{n+1} D_k(A) D_{kj}(A; b)}{\sum_{k=1}^{n+1} d_1 \ldots d_{k-1} d_{k+1} \ldots d_{n+1} |D(A)|^2} \qquad (3)$$

where

$$D_k(A) = \begin{vmatrix} a_{k_1 1} & a_{k_1 2} & \ldots & a_{k_1 n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k-1,1} & a_{k-1,2} & \ldots & a_{k-1,n} \\ a_{k+1,1} & a_{k+1,2} & \ldots & a_{k+1,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \ldots & a_{n+1,n} \end{vmatrix}$$

and $D_{kj}(A; b)$ is obtained in the same way as above.

We can remark here that the equalities (2) can be written in a more simple form as:

$$z_{0j} = \sum \Lambda_{k_1 k_2 \ldots k_n} \frac{D_j(k_1, k_2, \ldots, k_n; A; b)}{D(k_1, k_2, \ldots, k_n; A)}, \ j = 1, 2, \ldots, n \qquad (4)$$

and

$$\Lambda_{k_1k_2\dots k_n} = \frac{d_{k_1}d_{k_2}\dots d_{k_n}|D(k_1,k_2,\dots,k_n;A)|^2}{\sum d_{k_1}d_{k_2}\dots d_{k_n}|D(k_1,k_2,\dots,k_n;A)|^2}$$

It is easy to see that $\Lambda_{k_1k_2\dots k_n} > 0$ and $\sum_{k=1}^{n+1}\Lambda_{k_1k_2\dots k_n} = 1$ and in the real case, the equalities (3) can be rewritten as:

$$x_{0j} = \sum_{k=1}^{n+1}\Lambda_k\frac{D_{k,j}(A;b)}{D_k(A)},\ j = 1,2,\dots,n$$

where $\Lambda_k > 0$ and $\sum\Lambda_k = 1$.

Theorem 0.3 *Let (1) be an inconsistent system. If the matrix A of the system has the "H-property", there exists a subsystem of (1),*

$$f_{k_j}(z) = 0,\ j = 1,2,\dots,p\ nd\ n+1 \le p \le 2n+1 \qquad (5)$$

so that:

$$A_0(A,b) = onv\{z^c\}$$

where $onv\{z^c\}$ is the convex hull of all Cramer solutions z_0 of the nn subsystems of (5)

Proof For $z_0 \in A_0(A,b)$, if we apply Theorem 4.2, we conclude that relation (4) is true. Let $\Upsilon \in^n$, $\Upsilon = (\Upsilon_1,\Upsilon_2,\dots,\Upsilon_n)$ be a Cramer solution of the system $f_{k_j} = 0$ $(j = 1,2,\dots,n)$), i.e.:

$$\Upsilon_j = W_j(k_1,k_2,\dots k_n;A;b) := \frac{D_j(k_1,k_2,\dots,k_n;A;b)}{D(k_1,k_2,\dots,k_n;A)},$$

$$j = 1,2,\dots,n$$

The system $f_{k_j} = 0$ $(j = 1,2,\dots,n)$) is a subsystem of the system (13) and we have also:

$$z_{0j} = \sum\Lambda_{k_1,k_2,\dots,k_n}W_j(k_1,k_2,\dots k_n;A;b)$$

so that

$$z_0 = \sum\Lambda_{k_1,k_2,\dots,k_n}W(k_1,k_2,\dots k_n;A;b)$$

where

$$W(k_1,k_2,\dots k_n;A;b) :=$$
$$(W_1(k_1,k_2,\dots k_n;A;b),W_2(k_1,k_2,\dots k_n;A;b),\dots,$$
$$\dots,W_n(k_1,k_2,\dots k_n;A;b))$$

and thus, $z_0 \in onv\{z^c\}$. Conversely, if $z_0 \in onv\{z^c\}$, then z_0 is an open convex combination of Cramer solutions $W(k_1,k_2,\dots k_n;A;b)$ of subsystems having the form $f_{k_j} = 0$ $(j = 1,2,\dots,n)$). It follows that (4) holds, and thus, by Theorem 4.2, $z_0 \in A_0(A,b)$.

In [1] we saw that the least square solution and the best approximate solution of the system (1) (with A having the "H-property") are in fact two special cases of the Pareto minimal solutions of the same system. In other terms, using the notations defined above, this means that for a certain choice of the weights $\Lambda_{k_1,k_2,\dots,k_n}$ in the formula (4) we can obtain the two mentioned approximate solutions. We will look now on the least square solutions of the system (1), characterizing them.

Obtaining the least square solution of the system (1) means minimizing the function

$$(f) = \sum_{k=1}^{m}f_k(z)^2 = \sum_{k=1}^{m}f_k(z)\bar{f}_k(z)$$

Hence, here we have to solve the system:

$$\frac{\partial f}{\partial\bar{z}_j} = 2\sum_{k=1}^{m}f_k(z)\bar{a}_{kj} = 0,\ j = 1,2,\dots,n \qquad (6)$$

But system (6) is similar to system with $d_1 = d_2 = \dots = d_m = 1$ and we have:

Theorem 0.4 *If the matrix A of the system (1) has the "H-property", then $z_0 \in A_0(A,b)$ is the least square solution for the system (1) if and only if $d_1 = d_2 = \dots = d_m = 1$ that means that:*

$$z_{0j} = \frac{\bar{D}(k_1,k_2,\dots,k_n;A)D_j(k_1,k_2,\dots,k_n;A;b)}{|D(k_1,k_2,\dots,k_n;A)|^2},$$

$$j = 1,2,\dots,n \qquad (7)$$

In a similar way it is possible to characterize the best approximation solution of the system (1). Formula (7) is a synthetic form of the least square solution of an inconsistent system.

References:

[1] G. Caristi, *About Pareto Minimum Points of a System of Equations*, Applied Mathematical Science, Vol. 5, 2011, no.8, pp 351-358.

[2] M. Fekete, J.L. Walsh, *On restricted infrapolynomials*, Proceedings of the National Academy of Sciences, U.S.A., 37(1951), 95-103.

[3] I. Marusciac, *Infrapolynomials and Pareto optimization*, Mathematica (Cluj), 22(45), (1980), no. 2, 297-307.

[4] E.I. Remez, *Basis of the Numerical Methods in Tchebycheff's Approximation*, Izd. Kiev, 1969.

Effects of High-K Dielectric with Metal Gate for Electrical Characteristics of Nanostructured NMOS

Norani Bte Atan, Ibrahim Bin Ahmad, Burhanuddin Bin Yeop Majlis and Izzati Binti Ahmad Fauzi

Abstract**—**This paper presents a systematic study of various high-K materials on metal gate MOSFET for 18nm NMOS. From this study, we find the suitable combination materials between the high-K and metal gate, and how it is a good affected on the electrical characteristics of 18nm NMOS. The device shows a good improvement on I_{on}/I_{off} ratio, where the higher ratio that means this device it is suitable for low power applications. The virtually designed and fabricated of the devices was performed by using Athena module. While electrical characteristic performance was stimulated by using Atlas module of SILVACO software. Physical models used for simulation from Al_2O_3, HFO_2, and TiO_2 as gate dielectric with TiSix as metal gate, which provide the higher physical thickness and reduce the leakage current. Excellent dielectric properties such as high-K constant, low leakage current, threshold voltage and electrical characteristics were demonstrated. From the simulation result, it was shown that HfO2 is the best dielectric material with metal gate, TiSix.

Keywords**—**18 nm NMOS, High-K dielectric, Metal gate, Silvaco.

I. Introduction

New trend technologies, many industries have relied on a progression of smaller, denser, faster, cheaper and good quality of MOSFET. In order their main target is to reduce the production cost and at the same time can produce in big quantities of MOSFET. With increasing global competition, modern industries have to make their production process more efficient to compete. To do this, more advanced technologies have to be used. Scaling the MOSFET into nanometer regime, it is the best approach to solve many problems. It is a good and important challenging for future electronic technologies; refer to prospects of the scaling regime beyond 2011 ITRS [1].

Since MOSFET can be scaled down to smaller dimension which produce higher performance, at the same time gate length and oxide thickness also reduce. As the thickness scales of SiO_2 below 2nm, leakage currents due to tunneling increase drastically, leading to high power consumption and reduce device reliability. Therefore, replacing the SiO_2 with a high-K material allows increased gate capacitance [2]. The electrical characteristics of the device performance are analyzed with several of high-K materials and the gate oxide thickness is scaled to get same Equivalent Oxide Thickness (EOT). Recently, many researchers are focused on metal oxide materials with high-K values that have the ability to be integrated in MOSFET process flow. There are many high-K materials that are being studied nowadays such as Al_2O_3, HfO_2, and TiO_2 [3]. The best characteristics of gate dielectric should have high dielectric constant, large band gap with a favorable band alignment, low interface state density and good thermal stability. Among the high-K materials are compatible with silicon, and also materials have too low or high dielectric constant may not be adequate choice for alternative gate dielectric [4]. In this paper, we compare the electrical characteristics results for Titanium Silicide (TiSix) fabricated on Al_2O_3 (k~9), HfO_2 (k~25) and TiO_2 (k~85) gate dielectric. Thus the performance for all high-K materials with TiSix are explored and presented in the following section.

This work was supported in part by the Ministry of Higher Education Malaysia, The author is with the Centre of Micro and Nano Engineering (CeMNE), College of Engineering Universiti Tenaga Nasional (UNITEN),43009 Kajang Malaysia (e-mail: norani@ uniten.edu.my).

II. Materials and methods

This paper is presenting the fabrication of 18nm NMOS. The specification of the sample used in this experiment was p-type (boron doped) silicon substrate with doping concentration of $7x10^{14}$ atoms cm^{-3} and <100> orientation. The next procedure is to develop the P-well with growing a 200 Å oxide screen on the top of bulk silicon. This method using dry oxygen at high temperature 970°C followed by Boron as dopant with a dose of $3.75x10^{13}$ atoms/cm^3. The oxide layer as a mask was etched after the completed process of P-well doping. The next step was annealing process to ensure all boron atoms being spread uniformly in the wafer at 900°C with nitrogen and followed by 950°C with dry oxigen. The next step was preparing the isolate neighboring transistor or Shallow Trench Isolator (STR) of 130 Å thicknesses [5]. After that, the wafer was oxidized with dry oxygen for 25 minutes at 900°C. There were two important process involved in developing of STI such as Low Pressure Chemical Deposition (LPCVD) and Reactive Ion Etching (RIE). LPCVD process was applying with 1000 Å nitride layer was deposited on top of the STI oxide layer, followed by a photo resistor was deposition on the wafer. Then, RIE process was etching the unnecessary part on the top of STI area. Chemical Mechanical Polishing (CMP) was implementing to etched extra oxide on the wafer. STI was annealed for 15 min at 900°C. A sacrificial oxide layer was developed and etched to eliminate defects on the surface [6]. The process of STI was completed.

ISBN: 978-1-61804-230-9

Between the channel regions, the deposition of high-K materials (Al_2O_3, HfO_2, and TiO_2) process with gate oxide thickness is scaled so that they have the same EOT with SiO_2 by analyzed electrical characteristics of the device. The length of the high-K material was scaled to get 18nm same as the gate length of transistor. Then, the implantation dose of boron on the N-well active area for the threshold voltage (V_{th}) adjustment process. Next, the Titanium Silicide (TiSix) deposited on the top of high-K materials (Al_2O_3, HfO_2, and TiO_2) and followed by halo implantation with adjusted of indium dose to obtain the optimum value of the NMOS device [7,8]. The next process was formation of the sidewall spacer. It used as a mask for source and drain implantation. In this case, implantation with arsenic dose and followed by phosphor dose respectively. It was to ensure the smooth current flow in NMOS device [9].

The next step was development 0.5 μm layer of Boron Phosphor Silicate Glass (BPSG). This layer acted as the Pre Metal Dielectric (PMD) [10]. Again, annealing process was done on the wafer to strengthen the structure under temperature 950°C. The last process of wafer was compensation implantation which using the phosphor dose. From the above experiment, the dosages quantities for boron, indium, arsenic, and phosphor were different which based on the high-K materials (HfO_2, TiO_2 and Al_2O_3). The last step was deposition of aluminum layer as the metal contact for source and drain. Therefore, under ATHENA module the design of 18nm NMOS structure already completed. There are many factors influence as the input process parameters on the threshold voltage 18nm NMOS such as Gate oxide thickness, Substrate implant dose, Pocket-halo implant tilt angle, Gate oxide diffusion temperature, V_{th} implant dose, V_{th} implant energy, Pocket-halo implant energy, Pocket-halo implant dose, S/D implant dose, S/D implant energy, Compensation implant dose and Compensation implant energy. Now, we will proceed the stimulation process under ATLAS module to measure the electrical characteristic such as I_D versus V_{DS}, I_D versus V_{GS}, I_{on}, I_{off}, and V_{th}.

III. Results and Discussions

The complete fabrication of 18nm NMOS has been modeled and simulated successfully in Silvaco Simulink. Figure 1 shows the complete 18nm gate length NMOS. Figure 2 shows clearly on the doping profile of one of the design structure with gate length 18nm NMOS.

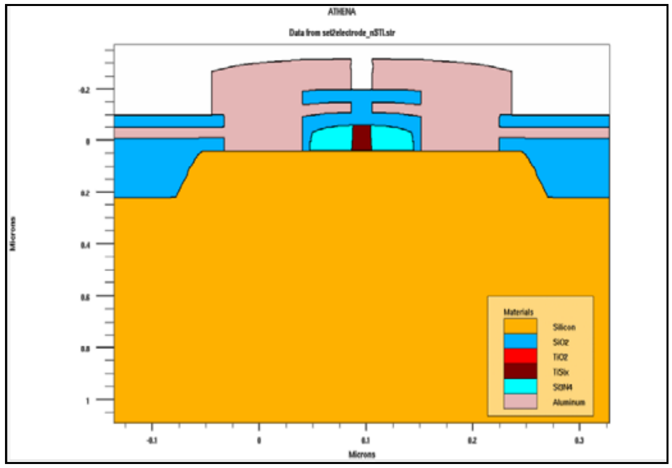


Figure 1. NMOS cross section after Fabrication Simulation

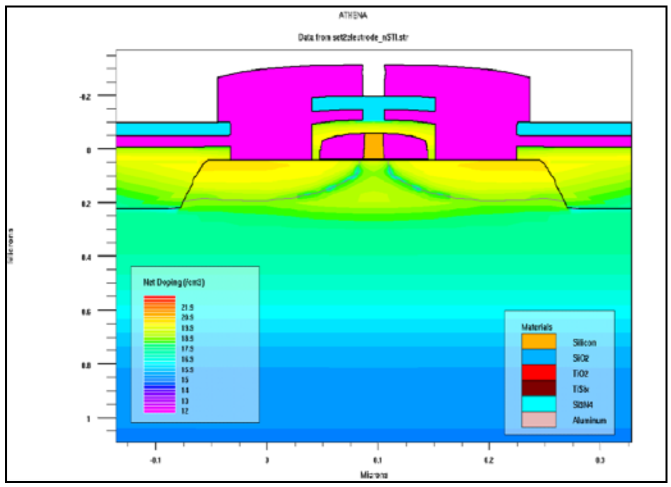


Figure 2. Doping profile of 18nm gate length NMOS with TiO2 dielectric and TiSix metal gate

Results of electrical characteristic simulation are obtained in Figure 3 for I_D –V_{DS} and Figure 4 for I_D – V_{GS} with different materials of high-K such as Al_2O_3 (k~9), HfO_2 (k~25) and TiO_2 (k~85). Voltage, V_{GS} = 2.6 volts is applied for I_D –V_{DS} graph with different voltage of V_{DS}. While, V_{DS} = 1.4 volts is supplied for I_D – V_{GS} graph with different V_{GS} voltage. The threshold voltage (V_{TH}), state on current (I_{on}) and state off current (I_{off}) can be extracted from I_D –V_{GS} curve.

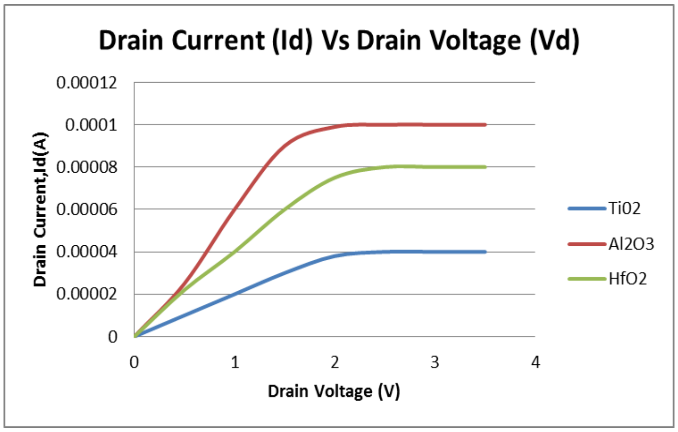


Figure 3. I_D Versus V_{DS} characteristic for different of high-K dielectric constants (k)

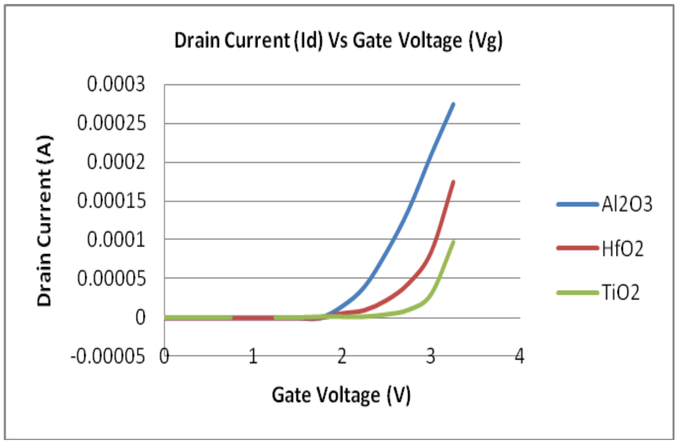


Figure 4. I_D Versus V_{GS} characteristic for different of high-K dielectric constants (k)

A good doping concentration is one of factor to ensure the transistor works well with fewer leakage current and enhance gate control [11]. There are four factors that influence in the threshold voltage counter measure such as Threshold voltage ajustment implant, Halo implant, Channel implant and Compensation implant. But for this research, only changing the various of dielectric material (Al_2O_3, HfO_2 and TiO_2) on the TiSix of transistor, so the Threshold voltage ajustment implant is best doping concentration to get a threshold voltage (V_{TH}) 0.302651 with 6.03036 x 10^{13} cm^{-2} boron for Al_2O_3. To maintain the same value of V_{TH}, due to the physically thicker dielectric layer, therefore the boron doping for HfO_2 and TiO2 were increased to 8.53256 x 10^{13} cm^{-2} and 9.73654 x 10^{13} cm^{-2} respectively. The increase of V_{TH} adjustment implant doping concentration was proportional with increasing the value of high-K dielectric, and at the same time the values of drain current (Ion) were decreased as shown in Figure 5.

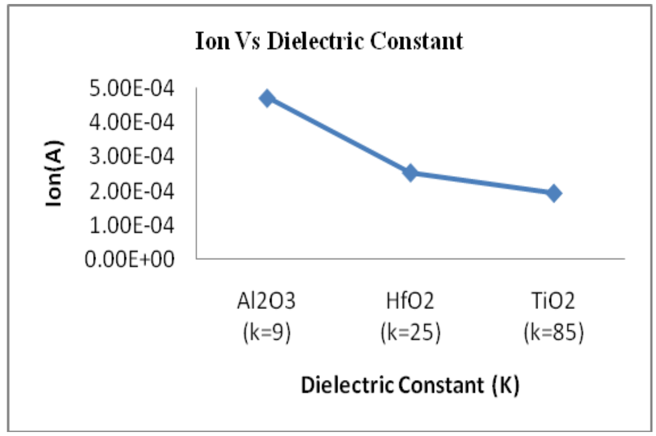


Figure 5. I_{on} current for Al_2O_3, HfO_2 and TiO_2 dielectric

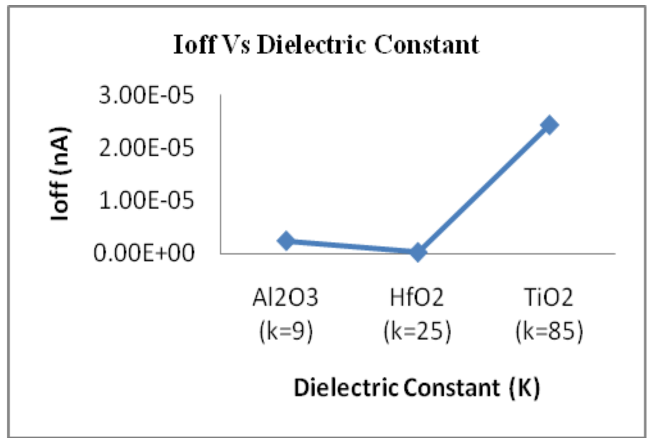


Figure 6. I_{off} current for Al_2O_3, HfO_2 and TiO_2 dielectric

Drain leakage current (I_{off}) or sub-threshold leakage current occurs when the gate voltage (V_{GS}) is lower than the threshold voltage (V_{TH}). In ideal case, when the transistor is turned off, V_{GS} = 0 volt and V_{DS} = V_{DD} (voltage supply), there is no current can through into the channel (I_{off} = 0). Refer to Figure 6, the leakage current for HfO_2 dielectric is lowest compared with Al_2O_3 and TiO_2 dielectrics. It is means, HfO_2 dielectric material most compatible with silicon and most stable oxide with the highest heat of formation [12,13].

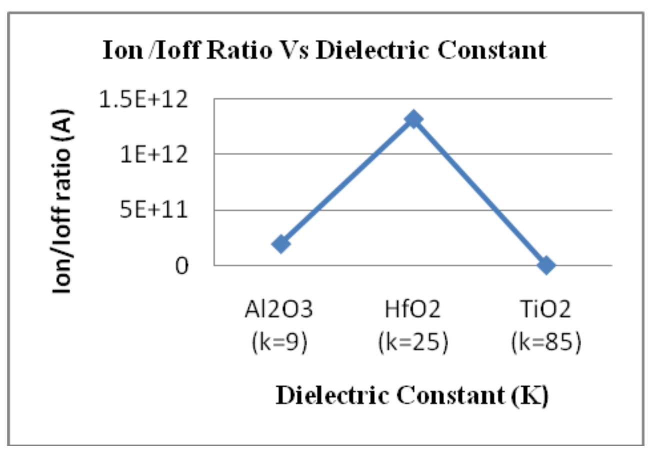


Figure 7. I_{on}/I_{off} current for Al_2O_3, HfO_2 and TiO_2 dielectric

Figure 7 shows the I_{on}/I_{off} current for different materials of high-K dielectric. HfO_2 dielectric gives the highest I_{on}/I_{off} ratio compared with Al_2O_3 and TiO_2 dielectrics. Hence, better performance of device can be obtained by using HfO_2 dielectric as gate oxide. This device is suitable for low power application [14].

Table 1. Simulated Results of Various Dielectric materials With ITRS 2011 Prediction

Parameter	**Simulation**			**ITRS 2011 Prediction**
	Al_2O_3	HfO_2	TiO_2	
V_{TH} (V)	0.302651	0.302651	0.302651	0.302
I_{on}	$4.7212x10^{-4}$	$2.5355x10^{-4}$	$1.9336x10^{-4}$	1.0×10^{-7}
I_{off}	$2.3652x10^{-15}$	$1.9123x10^{-16}$	$2.4316x10^{-14}$	$1.496 x10^{-6}$
$I_{on}/ I_{off\ ratio}$	$1.9961x10^{11}$	$1.3259x10^{12}$	$7.9519x10^{9}$	$6.6845x10^{-2}$

Table 1 shows the simulated results for Al_2O_3, HfO_2 and TiO_2 dielectric materials with TiSix as metal gate for 18nm NMOS. I_{on} results from the simulation are bigger value compared prediction value. While the simulation results for I_{off} are lower than prediction value. Therefore, all the above high-K materials are suitable with metal gate and compatible with silicon of transistor. But the best choice is HfO_2 as dielectric of transistor.

IV. Conclusion

NMOS structure with 18nm were successfully designed and stimulated to study the various of dielectric materials on metal gate of device performance. The performance of the three dielectric materials, Al_2O_3, HfO_2 and TiO_2 with TiSix as metal gate were compared and it was found HfO_2 is the best dielectric material for the future nano scale MOS devices technology. It based on the highest value of I_{on}/I_{off} ratio, and lowest value of sub-threshold leakage current (I_{off}). It is suitable for low power application.

Acknowledgment

The authors would like to thank to Ministry of Higher Education (MOE), Institute of Microengineering and Nanoelectronics (IMEN) Universiti Kebangsaan Malaysia (UKM), and Centre of Micro and Nano Engineering (CeMNE) Universiti Tenaga Nasional (UNITEN) for financial, facilities and moral throughout the project.

References

[1] ITRS 2011.www.ITRS2011.net.

[2] M.K.Bera and C.K. Maiti, "Electrical Properties of SiO_2/TiO_2 High-K Gate Dielectric Stack", Materials Science in Semiconductor Processing, pp.909-917, 2006.

[3] Hei Wong and Hiroshi Iwai, "On the Scaling Issues and High-k Replacement of Ultrathin Gate Dielectric for Nanoscale MOS Transistor," Microelectronic Engineering 83, pp.1867-1904, 2006.

[4] Byoung Hun Lee, Laegu Kang, We-Jie Qi, Renee Nieh, Yongjoo Jeon, Katsunori Onishi and Jack c.Lee, "Ultrathin Hafnium Oxide Low Leakage and Excellent Reliability for Alternative Gate Dielectric Application", IEEE Electron Device, pp. 133-136, Dec 1999.

[5] J.W. Sleight, I.Lauer, O.Dokumaci, D.M. Fried, D.Guo, B Harun, S Narasimha, C Sheraw, D Singh, M Steigerwalt, X Wang, P Olaiges, D Sadana, C Y Sung, W Haensch, M Khare, " Challenges and Opurtunities for High Performance 32nm CMOS Technology," IEDM Tech Digital, 2006.

[6] F.salahuddin, I.Ahmad, F.A Hamid and A.Zaharim, "Analyze and Optimize the Silicide Thickness in 45nm CMOS Technology Using Taguchi Metho," ICSE Proc.2010.

[7] H.A.Elgomati, B.Y.Majlis, I.Ahmad, F.Salahuddin ,F.A.Hamid,A.Zaharim, T.Ziad Mohamad and P.R.Apte, "Investigation of the Effect for 32nm PMOS Transistor and Optimizing Using Taguchi Method," Asian Jounal of Applied Science, 2011

[8] Shuan-Yuan Chen et al., "Using simulation to Characterize High performance 65nm Node Planar," International Symposium on Nano Science and Technology, Taiwan, 20-21 Nov 2004.

[9] F.salahuddin, I.Ahmad, and A.Zaharim, "Impact of Different Dose and Angle in Halo structure for NMOS Device," ICMST 2010, pp. 81-85, 26-28 Nov 2010.

[10] Sarcona G.T., M. Stewart, and M.K. Hatalis, " Polysilicon Thin Film Transistor Using Self-Aligned Cobalt and Nickel Silicide Source and Drain Contacts," IEEE Electron Device let, vol. 20, 1999

[11] Ashok K. Goel et. al., (1995), "Optimization of Device Performance Using Semiconductor TCAD Tools", Silvaco International Product Description, Silvaco International http://www.silvaco .com/products/descriptions.

[12] Hubbard, K.J., Schlom, D.G, "Thermodynamic stability of binary oxodes in contact with silicon", J.Mater.Res, 11, 2757-2776. 1996

[13] Chattererjee, S., Kuo, Y., Lu, J., Tewg, J.Y., Majhi, P., "Electrical reliability aspects of HfO2 high-K gate dielectric with TaN metal gate electrodes under constant voltage stress", Microelectronics Reliability, 46, 69-76.

Norani Bte Atan received her B.Eng.(Hons) in Electrical Engineering from University Teknology Mara (UITM) and MSc in Electrical and Electronic Engineering from Universiti Tenaga Nasional (UNITEN) in 1995 and 2008 respectively. She is currently pursuing the Ph.D degree in electrical engineering at Institute Microengineering and nanoelectronics (IMEN), Universiti Kebangsaan Malaysia (UKM). She has 7 years of working experience in the industries as a R&D design senior engineer at Matsushita Television and Network (M) Sdn. Bhd. Shah Alam, Selangor, Malaysia. She has involves in electrical design and radio frequency tuning for colour television production for international global exports. Responsible for the quality and cost reduction design for Panasonic TVs brand on various models, and to release them for mass production. Dealing with vendors for components and devices test, and conducts verification on TV performance based on customer's complaints and the up-to date technology requirement.

Her research interests include nano device integration and device modelling, Artificial Intelligence, Power Electronic and Power Quality. She is currently working at College of Engineering, Universiti Tenaga Nasional (Uniten), Malayasia as a lecturer. She is currently doing Ph.D. degree with Universiti Kebangsaan Malaysia (UKM), Malaysia. Her Ph.D research, which is supervised by Prof. Dr. Ibrahim Bin Ahmad and Prof Dato' Dr Burhanuddin Bin Yeop Majlis

She has been actively involved in research and short courses projects. She has completed various projects funded by MOSTI. She is currently an Engineer registered with the Board of Engineers Malaysia (BEM) and a Member of the Institute of Engineers Malaysia (IEM).

Ibrahim Ahmad received the B.Sc. degree in Physics in 1980 from Universiti Kebangsaan Malaysia (UKM). He received the M.Sc. degree in NuclearScience and Analytical Physics from UKM and University of Walesrespectively, in year of 1991 and 1992. He received the Ph.D. degree in Electrical, Electronic and System Engineering from Universiti KebangsaanMalaysia in 2007.

He was a Nuclear Science Officer at Nuclear Energy Unit (MINT) in charge of Radio scope production for medical and industry from1987 to 1992. From 1993 to 1996, he worked on Semiconductor Technology Division at Malaysian institute for Microelectronics Research Center & System (MIMOS), Kuala Lumpur. He joined the Department of Electrical, Electronic and System Engineering, University Kebangsaan Malaysia (UKM) as a lecturer in 1997 to 2002, and as Associate Professor from 2002 to 2007. He involved in several management and technical positions with MINT, MIMOS, Emisis Smartkom Sdn. Bhd. K.Lumpur, Bumi Hibiya Sdn. Bhd. K.Lumpur and UKM.

He is currently a Professor with the Department of Electronics and Communication Engineering, Universiti Tenaga Nasional, Malaysia. He published over 150 research papers in Journals and conferences. He is a senior member of the Institute of Electrical and Electronics Engineers (Senior MIEEE); Member of Institute of Physics Malaysia (MIPM) and Member of Malaysian Association of Solid States Science (MASS).

Burhanuddin Yeop Majlis is a professor of microelectronics and now the Director of Institute of Microengineering and Nanoelectronics(IMEN). He received the Ph.D degree from University of Durham, UK in 1988. He received the M.Sc. degree from University of Wales, UK in 1980, and B.Sc.(Hons) from UKM in 1979. He is senior member of the Institution of Electrical, Electronics Engineers (SMIEEE) and fellow of Malaysian Solid State Science and Technology Soceity (FMASS). He received Johan Mangku Negara (J.M.N.) from Yang di-Pertuan Agong in 2008. Received Darjah Dato' Paduka Mahkota Perak (D.P.M.P.) from the Sultan of Perak.

Izzati Binti Ahmad Fauzi is a under graduate student at Department of Electronics and Communication Engineering, Universiti Tenaga Nasional (UNITEN), Selangor, Malaysia. She is currently doing final year project under Centre of Micro and Nano Engineering (CeMNE), Universiti Tenaga Nasional supervised by Norani Bte Atan.

Studying Models issues on E-Commerce cashing

Nicolae Constantinescu and Oana Ticleanu, University of Craiova, str. A.I. Cuza no 13, Romania

Abstract—E-commerce is defined as any form of presenting, selling and marketing using Internet technologies. However, reasonable steps should be taken to protect the personal information and privacy of purchasers since such a system deals with personal data (used for identification) and credit cards' information. The most important problem that may appear when compromising the database is bank fraud. This article is an overview of security and privacy issues exploring the current state of e-commerce and, in the mean time, describing the today's most popular techniques which make such a transaction safer.

Keywords—e-commerce, e-cash, security issue, privacy, authentication.

The security issues for m-commerce are the same as for e-commerce since all the main differences between them are due to wireless technology. M-commerce can be considered as a subset of e-commerce. The advantages and disadvantages of m-commerce are the following [1]:

- **Advantages**
  - Ubiquity - the device used is mobile, so the user can access m-commerce in real time at any place
  - Security - mobile devices can offer a certain level of inherent security
  - Localization - user can be localized through GPS (Global Positioning System), GSM (Global System for Mobile) or UMTS (Universal Mobile Telecommunications System) network technology
  - Convenience - mobile devices are much smaller and lightweight
  - Personalization - since mobile devices are usually not shared between users they are adjustable to its user needs and wishes
- **Disadvantages**
  - Mobile devices provide limited capabilities
  - Mobile devices are heterogen leading to the integration of standard application platforms by the manufactures
  - Mobile devices are more prone to theft and destruction.

E-commerce popularity increases every day. In 2009, the number of transactions registered reached 1.2 million, twice the number in 2008. So there was registered a 75% growth of e-commerce reaching EUR 92.8 million, according to data from RomCard.

Clients expect online merchants to make payments a simple and secure process. In order to buy products (goods or services) clients must first create an account by filling up a form with personal data and credit cards' information. Most systems use account defined by usernames and a passwords. Most of the clients have confidence in online stores even if researchers have demonstrated many times that some sellers do not provide a high security level for the client's information.

This work was partially supported by the grant number 43C/2014, awarded in the internal grant competition of the University of Craiova and CNCS IDEI PCE 47/2011

TABLE I
TRANSACTIONS VALUE IN ASIA PACIFIC REGION (US BILLION)

	2006	2007	2008	2009	2010	2011
Australia	9.5	13.6	20.4	26.4	28.7	31.1
China (without Hong Kong)	2.4	3.8	6.4	11.1	16.9	24.1
India	0.8	1.2	1.9	2.8	4.1	5.6
Japan	36.8	43.7	56.6	69.9	80.0	90.0
South Korea	9.6	10.9	12.4	14.0	15.9	17.9
Asia-Pacific	59.1	73.3	97.7	124.1	145.5	168.7

The non-stop availability of online stores and the completeness of their inventory provide the possibility to compare products within or between stores, to access vendor return policies and to find warranty information, but also to read other users' opinions about the products.

Moreover, developing countries are given increased access to the global marketplace, where they compete with and complement the ones more developed. Most of them are already participating in e-commerce. To facilitate e-commerce growth in these countries the information infrastructure must be improved:

- High Internet access costs which include connection service fees, communication fees, and hosting charges for websites with sufficient bandwidth;
- Limited availability of credit cards and a nationwide credit card system;
- Underdeveloped transportation infrastructure which leads to slow and uncertain delivery of goods and services;
- Network security problems and insufficient security safeguards;
- Lack of skilled employees and key technologies;
- Content restriction on national security and other public policy grounds, which greatly affect business in the field of information services, such as the media and entertainment sectors;
- Cross-border issues, such as the recognition of transactions under laws of other countries, certification services, improvement of delivery methods and customs facilitation;
- Low cost labor leads to the fact that a shift to a financial intensive solution (including investments on the improvement and development of the physical and network infrastructure) is not visible.

In the following section we discuss how an e-commerce system works. Section II presents the security issues that must be considered in such systems. Section III provides information about attacks that may brake an e-commerce system and also the current methods used for avoiding these

ISBN: 978-1-61804-230-9

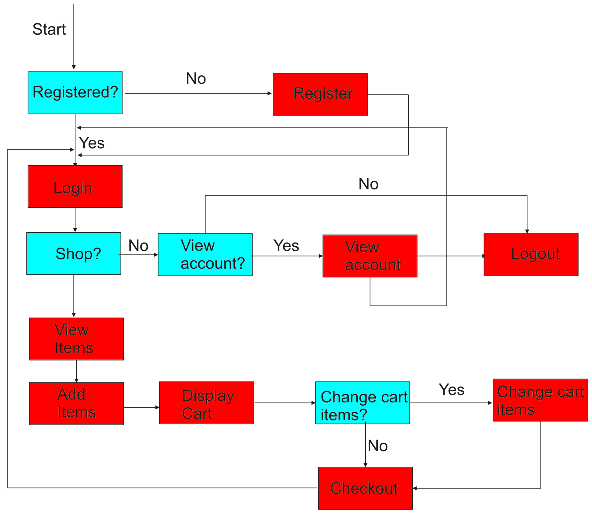


Fig. 1. Client's steps

attacks. We explain the term ”restrictive blind signature” which was first introduced by Brands in [2] together with an e-cash system, and the system he described which provides user anonymity. We also describe an attack agains e-coin payment systems based on blackmailing The last section presents our conclusions on these study.

I. Electronic commerce systems

A complete definition of e-commerce is given in [3]

> ”*E-commerce is the use of electronic communications and digital information processing technology in business transactions to create, transform, and redefine relationships for value creation between or among organizations, and between organizations and individuals.*”

Because of its advantages e-commerce has gain a semnificative growth, mainly in Asia Pacific Region. The transactions value are illustrated in table I

In a typical e-commerce system clients first visit a web site and browses a catalog with products (goods or services). After choosing at least one item clients can make a purchase. All the clients' available operations are presented in figure 1. An e-commerce system has three participants: the client who uses a browser to locate the site, the merchant (seller) who operates that site (usually the seller buys e-commerce software for his site from different software vendors), and the transaction partner.

Clients can be one of the following:

- People which have access to the Internet and disposable income, enabling widespread use of credit cards
- People who prefer purchasing products over Internet instead of physically buying them from regular stores.

The seller consists in:

1) A Web site with e-commerce capabilities
2) A corporate intranet so that orders can be purchased efficiently
3) IT employees for managing information and maintaining the system

Transaction partners can be:

1) Banking institutions that provide services like processing credit card payments and electronic financial transfers
2) National and international freight companies which are used to provide the transport of physical items. In B2C e-commerce these companies must offer cost-efficient services.
3) Authentication authority that is considered as being a trusted third party to ensure the integrity and security of the transactions.

There are sellers who use the freight companies as the transaction partner not only for transporting the products but also for physically transporting the money access. Such a process works like this: after the order is completed, a physical person (seller's employee) packs the product and contacts the delivery firm; when receiving the package the client pays it to the delivery firm which physically transports the money back to the seller.

This type of e-commerce is mostly preferred by the sellers who provide products only for their country. This is because the transportation of money is not a virtual process and can be very expensive if the distance is long. Since the client does not offer credit card information and the only personal information provided usually includes only the real name and the address – address that may not be the same with the one from the identity card of the client – where the packet must be delivered, such a system has a high security level.

Even so there some important issues that may appear. Most of them are affecting only the seller: when the package arrives, the client may not accept it (he changed his mind) and the delivery firm sends it back to the seller which must pay the delivery charges. The sellers who provide services and products for more than one country do not use this system because the delivery costs can be even higher than the package itself.

Further on, we consider a system where the money is accessed only virtually, and the third participant is a financial institution like a bank. The steps that a client follows to purchase one or more items are presented in figure 1. In order to login the client must be registered. The login operation can be considered as the authentication of the user. If he is not in the seller's database, he must register and then he can login. After the user logs in and purchases at least one item, the seller will receive the money through the electronic payment system used.

A. Electronic Payment

The most important part of an e-commerce system is the payment. Further on we will consider only the systems which imply an electronic payment, excluding the ones that use a delivery firm fro this step. The first mass mailing of credit cards was executed in 1960 in Fresno, California by BankAmerica. Then, in 1967, Westminister Bank installs first automated teller machine (ATM) at Victoria, London Branch in 1967. In the four decades that have passed since their appearance, technology development had a real impact on electronic payment systems.

ISBN: 978-1-61804-230-9

A real progress of electronic payment system came with the development of EFT (Electronic Fund Transfer) technology. This is an e-commerce technology which allows the transfer of funds from one person's or organization's bank account to another. This important evolution caused the appearance of funds' online remittance in electronic funds transfer and banking. However, the electronic fund transfer is one of the processes which already existed before the appearance of Internet. But, there is an important difference between the pre-Internet transfer systems and the online payment systems consisting in the fact that the former was almost exclusively used for proprietary networks, while the latter is used over a publicly accessible electronic medium.

These changes, naturally, have affected the definition of electronic payments, which is evolving depending on the needs of each period. A general definition of electronic payment is given by 1.

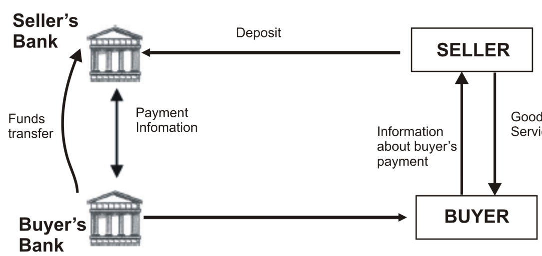


Fig. 2. Traditional Payment System

Definition 1: Electronic Payment. Any payment to businesses, banking institutions or public services from citizens or businesses, which are executed through a telecommunications or electronic networks using modern technology.

The payment is executed payer (consumer or business) without any intervention of other physical person. Furthermore, the payment transaction is made from distance without the physical presence of the payer. Such a transaction does not include cash.

There are different e-commerce categories (B2B, B2C, C2B, C2C). The main difference between them involves the payment value. In [6] the payments are classified in:

- Micro Payments (less than \$10) which mainly used in C2C and C2B e-commerce.
- Consumer Payment (between \$10 and \$500) which is mainly used in B2C e-commerce.
- Business Payment (more than \$500) which is mainly used in B2B e-commerce.

In [7] is presented another classification where the payment transactions can be:

- Tiny value transactions (below \$1).
- Medium value transactions (between \$1 and \$ 1.000)
- Large value transactions (above \$ 1.000)

To better understand the electronic payment system we shall first explain the traditional payment system. The traditional one consists in transferring cash (or payment information such as cheque and credit cards) from buyer to seller. For the cash payment there is required a buyer's withdrawals from his own bank account, a cash transfer to the seller, and last but not least the seller's deposit of payment to his own account. Non-cash payment mechanisms are settled by crediting and debiting the appropriate accounts between banks based on the information provided by the payment cheques or credit cards (figure 2).

Suppose a client goes to a shop and wants to pay with his credit card. This is a real example for using a traditional payment system. The client proves his identity by presenting the credit card (a cautious seller might always require identity papers). To purchase the goods he wants he must passes his bank card to a machine equipped with a card reader. To assure the buyer's bank that the client agrees on hid account being debited and that the credit card indeed belongs to that client, the seller usually offers a paper-based signature from the client. Typing his PIN (Personal Identification Number) also represents the buyer's approval for the bank transfer. Most of the sellers use both these methods together. All these steps are made face-to face. The two banks exchange payment information, here consisting in credit card information, and the buyer's bank transfers the cash to the seller's one. This is an example of non-cash payment and the two banks need an agreed method of exchange payment instructions. This is named payment clearing [8]. In real markets, the clearing process involves some type of intermediaries such as credit card services or cheque processing clearing companies. Schematically most payment systems are based on similar process.

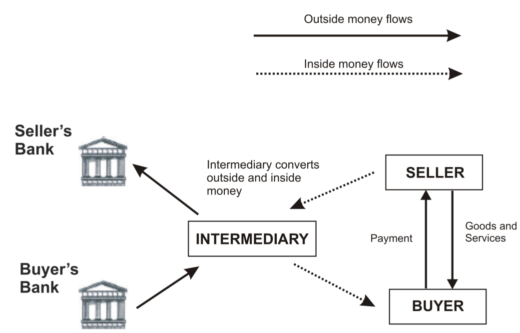


Fig. 3. Electronic Payment System

Such a traditional system has two disadvantages:

1) an online communication with the bank is needed at each payment,
2) the loss of anonymity for the client is almost totally since the bank knows when and where the client spends his money.

An electronic fund transfer is a financial application of EDI (Electronic Data Interchange), which sends credit card numbers or electronic cheques via secured private networks between banks and major corporations. To use such an application for clear payments and settle accounts, the online payment service must have the capability to process orders, accounts and receipts. A real impact on this area was determined by the appearance of digital concurrency concept developed in cryptography. The digital concurrency consists in encoded string of digits and can be carried on smart card or stored

ISBN: 978-1-61804-230-9

on a computer disk. A digital coin can be compared with a cheque since it is a floating claim on a bank that is not linked to any particular account.

E-coins systems involve:

- During withdrawal the financial institution issues an 'electronic coin' c, typically consisting of a (random) number n and a signature $sig(s_{fi}; n)$ on this number, where s_{fi} denotes the financial institution's secret key.
- At payment the client hands the coin to the seller who accepts it if the financial institution's signature on it is valid.
- The seller's account is credited only at deposit time: seller sends the coin to the financial institution who verifies the validity of the signature and checks that the coin has not already been deposited. If both verifications succeed, the seller's account is credited with the value of the coin.

Such a system provides anonymity and convenience. In figure 3, the intermediary acts as an electronic bank, which converts outside money (e.g. EUR or US \$), into inside money (e.g. coins or e-cash), which is circulated within online markets. Digital concurrency can be considered a private monetary system. Digital currency has already given rise to several types of new businesses such as: software vendors for currency server system; hardware vendor for the smart cards readers and other interface devices; technology firms for security, encryption and authentication and new banking services interfacing accounts in digital currency and conventional currency.

In [5], [4] there are described different types of payment. Most of the classifications are made considering what information is being transferred online. In [12] the author presents six types of electronic payment systems:

Fig. 4. Credit Card Payment Form

1) PC-Banking
2) Credit Cards
3) Electronic Cheques
4) Micro payment
5) Smart Cards
6) E-Cash

Other classifications are presented in [11] and [9] where there are considered three types on electronic payment systems (Digital Token, Smart Card, Credit) and two types (Electronic Cash, Electronic Debit-Credit Card), respectively.

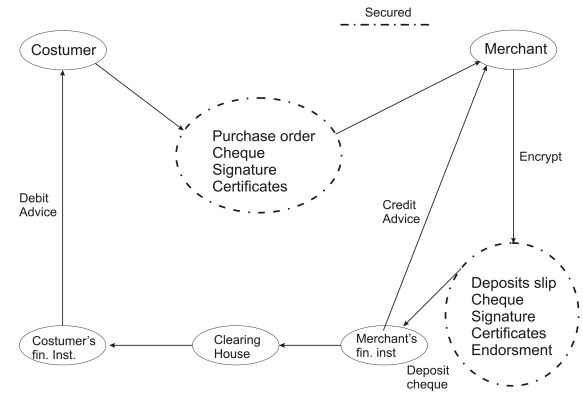


Fig. 5. Electronic Cheque Transaction

Considering all these classifications we can divide the electronic payment systems into four general types [10]:

1) Online Credit Card Payment System
2) Electronic Cheque System
3) Electronic Cash System
4) Smart Card based Electronic Payment System

The first type, Online Credit Card Payment System, extends the functionality of existing credit cards for use as online shopping payment tools. This is the most popular from the four types, especially in the retail markets [13]. The main advantages of these systems are privacy, integrity, compatibility, good transaction efficiency, acceptability, convenience, mobility, low financial risk and anonymity, advantages that the traditional systems do not have. When using such a system consumers simply send their credit card details to the service provider involved and the credit card organization will handle this payment like any other (figure 4). The main concerns of using such a system for merchant include lack of authentication, repudiation of charges and credit card frauds. Most of the clients are concerned because of the necessity to reveal credit information at multiple sites and repeatedly having to communicate sensitive information over the Internet. These systems where divided into three ([11]):

1) payment using clean credit card details
2) payment using encrypted credit card details
3) payment using third party verification

Electronic Cheque Payment Systems are mainly used by businesses. To better understand these systems we must explain the term "electronic cheque". E-cheques are used for electronic payments through an intermediary which will debit the client's account and credit the merchant's one. They are generated and exchanged online. The main difference between e-cheques and EFT is due to the fact that for electronic chequing, electronic versions of cheques are issued, received and processed, while for EFT automatic withdrawals are made for monthly bills or other fixed payments such that charges are not any more an issue.

Such a payment system is very simple: an account holder issues an electronic document containing the name of the financial institution, the payer's account number, the name of the payee and the amount of the cheque. To be valid this document also has a digital signature to authenticate

ISBN: 978-1-61804-230-9

the cheque. Unlike the credit card payment systems, the electronic cheque payment systems do not require consumers to reveal account information to other individuals, they do not require consumers to continually send sensitive financial information over the web being less expensive than credit cards. Another advantage is that they are much faster than paper based traditional cheque. The main disadvantages are the fact that they are relatively expensive and they are limited use only in virtual world. These systems also protect the users' anonymity. This is considered and advantage for the other systems, but for electronic cheque systems it is not suitable for the retail transactions by consumers, for the government and B2B operations since these transactions do not require anonymity.

An electronic cheque transaction has three phases, each consisting in one or more steps:

1) First phase: the consumer makes a purchase
    a) the costumer fills a purchase order form, attaches an electronic cheque, signs it with his private key (using his digital signature hardware), attaches his public key certificate, encrypts it using his private key and sends it to the merchant.
2) Second phase: the merchant sends the electronic cheque to its financial institution for redemption.
    a) the merchant decrypts the information using his private key, checks the costumer's certificates, signature and cheque, attaches his deposit slip, and approves the deposit attaching his public key certificates. All this information is encrypted and sent to his financial institution.
3) Third phase: the merchant financial institution approaches the clearinghouse or consumer financial institution to cash out the electronic cheques
    a) the merchant financial institution verifies the signatures and certificates and sends the cheque for clearance (note that the financial institutions and clearing houses normally have a private secure data network).
    b) when the cheque is cleared, the amount is credited to the merchant's account and a credit advice is sent to him
    c) the costumer also gets debit advice periodically.

The entire process is illustrated step by step in figure 5.

Electronic cash refers to money or scrip which is exchanged only electronically. The main advantage of e-cash is their portability since they can be freely transferable between two parties in all forms of e-commerce transactions. This property is not possessed by credit cards since a credit card transaction needs the recipient must have a merchant account established with a financial institution.

The most important characteristics of e-cash are:

- value – e-cash must have monetary value being backed by bank credit;
- storable and retrievable – the cash can be stored in transportable devices (e.g. smart cards) such that the users can exchange them even if they are traveling;

TABLE II
ADVANTAGES AND DISADVANTAGES OF E-CASH PAYMENT SYSTEMS

Advantages	
authority	
privacy	the merchant cannot access consumers bank account information
good acceptability	
low transactions cost	
convenience and	companies do not have access to consumers account information
good anonymity	while distributors do not know how e-consumers spend the e-cash
Disadvantages	
poor mobility	they can be use only with computers with e-cash purse system
poor transaction efficiency	for comparison is needed a large database
high financial risk	people are solely responsible for the lost or stolen

- interoperability – e-cash is exchangeable as payment for other e-cash, paper cash, goods or services, etc.;
- security – e-cash is not easy to tamper during an exchange process (preventing and detecting duplication, double spending etc).

Based on these characteristics e-cash payment systems have several advantages presented in the table I-A. However, there are also some big disadvantages that caused the e-cash relatively low-scale.

E-cash payment systems were a real success in Hong Kong and Singapore. In Hong Kong the card system Octopus started as a transit payment system. In Singapore, FeliCa is an e-cash system used for the public transportation system.

The main problems that must be taken in consideration when designing an e-cash payment system are [14]:

- who has the right to issue electronic cash?
- can every bank issue its own money? If so how do you prevent fraud?
- who will monitor the banking operations to protect consumers?

E-cash structure is a a string of bits with certain values: reference number, digital signature etc.. In [15] the author described the values of each substring. In [14] there has been added an extra value consisting in a digital watermark for protection against illegal copy and forgery. The author has also modified the reference number for supporting tractability. The final structure is:

$$Cr \;\; Vl \;\; Ref \;\; Dig\_Sig \;\; Dig\_Wm$$

- Cr – concurrency: defines the issued currency
- Vl – value: determines the e-cash value
- Ref – reference number: allows the issuer or any other authorized party to trace e-cash movement
- Dig_Sig – digital signature: authenticates the issuer as an authorized party.
- Dig_Wm – digital watermark: inserts invisible data into the digital file.

The last type, smart cards based electronic payment systems are gaining more and more popularity. Smart cards can be

Fig. 6. Smart Card

defined as plastic cards with a credit card size embedded with memory chip or, the advanced ones, with microprocessors. So, these cards can store much more information than credit cards [16]. Credit cards can store only a single account number on the magnetic strip while smart cards can hold 100 times more data, including multiple credit card numbers and information regarding health insurance, transportation, personal identification, bank accounts etc.. This capacity makes them attractive alternatives to carrying a dozen or so credit and ID cards in a physical wallet.

They also have a considerably greater security than credit cards since they contain some kinds of an encrypted key that is compared to a secret key contained on the user's processor. Some smart cards allow users to enter a personal identification number (PIN) code.

Unlike traditional electronic cash system, smart cards based electronic payment systems do not need to maintain a large real time database. They also have advantages, such as anonymity, transfer payment between individual parties, and low transactional costs. Smart cards are also better protected from misuse than credit cards since the smart card information is encrypted. Currently, there are two smart cards based electronic payment system – Mondex36 and Visa Cash. Unfortunately, they are incompatible in the smart cards and card reader specification. Because of that, financial institution avoid adopting either of them since they do not know which will become the leader. So, the main problems that must be considered in designing a smart card based payment system are establishing a standard smart card system, or making different system interoperable with one another.

In [11] the smart card systems were divided in relationship based smart cards and electronic purses. The latter are smart cards embedded with programmable microchips that store sums of money for people to use instead of cash for everything. For using such a purse there must be followed several steps: load it with money at an ATM or through the use of an inexpensive special telephone; use it, for example, to buy a sweets from a vending machine; the vending machine verifies the authenticity of the card and the amount of money; the value of purchase is deducted from the card's balance and added to e-cash box in the machine; the client can check its new balance at an ATM.

Smart cards based electronic payment systems where classified, depending on the technology used, in [17], [18] as:

- memory cards – store password or pin number
- shared key cards – store a private key such as those used in the asymmetric cryptosystems.
- signature carrying cards – stores a set of pregenerated random numbers.
- signature carrying cards – have a co-processor that can be used to generate large random numbers.

The random numbers from the last type of smart cards can then be used for the assignment as serial numbers for the electronic cash.

TABLE III
COMPARISON OF ELECTRONIC PAYMENT SYSTEMS

	Online Credit Card Payment	Electronic Cash	Electronic Cheque	Smart Cards
Actual Payment Time	Paid later	Prepaid	Paid later	Prepaid
Transaction information transfer	The store and bank checks the status of the credit card	Free transfer. No need to leave the name of parties involved	Electronic checks or payment indication must be endorsed	The smart card makes of both parties the transfer
Users	Any legitimate credit card users	Anyone	Anyone with a bank account	Anyone with a bank or credit card account
Payment receiver	Distributing Bank	Store	Store	Store
Limit on transfer	Depends on the limit of the credit card	Depends on how much is prepaid	No limit	Depends on how much money is saved.
Mobility	Yes	No	No	Yes
Online and offline transactions	Online transactions	Online transactions	Offline transfers are allowed	Offline transfers are allowed
Anonymity	Partially or entirely anonymous	Entirely anonymous	No anonymity	Entirely anonymous but the central processing agency can ask stores data about a user

II. SECURITY ISSUES

Nielsen made a global online survey credit cards are the most popular method of payments. The statistics show that the top three countries where credit cards are used for e-commerce payment are Turkey (91%), Irland (86%) and India 84%. Also, more than 60% of global online clients have recently used a credit card for online payment. The most popular card is visa since more than 53% of the clients used it.

The electronic payment system is the most important from an e-commerce system. Table IV illustrates the main concerns that the consumers when using such a system. According to presented data the main problem that make consumers think twice before using an e-commerce system is the security.

In most of e-commerce systems users must login using a username and a password. The method is designed to allow a client program to provide credentials when making a request. First, the username is appended with a colon and concatenated with the password. Using a colon is for separating the username from the password. The resulted string is then encrypted with an encryption algorithm and transmitted. When the receiver decrypts it, he obtains the username and the password [19]. The popularity of such an authentication process is due to its easy implementation and the fact that it is supported by the most used browsers. The most importanta disadvantage is the fact that such a system's security is based on the assumption that the connection between the client and the server is secure and can be trusted. An alternative to this method is represented by the digest access authentication which consists in negotiating credentials with an user. To validate the user identity the method uses MD5 to hash the password, preventing cryptanalysis [20].

In spite of the fact that their use has been roundly criticized on many occasions, passwords remain the most popular method for user authentication [21]. The main problems are human since users do not choose and store them properly. Most of the people choose poor passwords based on private data which can be easily guessed by others. In [22], [23] the author presents some facts that must be taken in consideration when choosing passwords:

1) users must not choose poor passwords (passwords that are too short, passwords that are related to personal information and can be easily guessed by relative or other person who knows them, passwords that are found in a dictionary and can be cracked by dictionary tools);
2) users must not share their passwords with anyone else;
3) users must be careful when storing passwords and avoid writing them down;
4) users must periodically change their passwords and they must not revert to using the previous one again at a later time.
5) users must not use the same password for a variety of systems since vulnerabilities may be caused to all the systems when one of them had been broken.

A workable password authentication requires at least two modes of authentication since users often forget or lose their passwords. The primary authentication mode implies using the password. If the user forgot the password he can use the emergency mode. For most of the sites the emergency mode implies sending the password or instructions for resetting it to an e-mail account. Another common emergency mode is based on ”life questions”. When the user created his account he had answered to several ”life questions”. For resetting the password he has to give the same answers as the ones given at the beginning.

Online buyers have many concerns regarding e-commerce systems. The top security concerns are:

- Receiving different products from the ones purchased or, even worse, receiving nothing at all.
- Email addresses given to an unauthorized party.
- The steal of personal data and credit cards' information which can cause financial frauds.
- Email scams known as ”phishing” or ”spoofing” – messages sent to consumers from disreputable sources which are disguised as messages from trusted merchants or financial institutions.

TABLE IV
MAIN CONCERNS FOR ELECTRONIC PAYMENT SYSTEMS

Factors	Percentage
Concern about security	70
Difficulties to enter information	9
Do not have credit cards/smart cards etc.	7
Do not like interest charge	6
Purchase value too small	4
Exceeded personal limit	4

A common problem is represented by the email spams which represent an unsolicited commercial communication. Many users receive advertisements about products from different companies. The companies do not pay for this kind of publicity so they need the user's approval for sending such emails. The ones sent without an approval are considered spams. Even if the company has the user's approval it must always respect at least the following:

- to be clearly identifiable as such;
- the natural or moral person in the name of which they are made to be clearly identified;
- the promotional offers, such as reductions, prizes and gifts, to be clearly identifiable, and the conditions to met for obtaining them must be easy accessible and clearly presented;
- competitions and promotional games must be clearly identifiable as such, and the participation conditions must be easy accessible and clearly presented;
- any other conditions stated by the legal dispositions.

The consumer consent can be obtained in any form and can be proved with any mean and the company must be able to prove that it has the consumer content.

The commercial communications must contain at least the following information, regarding the person for which are made:

- name;
- personal numerical code or registration code;
- postal address;
- phone and fax numbers;
- e-mail address.

The user has the right to revoke its approval to receive such advertisements by simply notifying the sender. The company has an obligation to implement a free procedure, accessible also by electronic means, through which the receiver would be able to revoke easily its consent. The cancellation of the consent by electronic means must be implemented in at most 48 hours from the beginning of the procedure.

The following represent the basic precautions that a client must consider:

- Knowing as much as possible about the seller (the seller's reputation can be inspected from previous purchases, from referrals or from reviews by other online shoppers).

- Understanding the retailers' refund policies. A client has to know the required timeframe he must contact the retailers and return the product and if he will receive a full refund. If the retailers do not have refund polices clients can use buyer protection which provides covering or refunding the payment if there is a transaction problem.
- Choosing a secure password for the account information.
- Using a secure checkout and payment process. To verify if the sites encrypt the personal and financial information the client must use safe browsers (browsers which show the icon of a locked padlock at the bottom of the screen to indicate the encryption). It is recommended to use the same account for more than one purchase to avoid filling up several forms with personal and financial data.

All the above recommendations do not ensure the client that all the concerns mentioned earlier will be eliminated but they represent the minimal precautions that a user has to consider. We further discuss only the security issues of an e-commerce system. It includes storing and using safely the personal and financial data of the user but other facts, like receiving different products, depend on the seriously and the reputation of the seller.

The most common e-commerce frauds imply a user authentication attack. The term "authentication" refers to a large class of electronic applications whose functions may range from pure identification and authorization to legal recognition. Referring to specific authentication techniques, electronic signature is often confused with digital signature. The term "electronic signature" has been defined by UNCITRAL as

> *"a signature in electronic form in, or attached to, or logically associated with, a data message, and used by or on behalf of a person with the intent to identify that person and to indicate that person's approval of the contents of the data message."*

Further, "digital signature" has been defined in ICC's General Usage for International Digitally Ensured Commerce (GUIDEC) as

> "*a transformation of a message using an asymmetric cryptosystem such that a person having the ensured message and the insurer's public key can accurately determine:*
> - *whether the transformation was created using the private key that corresponds to the signer's public key,*
> - *whether the signed message has been altered since the transformation was made."*

There are currently three security standards for IT systems for system security: the Common Criteria Redbook [25], the NIST's paper: "Underlying Technical Models for Information Technology Security" [24], and the Open Source Security Testing Methodology Manual [26]. Information about providing non-repudiation, encryption algorithms, and intrusion detection systems can studied in depth in [27].

In e-commerce the Common Criteria Redbook cannot be used in the design phase of the system. The CC is intended to be used to evaluate system security only after the system was already implemented and because of that it is not very useful in this area. Unlike CC, NIST's paper can be used as a security design standard in IT systems. Even so the paper does not provide specific details for the e-commerce systems. The Open Source Security Testing Methodology Manual can also be used after the system is implemented the difference between this manual and the CC document being the fact that the latter one does not emphasizes blocking malicious user requirements.

TABLE V
NIST'S MODEL SERVICES

Prevention Services	
Protected communication	ensure integrity, availability and confidentiality of data
Authentication	verify the veracity of the user's identity
Authorization	manage the actions of the allowed user for a given system
Access control enforcement	enforce the security policy when the user has been validated for access
Non-repudiation	ensure the property of the system that senders cannot deny sending data and receivers cannot deny receiving it
Transaction privacy	protect the system against losing privacy of an individual
Recovery Services	
Audit	provide auditing security relevant events
Intrusion detection and containment	detect insecure situations and security breaches
Proof of wholeness	determine the compromised integrity caused by data corruption
Restorability to a 'secure state'	return to a secure state when a security breach occurs
Supporting Services	
Identification	uniquely identify users data and processes
Cryptography key management	securely management of cryptographic keys
Security administration	administrate the security features to provide needs of a specific installation
System protections	security functional capabilities for providing the implementation quality

The main concepts of security are defined below [24]:

1) Confidentiality – the security objective that generates the requirement for protection from intentional or accidental attempts or perform unauthorized data reads (it covers data in storage, during processing and transmission).
2) Integrity – this objective has two phases:
   a) Data integrity – data has not been altered in an unauthorized manner while storage, during processing and transmission.
   b) System integrity – the system is able to perform functions free from unauthorized manipulation.
3) Availability – the security objective that generates the requirement for protection against intentional or accidental attempts to perform unauthorized deletion of data or, otherwise cause a denial of service or data.
4) Accountability – actions of an entity can be traced uniquely to the respective entity.
5) Assurance – it represents the basis for confidence that the foru security measures mentioned above work, pro-

ISBN: 978-1-61804-230-9

tecting both the system and data.

NIST's model classifies the services in:

- **PREVENT** – the services must focus on preventing security problems
- **RECOVER** – the services focus on detecting and recovering from a security breach
- **SUPPORT** – the services represent the basis for most data technology security capabilities

These services are described in table V.

III. Attacks

In [28] the authors present some types of authentication attacks which have been applied on e-commerce systems. Relying on different authentication methods (see [29], [30], [31], [32], [33], [34]) these attacks are classified in:

1) ID spoofing attacks
2) Sniffing attacks
3) Brute-force attacks
4) Dictionary attacks
5) Credential decryption attacks
6) Replay attacks.

ID spoofing attacks allow a malicious user (process) to impersonate someone else. Such attacks are possible when the user identity is validated using some static information that has already been definite. Most of these attacks impersonate in fact the IP address of a local system. Usually e-commerce system use passwords or certificates to identify users and because of that these attacks are rare in this domain. Even so, there are situations when requiring passwords or certificates is not possible. To provide the same security level there must be enabled the least possible access to the least possible number of users, processes or hosts after successful authentication.

Sniffing attacks allow an intruder to capture information from the communication channel. Named also man-in-the-middle attacks, this kind of attacks can be avoided by securing the communication channel. A transmission between a seller and a client can contain not only authentication data (passwords and usernames) but also personal data and credit cards' information. E-commerce system, and not only, use encryption algorithm to maintain security even if the smallest piece of information is leaked.

Brute-force attacks allows an intruder to try to find out passwords and usernames through exhaustive search. These attacks are possible only when the intruder has gained access to data from the system's database. Since all the data saved in the database is encrypted the intruder must use different encryption algorithm to obtain the plaintext passwords. Most of these attacks are made by programs or machines since the amount of operations is very large. The most important precaution must be taken by the users which must avoid poor passwords. Strong passwords provide increasing the operations' number needed for finding out the plaintexts. Besides that, a very important measure must be taken in the system's design phase: providing strong access polices in order to avoid information access for unauthorized persons (processes).

Dictionary attacks can be considered a smarter version of the brute-force attacks. To apply such an attack the intruder must have access to a valid username. Unlike the brute-force attacks, such an attack can be done without access to any other information beside the valid username. The idea is to try different combinations in order to guess the right password. Many attacks of this type were successful mainly because people tend to choose short and easy to remember passwords. To avoid dictionary attacks, e-commerce systems have an automatic setting which provides a limit of unsuccessful authentication. A human being can try maximum 50 passwords. Usually after 5 unsuccessful attempts the user uses the emergency mode. A program can try more than 200 successive unsuccessful attempts. So, using such a setting will immediately block the dictionary attack program.

Credential decryption attacks attempt to break the encryption algorithm. These attacks are usually used as together with sniffing attacks, brute-force attacks and dictionary attacks. The systems that are vulnerable to such attacks are the ones that do not have strong credential polices and strong cryptographic algorithms. The encryption/decryption algorithm must provide a high security level and must be correct implemented.

Replay attack are the attempting of an intruder to authenticate himself through valid authentication sequences trapped from the communication channel. To avoid such an attack there must be taken the same measures as the ones for the sniffing attacks. So, the transmitted data must be encrypted using a very strong cryptographic system. It can also be used a time stamp for all the messages.

Side-channel attacks occur when devices like smart cards leak information during cryptographic processes. In [37] the author describes such an attack where an intruder stores secret data leaked during cryptographic procedures. The side-channel attack includes timing attack, simple power analysis, and differential power analysis attack [28]. The designers which use smart cards or other external devices for authentication must take in consideration the side-channel attacks.

For further reading about these attacks and measures against them see [34], [35], [36].

A. Brands' System

We chose to present Brands' system because he was first to introduce the term "restrictive blind signature" together with an e-cash system providing double-spender identification without the use of cut-and-choose techniques. Also, it is one of the few systems which has not been broken even if its security cannot be proven. The system underlies on the concept of restricting blinding process while the blinded message still contains the original data. So the restrictiveness property means that the signer is assured that the verifier performs the blinding in the prescribed manner.

The system uses the digital concurrency (section I-A) concept or e-coins. The system works like this: at withdrawal the client includes his identity in the data which together with the blindly issued signature by the financial institution will form the e-coin. Because the financial institution cannot see the final coin, it cannot know that it contains the client's identity; during payment, the seller sends a challenge c to the client and he has to respond to it; through his response the client provides additional information about the internal structure of the coin, information that does not compromise the client's anonymity; if the client tries to double-spend the coin he has to send another response to another challenge c' received from the seller; form the two different responses the seller is able to extract the client's identity from the coin. With restrictive blind signatures, double-spending is traceable, but only after it has occurred.

For a mathematical description of Brands' system we use the following notations: U is the client, S is the seller and B is the bank (financial institution).

Algorithm 1 Generating Parameters

B chooses G_p where p is a prime
B generates random generators g, g_1, g_2 of G_p
B generates one-way hash functions H, H_1
B generates a secret key $x \in_R Z_p^*$
B computes $h = g^x$
B makes public (h, g, g_1, g_2, H, H_1)

Note that in his paper brands chooses $G_p = Z_p^*$.

Algorithm 2 Account Login

U chooses a secret value $u_1 \in Z_p$ such that $g_1^{u_1} g_2 \neq 1$
U computes $i = g_1^{u_1}$ and makes it public
U proves to B that he knows u_1
U computes $m_b = (ig_2)^s g^t$
U chooses $t = 0$ and so $m_b = g_1^{u_1 s} g_2^s$
B computes $z = (ig_2)^z$ and sends it to U
U computes $h_1 = g_1^x$ and $h_2 = g_2^x$
U computes $z = h_1^{u_1} h_2$

In the algorithm 2 is used the Chaum-Pedersen signature on the message ig_2. This proves that U's identity is contained in the message. s and t represent the Chaum-Peterson signature, and t has the value 0 because U must know at payment time a certain value [38]

$$m_b = g_1^{y_1} g_2^{y_2}$$

Note that the challenge d sent by S to U is based on the coin and on specific information to this payment.

To exemplify a double spending we suppose that d represents i_s and $date/time$. If U tries to deposit the same coin twice the responses of U are equal in both transcripts. If the coin was already double spent by U the two responses (r_1, r_2) and (r_1', r_2') are different since the challenges from payment were also different. Then B computes:

$$\frac{r_1 - r_1'}{r_2 - r_2'} = u_1$$

Next he computes

$$g_1^{u_1} = i$$

So he has obtained the account number of u_1 which proves the double spending.

Algorithm 3 Withdrawal Algorithm

B generates $w \in_R Z_p$
B computes $a = g^w$ and $b = (ig_2)^w$
B sends a, b to U
U chooses a secret random blinding factor s
U computes $A = (ig_2)^s$ representing the blinded message
U computes $z' = z^s$
U chooses random values $x_1, x_2, u, v \in Z_p$
U computes $B = g_1^{x_1} g_2^{x_2}$
U computes $a' = a^u g^v$
U computes $b' = b^{su} A^v$
U computes $c' = H(A||B||Z'||a'||b')$
U computes $c = c'/u$
U sends c to B
B computes $r = cx + w$
B sends r to U
U verifies if $g^r = h^c a$ and $(ig_2)^r = z^c b$
if the above are true U computes $r' = ru + v$
the signature is $sig(A, B) = (z', a', b', r')$

Algorithm 4 Payment

U sends S the values $(A, B, sig(A, B))$
S verifies if $A \neq 1$
if the above is true S computes d which represents a challenge
U responds with r_1, r_2
S verifies $sig(A, B)$ and the correctness of r_1, r_2

B. *Perfect Blackmailing Attack*

The blackmail attack was first presented in [39] and consists in escorting coins from B. To describe the attack we note f a one-way function and n a RSA modulus. Thus, $n = pq$ where p, q are large primes. Like in RSA system n is made public while p, q remain secret. To issue coins B follows the steps:

- U chooses a random value x and a blinding factor r
- U computes $f(x)$ and
$$B = r^3 f(x)$$
- B computes
$$D = B^{1/3} (mod\ n)$$
- B debits the account of U and sends D to U
- U computes
$$C = D/r(mod\ n) = f(x)^{1/3} (mod\ n)$$

Note that C can be considered the unblinding process of D. The coin is represented by the pair (x, C). The validity of the coin is verified if

$$f(x) = C^3$$

Suppose we have an intruder E who wants to obtain q coins from B. The algorithm is described in 5.

After completing the algorithm (x_j, C_j) are valid coins for $j \in \{1, \dots, q\}$.

ISBN: 978-1-61804-230-9

Algorithm 5 Blackmailing Attack

E randomly chooses

$$\{x_1, x_2 \dots x_q\}$$

and

$$\{r_1, r_2 \dots r_q\}$$

E computes

$$\{B_j | B_j = r_j^3 f(x_j)(mod\ n)\}$$

where $j \in \{1, \dots, q\}$
E sends B_j to B
E forces B to compute

$$\{D_j | D_j = B_j^{1/3}(mod\ n)\}$$

where $j \in \{1, \dots, q\}$
B is forced to publish this set in a public paper
E buys the public paper and computes

$$\{C_j | C_j = D_j / r_j(mod\ n)\}$$

where $j \in \{1, \dots, q\}$

IV. Concluding Remarks

The importance of e-commerce has increased significantly in the last decade since it empowers people to easily compare products, prices and delivery options which has made shopping more enjoyable, less expensive and less time consuming. Clients make their preferences on e-commerce systems depending on the electronic payment system used. We have presented the four types of these systems emphasizing their advantages and disadvantages.

Credits cards based systems are the most dominant payment all over the world. This is especially true about the developed and fastest developing countries. The security of an e-commerce system must be treated with high consideration because any weakness can cause serious frauds, especially financial ones. NIST proposed for the first time the main concepts that must be taken in consideration when designing an e-commerce system. NIST also classified the services of such a system in three main classes: PREVENT, RECOVER, SUPPORT. The most extensive class is the PREVENT services one. Here the authentication services have a great importance since compromising such a system is usually caused by authentication problems. We have explained the most popular six attacks against an authetincation and proposed methods to avoid them.

Blind signatures are often used for authenticating a user in an e-commerce system because they provide anonymity. A very important system is Brands' system which uses a restrictive blind signature that provides double-spending identification. The main difference between restrictive blind signature and the common blind signature is the fact that the former provides containing the original data in the blinded message. In spite of the fact that the system's security cannot be proven, Brands' system was never broken. The attack we described shows how an entity can obtain coins through blackmailing a financial institution.

Users must not underlie only on the security measures taken by the e-commerce system. We have presented some fact that users must take in consideration when using e-commerce systems. This paper tries to emphasize the issues of e-commerce systems and to warn clients and not only about the security problems that may appear. As future work we take in consideration implementing a safe authenticating method which may be used in an e-commerce system.

References

[1] Scarlet Schwinderski-Grosche and Heiko Knospe, Secure M-commerce, *IEE Electronics and Communication Engineering Journal.* 14(5), 2002, pp. 228–238.

[2] S. Brands, Untraceable off-line cash in wallets with observers, Advances in Cryptology - CRYPTO '93, vol. 773 of Lecture Notes in Computer Science, *Springer-Verlag Berlin*, 1993, pp. 302–318.

[3] Emmanuel Lallana, Rudy Quimbo, Zorayda Ruth Andam, *ePrimer: An Introduction to eCommerce*, Philippines, DAI-AGILE, 2000, 2.

[4] N. Asokan, P.A. Janson, M. Steiner and M. Waidner, The state of the art in electronic payment systems, *Computer*, 30(9), September 1997 pp. 28–35.

[5] H. Bürk and A. Pfitzmann, Digital payment systems enabling security and unobservability, *Computers & Security*, 8(5), August 1989, pp. 399–416.

[6] Daniel Amor, *E-Business (R)evolution*, New York: Prentice Hall, 2002.

[7] F. Cavarretta and J. de Silva, "Market Overview of the Payments Mechanisms for the Internet Commerce", http://www.mba96.hbs.edu/fcavarretta/money.html, 1995.

[8] J.M. Peha and I.M. Khamitov, PayCash: a secure and efficient Internet Payment System, *Electronic Commerce Research and Applications*, 3(4), 2004, pp 381–388.

[9] Dennis Abrazhevich, Classifications and Characteristics of Electronic Payment Systems, *Lecture Notes in Computer Science*, 21(5), 2001, pp. 81–90.

[10] M.M. Anderson, Electronic Cheque Architecture, Version 1.0.2, *Financial Services Technology Consortium*, September 1998.

[11] Ravi Kalakota and B.Andrew Whinston, Frontiers of Electronic Commerce, *Singapore: Pearson Education*, 1996, pp. 295.

[12] C.S.V. Murthy, E-Commerce: Concepts, Models and Strategies, *New Delhi: Himalaya Publishing House*, 2002, pp. 626.

[13] C. Kenneth Laudon and Carol Traver, *E-Commerce: Business, Technology, Society (3rd Edition)*, Prentice Hall, 2006.

[14] Singh Sumanjeet, Emerge of payment systems in the age of electronic commerce: state of art, *Global Journal of International Business Research*, 2(2), pp. 17–36 2009.

[15] David Wright, Comparative Evaluation of Electronic Payment System, *INFO* 2002.

[16] Rajesh Chakrabarti and Vikas Kardile, *E-Commerce: The Asian Manager Handbook*, New Delhi: Tata McGraw Hill, 2002.

[17] Parag Diwan and Sunil Sharma, *E-Commerce: A Managerial Guide to E-Business*, New Delhi: Excel Books, 2001.

[18] Parag Diwan and Dharmvir Singh, *Computer Networks Driven E-Commerce Technologies*, New Delhi: Amexcel Publisher Pvt. Ltd, 2000.

[19] RFC 1945, *Hypertext Transfer Protocol - HTTP/1.0.*

[20] M.J.B. Robshaw, On Recent Results for MD2, MD4 and MD5, *RSA Laboratories Bulletin Nr. 4*, November 1996.

[21] DTI, Information security breaches survey 2006, *Department of Trade and Industry; April 2006* [URN 06/803].

[22] S. Furnell, Authenticating ourselves: will we ever escape the password?, *Netw Secur*, pp. 8–13, March 2005.

[23] S. Furnell, *Cybercrime: vandalizing the information society*, Addison Wesley Professional, 2001.

[24] National Institute of Standards and Technology, Special Publication 800-33, Underlying Technical Models for Information Technology Security, 2001, ccsrc.nist.gov/publications/nistpubs/800-33/sp800-33.pdf

[25] Common Criteria, Common Criteria for Information Technology Security Evaluation, 1999, www.commoncriteria.org.

[26] P. Herzog, *The Open Source Security Testing Methodology Manual*, version 1.5, 2001, www.ideahamster.org.

[27] International Organization for Standardization, ISO Technical Programme, *It Security Techniques Sub-Committee 27.*, www.iso.ch/iso/en/stdsdevelopment/techprog/workprog/TechnicalProgrammeSCDetailPage.TechnicalProgrammesSCDetail?COMMID=143
[28] Victor D. Sawma and Robert L. Probert, *E-COMMERCE AUTHENTICATION An Effective Countermeasures Design Model*, ICEIS (3), pp. 447–455, 2003.
[29] G. Agnew,Cryptography, Data Security, and Applications to E-commerce, *Electronic Commerce TacTechnology Trends Challenges and Opportunities, IBM Press*, pp. 69-85, 2000.
[30] M. Bellare, R. Canetti and H. Krawczyk, A Modular Approach to the Design and Analysis of Authentication and Key Exchange Protocols, *Proceedings of the thirtieth annual ACM symposium on Theory of computing, ACM NY*, pp. 419 – 428, 1998.
[31] K. Lee Chang and T. Kim, Open Authentication Model Supporting Electronic Commerce in Distributed Computing, *Electronic Commerce Research, Vol. 2, Kluwer Academic Publishers*, pp. 135–149, 2002.
[32] W. Ford W. and M. Baum, *Secure Electronic Commerce*, Prentice Hall, 1997.
[33] S. Hawkins, D. Yen and D. Chou, Awareness and Challenges of Internet Security, Information Management & Computer Security, Vol. 8(3), *MCB University Press*, 2000.
[34] S. Wilson, Certificates and trust in electronic commerce, *Information Management & Computer Security, Vol. 5(5), MCB University Press*, pp. 175–181, 1997.
[35] R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*, John Wiley & Sons, 2001.
[36] J. Viega and G. McGraw, *Building Secure Software*, Addison-Wesley, 2002.
[37] C. Kocher, Cryptanalysis of Diffie-Hellman, RSA, DSS, and Other Systems Using Timing Attacks, *technical report*, Dec. 7, 1995.
[38] B. Pfitzmann, *Security (Sicherheit) – additional course material*, Course material, Universität des Saarlandes, Germany, http://www-krypt.cs.uni-sb.de, 1999.
[39] S. von Solms and D. Naccache, On blind signatures and perfect crimes, *Computers and Security, 11(6)*, pp. 581–583, October 1992.

Nicolae Constantinescu Nicolae Constantinescu is Associate Professor at University of Craiova, former CNRS researcher at Paris Orsay University. He is studying on mathematical models on Authentication and key agreement for government level security.

Oana Ticleanu Oana Ticleanu is PhD Student in Applied Mathematics on Cryptography. She is Studying in mathematical models on authentication.

Radiation damage evolution: nonlinear dynamics and Hopf bifurcation

P. Selyshchev, I.Velychko

***Abstract*—**It is considered the system of nonlinear differential equations that describes evolution of radiation damage of materials under irradiation. The variables of the system which describe material under irradiation are densities of different radiation defects, defect characteristics and temperature of the material. Nonlinear feedbacks between different elements of the model are taken into account. The qualitative analysis of this non-linear dynamical system is carried out. It is shown that material temperature and defect concentration as constant valued function of the time becomes unstable under certain opportune conditions of irradiation. A Hopf bifurcation arises and self-oscillations develop. The field of the parameters (defect generation rate and environment temperature) at which self-oscillations develop is found. Bifurcation curve and self-oscillation period for parameters belonging to its vicinity is obtained.

It is given physical interpretation of obtained results. Frequency of the self-oscillation and conditions of their development is rated. Numerical calculations show that period of self-oscillations ranges from 10^2 to 10^3 s depending on irradiation conditions and characteristic of the materials of the sample. Results of the experiments and simulation are compared.

***Keywords*—**Irradiation, radiation defects, nonlinear system, feedbacks, limit cycle.

I. Introduction

NUCLEAR facilities are operated under extreme thermal and radiation conditions. It places high demands on the control of their state and their operating conditions. The materials of nuclear core are situated in the most extreme thermal and radiation conditions. First of all, it is nuclear fuel, fuel assemblies and reactor vessel. Fuel irradiates and heats itself inside. Fuel assemblies and reactor vessel can be subjected to internal action or internal and external actions depending on kind of radioactive substance and kind of irradiation respectively [1].

The irradiated material is a typical example of an open non-equilibrium system with nonlinear feedbacks between its elements. Fluxes of energy, matter and entropy maintain states which are far away from thermal equilibrium. The connections that arise under irradiation are inherently nonlinear. They form the mechanisms of feed-backs that drive evolution of radiation damage. Well known that processes of self-organization can develop in such systems [2].

The central part of our approach is the nonlinear feedback that is a mechanism of self-organization. It is usually expected that the stationary thermal state is always realized under stationary external conditions. But for nonlinear system the steady state can become unstable with respect to the development of non-stationary states, for example self-oscillations. Hopf bifurcation occurs due to the nonlinear feed-backs of the system under certain opportune conditions. Radiation defect concentration and material temperature can reach much higher values than the expected steady-state values. It can lead to change of typical operation and an accident.

Let us consider an irradiated sample in an environmental. Environmental temperature is fixed. The sample consists of a central part and an external layer which surround the central part. The central part of the sample is radioactive crystalline material which irradiates itself. It is heated and radiation defects are created inside it as result of irradiation. It is assumed the irradiation is fixed so heating and defect generation is fixed too. Rate of defect generation and heating are constant. Substance of the external layer transfers heat from the central part to environmental. Defects aren’t created in the external layer. It takes place under irradiation with ionizing particles because range of ionizing particles is usually much smaller then thickness of the layer.

Let small increase of the temperature arise as result of small fluctuation. Increase of temperature results in increase of defect annealing. When defect annealing increases, energy that stores by radiation defects is released into heat and further temperature increase go on. The positive feed-back is formed. Temperature grows quickly and concentration of defects drops. Heat transfer increases. But defect annealing and release of energy drop. The sample cools, heat transfer decreases and radiation defects are accumulated slowly. After that all processes is repeated. Self-oscillations of sample temperature and material temperature are developed. Theoretical approach to describe self-oscillations in thin plates under irradiation is developed in [3],[4]. Here self-oscillations in layered plates are considered.

A theoretical research of the self-oscillations is developed via Poincare formalism within the framework of dynamical system on the plane.

This work was supported from NRF of South Africa.

P. Selyshchev is with the Department of Physics University of Pretoria Private bag X20 Hatfield 0028 South Africa (phone: 012-420-6641; fax: 012-362-5288; e-mail: pavel.selyshchev@up.ac.za).

I. Velychko, was with Kiev Institute for Nuclear Research, Kiev, Ukraine. He is now as software developer with DataArt Solutions Inc., Kiev, Ukraine (phone: +380636639060, e-mail: igor.s.velichko@gmail.com).

ISBN: 978-1-61804-230-9

II. Model and Basic Equations

Let the sample have the shape of a plane-parallel plate consisting of three plane-parallel plates. The material of internal plate is crystalline and radioactive. Generation of radiation defects and radiation heating takes place in the internal plate only. The internal plate's thickness is *l*. Thicknesses of all the external plates are the same and are equal to l_1. Environment temperature is kept constant and equal to T_e. Due to irradiation the internal sample heats with rate *Q* and defects are created in it with rate *K*. The defects are absorbed by dislocations. The dislocation density is ρ_d. When defect is absorbed, some energy releases. It is approximately equal to the energy of the defect formation θ.

So the change of the temperature of the internal plate *T(x,t)* is described by

$$c_i \frac{\partial T}{\partial t} = -divJ + Q + \theta\beta(T)n \qquad (0 \le x \le l/2) \tag{1}$$

The change of the defect density inside the sample n(x,t) is described by

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2} + K - \beta(T)n \qquad (0 \le x \le l/2) \tag{2}$$

Here $\beta(T) = \rho_d D(T)$ is inverse lifetime of defects, $D(T) = D_0 \exp(-E_m/T)$ is diffusion coefficient of defects, E_m is migration energy of the defect. The change of temperature of the external plate is described by

$$c_e \frac{\partial T}{\partial t} = -divJ \qquad (l/2 \le x \le l_1 + l/2) \tag{3}$$

Boundary conditions are

$J\big|_{x=0} = 0$, $J\big|_{x=l/2-0} = J\big|_{x=l/2+0}$, $T_{x=l/2-0} = T_{x=l/2+0}$,

$J\big|_{x=l/2+l_1} = -h'\left(T\big|_{x=l/2+l_1} - T_e\right)$.

Where the heat flux is $J = -\kappa\nabla T$, κ is thermal conductivity of the corresponding plate. Value *h'* is the heat transfer coefficient between the external plates and the environment, c_i and c_e are the heat capacities per unit of mass for internal and external plates. We use symmetry and take into account that the flux of defects on boundary is equal to zero because the absorption of defects by the plate surface compared to their absorption by the internal sinks is neglected. The nonlinear terms (the third in right side of (1) and (2)) connect these equations and describe the nonlinear feedback between the defect density, the rate of annealing and the temperature.

If the internal plate is so thin that its temperature and defect density inside it are approximately constants we can average the equations (1) – (2) and find the heat flux from the equation (3). So the average defect density *n* and average internal plate temperature *T* are described by the system of equations.

$$\frac{dT}{dt} = \frac{1}{c_i}(Q + \theta\beta(T)n - h(T - T_e)) \tag{4}$$

$$\frac{dn}{dt} = K - \beta(T)n \tag{5}$$

Here $h = \dfrac{\kappa_e h'}{2l(\kappa_e + h' l_1)}$. The heating rate is proportional to the intensity of the irradiation, and therefore it is proportional to rate of defect generation: $Q = \xi\theta K$. Parameter ξ is ratio of energy of irradiation which transforms into heating and energy of irradiation which transforms into defect generation.

III. Stationary regime and self-oscillations

The system (4) - (5) is nonlinear as result of the exponential dependence of β on the temperature.

There is the only one possible stationary solution of (4) - (5), that describes the stationary homogeneous density of defects under irradiation,

$$T_s = T_e + \frac{2l(\kappa_e + h' l_1)\theta(\xi + 1)K}{\kappa_e h'} \tag{6}$$

$$n_s = K / \beta(T_s) \tag{7}$$

The stationary solution (6) – (7) takes place physically if it is stable. To exam stability let us consider the evolution of its small perturbations δn and δT. The damping decrement of the small perturbations satisfies equation

$$\lambda^2 + p\lambda + q = 0 \tag{8}$$

Where

$$p = \frac{\kappa_e h' c_i}{2l(\kappa_e + h' l_1)} - K\theta E_m / c_i T_s^2 + \beta(T_s) \tag{9}$$

$$q = \frac{\kappa_e h' \beta(T_s)}{2l c_i(\kappa_e + h' l_1)} \tag{10}$$

The value of *q* is positive for all physically admissible values. The value of *p* has variable sign. If $K \to \infty$ and $T_e \to \infty$, then $p > 0$ and therefore $\mathrm{Re}\lambda < 0$. So the stationary distribution is stable. With decreasing values of *K* and T_e condition $p > 0$ can be broken.

So for as for any parameters there is a loop without contact which covers the stationary point and all phase trajectories of the system (4) - (5) go in the loop, there is a limit cycle of the system (4) - (5). So self-oscillations of temperature, defect density and heat transfer are developed.

The self-oscillations are developed if inequality

$$2l(\kappa_e + h' l_1)\theta E_m \kappa_e h' K \geq$$
$$\left(\kappa_e h' + 2l(\kappa_e + h' l_1)c_i \beta(T_s)\right)\left(\kappa_e h' T_e + 2l(\kappa_e + h' l_1)\theta(\xi + 1)K\right)^2 \quad (10)$$

is satisfied.

Let all parameters are constants except the environmental temperature (T_e) and the defect production rate (K). To obtain the stability diagram (see Fig.1 and Fig.2) let us divide the space of these parameters into two parts. For parameters from the first part the stationary distribution of defects takes place under irradiation. The second part is field of instability of the stationary distribution. For parameters from this field the self-oscillations are developed. The parametric equations for bifurcation curve are the following

$$K = T^2\left(\kappa_e h' + 2l(\kappa_e + h' l_1)c_i \beta(T)\right) \Big/ 2l(\kappa_e + h' l_1)\theta E_m \quad (12)$$

$$T_e = T - T^2(\xi + 1)\left(\kappa_e h' + 2l(\kappa_e + h' l_1)c_i \beta(T)\right) \Big/ \kappa_e h' E_m \quad (13)$$

where sample temperature T is parameter.

The bifurcation value of environment temperature is limited from above, since the second term in (9) for large values of T begins to dominate. The necessary condition for the self-oscillations is irradiation, but their parameters are first of all determined by feed-backs of the system.

A period of oscillations near the bifurcation curve is

$$\tau = 2\pi \sqrt{\frac{2l(\kappa_e + h' l_1)c_i}{\kappa_e h' \beta}} \quad (14)$$

Thus, the period of oscillation is the square root of the product of the defect lifetime and the characteristic time of sample cooling.

IV. Results and Discussion

Oscillations of different properties under stationary irradiation are observed often enough and well known.

Jumps and oscillations of conductivity which is indicator of defect density were observed in metals, semiconductors and ceramic materials [5]-[9].

Regular electrical oscillations with a period is approximately equal to 100 hours and an amplitude is equal to 30% of the mean value was observed during irradiation of aluminum and sapphire samples with 1 keV neutron flux ($2 \cdot 10^{16}$ - $7.2 \cdot 10^{16}$ $m^{-2} \cdot s^{-1}$) at 615 C [5].

The oscillations of resistivity with a frequency of about 10^{-5} s^{-1} was observed during irradiation of a copper sample with 2.2 MeV electrons (22.4 $mA \cdot m^{-2}$) at a temperature of 103 K [6].

Also periodic variation of the electrical resistivity during irradiation gold specimen by 2.3 MeV electrons (11.5 $mA \cdot m^{-2}$) at a temperature 197 K is given in [6] too. The oscillations with a frequency of about $2 \cdot 10^{-4}$ s^{-1} are modulated by sinusoidal wave with frequency of about 10^{-5} s^{-1}.

Periodic variation of microhardness of Nimonic 90 containing γ-precipitates as a function of dose was observed in [9] during irradiation with 50 keV helium ions at temperature near 300K and defect generation rate of the order of 10^{-2} dpa/s.

Not monotone dependence of the relative change of microhardness of 79 Permaloy on dose was observed in [10].

It is well known the cases of non-monotonic dose dependence of the radiation creep [11], [12]. Creep of loaded steel samples during irradiation by neutrons ($2 \cdot 10^{18}$ $m^{-2} \cdot s^{-1}$) at 350 C and defect generation rate $1.7 \cdot 10^{-7}$ dpa/s looks like damped oscillations with a period of about 10^7 s.

Variations of temperature of organic crystals CH_4 at the reactor irradiation were observed for the environment temperature below 120 K [13].

Finally, oscillations of the void size were observed in nickel samples irradiated with 180 keV nickel ions at 750 C and $5 \cdot 10^{-2}$ dpa/s. Period of the oscillations was of about 10^3 s. [14].

As a rule these oscillations are not strictly periodic. This can be explained by change of all elements of microstructure during irradiation. All of them are connected. Characteristic times for change of different elements are quite different. During some time period this allows to consider them separately and oscillations can look like as periodic function. But slow change of parameters of microstructure leads to change of period and amplitude of oscillations. Thus in general case the development of microstructure under irradiation cannot be described only by the stationary or periodical functions.

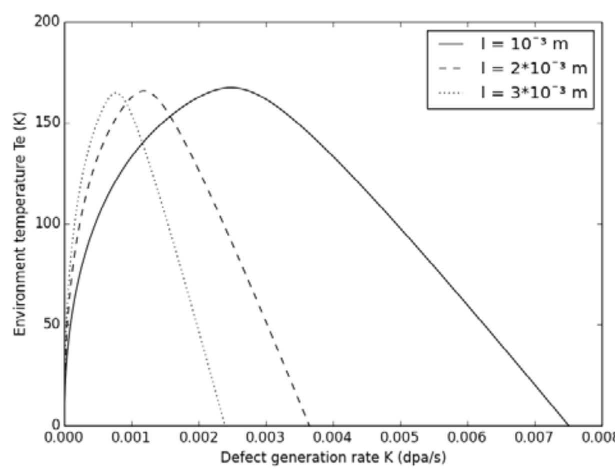


Fig 1. The region of instability for Si-Al-Si sample for different internal plate thicknesses ($l_1 = 10^{-3}$ m, $h' = 3000$ $W \cdot m^{-2} \cdot K^{-1}$, $\kappa_e = 149$ $W \cdot m^{-1} \cdot K^{-1}$, $\rho_d = 10^{14}$ m^{-2}, $\xi = 9$, $D_0 = 3 \cdot 10^{-5}$ $m^{-2} \cdot s^{-1}$, $\theta = 0.66$ eV ($1.05 \cdot 10^{-19}$ J), $E_m =$ 0.62 eV ($9.93 \cdot 10^{-20}$ J)).

Computer simulation of self-oscillations for our simplest model shows that the frequency of oscillation is about 10^{-5} – 10^{-2} s^{-1}. The frequency of oscillation depends on the pre-exponential factor of the diffusion coefficient and practically

ISBN: 978-1-61804-230-9

don't depend on the energy of the defect migration and the energy of the defect formation. In materials with higher density sinks the region of instability is less and the frequencies of the self-oscillations are higher.

Stability diagrams for different metals are similar. The highest environment temperatures at which self-oscillations develops in different metals are about 100 – 200 K, at defect generation rate about 10^{-3} dpa/s. Temperature of the sample for these parameters is about 300 K.

If ratio ξ increases, the region of instability expands and the frequency of the self-oscillations increases too. The region of the instability increases for more complex systems, for which one should take into account formation of secondary defects, e.g. complexes of the defects, voids and so on. During development of the self-oscillations the temperature of the sample and the defect density can exceed the steady-state value in several times.

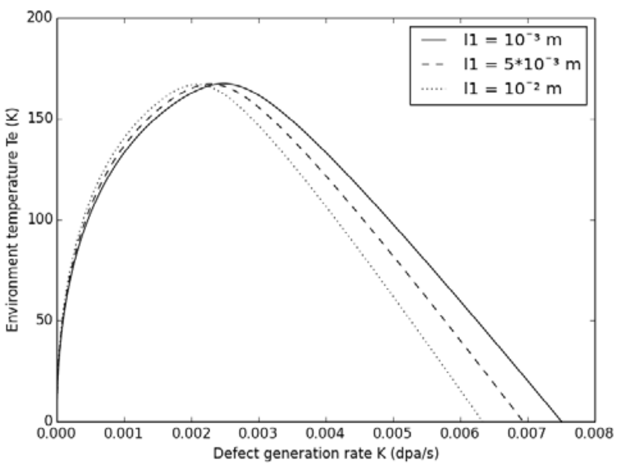


Fig 2. The region of instability for Si-Al-Si sample for different external plate thicknesses ($l = 10^{-3}$ m, $h' = 3000$ W·m^{-2}·K^{-1}, $\kappa_e = 149$ W·m^{-1}·K^{-1}, $\rho_d = 10^{14}$ m^{-2}, $\xi = 9$, $D_0 = 3\cdot10^{-5}$ m^{-2}·s^{-1}, $\theta = 0.66$ eV ($1.05\cdot10^{-19}$ J), $E_m = 0.62$ eV ($9.93\cdot10^{-20}$ J)).

When heat conductivity of the external plates is high enough or thickness of the external plates is small enough, development of self-oscillations doesn't depend on their properties (except heat exchange between surface and environment). This allows changing characteristics of self-oscillations by forming thin surface layers on the irradiated sample.

The period of oscillation depends on the properties of internal and external layers. If the heat capacity or the thickness of internal layer is reduced, the period of self-oscillations decreases. The reduction of the thickness of the external layer leads to reduce of the period. The period increases with decreasing the heat conductivity of the external layer and with increasing of the heat transfer coefficient.

Region of existence of self-oscillations becomes narrower and moves to higher rates of defect generation with decreasing thickness of the internal layer (see Fig.1). Self-oscillations develop if the rate of defect generation belongs to narrow range near

$$K = h' E_m \big/ 4(\xi + 1)^2 \theta \tag{15}$$

and the environmental temperature do not exceed

$$T_e = E_m \big/ 4(\xi + 1) \tag{16}$$

When the thickness of the internal layer grows or heat exchange with environment decreases, the region of existence of self-oscillations shrinks and disappears and self-oscillations do not develop (see Fig.2).

V. Conclusion

The avalanche increase of the temperature and defect density can occur under certain opportune stationary conditions of irradiation. The reason of Hopf bifurcation and self-oscillations is nonlinear coupling between temperature and defect density. During the non-stationary evolution defect density and temperature can reach much higher values than the expected steady-state values.

References

[1] G. S. Was, *Fundamentals of Radiation Materials Science*. Berlin Heidelberg New York: Springer 2010.

[2] P.A. Selyshchev, *Self-organization phenomena in materials under irradiation*, Moscow: NITs "Regulyarnaya i haoticheskaya dinamika" 2008.

[3] P. A. Selishchev, V. I. Sugakov "Auto-oscillations of temperature and defect density in thin plates under irradiation". *Fizika tverdogo tela*, -, vol.30, pp.2611-2619, 1988.

[4] P. A. Selishchev, V. I. Sugakov "Auto-oscillation process in impure crystals under irradiation" in *Effects of Radiation on Materials*, R. E. Stoller, Arvind S. Kumar, and David S. Gelles, Ed. ASTM, Philadelphia, 1992, pp.594-599.

[5] E. H. Farnum, F. W. Clinard, W. F. Sommer, J.C. Kennedy, and T. Shikama, "Search for radiation-induced electrical degradation in alumina during spallation-neutron irradiation" *J. Nucl. Mater*.,vol. 212-215, pp. 1128-1132, Sep. 1994.

[6] W. Schule "Radiation-enhanced diffusion due to intersttials and dynamic crowdions", *J. Nucl. Mater*., vol.233-237, pp.964-968, 1996.

[7] P. Sen, G. Aggarwal, Tiwari U. "Dissipative structure formation in could-rolled Fe and Ni during heavy ion irradiation", *Phys. Rev. Letters*, vol.80, pp.1128-1132, 1998.

[8] C. Kinoshita, S. J. Zinkle" Potential and limitations of ceramics in terms of structural and electrical integrity in fusion environments", *J. Nucl. Mater*., vol.233-237, pp. 100-110, 1996.

[9] K. Varatharajan and R.V. Nandedkar "Microhardness-microstructure study of aged nimonic 90irradiated with helium" in *Effects of Radiation on Materials*. R.E.Stoller Edr. - Philodelphia, 1989, pp.263-270.

[10] D. I. Tetelbaum, E. V. Kurilchik, and N. D. Latisheva "Lonng-range effect at low-dose ion and electron irradiation of metals", *Nucl.Inst. and Meth. in Phys. Research B*, vol.127-128, pp. 153-156, 1997.

[11] M. B. Toloczko, F. A. Garner, and C. R. Eiholzer "Irradiation creep and swelling of the US fusion heats of HT9 and 9Cr-1Mo to 208 dpa at ~ 400^{o}C", *J. Nucl.Mater*., vol.212-215, pp.604-607, 1994.

[12] W. Schule and H. Hausen "Neutron irradiation creep in stainless steel alloys", *J. Nucl.Mater*., vol.212-215, pp. 388 - 392, 1994.

[13] J. M. Carpenter "Thermally activated release of stored chemical energy in cryogenic media", *Nature*, vol.36, pp. 358-360, 1987.

[14] J.K. Steele and Potter D.I. "The disappearance of voids during 180 keV Ni^+ bombardment of nickel", *J. Nucl.Mater*., vol.218, pp. 95-107, 1995.

Qualitative information method of an assessment of scientific and innovative projects during implementation of the industrial and innovative program in Kazakhstan

G. Mutanov, and Zh. Yessengalieva

***Abstract*—**During an era of industrial and innovative development of Kazakhstan new methods of an assessment of scientific and innovative projects are very necessary. In this article the qualitative information method for an estimation of scientific and innovative projects is developed. Criteria for assessing innovativeness and competitiveness are specified. The qualitative information method consists of three stages: assessment by criteria of innovativeness and competitiveness of scientific and innovative projects, application of genetic algorithm, an evaluation of economic efficiency of the project. This method is designed to be utilized by expert commissions needing to select appropriate innovative projects.

***Keywords*—** assessment, economic efficiency, genetic algorithm, innovative projects.

I. Introduction

Any scientific and innovative project is a complex system of actions that are interdependent but are interconnected by resources, time, and performers and are aimed at achieving specific targets in priority areas of development of science and technology [1].

Scientific and innovative projects are, in effect, long-term investment projects characterized by a high degree of uncertainty as to their future outcomes and by the need to commit significant material and financial resources in the course of their implementation [2].

Uncertainty is inherent to all stages of scientific and innovative project cycle: the initial phase of developing an idea, when selecting a project, and again when implementing it [3]. Moreover, it may well happen that novelties that have successfully passed the testing phase and have found a manufacturing application, are then rejected by the market and their production must be stopped [4].

Even the most successful scientific and innovative projects are not foolproof. At any time of their life cycle they are vulnerable to the advent of a more promising novelty offered by a competitor [5]-[7].

It is also quite characteristic of scientific and innovative project – as compared to an investment project – that modified alternative options can be developed during any stage of its life cycle. In case of a long-term investment project, only one option is selected to be implemented, while the scientific and innovative project requires that reevaluations and revisions be carried out at every implementation stage using numerous benchmarks and milestones. In fact, any innovation is characterized by its alternative nature, uncertainty[6], and availability of many options at all phases. Therefore it is quite a challenge to forecast innovation behavior[5], since this task entails assessing the integral performance index, projected future competitiveness, and adaptation to the market.

Practice shows that while 10 projects have been thoroughly vetted and launched, 4 or 5 of them result in total failure, 3 to 4 result in setting up viable companies that do not yield any tangible profit, and only 1 or 2 bring really good outcomes. It is due to the success of such projects that investors – on the average – get a high rate of return (venture funds, etc.) [8],[9].

As compared to investment projects, the scientific and innovative projects have the following specific features:

– Higher uncertainty as to future costs, period of achieving the intended targets, and future revenues; all these factors affect the accuracy and reliability of preliminary financial and economic assessments and suggest that additional criteria should be used for project appraisal and selection [6];

– When developing an innovation project, the time factor has to be taken into account to a greater extent;

– Scientific and innovative projects have certain advantages as compared to investment projects, in that they can be terminated without significant financial losses[9];

– Spin-off results of the research involved in innovation projects can be of commercial value, above and beyond the value of the project itself [10].

G. Mutanov, al-Farabi Kazakh National University, Almaty, Kazakhstan, 050040 (phone: +7727-377-3320; fax: +7727-377-3344; e-mail: Rector@ kaznu.kz).

Zh. Yessengaliyeva, al-Farabi Kazakh National University, Almaty, Kazakhstan, 050040 (corresponding author to provide phone: +7727-377-3320; fax: +7727-377-3344; e-mail: Zhanna_Serzhan_7@mail.ru).

ISBN: 978-1-61804-230-9

Thus, the scientific and innovative project must be considered as a complex of interrelated goals and objectives, each with an implementation plan, and it is therefore necessary that a more detailed analysis be carried out of all project stages including operational management, and that strict control be exercised over its implementation.

II. Materials and Methods

During an era of industrial and innovative development of Kazakhstan an assessment [11] of the scientific and innovative project is an important procedure at the research and development stage [12]. It is a continuous process that implies a possible suspension or termination of a project at any point of time when new information is obtained.

In this section, we present qualitative information method of an assessment of scientific and innovative projects, whether a project is feasible or not, the project must go through several stages of examination, which are shown in Fig.1.

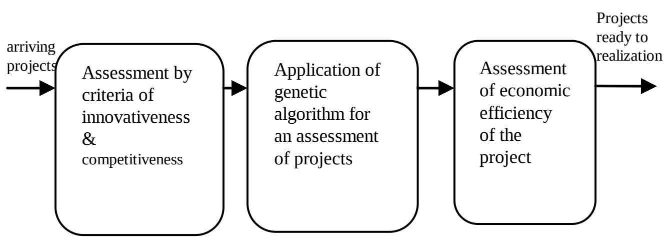


Fig.1 Scientific and innovative project appraisal stages

Here we propose a method of assessment of scientific and innovative projects referred to the scientific, technical, and industrial sector, with a system of target indicators[13],[14].

In first stage we used a methodological approach based on expert assessment of innovation and competitiveness indicators for projects, accompanied by a graphic model of project innovativeness and competitiveness assessment. Adequacy of the criteria for the complex index is determined by assigning weights to each criterion and using an additive–multiplicative method of calculation. Project assessment[15], [16], based on the graphic model for assessing project innovativeness and competitiveness, should be carried out in three steps: a) selecting optimal criteria, b) determining weight coefficients, and c) positioning projects in the matrix. This method is in detail described in work [17].

From the perspective of the market, scientific and innovative projects are the objects of two interacting segments: science and business. Therefore, they should be formalized as two-dimensional objects: innovativeness (I) and competitiveness (K). The main distinctive feature of these indicators is that they are considered as an assessment of project viability and attractiveness to investors and depend on numerous criteria. A set of these criteria is presented in Table 1, and Table 2.

Table 1. Criteria of innovativeness

#	Criteria of innovativeness
1	Relevance of research, uniqueness and prospects of the project (absence of analogs).
2	Scientific novelty, scientific and technical level, technological level (new technology), prospects of solutions proposed in the project.
3	Assessment of applied methodology for scientific researches and technological level of the project.
4	Degree of a scientific and methodological, technical and technological readiness of the project.
5	Assessment of expected results of scientific researches/projects.
6	Assessment of possibility of performance/creation of future research and development on the basis of this scientific research.
7	Assessment of probability of scientific/technical/technological success of the project.
8	Assessment of advantage of the project in comparison with existing analogs in the world.
9	Assessment of compliance of a quoting of the used scientific works and literature, scientific and developmental development.

Table 2. Criteria of competitiveness

#	Criteria of competitiveness
	group of marketing and financial criteria
1	Assessment of economic validity of the required amount of financing according to the estimate of expenses and the planned schedule of operation/time of development of the project.
2	Assessment of probability of commercial success of scientific/technical/technological result of the research/project.
3	Assessment of investment appeal of scientific/technical/technological result of the research/project.
4	Assessment of the potential annual size of profit of the offered project.
5	Assessment of compliance of the project to accurately certain requirements of the consumer market and society.
	group of production and scientific and technical security
1	Assessment of personnel security with professional qualification and experience of works in the field of the offered research/project.
2	Assessment of material security of the research/project.
3	Assessment of degree of patent and license security of the research/project.
4	Assessment of the scientific and practical importance of proposed solutions of the research/project.
5	Assessment of the social and economic importance and level of ecological effect on society of the offered research/project.
6	Assessment of level of import substitution, increase of an export potential of the country of proposed solutions of this research/project.

In second stage of qualitative information method we use genetic algorithm for an assessment of projects. Research of application of the theory of genetic algorithms[18] to procedure of an assessment of scientific and innovative projects allowed considering influence of last skilled projects on assessment examination, as from the point of view of investment expenses, and innovative appeal.

It is necessary to understand set of the phenomena forming new qualities, the enterprises/countries promoting economic development as innovative appeal as a whole.

In this regard procedure of an appraisal of projects with application of genetic algorithm is offered. The result of modeling of function of fitness is given below. The index of innovative appeal is thus expressed in a percentage ratio on a formula (1):

$$f(x) = (\sum_{i=1}^{n} I_j + \sum_{k=1}^{m} K_j) * IA_j \rightarrow \max \qquad (1)$$

$0 \le IA_j \le 0{,}4$ *(the worst projects)*
$IA_j = 0{,}5$ *(offered new projects)*
$0{,}6 \le IA_j \le 1$ *(elite projects)*

where I – innovativeness of j-project,
K – competitiveness of j-project,
IAj – an index of innovative appeal of j-project, the characterizing percent of successful introduction of an elite element of population,
n – quantity of indicators of innovativeness,
m – quantity of indicators of competitiveness.

The index of innovative appeal is the factor, allowing to consider degree of riskiness of investment of the project, based on knowledge of the "survived" projects.

In the presented algorithm the point-to-point crossover is used, where two chromosomes *A* and *B* are cut on three parts in incidentally chosen two points [19], exchange them, generating two new chromosomes:

$A=(A_1,A_2,A_3)$ and $B=(B_1,B_2,B_3)$
$A'=(A_1,B_2,A_3)$ and $B'=(B_1, A_2,B_2)$.

Set of genetic operators, such as selection, crossover, mutation and inversion, an assessment of suitability is called as generation iteration.

After each step of iteration on which all individuals are mutated and exposed to a crossover, for each of new chromosomes value of function of fitness is calculated. Than value for this chromosome, subjects with bigger probability it there is more it is selected for a crossover. After end of the set number of iterations the chromosome with the maximum value of function of fitness gets out as the decision at selection of the most favorable projects.

Thus, using elitism strategy we lead a qualitative set of projects, and we improve quality of a choice, thereby promoting purposeful distribution and financing of investments in scientific and innovative projects in the priority directions of national economy of Kazakhstan.

The third stage of qualitative information method is to determine the economic viability of the project, the methodology of which is described in the work [20]-[23].

The method of an assessment of efficiency of projects is based on following methods: Net present value (NPV) method of project evaluation; Profitability index (PI) method to estimate investment profitability; Internal rate of return (IRR); and Payback period (PB) [20]-[23].

Thus, readiness of the scientific and innovative project for realization is defined by developed qualitative information method which allows making an objective assessment of offered projects taking into account opinions of experts of former years.

III. Experimental Results

We made experiment on the proposed method on the basis of five scientific and innovative projects. Fig. 2 demonstrates graphic model for assessment of project innovativeness and competitiveness, as a result of first stage of qualitative information method. The proposed projects were positioned according to expert assessments received.

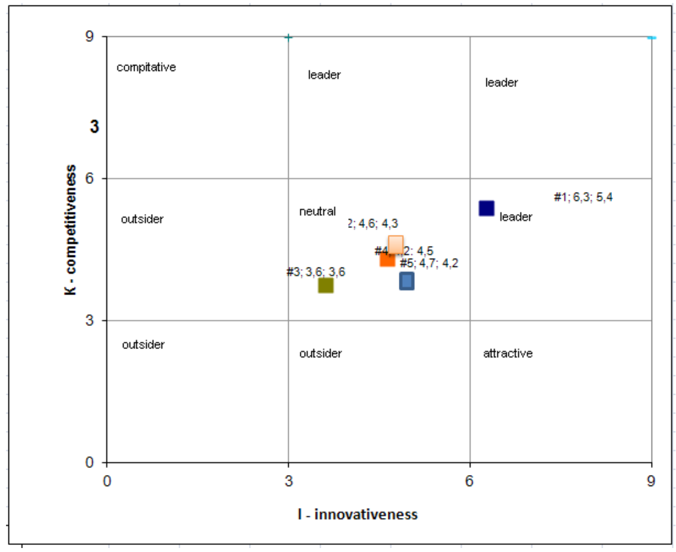


Fig.2 Example of project positioning in graphic model for assessing project innovativeness and competitiveness

The genetic algorithm offered at the second stage allows to consider experience of the previous realized projects which are a part of elite individuals that leads to decrease in degree of risk of investment. Experimental and theoretical researches of genetic algorithm are shown in the Table 3. The behavior neutral, in expert opinion, projects can lead to good results on the basis of available knowledge in a databank of projects that is distinctive feature of this qualitative information method in comparison with the carried-out works of other scientists in the field.

Table 3. Results of experiment

Offered new projects	Probability of a choice of the project	Share
Project #1	0,19	19%
Project #2	0,09	9%
Project #3	0,20	20%
Project #4	0,37	37%
Project #5	0	0%

According to Table 3, the diagram of scientific results is constructed.

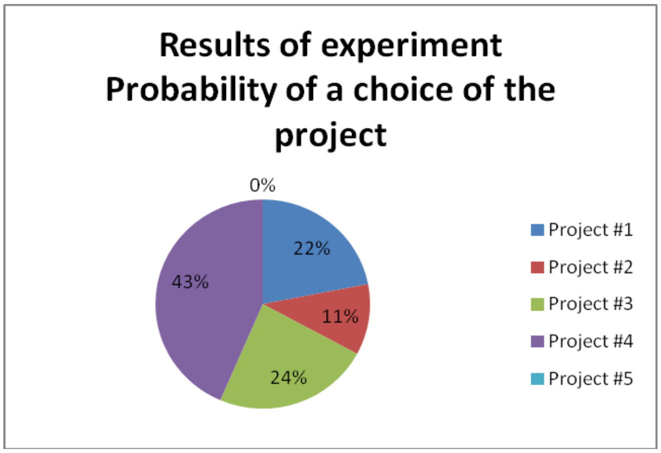


Fig.3 Results of computing experiment of genetic algorithm in a context of estimates of scientific and innovative projects

At the third stage following the results of the previous stages the project #1 and the project #4 are selected. The project examination procedure is the economic effect analysis of a project using project life cycles.

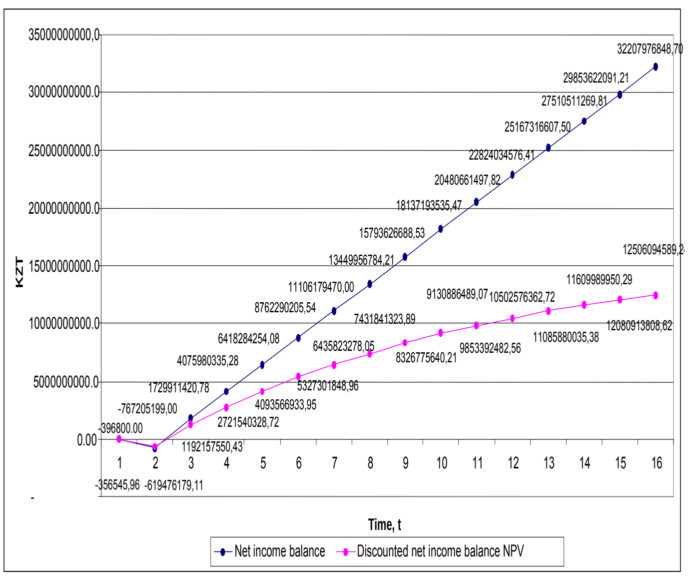


Fig. 4 A sample of project implementation as represented by the graphic model of feasibility and cost effectiveness of scientific and innovative project

Therefore, following the results of computing experiments on the standard method the projects has sufficiently high indices of effectiveness and can be accepted for implementation.

IV. Discussion and Conclusion

To achieve competitive advantage, firms must find new ways to compete in their niche to enter the market in the only possible way: through innovation. Thus innovation equally includes R&D results of production purposes, and results geared at improving the organizational structure of production [3]. Our results suggest that R&D efforts play an important role [17] in affecting product innovation. In order to protect the safety of financial investments in conditions of information uncertainty, as a solution, to select alternative scientific and innovative projects this qualitative information method is developed.

The method presented here can be used by experts of venture funds, institutes, and other potential investors to meaningfully assess the scientific and innovative projects.

Acknowledgment

We gratefully acknowledge the financial support of the Committee of Science of The Ministry of Education and Science of Republic of Kazakhstan which enabled this study. Grant number 1150 / GF 2012-2014.

References

[1] Bos-Brouwers HEJ. Corporate Sustainability and Innovation in SMEs: Evidence of Themes and Activities in Practice. *Business Strategy and the Environment* 2010; **19**: 417-35.

[2] Hoegl M, Gemuenden HG. Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science* 2001; **12**: 435-49.

[3] Laupacis A, Feeny D, Detsky AS, Tugwell PX. HOW ATTRACTIVE DOES A NEW TECHNOLOGY HAVE TO BE TO WARRANT ADOPTION AND UTILIZATION - TENTATIVE GUIDELINES FOR USING CLINICAL AND ECONOMIC EVALUATIONS. *Canadian Medical Association Journal* 1992; **146**: 473-81.

[4] Stafinski T, McCabe CJ, Menon D. Funding the Unfundable Mechanisms for Managing Uncertainty in Decisions on the Introduction of New and Innovative Technologies into Healthcare Systems. *Pharmacoeconomics* 2010; **28**: 113-42.

[5] Westergren UH. Opening up innovation: the impact of contextual factors on the co-creation of IT-enabled value adding services within the manufacturing industry. *Information Systems and E-Business Management* 2011; **9**: 223-45.

[6] Wouters M, Roorda B, Gal R. MANAGING UNCERTAINTY DURING R&D PROJECTS: A CASE STUDY. *Research-Technology Management* 2011; **54**: 37-46.

[7] Zakharchyn GM. INTEGRATED DIMENSION OF INNOVATIVENESS PROBLEMS WITHIN INFORMATION SOCIETY. *Actual Problems of Economics* 2012: 10-5.

[8] Alpkan L, Bulut C, Gunday G, Ulusoy G, Kilic K. Organizational support for intrapreneurship and its interaction with human capital to enhance innovative performance. *Management Decision* 2010; **48**: 732-55.

[9] Aubrecht C, Ozceylan D. Identification of heat risk patterns in the U.S. National Capital Region by integrating heat stress and related vulnerability. *Environment International* 2013; **56**: 65-77.

[10] Brooks N, Adger WN, Kelly PM. The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation. *Global Environmental Change-Human and Policy Dimensions* 2005; **15**: 151-63.

[11] Cronin JJ, Brady MK, Hult GTM. Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing* 2000; **76**: 193-218.

[12] Das S, Suganthan PN. Differential Evolution: A Survey of the State-of-the-Art. *Ieee Transactions on Evolutionary Computation* 2011; **15**: 4-31.

[13] Gandomi AH, Yang X-S, Alavi AH. Mixed variable structural optimization using Firefly Algorithm. *Computers & Structures* 2011; **89**: 2325-36.

[14] Gerbing DW, Anderson JC. AN UPDATED PARADIGM FOR SCALE DEVELOPMENT INCORPORATING UNIDIMENSIONALITY AND ITS ASSESSMENT. *Journal of Marketing Research* 1988; **25**: 186-92.

[15] Hoang H, Rothaermel FT. LEVERAGING INTERNAL AND EXTERNAL EXPERIENCE: EXPLORATION, EXPLOITATION, AND R&D PROJECT PERFORMANCE. *Strategic Management Journal* 2010; **31**: 734-58.

[16] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 2011; **28**: 2731-9.

[17] G. Mutanov , Zh. Yessengaliyeva, Development of method and models for assessing innovativeness and competitiveness of scientific-innovative projects. World Applied Sciences Journal 23(9): P.1192-1200, 2013. ISSN 1818-4952. ©IDOSI Publications, 2013 http://www.idosi.org/wasj/wasj23(9)2013.htm.

[18] Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 2012; **22**: 568-76.

[19] Marano GC, Quaranta G, Monti G. Modified Genetic Algorithm for the Dynamic Identification of Structural Systems Using Incomplete Measurements. *Computer-Aided Civil and Infrastructure Engineering* 2011; **26**: 92-110.

[20] Lambin EF, Meyfroidt P. Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences of the United States of America* 2011; **108**: 3465-72.

[21] Picazo-Tadeo AJ, Beltran-Esteve M, Gomez-Limon JA. Assessing eco-efficiency with directional distance functions. *European Journal of Operational Research* 2012; **220**: 798-809.

[22] Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. *Jama-Journal of the American Medical Association* 1996; **276**: 1339-41.

[23] Sueyoshi T, Goto M. Methodological comparison between two unified (operational and environmental) efficiency measurements for environmental assessment. *European Journal of Operational Research* 2011; **210**: 684-93.

Mutanov Galim graduated from Kazakh Polytechnic institute in 1979 from the specialty “Automation and teleautomatics”. He received post-graduate education, doctorate at the department of “Automated controlling systems” in Moscow State Mining university. Doctor dissertation was devoted to the theory and practice of automated control of ethnological processes with the use of elements of artificial intellect. This author became a Member of Econometric Society, and a Member of Royal Economic Society since 2013.

In July 2003 he was appointed the rector of EKSTU in honor of D. Serikbayev. Innovation, knowledge quality, integration into world educational space this is the most important direction for the university. At a short time constitutive innovative reforms were made that contribute to efficient innovative development of the whole region.

Prof. Mutanov is the president of Central-Asian department and a member of monitoring committee of International society of engineering pedagogics IGIP, (Sweden), the President of Kazakhstan society (KazSEE) in International federation of engineering education (IFEES), a member of the Committee of managers of the institute of information technologies in education UNESCO (Paris), holder of the title “European professor of a technical university, scholarship of the National Academy of Science of the Republic of Kazakhstan for outstanding scientists, who has made significant contribution to the development of the republic and actively carrying out scientific research on priority direction for Kazakhstan. G. Mutanov was awarded a silver medal of International bibliographic centre (Cambridge), a gold medal "The United Europe" (Oxford). The chairman of board of directors of JSC “Centre of engineering and technologies transfer”, the Chairman of social consulting Committee JSC “National innovative fund”, a member of regional political council of people’s-democratic party “Nur-Otan”, the Chairman of the Committee on budget and economics of East-Kazakhstan regional maslihat.

Disruptive innovation and its implications on Lebanese telecom industry

Wael Bakhit
LebaneseUniversity, Faculty of Economics and Management
Lebanese French University, Faculty of Management

Abstract **—**The level of innovation in the telecommunication industry progressed over the last decade. It provides new challenges that are assumed to disturb the sustainability of telecommunication companies (service of phone). Mobile phone companies are likely to be directly affected by rapid mutation occurred in consumers' spending habits closely linked to a variety of technological innovation (new products and services). Data growth and associated growth in new services and media will drive the bulk of new revenues for telecom operators. As this comes at the sacrifice of having to accept lower margins, operators need to decide how far to go and how. Through a survey we developed on the basis of the Lebanese telecom sector (Alfa / Touch), we test the reliability of innovation and its impact on profitability.

Keywords: Disruptive innovation, Telecommunication, Smartphones, Value Migration

I. INTRODUCTION

The mobile telecom industry is changing and the competitive landscape for mobile network operators has been disrupted. The industry is shifting from an environment characterized by reliability and scale of networks, to an environment where choice and flexibility of services is more prominent. This has changed the basis of competition and represents the shift from “mobile telephony” to “mobile computing” (VisionMobile, 2012a). Today, this sector plays a key role in the Lebanese economy even if it’s performing below potential despite growth. According to Business Monitor International (BMI, 2013), Lebanon is one of a few countries in the Middle East and North Africa Region with mobile phone penetration rate of less than 100% at the end of 2012 despite having a relatively small population and high urbanization level. It attributed this trend to the lack of competition in the market, which has kept tariffs relatively high and delayed the roll out of network services to underserved areas. It forecast mobile phone penetration in Lebanon to reach 100.8 subscriptions per 100 inhabitants in 2013 and to rise to 110.1 subscriptions per 100 inhabitants in 2017. BMI concludes in (2013) expecting the mobile phone sector in Lebanon to post steady growth between 2013 and 2017 despite relatively uncompetitive tariffs and poor quality of service. It considered that Lebanon's young and well-educated population will continue to drive demand for advanced communication services. (BMI, 2013)

Therefore, mobile network operators launched third-generation (3G) mobile networks in 2001 and the high-speed wireless data transfer enabled the mobile network operators to distribute more services (Steinbock, 2005). The mobile network operators worldwide attempted to control the services in closed ecosystem *portals* (VisionMobile, 2011a). The portals gave the mobile network operators the opportunity to build empires and lock-in the customers by restricting them from accessing potential competitors services. The development of 3G technologies provided a foundation for the upcoming smartphone era and in 2007 Apple launched the iPhone (Sharma, Operator’s dilemma (and opportunity): the 4th wave, 2012a). Smartphones had earlier been used by a small segment of primarily business people, but the iPhone managed to target mainstream customers. According to Ahonen (2011) the introduction of iPhone transformed the industry and he states the demarcation of time in the mobile telecom industry as “before iPhone” and “after iPhone”.

The smartphone era has enabled dozens of new services for phones (Sharma, 2012a), which were boosted by touchscreens (Salz, 2011). Content of smartphones is controlled by platforms such as Google Android and Apple iOS, and not by the mobile network operators. It has caused the mobile network operators portals to decline rapidly and made them loose some of their interaction with end-users (VisionMobile, 2011a). The mobile network operators have previously been able to generate large profits and maintain high margins from the traditional voice and messaging services. But in the smartphone era the profits from voice services have are stopped growing and even declined for some mobile network operators. The usage of data access on the other hand has increased rapidly due to the smartphone usage (Sharma, 2012a).

Indeed, the innovation in mobile telecom industry is accelerating at a breakneck pace. New telecommunications technology significantly reduces the barriers to entry in the market and eliminating middlemen, allowing businesses to interact directly with their customers around the globe. The advent of the converged voice, data and video technologies mean that media, entertainment, computer and telecommunications organizations will all be merging and interoperability of their activities. In the last five years we have witnessed to a particular growing body of research regarding the importance of innovation called disruptive and its impact on financial performance. Mobile telecom has become critical to drive technological growth, and it impacts how humans communicate and interact in everyday life (Sharma, 2012a). The mobile network operators have been affected by disruptive innovation as it will be described by Christensen (2007). The objective of our study is to show how the Lebanese mobile operators, faced with the technological innovations and free services on the telecom market, can profit in a constant manner and ensure their continuity. We believe that understanding the impact of disruptive innovation for Lebanese mobile telecom industry is the more judicious approach for the purposes of this study.

- What are the problems of the telecommunications to Lebanon and its challenges
- Can the LMO ensure their competitive place in the presence of free applications on the telecom market?
- What are the solutions and the policies adopted by these companies to protect their market share?

A. Definition of Disruptive innovation

Given the complexity of innovation activity, it seems difficult to find a universal definition. According to Schumpeter (1935), the realization of an invention and the implementation of corresponding innovation are economically and sociologically two entirely different things. As such, Alter (2002) describes the invention such as the creation of a technical or organizational novelty, concerning goods, services, or devices, while innovation represents the entire social and economic process bringing the invention to be ultimately used or not. The term 'innovation' applies both to the result of a creative process (which is new), and this same process (Mayrhofer, 2011). According to Cantwell (2010), innovation can be explained as the introduction of new products and processes (process). The main difference between product innovation and innovation process (processes): the first relates to the product or service marketed, particularly in terms of functionality, the second characterizes the manner in which this offer is developed and distributed, particularly in terms of costs and qualities (Johnson, 2011).

Hence, Christensen explains (2013) that disruptive innovations are products or services with business models that introduce performance packages that are inferior to what mainstream customers value. During the early development of a disruptive innovation it only serves niche segments. Both the disruptive innovation and the established offerings improve; nevertheless the disruptive innovation improves enough over time to satisfy the mainstream customers and eventually replaces the established offerings and incumbents that exceed the demanded performance, see Figure 1 (Christensen C. M., 2013).

B. Problem of the research

Disruptive innovation has completely reshaped numerous companies and industries, and caused companies to fail while other flourish. Many firms need to periodically engage in processes of disruptive innovation for long-term survival (Christensen &Raynor, 2003).

Disruptive innovation has drawn an unusual amount of attention from both scholars and practitioners, which is rare (Danneels, 2004).

The effects of disruptive innovation have been described by a number of authors (Bower & Christensen, 1995; Bower & Christensen, 1996; Christensen et al., 2001; Adner, 2002; Gilbert & Bower, 2002; Charitou&Markides, 2003; Christensen &Raynor, 2003; Danneels, 2004; Schmidt, 2004; Adner&Zemsky, 2005; Utterback&Acee, 2005; Christensen, 2006; Danneels, 2006, Govindarajan&Kopalle, 2006; Markides, 2006; Tellis, 2006; Dan &Chieh, 2008; Sandstrom et al., 2009; Ansari &Krop, 2012; Wessel & Christensen, 2012; Christensen, 2013). The focus is on issues as definitions, what causes disruptive innovation and how it can be classified, foreseen or handled. There is no extensive research on the overall effects of disruptive innovation at an industry level. In our study, the purpose is to shed the light in the relationship between disruptive innovation and value migration.

Knowledge about the effects of disruptive innovation can help managers understand the consequences of their strategic decision-making. Some question the ability of making predictions of disruptive innovation in advance (Thomond&Lettice, 2002), but for example Christensen (2006) and Govindarajan&Kopalle (2006) emphasize that predictions are possible. Hence, findings about the effects of disruptive innovation can result in more cost-efficient investments, which is positive from both a company and a societal perspective. Thus new knowledge adds to the theoretical field of disruptive innovation, which can have practical relevance. The practical significance can also be accentuated by the large interest in disruptive innovation by practitioners (Danneels, 2004).

Gathering information on mobile telecom industry in Lebanon, we have identified three main categories of the mobile telecom value network:

- *Mobile networks operators*
- *Infrastructure providers*
- *OTT players*

Where mobile network operators play a critical and dominant role of the mobile telecom since its inception dominant. Close to 94 % of the value of the mobile telecom flowed through the mobile network operators and they captured 97 % of the profits in 2011 (BMI, 2013). The sector of mobile communications in the Lebanon is one of the main contributors to the Lebanese economy with 2% GDP. According to the Ministry of telecommunications, the number of subscribers to mobile telephony to the Lebanon crossed the 3 million in May 2011. Despite these important developments, the Lebanon is the least competitive country in the Arab world on the mobile telephony market. Thus, mobile telephony to the Lebanon recorded an increase of subscribers.BMI (2013) forecast the number of mobile phone subscriptions to grow at a compound annual rate (CAGR) of 2.9% during the 2013-17 period and to reach 4.9 million in 2017.

Mobile telecom has become critical to drive technological growth, and it impacts how humans communicate and interact in everyday life (Sharma, 2012a). The mobile network operators have been affected by disruptive innovation as described before in this paper. However, no complete analysis with the lessons from the mobile network operators in relation to the disruptive innovation in Lebanon is available yet.

Indeed, the Lebanon is lagging behind in terms of mobile phone, which the penetration rate has doubled since 2008 to around 60% in summer 2013, and it remains much lower than that observed in other countries in the region like Saudi Arabia (177%). But, despite this recent growth attributed to a significant reduction of tariffs, mobile, and despite the growth of the Internet and the free chat applications to the Lebanon Lebanese telecoms market is one of the most expensive in the world.We propose to explain how the Lebanese Mobile Operators (LMO), faced with the technological innovations and free services on the telecom market.

II. LITERATURE REVIEW

According to Christensen (2013), disruptive innovations are products or services with business models that introduce performance packages that are inferior to what mainstream customer's value. During the early development of a disruptive innovation it only serves niche segments. Both the disruptive innovation and the established offerings improve; nevertheless the disruptive innovation improves enough over time to satisfy the mainstream customers and eventually replaces the established offerings and incumbents that exceed the demanded performance, see Figure 1 (Christensen C. M., 2013).

Eventually almost all products improve beyond the needs of the mainstream customers. This triggers a shift in the basis of competition to focus on price, flexibility, convenience or customization, and disruptive competitors start to replace established offerings (Christensen et *al.*, 2001). Christensen & Raynor (2003) describes disruptive innovation as a process rather than an event. Many disruptive innovations fail because they are part of a value network that cannot be adapted to support the disruption.

Most waves of disruptive innovation are captured by others than the earlier leaders of an industry(Christensen, 2013). One example is the computer industry, where IBM dominated the mainframe computer market, but missed the emergence of minicomputers by years (Christensen & Raynor, 2003). The mainframe computers were available to a few experts at universities only, but minicomputers made the technology available to a much larger population. The minicomputers have been preceded by desktops, then laptops and now smartphones. One reason is that leading companies listen too carefully to their customers, so they miss the emergence of innovation that is not valued by their customers initially. However, there are also exceptions when leading companies manage to stay on top (Christensen & Raynor, 2003).

The concept of disruptive innovation has been broadened by a number of authors and there is a debate on what exactly can be defined as disruptive innovation. Many authors classify and use disruptive innovation in a broader sense (Danneels, 2004; Dan &Chieh, 2008). (Christensen, 2006) recognizes that disruptive innovation has been improved by other authors and sees the building of theory on disruptive innovation as an ongoing process.

Christensen &Raynor (2003) divides disruptive innovation into low-end and new-market disruption. New-market disruptive innovation creates a new value network, i.e. expands the market to new customers. Low-end disruption on the other hand, target the most over served and least-profitable customers at the low end of the original value network. Markides (2006) agrees that disruptive innovation can enlarge the industry, by attracting new customers and making existing customers consume more. Furthermore, he states that disruptive innovation can significantly change customers' behaviors and habits.

Govindarajan&Kopalle (2006) on the other hand, refers to high-end and low-end disruptive innovation. High-end disruptive innovation is disruptive innovation with a higher-per unit margin than established offerings, but with different performance features that mainstream customers do not value at the time of introduction, so it serves a small niche before it disrupts the market.

According to Slywotzky (1996) *business designs* go through life cycles, from growth to economic obsolescence. Value migration occurs when value moves from outdated business designs to new ones that better satisfy the customers' priorities. The driving force of value migration is the changing pattern in what customers want, need and are willing to pay for.

Slywotzky&Baumgarter (2006) defines the term business design as the blueprint of the way a company does business. For example how a company selects its customers, what value proposition it offers, which profit model it uses, which activities it engages in, what strategy it uses to capture profits and which organizational architecture it uses to implement decisions.

Three phases of value migration can be identified; these are the *value inflow*, *value stability* and *value outflow* phase, see Figure 3 below (Slywotzky, 1996). The phasesdescribe how well a business designs can create value, which matches customers'priorities in relation to its competitors, and as a result generate high returns. The modelcan be used to describe value migration within a company, between companies andbetween industries.

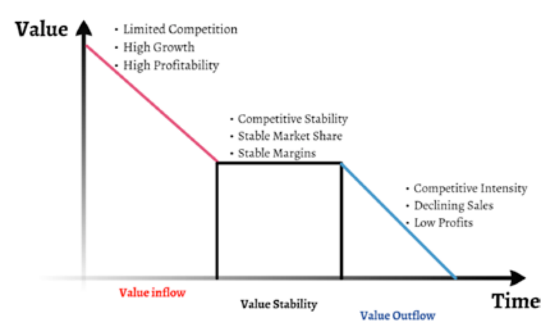


Figure 1: The three phases of value migration in accordance with Slywotzky (1996)

The *value inflow* phase provides several opportunities to abstract value from a surrounding with high growth, limited competition and profitability (Slywotzky, 1996). Value can be absorbed from other parts of for example an industry if the business design is superior in meeting customers' priorities. A shift in value migration can be initiated when a company employs a new business design, which responds to customer priorities that established competitors have failed to see or neglected. Companies with business designs in the value inflow phase are often characterized by excitement, confidence and a

ISBN: 978-1-61804-230-9

capability to attract top talent.

In the second phase, *value stability,* business designs match customer priorities well and a competitive balance predominates the market (Slywotzky, 1996). Companies can grow by continuing to serve customer's priorities and improving operational efficiencies, but only at a low or moderate pace. Market shares and margins remain steady. Focus is on improving activities that have led to success in the past.

In the final phase, *value outflow,* the competition is intensifying, the opportunities to abstract value are decreasing and profits decline (Slywotzky, 1996). Value migrates from obsolete business designs to new ones, which are better able to satisfy customers' evolving priorities. Moreover, in-bound talent, customers and resources leave at an accelerating rate. In the value outflow phase, focus should be on redesigning obsolete parts of a business design.

The length of the phases varies depending on at which rate new business designs that better respond to the customer priorities emerges (Slywotzky, 1996). However, shorter product life cycles, international competition and well-informed customers have made the phases shorter. A company can only exist in one phase at the time, and only move from value outflow to value stability, or from value stability to value inflow, if it applies a new business design.

Performance is generally demanded early in business designs life cycles (Slywotzky, 1996). However, as products or services mature and competitors match the performance, the consumers' priorities often shift to cost-efficiency as long as the business designs remain the same. The reason is that what was initially new is regarded as something that all products or services must have at a later phase, i.e. commoditization.

Transition of phases

Business design phase transitions are typically subtle with no sharp transition points, so managements can easily miss them (Slywotzky, 1996). Managers have tendencies to dismiss downturns as seasonal effects or special circumstances. The above-mentioned limitations can lead to unexpected collapses of companies that have not adapted their objectives in line with the business design life cycle. Furthermore, value can migrate towards several new types of business designs simultaneously and it is even harder to be prepared for. Flexible organizations with an ability to detect early signals of transitions and adapt to them, have better preconditions of handling transitions.

As mentioned, transitions normally occur when new business designs, which better meet customer priorities, become available and customers' priorities change (Slywotzky,1996). However, unexpected external shocks can also trigger business design phase transitions. For example trade restrictions, aggressive pricing, war, regulations and innovation can result in value migration from one business design to another. Companies can sometimes reduce the damage or even benefit from these external events if they understand them at an early stage.

Bargaining power in relation to value migration

Bowman &Ambrosini (2000) argues that the realization of value is determined by the bargaining relationships between the sellers and buyers. For instance, the level of differentiation, switching costs, presence of substitutes, strength of distribution channels and supplier competition can determine a supplier's bargaining power (Porter, 2008), which in turn decides its ability to capture value (Bowman &Ambrosini, 2000). Cox (2001) comes to the same conclusion, i.e. that value migrates in the direction of power. If the buyer is dominant, the seller has few alternatives for its services and products, thus the value flows to the buyer.

III. METHODOLOGY

A. The hypothesis of the study

Upon the theoretical literature discussed previously, it is clear that new trends in Telecoms Innovation lead to improvement in the financial performance in Telecom Companies. Therefore in order to achieve our ultimate goal (say whether an innovation, in particular by smartphones and its content, is disruptive at the mobile network operators and so on the industry level?) this is to understand the impact that applying of a variety of "intelligent" user-driven innovation tools in telecom services on customer satisfaction. The hypotheses are: H1: Decrease in price of product/service would lead to increase in customer satisfaction. H2: Qualities of services would increase the number of satisfied customer. H3: Innovation of products/services would increase the number of satisfied customer hence enhances customer satisfaction and financial revenue by a telecom industry.

To assess the hypotheses, questionnaires were designed in attempt to answer these questions, which formulated and divided in to three aspects to serve the objectives of this paper: 1) Does a pricing strategy in which a Telecom company offers relatively low prices for mobile calls led to improvement in demand for many other services provided? 2) Does a Business strategy that emphasizes on client profiles (age, occupation) to meet various expectations and needs led to increase the number of satisfied customer? 3) Do the telecom companies have the needed to follow up new technologies to limit losses on customers in a competitive market? Is it sufficient?

The iPhone platform was launched in 2007, which was an event resulting from disruptive technologies as discussed previously. Given that, we are going to focus on data starting from 2007 (see Tab 1 and Fig 2). In our case, the mobile network operators did not provide net income or ARPU for their mobile segments, thus we replace estimating ARPU by viewing the mobile segments revenue as number of subscribers between the years 2000 to 2012.

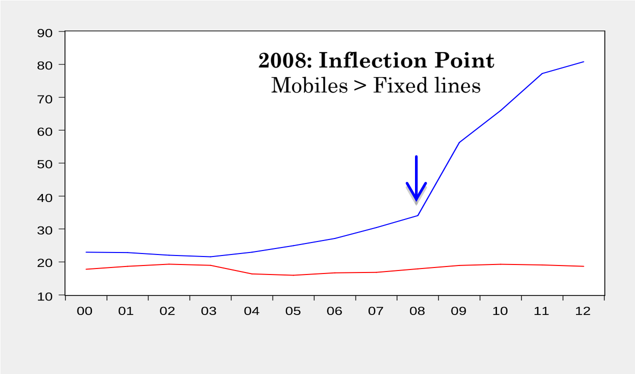


Figure 2: Global FixedTelephoneLines vs. Mobile Subscriptions, 2000 – 2012.

All prices are set by the MOT. Previous governments followed strategies to limit subscriber numbers with high tariffs. Monthly ARPU levels were over US$60 until end of 2009. It is especially in April 2009 when government took a radical step consists of lowering tariffs and increasing the maximum number of subscribers for both operators, arguing that total revenue would increase. This had a startling affect on subscriber growth and typically generates significantly higher ARPU.

Table 1: Revenues and ARPUS

Cellular	**2008**	**2009**	**2010**
Cellular revenues ($ 000)*	1 361 000	1 460 000	1 573 743
Monthly ARPU ($)*	84,8	62,4	45.0

Internet			
Internet revenues ($ 000) *	48 685	62 063	74 680
Monthly ARPU ($)*	14.0	15.0	16.0

* Estimated

Source: The ArabAdvisors Group

In the information and data gathering process we have solely gathered information related to the mobile segment of the network operators. Some of the network operators provide fixed - line services that also may have been affected by the disruptive innovation. A main issue, consubstantial to the argument, it would remain to show a relationship between financial performance and usage of disruptive innovations in telecom industry.

B. *A governance Structure of Telecom (Lebanon)*

Lebanon's telecommunications sector is entirely owned by the government via two main licensed and working operators by 2014 (including granted and expected):

1) ALFA, aMobile Interim Company, run by **Orascom** Telecom's Alfa,and established in1994. 2) Touch, a Mobile Interim Company run by **Zain**telecom's MTC Touch, established in1983 (Fig 4).

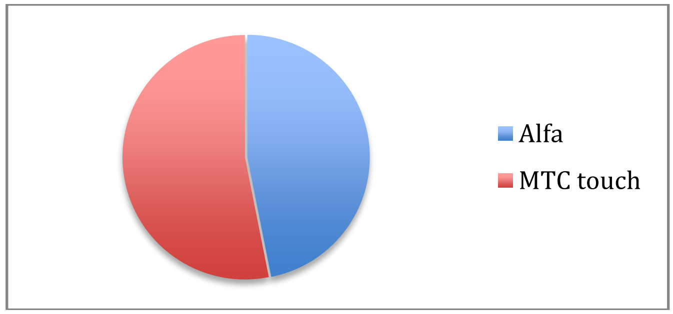


Figure 3: Lebanese operator share of mobile subscribers, 2010

C. *Companies Reports*

By mid-2013, Alfa announced an **increase of 296%on mobile phone subscriptions** and 446% on Internet services (over the past 13 months).

According to Touch Company reports, **mobile phone subscriptions have grown more than eightfold** over the past 18 months (from 100 000 to 800 000), and an average increase of 145% up to 270 MB is estimated on personal consumption. **CDL: 10-05-2013**. (See Table 2)

Table 2: Distribution of Mobile subscribers in Lebanon

	Dec. 2009	Dec. 2010	% Variation
*Subscribers*Alfa	210 750	241 142	+14,4
Prepaid phone card – Alfa	856 802	1 101 243	+28,5
Total Alfa	**1 067 552**	**1 342 385**	**+25,7**
Subscribers Touch	212 234	238.528	+12,4
Prepaid phone card - Touch	1 110 531	1 282 751	+15,5
Total Touch	**1 322 765**	**1 521 279**	**+15,0**
Total	**2 390 317**	**2 863 664**	**+19,8**

Source: MOT.

Increasingly competitive between companies operating in the Internet industry required effective productivity enhancement from the mobile operators, government and ISPs in a way to *draw upon up*-to-*dateastechnology*evolves, *more likely* to keep Internet *at the power of innovation to communicate could be.* Mobile broadband, ADSL, 3G licenses and other international Internet bandwidth are already advertised and lead to connect the large majority users in Lebanon. 3G service officially launched in Lebanon by Alfa and Touch on November 1 2011[1] (CDL/ 21-10-2011). The 4G networks are as well being installed in Beirut, Dbayeh, Jounié&Kesrouan by end of 2013 and must be implemented in Tripoli, Saida&Zahlé in coming months of 2014, according to CDL (31-05-2013). This technology is

[1] Local ISP Cedarcomclaimed on involving public sector on the third-generation mobile Internet technology in Lebanon, calling to grant the right to use 3G frequencies as specified in the TelecommunicationsAct – 431 - "ART shouldexclusivelydefineLicensees."

The Council of State react by sayingthat the application of Law 431 has been suspendedunder a previousjudgement, addingthat, Alfa and Touch are anywayboth state-ownedcompanies and therefore do not requirelicenses.

the latest in the world of mobile data transfer, which theoretical data downlink speeds can go up to 100 Mbps. Cell phone Plans in Lebanon are offered with a wide range of $99/10GB to $249/100GB – per month).

Rapidly growing mobile Internet usage surpasses more highly monetized desktop Internet usage. Moreover it is known that eCPMs is 5x lower on Mobile than desktop. Thus, because of innovation, mobile ARPU can raise rapidly straining Revenue Growth in the telecom service sector.

Table 3: Telephone subscribers and Internet users for Lebanon 2000-2012.

	Mobile cellular Subscriptions/ 100 inhabitants	Fixed telephone Subscription/100 inhabitants	Internet users/100 inhabitant	Fixed (wired)-broadband internet subscribers per 100 inhabitants
2000	22,96*	17,80	7,95***	0%
2001	22,84*	18,64	6,78***	0%
2002	22,05*	19,31	7,00*	1,00
2003	21,56*	18,97	8,00*	1,90
2004	22,95*	16,35	9,00*	2,08
2005	24,92*	15,92	10,14*(a)	3,26
2006	27,12*	16,70	15,00*	4,66
2007	30,44**	16,85	18,74	4,64*
2008	34,09*	17,92*	22,53(b)	4,66*
2009	56,28	18,93	30,14*(c)	4,64*
2010	65,97	19,30	43,68* (a)	6,98*****
2011	77,19	19,09	52,00*** * (d)	8,28*****
2012	80,81	18,66	61,25*	9,71*****

Source: MOT. * ITU estimate. **BMI. ***Lebanese Broadcasting International. ****Presidency of the Council of Ministers. ***** TRA.

Notes:

(a) Estimate based on population aged 6+.

(b) TRA estimates the number of Internet users based on the number of Internet subscriptions (3 users for every subscription).

(c) Estimate based on population aged 15+.

(d) Population age 15+.

ITU: International Telecommunication Union.
TRA: Telecommunications Regulatory Authority.
MOT: Ministry Of telecommunications.
CDL: Commerce du Levant.

Upon the stated statistics and facts in previous sections a hypothesis for this study has been elaborated in order to be tested for validity.

Due to the unattainability of financial statements from the telecom industry in Lebanon, a survey was conducted with both end users and providers. Moreover, a study was performed to investigate various disruptive and conventional means which might have an impact on the financial performance of telecom industry. The survey consisted of two part questionnaires, the first one was conducted with customers selected randomly of n=100, whereas the second was performed with n=20 employees from the two telecom providers, ‘Alfa’ and ‘Touch’ at the providers’ offices located in North Lebanon. All items were measured with a five-likert scale, ranging from 1(strongly disagree) to 5 (strongly agree).

The questions were designed in an easily understandable manner and performed on a number of days from different locations in the city to govern the credibility and variety of sources. Moreover, the secondary data applied were mainly obtained through websites, research articles, and journals.

The proposed hypothesis is tested based upon the gathered data, which dealt with questions related to financial performance and the role of disruptive of innovations in its development. Furthermore, the disruptive innovations in this study are categorized in to several aspects as usage of mobile phones for internet browsing, download and upload, VoIP, and social media communication which is replacing to a certain degree the conventional means of communication. Whereas the conventional means are categorized in to calls, sms and mms services.

IV. RESULTS

The results attained from the first survey are indicated in Table 4, which shows the demographics of 53%male and 47% female with a majority of surveyed are youths with age ranging between 16 and 40 (representing the biggest sector of end users). Moving away from demographic factors, the results show that 89% of customers own smart phones, and 90% are subscribed for broad band internet. These results show the change of customer needs from past years where they are highly demanding advanced communication services. These numbers indicate how customers are migrating from conventional to disruptive innovations through enlarging consumption of disruptive services, where (Markides, 2006) tends to be on the same path

Table 4: Demographic

Characteristics	**Category**	**N**	**%**
Gender	Male	53	53
	Female	47	47
Age	16-25	36	36.4
	26-40	41	41.4
	41-60	14	14.1
	61-80	8	8.1
Subscription to broadband services	Yes	92	92
	No	8	8
Company	Alfa	53	53
	Touch	41	41
	Both	6	6
Usage ofConventionalterms(SMS, Calls)	More than 30 minutes daily	26	26
	Less than 30 minutes daily	74	74
Usage of Disruptive Innovations	More than 30 minutes daily	82	82
	Less than 30 minutes daily	18	18

The proposed hypothesis is tested through a two parts survey done with end users of telecom services n=100, in addition to aquestionnaires distributed on the two service providers Alfa and Touch with n=20.

The questionnaire for this study is designed according to likert scaling technique, which ranged from strongly agree to strongly disagree. Among the surveyed customers 90% of them are subscribed for broadbandInternet.

The disruptive innovations in this study are categorized in to several aspects as usage of mobile phones for internet download and upload, VoIP, and social media communication which is replacing to a certain degree the conventional means of communication.

Whereas the conventional means are categorized in to calls, sms and mms services.

Table 5: Linear Regression Analysis

Predictor	Beta	p value	Condition index
Calls	**.191**	.040	6.456
Upload	**.208**	.038	6.983
Social network services	**.283**	.004	4.719
VoIP	.084	.337	5.530
Email	**.201**	.028	9.796

The Cronbach's alpha values which revealed calls, upload, social, VoIp, and email expressed the assessment of reliability with the following scores being 0.546,0.615,0.633, 0.511, and 0.524 respectively. Furthermore, reliability level for these variables is 0.674, and linear regression equation was used for providing better analysis for independent variables stated above.

Reading the results in Table 5, we get the regression equation for Telecom sector (with correlation parameters on the variable in the following equation):**Total Telecom revenue=0.19***Calls+**0.2***Upload+**0.28***Social network+**0.2***E-mail

The coefficient of determination; R square is 0.474, thus indicating that 47.4% of variance in the variable customers' performance is explained by the model. All the values present in Table 5 are statisticallysignificantsincethey have a value smallerthan 0.05, exceptVoIPwhichwasn'tproven in thisstudy (Siggreaterthan 0.05). Furthermore, the condition index has been calculated in order to check for the colinearityproblem. All the CI values for the five variables are below 15; whichrevealsthatthisstudy has no seriousproblemwithcolinearity. The linearregression and collinearityresults are shown in(Table5).

The second part of the survey concerning the providers revealed that the widest area for development is the category related to broadband servicesas shown in figure4. The gathered results thus meet the previously studied factors showing the shift of customers from conventional modems to up to date services, which they consider to be more compatible with their capabilities and interests thus resulting in more profits. These results show moderate similarity with (Christensen C. M., 2013) and (Sharma, 2012a) who claimed that with time disruptive innovation will improve and replace other services, thus introducing more profits from data access. Moreover, the profit from disruptive innovation will lead to a great impact on the era. Figure 4 shows that the quality of services plays a role in the future telecom market, which might boost the customers' consumption and satisfaction thus leading to higher demand for services.

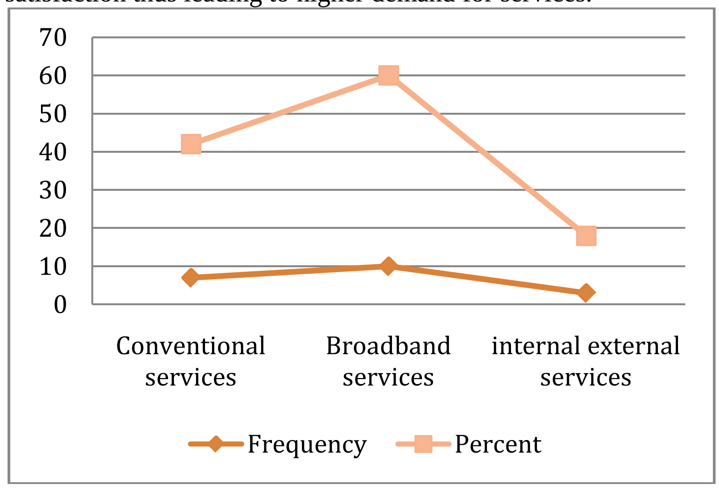


Figure 4: Providers' Frequencies of Future Migration

V. CONCLUSION

In an era of increased competition and customer demands there are various factors affecting the telecom industry and prohibiting it from flourishing. Concerning the case of telecom industry in Lebanon, there is lack of competition,high tariffs, and not very good connexionsince there are only two service providers, Alfa and Touch. However, the Lebanese customers are in frequent demand for advanced communication services, which made Internet an essential mean of communication. Customers are nowadays migrating to data, thus pressuring on the industry to improve such services and providing a good value bundle offer.

In conclusion, reading the results in the regression equation stated previously shows that the highest revenue is still coming from Social network and upload and download services whereas no significant response was found on VoIP (where the use of this service is prohibited in Lebanon). Therefore, the most revenue is covered by the usage of data such as social networks and uploads. This fact has to be detected by the organizations as signals, and then adapt these signals in order to reach targeted results in the telecom industry and meet other neighboring countries instead of lagging behind them.

Upon previously stated hypothesis and literature review, the tariffs of services affect customer

For future recommendations, the presence of more competition in telecom industry will enable customers with a satisfying service and encourage them for more consumption thus higher revenues for the telecom industry.

ISBN: 978-1-61804-230-9

REFERENCES

[1] Adner, R. (2002). *When Are Technologies Disruptive? A Demand-Based View of the Emergence of Competition* (Vol. 23). Strategic Management Journal.

[2] Adner, R., & Zemsky, P. (2005). *Disruptive Technologies and the Emergence of Competition* (Vol. 36). The RAND Journal of Economics.

[3] Ahonen, T. (2011). *Insiders Guide to Mobile: The Customers, Services, Apps, Phones and Business of the Newest Trillion Dollar Market.* TomiAhonen Consulting.

[4] Ahonen, T. (2012). *Phone Book 2012: Statistics and facts on the Mobile Phone Industry.* TomiAhonen Consulting.

[5] Ahonen, T. (2012). *Phone Book 2012: Statistics and facts on the Mobile Phone Industry.* TomiAhonen Consulting.

[6] Alter, N. (2002). *Les logiques d'innovation: Approche pluridisciplinaire.* Paris: La découverte.

[7] Ansari, S., & Krop, P. (2012). *ncumbent performance in the face of a radical innovation: Towards a framework for incumbent challenger dynamics* (Vol. 41). Research Policy.

[8] BMI. (2013). *Lebanon This Week* (Vol. 320). Beirut: Research Economic Research & Analysis Department: Byblos Bank.

[9] Bower, J. L., & Christensen, C. M. (1996). *Customer Power, Strategic Investment, and the Failure of Leading Firms* (Vol. 17). Strategic Management Journal.

[10] Bower, J. L., & Christensen, C. M. (1996). *Customer Power, Strategic Investment, and the Failure of Leading Firms* (Vol. 17). Strategic Management Journal.

[11] Cantwell, J. (2010). *"Innovation and information technology in the MNE".* (T. O. Business, Éd.) Oxford University Press.

[12] Charitou, C. D., & Markides, C. C. (2003). *Responses to Disruptive Strategic Innovation.* MIT Sloan Management Review.

[13] Christensen, C. M. (2013). *Disruptive Innovation.* http://www.claytonchristensen.com/key-concepts.

[14] Christensen, C. M. (2003). *The Innovator's Solution: Creating and Sustaining Successful Growth.* Cambridge MA: Harvard University Press.

[15] Christensen, C. M. (2006). *The Ongoing Process of Building a Theory of Disruption.* The Journal of product innovation Management.

[16] Christensen, C. M., & Raynor, M. E. (2003). *The Innovator's Solution: Creating and Sustaining Successful Growth.* Cambridge MA: Harvard University Press.

[17] Christensen, C. M., E., R. M., & M., V. (2001). *Skate to where the money will be.* Harvard Business Review.

[18] Daannels, E. (2004). *Disruptive Technology Reconsidered: A Critique and Research Agenda.* The Journal of product innovation Management.

[19] Dan, Y., & Chieh, H. (2008). *A Reflective Review of Disruptive Innovation Theory.* PICMET 2008 Proceedings, Cape Town.

[20] Dan, Y., & Chieh, H. (2008). *A Reflective Review of Disruptive Innovation Theory* (éd.). PICMET 2008 Proceedings, Cape Town.

[21] Danneels, E. (2006). *Dialogue on the Effects of Disruptive Technology on Firms and Industries* (Vol. 23). The Journal of product innovation Management.

[22] Gilbert, C., & Bower, J. L. (2002). *Disruptive Change – When Trying Harder Is Part of the Problem* (Vol. 80). Harvard Business Review.

[23] Govindarajan, V., & Kopalle, P. (2006). *The Usefulness of Measuring Disruptiveness of Innovations Ax Post in Making Ex Ante Predictions* (Vol. 23). Journal of product innovation Management.

[24] Johnson, G. W. (2011). *Stratégique.* Paris: Pearson Education France.

[25] Markides, C. (2006). *Disruptive Innovation: In Need of Better Theory* (Vol. 23). The journal of product innovation management.

[26] Markides, C. (2006). *Disruptive Innovation: In Need of Better Theory.* The journal of product innovation management.

[27] Mayrhofer, U. U. (2011). *Management international: Des pratiques en mutation.* Paris: Pearson Education France.

[28] Salz, P. A. (2011). *The Netsize Guide: Truly Mobile, Gemalto.*

[29] Sandstrom, C., Magnusson, M., & Jörnmark, J. (2009). *Exploring Factors Influencing Incumbents' Response to Disruptive Innovation* (Vol. 18). Journal Compilation.

[30] Schmidt, G. (2004). *Low-end and High-end Encroachment Strategies for New Products* (Vol. 8). International Journal of Innovation Management.

[31] Schumpeter, J. (1935). *Théorie de l'évolution économique.* Paris: Dalloz.

[32] Sharma, C. (2012a). *Operator's dilemma (and opportunity): the 4th wave.* Mobile Future Forward.

[33] Sharma, C. (2012b). *State of Global Mobile Industry Annual Assessment 2012.* Mobile Future Forward.

[34] Steinbock, D. (2005). *The Mobile revolution - The making of mobile services worldwide.* Kogan Page Limited.

[35] Tellis, G. J. (2006). *Disruptive Technology or Visionary Leadership?* (Vol. 23). The Journal of Product Innovation Management.

[36] Thomond, P., & Lettice, F. (2002). *Disruptive Innovation Explored.* Concurrent Engineering Conference Proceedings.

[37] Utterback, J. M., & Acee, H. J. (2005). *Disruptive Technologies: An Expanded View* (Vol. 9). International Journal of Innovation Management.

[38] VisionMobile. (2011a). *Developer Economics 2011.*

[39] VisionMobile. (2012b). *Mobile Industry Dynamics 2012.*

[40] VisionMobile. (2012a). *The Telco Innovation Toolbox: Economic Models for Managing Disruption and Reinventing the Telco 2013.* In association with Ericsson.

[41] Wessel, M., & Christensen, C. (2012). *Surviving Disruption.* Harvard Business Review.

Role of Innovation in SMEs Performance: A Case of Malaysian SMEs

Kamariah Ismail, Wan Zaidi Wan Omar, Khairiah Soehod, Aslan A. Senin and Ch. Shoaib Akhtar

Abstract**—**Innovation has been termed as the most crucial of the elements in today's globalized and competitive environment. Companies focusing on innovation achieve not only competitiveness but also are able to sustain them for a longer period of time. The present study is conducted in the context of Malaysian SMEs using a sample of 870 SMEs from both manufacturing and services sectors. These SMEs have been found engaged in variety of businesses and to some extent have also been involved in the process of innovation. The results indicate that the SMEs though are aware of the innovation and its relevance to organizational performance and growth yet are unable to focus on innovation processes as much as they should to gain competitive position.

Keywords**—** Innovation, Performance, Growth, SMEs, Malaysia.

I. Introduction

Innovation is an important tool that provides opportunities to new inventions and building of new markets (Kuhn & Marisck, 2010). Moreover, there is a remarkable increasing interest trade and industry growth based on innovation and creation of competitive advantage (Birkinshaw, 2011; Clawson, 2009; Grant, 2010; Hamel, 2002; Kim & Mauborgne, 2005). Furthermore, due to mounting competition, the capability to control the innovation and manage the innovation processes has become extremely important to governments and organizations alike because of the impact, limited and inadequate resources will have on the future growth. Kanter (2006) terms innovation as a natural renewable source accessible to all restricted only by human effort.

Remaining competitive in today's modern world require organizations to pursue innovation (Teece, 2007). Hence the million dollar question in this regard relates with 'how to innovate' which still draws researchers attention. Clausen, et al (2013) recently focused on this challenge by attending to the question and used prior theory to identify four modes of innovation i.e. open exploration, closed exploration, open exploitation, and close exploitation. It has been suggested by researchers that SMEs have limited innovation as compared to larger enterprises. However, this misconception has been negated by Kaufmann and Todtling (2002) who highlight that SMEs are more innovative due to their heterogeneous character but are restricted in innovative capacity due to their financial and human resources. Similarly, Rosenbusch et al (2011) have identified several factors that affect the relationship between innovation and SME performance. They are of the view that new SMEs benefit more from innovation than the mature organizations mainly due to their flexibility to accept change in their environment or industry.

Keeping in view the importance of innovation in SMEs performance, the present study focuses on the Malaysian SMEs. Malaysia is one of the fastest growing economies of the world and is based largely on SMEs that contribute almost 33 percent to the national GDP. The importance of SMEs growth and their sustainability cannot be overemphasized for Malaysian economic growth and development especially in this competitive global environment. Despite the government focus on SMEs, researchers question the role of innovation in maintaining the growth of the Malaysian economy and the role of innovation in the overall performance of the SMEs. This concerns stems from the Hill et al (2012) who feared about country being victim of middle income trap evidenced from the visible downfall in the growth rate since late 1990s. Thus, the main focus of the present study is to evaluate the role of innovation towards SMEs performance and growth.

Kamairah Ismail is the Director of Technology Entrepreneurship Center, Universiti Teknologi Malaysia (UTMTEC), Johor Bahru, Johor, Malaysia (phone: +60197726991; e-mail: m-maria@utm.my).

W. Z. Wan Omar, is with the department of aeronautical engineering, Universiti Technologi Malaysia, Johor Bahru, Johor, Malaysia (e-mail: wanzaidi@fkm.utm.my).

Khairiah Soehod is senior lecturer at the faculty of management, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia (khairiah@fppsm.utm.my).

Aslan A. Senin is senior lecturer at the faculty of management, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia (aslant@utm.my).

Ch. Shoaib Akhtar is PhD scholar at Department of Management, Faculty of Management, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia (e-mail: sachaudhry2@live.utm.my).

II. Review of Literature

A. Innovation

Wood (2008) opines that creativity of an organization is mainly concerned with the establishment of valuable and useful new product, service or idea and the methods by which individuals work together in a complicated social system. Innovation is basically concerned and defined with the adoption of a product, service and methods that are new for organizations and adopted by them.

Crossan and Apaydin (2010) highlights that innovation is creation or acceptance , adaptation and utilization of a value - added novelty in trade and industry spheres, regeneration and expansion of product , services and markets, making of new ways of product development and establishing new management system. Similarly, Orlikowski (2010), Tsoukas et. al (2002) and Wierdsma (2004) stated that innovation is the process of development of new outcome by adopting new ways of working and product development. Moreover, this new method of working is concerned with the improvement and better performance of an organization that result in production of a new service, process and product. Innovation is generative renewal and competence of an organization to perform in correspondence to environment. Innovation is seen as a most important part of an organizational life that emerges in daily activities and interaction of the organizational members while they carry out their work and goals.

Innovation is considered as every day issue for members of organizations in defining their problems, responding to unforeseen events, creation of solutions and development of new ways and procedures to organize work, through the use of experience, skills, motivation and the knowledge accumulated is converted in to production of an innovative product or service (Tsoukas & Vladimirou, 2002; Wierdsma 2004, Kocher et al. 2011, Miettinen et al. 2009). Boer et al. (2005) explains that the organizational practices of the innovation are maintained, established and uses the standard set of actions or systems like designing of an idea or thought, evaluation and managerial efforts and practices for innovation like flexible roles, rotation , for time being projects teams, self organizing groups. Moreover, the formal practices help out and encourage the employees to participate not only in innovation and learning activities but also be a part of designing activities (Wilhelmsson & Döös, 2009, Kianto, 2008).

Tsoukas and Vladimirou, (2011), Jensenet al. (2007) stresses on dualistic nature of innovation between exploration and exploitation, individual and collective, STI (Science, Technology and Innovation) and DUI (Doing, Using and Interaction) mode of innovation and on organizational level between suppleness and competence. Eisenhardt et. al. (2010) suggests that either the complementary processes, harmonization, meta level collective orientation or gather them in constant dialogic relationship in order to have control on both sides. Moreover, Foss et al. (2010) opines that it is not clear that how the activities of exploration/ exploitation and the process of knowledge usage, creation and integration occurs, it is not only concerned with the individual and team level, thus, we can get the complete picture of how the innovations are performed in an organizations.

Wierdsma, (2004) and Yuan and Woodman (2010) state that innovation is seen as rising trend in day to day work of organizational members and on individual level the exploration and generation of an idea is performed by the individual actions and via social interaction. If we talk about in terms of individual it means that capability to express skills and insight that is creation, encouragement and endorsement of new idea in to action. In order to improve the performance of an individual or group of an organization the thought or idea can be taken as a collective practice in order to get the best play of day by day increasing innovation and renewal demands that is the main player between the individual and organizational knowledge.

.

B. Growth

Researchers normally evaluate a company's ability based on its performance (Bonn, 2000; Rosli & Sidek, 2013) or growth (Dobbs & Hamilton, 2006). The concept of firm growth is associated with the law of proportionate effect, which states that the firm growth is proportional to the firm's current size. However, this law fails to hold in relation to the age of firm and show a negative relationship between growth and firm size and age (Hutchinson & Xavier, 2004). The empirical research indicates that there is positive relationship between innovation and growth of the firms if there is a constant supply of finances (Hyytinen & Toivanen, 2003). In the presence of innovation, the overall firm performance would enhance (Rosli & Sidek, 2013; Damanpour, 1991; Lin & Chen, 2007, Van Auken, et al. 2008; Li, et al. 2010; Salim & Sulaiman, 2011). This shows that innovation is critical for the growth of the organization in terms of its sales, market penetration, profitability and sustainability of organizations especially for small and medium enterprises.

III. METHODOLOGY

For the present study, sample consisted of SMEs belonging to both manufacturing and services sectors. A total of 870 SMEs from across the country participated in the study that spanned over a period of twelve months. The criterion followed in selection of SMEs was based on number of employees not exceeding 150 full time employees. The instrument of the study was based on the innovation (process, product, administrative and marketing) and the growth of the company (sales turnover). A team of 20 specially trained enumerators was engaged to collect the data. The questionnaire was developed for the study and was based on the innovation survey. The reliability of the instrument was found to be 0.827. The scale of the study used five point likert rating. The demographic details of the respondent companies indicated that majority of these SMEs were operational since 2000 (75%) while rest of them were established during 1990s. Regarding the ownership of the companies, it was found that sole proprietorship accounted for 32% of the sampled SMEs, partnership accounted for 8% while the rest were private limited. The main activities of these sampled SMEs were found to be concentrated around travel & Tours (30%), scientific activities (12%), computer and electronics (16%) and the rest were distributed in the others miscellaneous activities. All of the sampled SMEs were involved in one of the innovation processes that is either they have been involved in product innovation, innovation related to processes, structural innovations or marketing innovations.

IV. RESULTS AND DISCUSSION

First the significance of innovation to the company was assessed for both manufacturing and service sector SMEs. The

ISBN: 978-1-61804-230-9

result is presented in Table 1.

Table 1 Significance of Innovation for Growth of Firm in opinion of Managers

To replace products		To improve product	To extend product range	To increase market share
Response	%	%	%	%
Not Relevant	44.1	25.1	26.7	25.8
Low	11.9	8.0	12.1	9.6
Medium	19.1	13.6	17.2	15.2
High	25.0	53.3	43.9	49.4

The results indicate that for the growth of the company, manager's view replacing products as having no significance (44.1%), but they are of the view that to improve the product quality, increasing the market share and to extend existing the product lines, innovation is of utmost importance for the company.

To assess whether there exist any differences between manufacturing and service SMEs, t-test was applied (Table 2). The result indicates that there exist a significant difference of opinion between manufacturing companies and service oriented companies. The results indicate that manufacturing companies are aware and are acknowledge and apply the innovation in their processes.

Table 2 t-test for Manufacturing vs. Services SMEs

	Sector	Mean	Std. Dev.
To replace products	Manufacturing	1.5461	1.28777
	Services	1.0696	1.19622
To improve product quality	Manufacturing	2.3775	1.04685
	Services	1.6917	1.32562
To extend product range	Manufacturing	2.2247	1.11048
	Services	1.5143	1.26809
To open up new markets or increase market share	Manufacturing	2.2989	1.13428
	Services	1.6289	1.27987
N	870		

To assess whether the sampled SMEs were involved in any activities related to R&D, managers were asked about their response on a dichotomous scale. The result is shown in Table 3.

Table 3 R&D Activities of SMES

Response	In-house R&D	Outside Acquisition of R&D	External knowledge	Training
	%	%	%	%
No	59.4	88.4	84.9	49.7
Yes	40.6	11.6	15.1	50.3
N	870			

The managers of SMEs were asked to respond to whether they have been involved in in-house or outside the company R&D, acquisition of external knowledge for R&D activities or training their employees in this regard. The results indicate that on all accounts SMEs surveyed were less involved in R&D activities except for the training of employees.

V. CONCLUSION

The present study was conducted to find out the role of innovation in the development and growth of small and medium sized firms. The results indicate that the Malaysian SMEs management is aware of the role innovation plays in the growth of the firms. However, being resource starved these SMEs are not in a position to either enter R&D activities or acquire new and advanced technologies, although, these companies are engaged in developing the skills and capacities of their employees through various trainings. The results also suggest that the manufacturing companies are more involved in research and development activities than their counterparts in the services industry. The results of the study do indicate that the innovation is essential if companies want to grow and become more competitive in relation to their national and international competitors. This would not only help the SMEs to gain market share but would also help them sustain themselves in the longer run especially with the coming of the ASEAN Economic Community single market commencing in 2015. The survival of Malaysian SMEs will depend on their ability to innovate as they will face enormous challenge from other member countries SMEs. It is suggested that the future studies should take into account external factors like sources of government funding and incentives systems, external collaborative linkages between SMEs and research and academic institutions that could help build the capacities and capabilities of the SMEs to attain growth and competitive advantage.

ACKNOWLEDGMENT

The authors wish to acknowledge the Ministry of Science, Technology and Innovation (MOSTI) Malaysia for their support.

REFERENCES

[1] Woodman, R. W. (2008). Creativity and organizational change: Linking ideas and extending theory. In J. Zhou & C. Shalley (Eds.), *Handbook of organizational creativity* (pp. 283-300). New York, NY: Lawrence Erlbaum Associates.

[2] Crossan, M. M., & Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: A systematic review of the literature. *Journal of Management Studies*, *47*(6), 1154-1191.

[3] Kanter, R. M. (2006). Innovation: The classic traps. *Harvard Business Review*,*84*(11), 72-83.

[4] Kim, W. C., & Mauborgne, R. (2005). *Blue ocean strategy: How to createmuncontested market space and make competition irrelevant.* Boston, MA:Harvard Business School Press.

[5] Hamel, G. (2002). *Leading the revolution: How to thrive in turbulent times by making innovation a way of life*. Boston, MA: Harvard Business School Press.

[6] Grant, R. M. (2010*). Contemporary strategy analysis* (7th ed.). West Sussex, UK:John Wiley & Sons.

[7] Clawson, J. G. (2009). *Level three leadership* (4th ed.). Upper Saddle River, NJ:Prentice Hall.

[8] Birkinshaw, J., Bouquet, C., & Barsoux J.-L. (2011). The 5 myths of innovation. *MIT Sloan Management Review, 52*(2), 42-43.

[9] Kuhn, J. S., & Marisck, V. J. (2010). Action learning for strategic innovation in mature organizations: Key cognitive, design and contextual considerations. *Action Learning: Research and Practice, 2*(1), 27-48.

[10] Orlikowski, W. (2010) *Practice in research: phenomenon, perspective and philosophy.* In: Golsorkhi, D., Rouleau, L. Seidl, D. and Vaara, E. (eds). Strategyas Practice. Gambridge University Press. pp. 23 – 33.

[11] Tsoukas, H. and Vladimirou, E. (2002). *What is organizational knowledge?*Journal of Management Studies 38:7. November, 2001. pp. 973-993.

[12] Wierdsma, A. (2004) *Beyond Implementation. Co-creation in Change andDevelopment.* **In**. Boonstra (ed). Dynamics of Organizational Change andLearning. John Wiley&Sons, Ltd. pp. 227-257.

[13] Kocher, P-Y., Kaudela-Baum, S and Wolf, P. (2011). *EnhancingOrganizational Innovation Capability Through Systemic Action Research: ACase of a Swiss SME in the Food Industry.* Syst Pract Action Research 24, pp.17-44.

[14] Miettinen, R., Samra-Frederics, D. and Yanow, D. (2009**).** *Re-turn toPractice: An Introductory Essay.* Organization Studies. Vol. 30, No 12. pp.1309-1329

[15] Boer, H., Kuhn, J. and Gertsen, F. (2005). *Continuous innovation. Managing dualities through co-ordination.* Working paper presented at the 7.Fachtagung der Kommission Technologies and Innovations management in Berband derHochschullehrer für Betriebswissenschaft e.V. Erfurt, 17-29 October 2005.

[16] Wilhelmsson, L. & Döös, M. (2009**).** *Sustainable heritage in a rapidly changing environment.* In: Docherty, P., Kira, M. ans Shani (Rami) A.B.S. Creating sustainable work systems. Developing social sustainability.Routledge. UK

[17] Kianto, A. (2008). *Development and validation of a survey instrument formeasuring organizational renewal capability*. International Journal ofTechnology Management, Vol. 42. Nos. 1/2, pp. 69 - 88.

[18] Jensen, M. B., Johanson, B. Lorenz, E. and Lundvall, B. Å. (2007) *Forms of knowledge and modes of innovation.* Research Policy, Vol. 36. pp. 680-693.

[19] Eisenhardt, K. M., Furr, N. R. and Bingham, C. B. (2010). *Microfoundations of Performance: Balancing Efficiency and Flexibility in Dynamic Environments*.Organization Science. Vol 21, Nol 6. November-December. Pp. 1263-1273.

[20] Foss, N. J., Husted, K. and Michailova, S. (2010). *Governing Knowledge Sharing in Organizations: Levels of Analysis, Governance Mechanisms, and Research Directions.* Journal of Management Studies. Vol. 47, No 3. pp. 455-482.

[21] Wierdsma, A. (2004) *Beyond Implementation. Co-creation in Change and Development.* **In**. Boonstra (ed). Dynamics of Organizational Change andLearning. John Wiley&Sons, Ltd. pp. 227-257.

[22] Yuan, F. and Woodman, R. W. (2010). *Innovative behavior in the workplace:The role of performance and image outcome expectations.* Academy ofManagement Journal. Vol 53, No.2. 323-342.

[23] Rosli, M. M. & Sidek, S. (2013). Innovation and firm performance: evidence from Malaysian small and medium enterprises. Paper presented at the 20th International Business Information Management Conference (IBIMA). International Business Information Management Association, pp. 794-809.

[24] Dobbs, M. and Hamilton, R.T. (2006). Small Business Growth: Recent Evidence and New Directions, *International Journal of Entrepreneurial Behaviour and Research*, 13 (5), 296-322.

[25] Hutchinson, John; Xavier, Ana (2004) : Comparing the Impact of Credit Constraints on the Growth of SMEs in a Transition Country with an Established Market Economy, LICOS Discussion Paper, No. 150.

[26] Hyytinen, Ari; Toivanen, Otto (2003) : Do financial constraints hold back innovation and growth? Evidence on the role of public policy, ETLA Discussion Papers, The Research Institute of the Finnish Economy (ETLA), No. 820

[27] Salim, I. M., & Sulaiman, M. B. (2011). Impact of organizational innovation on firm performance: evidence from Malaysian-based ICT companies. *Business and Management Review*, *1*(5).

[28] Van Auken, H., Madrid-Guijarro, A., & Garcia-Perez-de-Lema, D. (2008). Innovation and performance in Spanish manufacturing SMEs. *International Journal of Entrepreneurship and Innovation Management*, *8*(1), 36-56.

[29] Lin, C. Y. Y., & Chen, M. Y. C. (2007). Does innovation lead to performance? An empirical study of SMEs in Taiwan. *Management Research News*, *30*(2), 115-132.

[30] Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of management journal*, *34*(3), 555-590.

[31] Li, Y., Zhou, N., & Si, Y. (2010). Exploratory innovation, exploitative innovation, and performance: influence of business strategies and environment. *Nankai Business Review International*, *1*(3), 297-316.

[32] Clausen, T. H., Korneliussen, T., & Madsen, E. L. (2013). Modes of innovation, resources and their influence on product innovation: Empirical evidence from R&D active firms in Norway. *Technovation*, *33*(6), 225-233.

[33] Hill, H., Tham, S. Y., & Zin, R. H. M. (Eds.). (2012). *Malaysia's development challenges: Graduating from the middle.* Routledge.

[34] Kaufmann, A., & Tödtling, F. (2002). How effective is innovation support for SMEs? An analysis of the region of Upper Austria. *Technovation*, *22*(3), 147-159.

[35] Muhammad, M. Z., Char, A. K., bin Yasoa, M. R., & Hassan, Z. (2009). Small and medium enterprises (SMEs) competing in the global business environment: A case of Malaysia. *International Business Research*, *3*(1), P66.

[36] Rosenbusch, N., Brinckmann, J., & Bausch, A. (2011). Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing*, *26*(4), 441-457.

[37] Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic management journal*, *28*(13), 1319-1350.

ISBN: 978-1-61804-230-9

Knowledge workers – drivers to organizational performance in a knowledge-based economy

Muscalu Emanoil, Stanit Alexandra, Constantinescu Liliana Mihaela

Abstract— This paper illuminates significant relationships between knowledge workers and organizational performance and the perceived ability of today's organizations to descover and retain their knowledge workers. Knowledge workers are a critical challenge for today's organizations as they face increasing global competition with its demands for even more such workers. Knowledge management designs initiatives that accelerate knowledge creation, acquisition, and particularly knowledge capture, sharing and retention, are receiving unprecedented levels of investment as a result. While many factors impact organization financial performance, this research indicates that successful knowledge workers retention is significantly related with higher reported organizational performance in a knowledge-based economy.

Keywords—knowledge, knowledge worker, performance, knowledge-based economy.

I. Introduction

Retaining employees whose knowledge has high competitive value is becoming a critical and well-recognized challenge[1]. Such employees are known as knowledge workers in that they "have high degrees of expertise, education, or experience, and the primary purpose of their jobs involves the creation, distribution or application of knowledge" [2].

This study investigates if in today's global knowledg economy the impact of an organization's strategic orientation toward knowledge management, the learning culture of an organization, and specific human resource practices have an impact in knowledge worker development or retention and achieving organizational performance.

II. Definition of knowledge

One of the central problems in defining knowledge work has been the difficulty of defining knowledge itself and distinguishing knowledge from information. Indeed, the terms 5. 'information worker' and 'knowledge worker' can be used interchangeably. There is a vast literature in which the concept of management of knowledge is hard to distinguish from the management of information. What distinguishes knowledge from information is the way in which knowledge empowers actors with the capacity for intellectual or physical activity. Knowledge is a matter of cognitive capability and enables actors to do and reflect. Information, by contrast, is passive and meaningless to those without suitable knowledge. Knowledge provides the means by which information is interpreted and brought to life.

Nonaka described knowledge as the fuel for innovation, but was concerned that many managers failed to understand how knowledge could be leveraged. Companies are more like living organisms than machines, he argued, and most viewed knowledge as a static input to the corporate machine. Nonaka advocated a view of knowledge as renewable and changing, and that knowledge workers were the agents for that change. Knowledge-creating companies, he believed, should be focused primarily on the task of innovation.

Davenport[3] says that the rise of knowledge work has actually been foreseen for years. He points to the fact that Fritz Machlup did a lot of the early work on both knowledge as well as knowledge work roles and as early as 1958 stated that the sector was growing much faster than the rest of the economy with knowledge workers making up almost a third of the workforce in the United States [3].

III. What is a knowledge worker?

Several experts have outlined conceptual definitions of knowledge work. For example, Peter Drucker focused on the differences between 'manual worker productivity' and 'knowledge worker productivity.' The key enablers of the latter include abstractly defined tasks, flexible application of knowledge, workers' autonomy, continuous innovation and learning into job roles, assessment based on quality of output and perceiving workers as organisational assets.

In Drucker's oppinion Knowledge worker productivity is the biggest of the 21st century management challenges. In the developed countries it is their first survival requirement. In no other way can the developed countries hope to maintain themselves, let alone to maintain their leadership and their standards of living."

Knowledge work is mostly unseen, and therefore difficult to measure. You can't watch knowledge being created in the same way as a physical, tangible product. With knowledge work, it's the final output that matters, and the steps along the way are often less important.

Knowledge workers are those employees who have responsibility for exploring and generating ideas and concepts rather than concentrating solely on implementing or managing existing processes or operations Generally speaking, knowledge workers have high degrees of expertise, education,

ISBN: 978-1-61804-230-9

or experience and the primary purpose of their jobs involves the creation, distribution or application of knowledge. Knowledge workers differ from manual workers because:

- knowledge work is less standardized and structured
- knowledge workers are used to a certain level of autonomy
- before certain ends result it may be difficult to know whether knowledge workers are working or not
- knowledge workers basically own their key production mean – brains knowledge workers need to be committed to and enjoy their jobs.

In all organizations knowledge workers are a large category of workers that continues to grow. They are also the most expensive workers in organizations and they are essential to realizing the business strategies of the organization.

Knowledge workers are usually responsible for exploring and creating ideas, rather than implementing and managing existing processes. New products, new designs, new models for doing business – these are typical outputs of knowledge work.

Because knowledge workers are expected to produce results that are different from traditional workers, you should also manage them and measure their performance differently. Have an open mind, and recognize the different needs and motivations of knowledge workers. This will make it much easier to find creative and effective ways to keep their productivity high.

What differentiates knowledge work from other forms of work is its primary task of "non-routine" problem solving that requires a combination of convergent, divergent, and creative thinking [4].Also, despite the amount of research and literature on knowledge work there is yet to be a succinct definition of the term [5].

Knowledge workers are employees who have a deep background in education and experience and are considered people who "think for a living." [6].

Tapscott sees a strong, on-going linkage between knowledge workers and innovation, but the pace and manner of interaction have become more advanced. He describessocial media tools on the internet that now drive more powerful forms of collaboration. Knowledge workers engage in ‘’peer-to-peer’’ knowledge sharing across organizational and company boundaries, forming networks of expertise. Some of these are open to the public. While he echoes concern over copyright and intellectual property law being challenged in the marketplace, he feels strongly that businesses must engage in collaboration to survive. [7].

IV. TYPOLOGY OF KNOWLEDGE WORKERS ROLES

Knowledge workers can be grouped into various categories, based on the amount of time spent on individual tasks or on the type of information or skills possessed. The fact that knowledge workers can be classified in different ways is indicative of the variety of jobs they hold.

Knowledge workers can be categorized according to the amount of time engaged in routine versus innovative behaviors. On one end of the scale, workers perform tasks that are primarily repetitive and routine in nature but occasionally use complex information to make independent decisions, often with regard to customer service issues. Employees at the spectrum's opposite end spend most of their time accessing information and making independent decisions with regard to that information.

A second way to categorize those whose work focuses on information and ideas is as follows: specialty knowledge workers, portable knowledge workers, and creation of knowledge workers. Specialty knowledge workers possess a significant amount of knowledge related to a specific company's products or services. These individuals can be thought of as housing vital corporate assets in their heads. Portable knowledge workers possess information of wide and immediate utility. They are familiar with knowledge that is in demand by a variety of organizations. Software programmers, librarians, and persons with business degrees are examples of portable knowledge workers. Creation of knowledge workers focuses the majority of their efforts on innovative behaviors, such as product design and development. Examples of creation of knowledge workers include scientists and information systems designers.

Knowledge workers bring benefits to organizations in a variety of important ways. These include:

- analyzing data to establish relationships
- assessing input in order to evaluate complex or conflicting priorities
- identifying and understanding trends
- making connections
- understanding cause and effect
- ability to brainstorm, thinking broadly (divergent thinking)
- ability to drill down, creating more focus (convergent thinking)
- producing a new capability
- creating or modifying a strategy

These knowledge worker contributions are in contrast with activities that they would typically *not* be asked to perform, including:

- transaction processing
- routine tasks
- simple prioritization of work

There is a set of transitional tasks includes roles that are seemingly routine, but that require deep technology, product, or customer knowledge to fulfill the function. These include:

- providing technical or customer support
- handling unique customer issues
- addressing open-ended inquiries

Generally, if the knowledge can be retained, knowledge worker contributions will serve to expand the knowledge assets of a company. While it can be difficult to measure, this increases the overall value of its intellectual capital. In cases where the knowledge assets have commercial or monetary value, companies may create patents around their assets, at which point the material becomes restricted intellectual property. In these knowledge-intensive situations, knowledge workers play a direct, vital role in increasing the financial value of a company. They can do this by finding solutions on how they can find new ways to make profits this can also be related with market and research. Davenport, (2005) says that even if knowledge workers are not a majority of all workers, they do have the most influence on their economies. He adds that companies with a high volume of knowledge workers are the most successful and fastest growing in leading economies including the United States.

Role	**Description**	**Typical knowledge actions (expected)**	**Existence of the role in literature**
Controller	People who monitor the organizational performance based on raw information.	Analyze, dissemination, information organization, monitoring	(Moore and Rugullies, 2005) (Geisler, 2007)
Helper	People who transfer information to teach others, once they passed a problem.	Authoring, analyze, dissemination, feedback, information search, learning, networking	(Davenport and Prusak, 1998)
Learner	People use information and practices to improve personal skills and competence.	Acquisition, analyze, expert search, information search, learning, service search	
Linker	People who associate and mash up information from different sources to generate new information.	Analyze, dissemination, information search, information organization, networking	(Davenport and Prusak, 1998) (Nonaka and Takeushi, 1995) (Geisler, 2007)
Networker	People who create personal or project related connections with people involved in the same kind of work, to share information and support each other.	Analyze, dissemination, expert search, monitoring, networking, service search	(Davenport and Prusak, 1998) (Nonaka and Takeushi, 1995) (Geisler, 2007)
Organizer	People who are involved in personal or organizational planning of activities, e.g. to-do lists and scheduling.	Analyze, information organization, monitoring, networking	(Moore and Rugullies, 2005)
Retriever	People who search and	Acquisition, analyze,	(Snyder-Halpern *et*

	collect information on a given topic.	expert search, information search, information organization, monitoring	*al.*, 2001)
Sharer	People who disseminate information in a community.	Authoring, co-authoring, dissemination, networking	(Davenport and Prusak, 1998) (Brown *et al.*, 2002) (Geisler, 2007)
Solver	People who find or provide a way to deal with a problem.	Acquisition, analyze, dissemination, information search, learning, service search	(Davenport and Prusak, 1998) (Nonaka and Takeushi, 1995) (Moore and Rugullies, 2005)
Tracker	People who monitor and react on personal and organizational actions that may become problems.	Analyze, information search, monitoring, networking	(Moore and Rugullies, 2005)

Note: From "Knowledge Worker Roles and Actions—Results of Two Empirical Studies," by W. Reinhardt, B. Schmidt, P. Sloep, and H. Drachsler, 2011, *Knowledge and Process Management*, 18.3, p. 160.

V. RETAINING THE KNOWLEDGE WORKER

The shortage of knowledge workers makes employers concerned with attracting and retaining these employees. In order to hire and retain knowledge workers, employers may offer higher salaries, attractive work environments, and continuing educational opportunities. Employers take actions designed to attract and retain knowledge workers by creating a free-agent community, respecting knowledge workers as new bosses, and providing growth opportunities. In a free-agent community, employees have the freedom to choose their work methods and work in the environments in which they function best. Treating knowledge workers as the new bosses means that management operates as a facilitator rather than as a controller of work. This gives knowledge workers the autonomy they need to complete their work as they see fit. Employers make work attractive and rewarding by providing growth opportunities, such as those that are associated with ongoing training and development, special assignments, and rotation of jobs and job responsibilities. In such ways, employers attempt to address the knowledge worker shortage.

VI. THE KNOWLEDGE WORKER AND ORGANIZATIONAL PERFORMANCE

Knowledge worker work performance influences success in today's competitive work economy, and businesses are focusing on increasing this performance. Management facilitates the knowledge worker's job performance by providing access to relevant information; environments that promote this information's desired use, continuing educational opportunities, and a balance between guidance and autonomy. Employers use costly technologies to facilitate access to and manipulation of information.

The term information technology refers to computer equipment and programs used to access, process, store, and disseminate information. Examples of information technologies include word processing, spreadsheet, and electronic mail programs, and a variety of other software programs designed to process information in specific ways. Information technologies are designed to reduce the amount of time employees spend on information access, management and manipulation and to increase the accuracy of these processes. Information technology is important because it helps make information accessible and manageable in a time when accessibility and manipulation of information are crucial to the world economy.

ISBN: 978-1-61804-230-9

VII. Conclusion

The increasing demand for employees who use their skills and talents to perform complex and non-repetitive work presents both challenges and opportunities. The challenges include attainment and maintenance of a well educated, highly skilled, and efficient workforce. Opportunities include chances for greater numbers of working age people to hold more rewarding jobs than previously possible and for employees to be judged according to their unique talents and abilities rather with regard to how quickly they complete repetitious tasks or how well they conform to pre-established work standards.

Hiring, retention, and performance of knowledge workers will remain important issues. As the shortage of persons qualified to perform knowledge work increases, employers will be challenged to find more effective ways to hire and retain these individuals. In order to improve productivity, employers will try to figure out how to promote teamwork among knowledge workers, how to best design the workplace, and how to keep knowledge workers from becoming overwhelmed with the information they need to do their jobs.

References

[1] Somaya, D., & Williamson, I. O. (2008). *"Rethinking the 'War for Talent'."* MIT Sloan, Management Review, Volume 49:4.W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] Davenport, T.H., & Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know*. Boston: Harvard Business School Press.

[3] Davenport, Thomas H. (2005). *Thinking For A Living: How to Get Better Performance and Results From Knowledge Workers*. Boston: Harvard Business School Press. ISBN 1-59139-423-6.

[4] Reinhardt, W.; Schmidt, B.; Sloep, P.; Drachsler, H. (2011). "Knowledge Worker Roles and Actions – Results of Two Empirical Studies". *Knowledge and Process Management* **18** (3): 150–174

[5] Pyöriä, P. (2005). "The Concept of Knowledge Work Revisited". *Journal of Knowledge Management* **9** (3): 116–127

[6] Cooper, Doug (2006). "Knowledge Workers". *Canadian Business* **79** (20): 59.

[7] Tapscott, Don; Williams, Anthony D. (2006). *Wikinomics: How Mass Collaboration Changes Everything*. New York: Penguin. ISBN 1-59184-138-0.

Numerical Solution of Compressible Viscous Airflow in Vocal Tract

Petra Pořízková, Karel Kozel, Jaromír Horáček,

Abstract—This study deals with the numerical solution of a compressible flow of a viscous fluid in a 2D and 3D channel for low inlet airflow velocity. The channel is a simplified model of the glottal space in the human vocal tract. The system of Navier-Stokes equations has been used as mathematical model of laminar flow of the compressible viscous fluid in a domain. The numerical solution is implemented using the finite volume method (FVM) and the predictor-corrector MacCormack scheme with artificial viscosity using a grid of quadrilateral cells. The numerical simulations of flow fields in the channel, acquired from a developed program, are presented for inlet velocity $\hat{u}_\infty = 4.12\mathrm{ms}^{-1}$ and Reynolds number $\mathrm{Re}_\infty = 4481$.

Keywords—CFD, compressible, viscous, vocal tract, acoustic energy

I. Introduction

A current challenging question is a mathematical and physical description of the mechanism for transforming the airflow energy in the glottis into the acoustic energy representing the voice source in humans. The voice source signal travels from the glottis to the mouth, exciting the acoustic supraglottal spaces, and becomes modified by acoustic resonance properties of the vocal tract [1].

Acoustic wave propagation in the vocal tract is usually modeled from incompressible flow models separately using linear acoustic perturbation theory, the wave equation for the potential flow [2] or the Lighthill approach on sound generated aerodynamically [3]. In reality, the airflow coming from the lungs causes self-oscillations of the vocal folds, and the glottis completely closes in normal phonation regimes, generating acoustic pressure fluctuations. In this study, the movement of the boundary channel is known, harmonically opening and nearly closing in the narrowest cross-section of the channel.

Goal is numerical simulation of flow in the channel which involves attributes of real flow causing acoustic perturbations.

II. Mathematical model

The system of Navier-Stokes equations has been used as mathematical model to describe the unsteady laminar flow of the compressible viscous fluid in a domain. The system is expressed in non-dimensional conservative form [4]:

$$\frac{\partial \mathbf{W}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} + \frac{\partial \mathbf{H}}{\partial z} = \frac{1}{\mathrm{Re}} \left(\frac{\partial \mathbf{R}}{\partial x} + \frac{\partial \mathbf{S}}{\partial y} + \frac{\partial \mathbf{T}}{\partial z} \right). \quad (1)$$

P. Pořízková is with the Department of Technical Mathematics, Czech Technical University in Prague, Karlovo náměstí 13, 121 35, Prague 2, Czech Republic, e-mail: puncocha@marian.fsik.cvut.cz.

K. Kozel and J. Horáček are with Institute of Thermomechanics Academy of Sciences, Dolejškova 5, Prague 8, Czech Republic.

Manuscript received April 19, 2005; revised January 11, 2007.

$\mathbf{W} = [\rho, \rho u, \rho v, \rho w, e]^T$ is the vector of conservative variables where ρ denotes density, (u, v, w) is velocity vector and e is the total energy per unit volume. $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ are the vectors of inviscid fluxes and $\mathbf{R}$, $\mathbf{S}$, $\mathbf{T}$ are the vectors of viscous fluxes. The static pressure p in inviscid fluxes is expressed by the state equation in the form

$$p = (\kappa - 1) \left[e - \frac{1}{2} \rho \left(u^2 + v^2 + w^2 \right) \right], \quad (2)$$

where $\kappa = 1.4$ is the ratio of specific heats.

The reference variables for transformation are inflow variables (marked with the infinity subscript): the speed of sound $\hat{c}_\infty = 343\ \mathrm{ms}^{-1}$, density $\hat{\rho}_\infty = 1.225\ \mathrm{kg\ m}^{-3}$, dynamic viscosity $\hat{\eta}_\infty = 18 \cdot 10^{-6}\ \mathrm{Pa \cdot s}$ (for temperature $\hat{T}_\infty = 293.15\ \mathrm{K}$) and a reference length $\hat{L}_r = 0.02\ \mathrm{m}$. General Reynolds number in (1) is computed from reference variables $\mathrm{Re} = \hat{\rho}_\infty \hat{c}_\infty \hat{L}_r / \hat{\eta}_\infty$. The non-dimensional dynamic viscosity in the dissipative terms is a function of temperature in the form $\eta = (T/T_\infty)^{3/4}$.

III. Computational domain and boundary conditions

The bounded computational domains $D_{2\mathrm{D}}$ and $D_{3\mathrm{D}}$ used for the numerical solution are shown in Fig. 1. The domains are symmetric, the shape of which is inspired by the shape of the trachea (inlet part), vocal folds, false vocal folds and supraglottal spaces (outlet part) in human vocal tract. The upper and the lower boundaries are the channel walls. A part of the upper and the lower boundaries (walls) changes its shape between the points A and B according to given harmonic function of time and axial coordinate. The gap width is the narrowest part of the channel in y direction and is oscillating between the minimum $g_{min} = 0.4$ mm and maximum $g_{max} = 2.8$ mm.

The boundary conditions are considered in the following formulation:

1) Upstream conditions: in 2D uniform velocity profile $u_\infty = \frac{\hat{u}_\infty}{\hat{c}_\infty}$ or in 3D constant flow rate $H^2 \cdot u_\infty$ is at the inlet; $\rho_\infty = 1$; p_∞ is extrapolated from domain.
2) Downstream conditions: $p_2 = 1/\kappa$; $(\rho, \rho u, \rho v, \rho w)$ are extrapolated from domain.
3) Flow on the wall: $(u, v, w) = (u_{wall}, v_{wall}, w_{wall})$ - velocity of the channel walls and for temperature $T = \kappa p/\rho$ is $\frac{\partial T}{\partial \vec{n}} = 0$.

The general Reynolds number in (1) is multiply with non-dimensional value $\frac{\hat{u}_\infty}{\hat{c}_\infty} H$ represents kinematic viscosity scale and for computation of the real problem inlet Reynolds number $\mathrm{Re}_\infty = \hat{\rho}_\infty \hat{c}_\infty \frac{\hat{u}_\infty}{\hat{c}_\infty} H \hat{L}_r / \hat{\eta}_\infty$ is used.

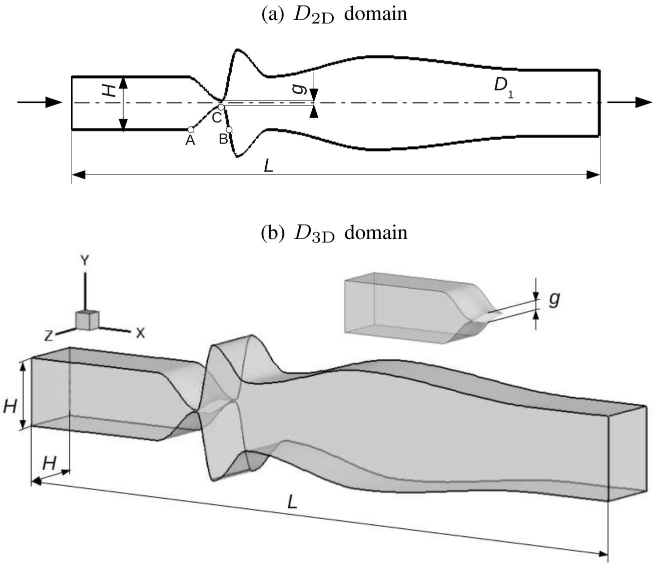


Fig. 1. The computational domains $D_{2\mathrm{D}}$ and $D_{3\mathrm{D}}$. $L = 8$ (160 mm), $H = 0.8$ (16 mm), $g = 0.08$ (1.6 mm) - middle position.

IV. Numerical solution

The numerical solution uses finite volume method (FVM) in cell centered form on the grid of quadrilateral cells, see e.g. [4]. The bounded domain is divided into mutually disjoint sub-domains $D_{i,j,k}$ (i.e. quadrilateral cells). The system of equations (1) is integrated over the sub-domains $D_{i,j,k}$ using the Green formula and the Mean value theorem. In the time-changing domain, the integral form of FVM is derived using Arbitrary Lagrangian-Eulerian (ALE) formulation. The ALE method defines homomorphic mapping of the reference domain $D_{t=0}$ at initial time $t = 0$ to a domain D_t at $t > 0$ [5]. The explicit predictor-corrector MacCormack (MC) scheme in the domain is used. The scheme is 2nd order accurate in time and space [4]:

$$\begin{aligned}
\mathbf{W}^{n+1/2}_{i,j,k} &= \frac{\mu^n_{i,j,k}}{\mu^{n+1}_{i,j,k}} \mathbf{W}^n_{i,j,k} \\
&- \frac{\Delta t}{\mu^{n+1}_{i,j,k}} \sum_{q=1}^{6} A_q \Big[\Big(\tilde{\mathbf{F}}^n_q - s_{1_q} \mathbf{W}^n_q - \frac{1}{Re} \tilde{\mathbf{R}}^n_q \Big) n_{1_q} \\
&+ \Big(\tilde{\mathbf{G}}^n_q - s_{2_q} \mathbf{W}^n_q - \frac{1}{Re} \tilde{\mathbf{S}}^n_q \Big) n_{2_q} \\
&+ \Big(\tilde{\mathbf{H}}^n_q - s_{3_q} \mathbf{W}^n_q - \frac{1}{Re} \tilde{\mathbf{T}}^n_q \Big) n_{3_q} \Big], \\
\overline{\mathbf{W}}^{n+1}_{i,j,k} &= \frac{\mu^n_{i,j,k}}{\mu^{n+1}_{i,j,k}} \frac{1}{2} \Big(\mathbf{W}^n_{i,j,k} + \mathbf{W}^{n+1/2}_{i,j,k} \Big) \\
&- \frac{\Delta t}{2\mu^{n+1}_{i,j,k}} \sum_{q=1}^{6} A_q \Big[\Big(\tilde{\mathbf{F}}^{n+1/2}_q - s_{1_q} \mathbf{W}^{n+1/2}_q - \frac{1}{Re} \tilde{\mathbf{R}}^{n+1/2}_q \Big) n_{1_q} \\
&+ \Big(\tilde{\mathbf{G}}^{n+1/2}_q - s_{2_q} \mathbf{W}^{n+1/2}_q - \frac{1}{Re} \tilde{\mathbf{S}}^{n+1/2}_q \Big) n_{2_q} \\
&+ \Big(\tilde{\mathbf{H}}^{n+1/2}_q - s_{3_q} \mathbf{W}^{n+1/2}_q - \frac{1}{Re} \tilde{\mathbf{T}}^{n+1/2}_q \Big) n_{3_q} \Big], \qquad (3)
\end{aligned}$$

where $\Delta t = t^{n+1} - t^n$ is the time step, $\mu_{i,j,k} = \int\int_{D_{i,j,k}} dx\ dy\ dz$ is the volume of cell $D_{i,j,k}$, $\mathbf{n}_q = (n_1, n_2, n_3)_q$ is outlet normal vector on face q (see Fig. 2), A_q is area of the face and vector $\mathbf{s}_q = (s_1, s_2, s_3)_q$ represents the speed of the face. The physical fluxes $\mathbf{F}$, $\mathbf{G}$, $\mathbf{R}$, $\mathbf{S}$, $\mathbf{T}$ on the face q of the cell $D_{i,j,k}$ are replaced by numerical fluxes (marked with tilde) $\tilde{\mathbf{F}}$, $\tilde{\mathbf{G}}$, $\tilde{\mathbf{R}}$, $\tilde{\mathbf{S}}$, $\tilde{\mathbf{T}}$ as approximations of the physical fluxes. The higher partial derivatives of velocity and temperature in $\tilde{\mathbf{R}}_q$, $\tilde{\mathbf{S}}_q$, $\tilde{\mathbf{T}}_q$ are approximated using dual volumes V'_q as shown in Fig. 2.

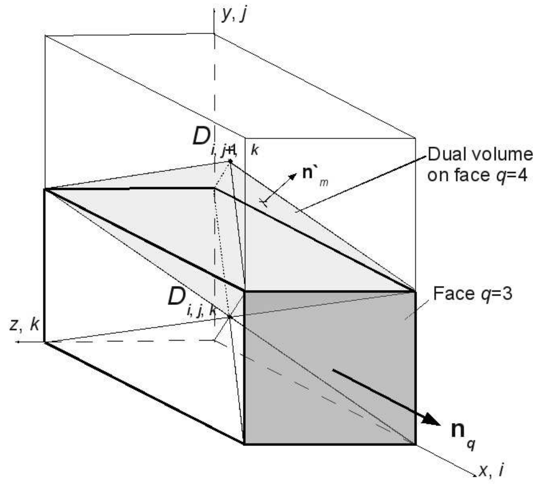


Fig. 2. Finite volume of cell $D_{i,j,k}$ and dual volume V'_q on face q

The last term used in the MC scheme is the Jameson artificial dissipation $AD(W_{i,j,k})^n$ [6], then the vector of conservative variables $\mathbf{W}$ can be computed at a new time level $\mathbf{W}^{n+1}_{i,j,k} = \overline{\mathbf{W}}^{n+1}_{i,j,k} + AD(W_{i,j,k})^n$.

The grid of the channel have successive refinement cells near the wall, the minimum cell size in y and z directions is $\Delta y_{min}, \Delta z_{min} \approx 1/\sqrt{\mathrm{Re}_\infty}$ to resolve capture boundary layer effects (see Fig. 3).

V. Numerical results

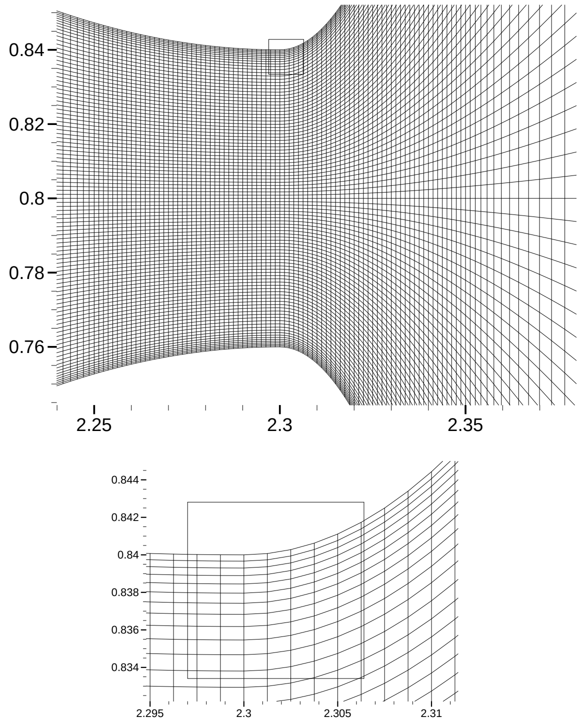
The numerical results were obtained (using a specifically developed program) for the following input data: inflow ratio velocity $\frac{\hat{u}_\infty}{\hat{c}_\infty} = 0.012$ ($\hat{u}_\infty = 4.116\ m/s$), Reynolds number $\mathrm{Re}_\infty = 4481$ and atmospheric pressure $p_2 = 1/\kappa$ ($\hat{p}_2 = 102942$ Pa) at the outlet. For solution in unsteady domain the wall oscillation frequency $\hat{f} = 100\ Hz$.

The computational domains contained 450×100 cells in $D_{2\mathrm{D}}$ and $200 \times 100 \times 40$ cells in $D_{3\mathrm{D}}$, detail of the mesh is shown in Fig. 3.

The unsteady numerical simulation of the airflow computed in domain $D_{2\mathrm{D}}$ over the fourth cycle of the wall oscillations is presented in Fig. 4 showing the unsteady flow field in three time instants during opening phase of one vibration period. Large eddies are developing in supraglottal spaces and a “Coandă” effect is apparent in the flow field pattern. The absolute maximum of Mach number $M_{max} = 0.535$ ($\hat{u}_{max} = 183.5\ ms^{-1}$) in the flow field was achieved during opening phase behind the narrowest channel cross-section. The flow becomes practically periodic after the first period of oscillations.

Figure 5 shows the acoustic pressure spectrum during three vibrations periods of the walls computed near the outlet on the axis of the channel $D_{2\mathrm{D}}$ from pressure field. Four acoustic resonances of the channel cavity at about $\hat{f}_0 = 100$, $\hat{f}_1 =$

(a) Mesh of the quadrilateral cells $D_{i,j}$ between A,B points of the domain D_{2D} and detail of the mesh near the wall



(b) Mesh of the quadrilateral cells $D_{i,j,k}$ in part of the domain D_{3D} and detail of mesh in the gap

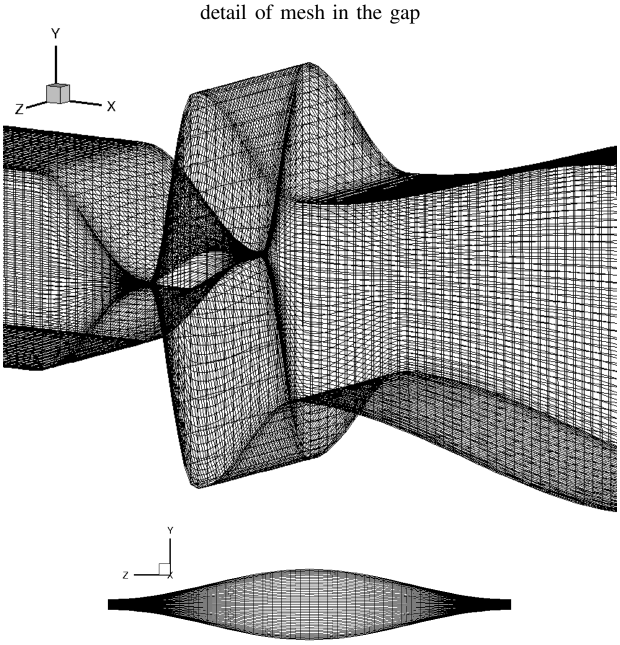
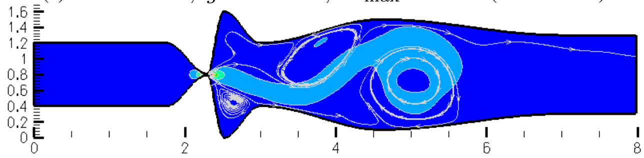
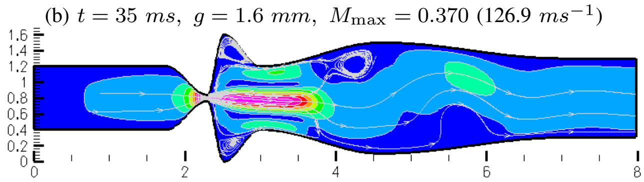


Fig. 3. Mesh in 2D and 3D domains.

(a) $t = 32.5\ ms,\ g = 0.4\ mm,\ M_{\max} = 0.236$ (80.9 ms^{-1})



(b) $t = 35\ ms,\ g = 1.6\ mm,\ M_{\max} = 0.370$ (126.9 ms^{-1})



(c) $t = 37.5\ ms,\ g = 2.8\ mm,\ M_{\max} = 0.097$ (33.3 ms^{-1})

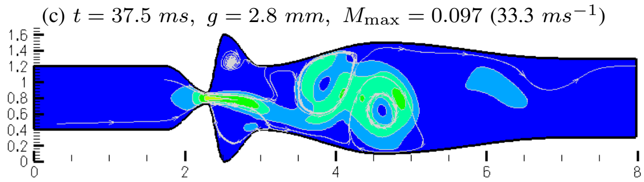


Fig. 4. The unsteady numerical solution of the airflow in D_{2D}- $\hat{f} = 100\ Hz$, $\frac{\hat{u}_\infty}{\hat{c}_\infty} = 0.012$, $\mathrm{Re}_\infty = 4481$, $p_2 = 1/\kappa$, 450×100 cells. Data computed during the fourth oscillation cycle. Results are mapped by iso-lines of velocity ratio and by streamlines.

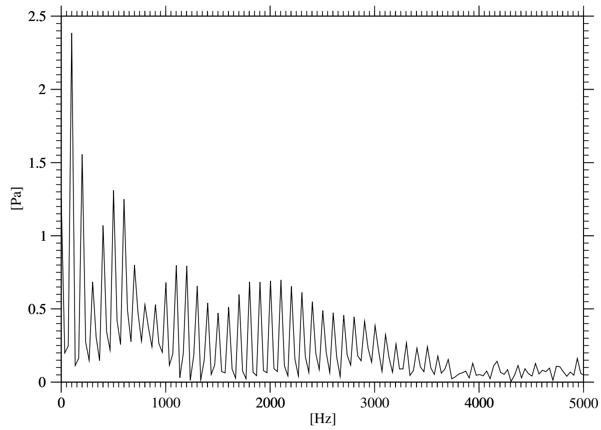


Fig. 5. The acoustic pressure spectrum in D_{2D} - $\frac{\hat{u}_\infty}{\hat{c}_\infty} = 0.012$, $\mathrm{Re}_\infty = 4481$, $p_2 = 1/\kappa$, 450×100 cells.

550, $\hat{f}_2 = 1150$, $\hat{f}_3 = 1950\ Hz$ can be identified in the spectrum envelope of the pressure. The first acoustic resonance $\hat{f}_1 = 550\ Hz$ (see spectrum peak at cc 500 Hz) corresponds to the first eigenfrequency $\hat{F}_1$ of a simple tube of the length of the complete channel closed at the inlet and open at the outlet ($\hat{F}_1 = \hat{c}_\infty/(4L \cdot \hat{L}_r) = 343\ ms^{-1}/(4 \cdot 0.16\ m) = 536\ Hz$). The acoustic resonances are more damped with increasing frequency, which can be caused by the fluid viscosity as well as by a numerical viscosity implemented in the numerical method (constants magnitude in $AD(W_{i,j})^n$) [7].

The application of the method for low Mach number at inlet ($M_\infty = u_\infty = 0.012$) in 3D domain D_{3D} is shown in Fig. 6. The results are mapped by isolines of the velocity ratio and by streamlines using slices of the domain in x, y, z directions. Fig. 7 shows convergence to the steady state solution using L_2 norm of momentum residuals which is good. From 2D computation experiences authors have decided to compute more iterations until the “Coandă” effect in flow field appears (for more see e.g. [7]). Then the steady numerical solution will be used as initial condition for simulation of the unsteady solution in domain D_{3D} with moving upper and lower walls to obtain acoustic characteristics.

ISBN: 978-1-61804-230-9

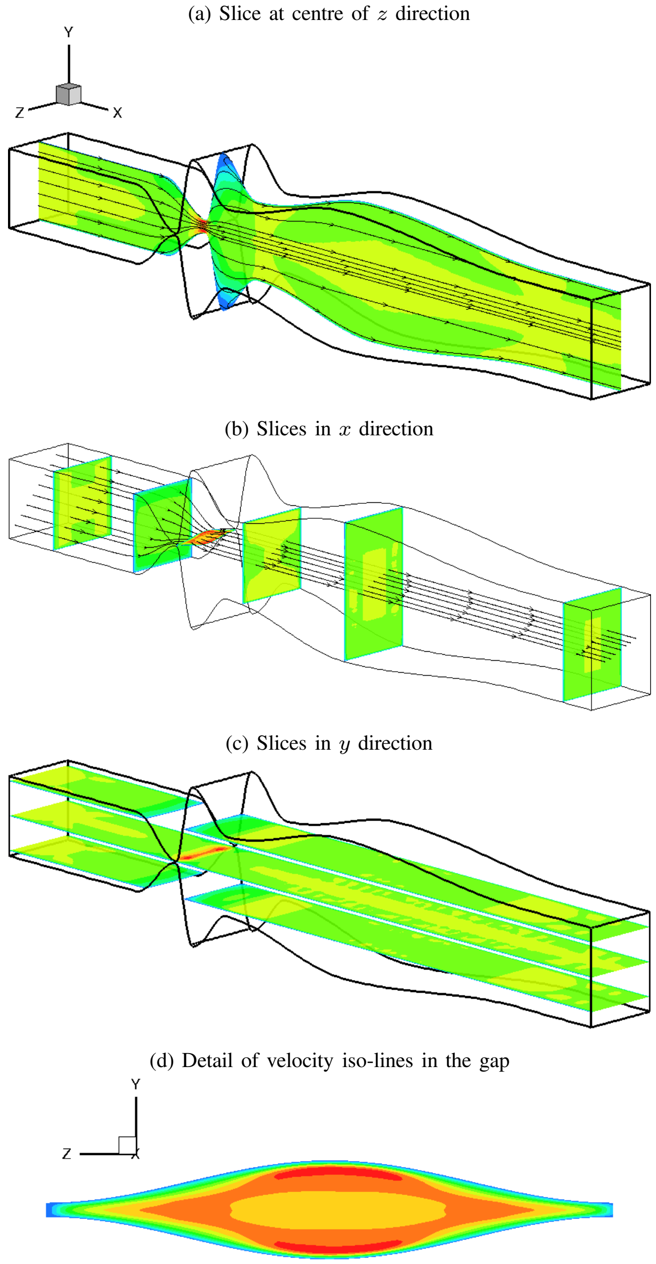


Fig. 6. The steady numerical solution ($\hat{f} = 0\ Hz$) of the airflow in D_{3D} - $\frac{\hat{u}_\infty}{\hat{c}_\infty} = 0.012$, $\mathrm{Re}_\infty = 4481$, $p_2 = 1/\kappa$, $200 \times 100 \times 40$ cells. Results are mapped by iso-lines of velocity ratio and by streamlines.

VI. Conclusion

In 2D unsteady simulations acoustic properties have been found. Is possible to detect a “Coandă phenomenon” and large-scale vortices in the flow field patterns. A similar generation of large-scale vortices, vortex convection and diffusion, jet flapping, and general flow patterns were experimentally obtained in physical models of the vocal folds by using Particle Image Velocimetry method in [8]. In [9] the compressible viscous flow is solved using discontinuous Galerkin finite element method (DGFM). Numerical solution by DGFM shows similar behavior of the flow jet and large-scale vortices. The analysis of the computed pressure revealed basic acoustic characteristics of the channel D_{2D}.

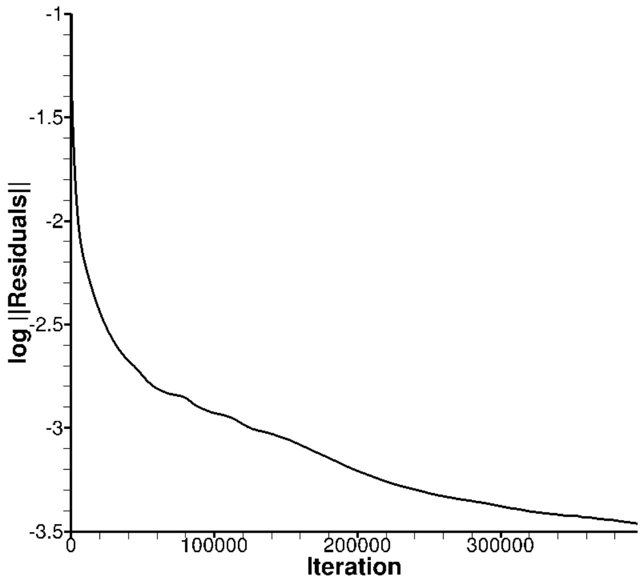


Fig. 7. The convergence to the steady state solution using L_2 norm of momentum residuals (ρu) in D_{3D} - $\frac{\hat{u}_\infty}{\hat{c}_\infty} = 0.012$, $\mathrm{Re}_\infty = 4481$, $p_2 = 1/\kappa$, $200 \times 100 \times 40$ cells.

The method described in this study will be used for 3D simulation of unsteady flow in domain with vibrating walls near the gap region to simulate airflow in human vocal tract. Completion computation of the unsteady flows in 3D channel with vibrating walls is expecting in short time.

Acknowledgment

This contribution was partially supported by GAČR P101/11/0207, 13-005-22S and P101/10/1329.

References

[1] Titze, I.R., Principles of Voice Production. National Centre for Voice and Speech, Iowa City, 2000. ISBN 0-87414-122-2.

[2] Titze, I.R., The Myoelastic Aerodynamic Theory of Phonation. National Centre for Voice and Speech, Iowa City, 2006. ISBN 0-87414-122-2.

[3] Zörner, S., Kalteenbacher, M., Mattheus, W., Brücker, C., Human phonation analysis by 3d aero-acoustic computation. In: Proceedings of the International Conference on Acoustic NAG/DAGA 2009, 1730-1732, Rotterdam, 2009.

[4] Fürst, J., Janda, M., Kozel, K., Finite volume solution of 2D and 3D Euler and Navier-Stokes equations. In: P. Neustupa, J. & Penel (Eds.), Mathematical fluid mechanics, 173-194, Berlin, 2001.

[5] Honzátko, R., Horáček, J., Kozel, K., Solution of inviscid incompressible flow over a vibrating profile. In: M. Beneš & M. Kimura & T. Nataki (Eds.), COE Lecture notes, **3**: 26-32, Kyushu university, 2006. ISSN 1881-4042.

[6] Jameson, A., Schmidt, W.,Turkel, E., Numerical solution of the Euler equations by the finite volume methods using Runge-Kutta time-stepping schemes, AIAA,81-125,1981.

[7] Pořízková, P., Kozel, K., Horáček, J., Numerical Simulation of Unsteady Compressible Flow in Convergent Channel: Pressure Spectral Analysis In: Journal of Applied Mathematics. 2012, vol. 2012, no. 1, p. 1-9. ISSN 1687-0042.

[8] Horáček, J., Šidlof, P., Uruba, V., Veselý, J., Radolf, V., Bula, V., PIV Measurement of Flow-Patterns in Human Vocal Tract Model. In: Proceedings of the International Conference on Acoustic NAG/DAGA 2009, 1737-1740, Rotterdam, 2009.

[9] Prokopová, J., Feistauer, M., Horáček, J., Kučera, V., Numerical simulation of airflow through the model of oscillating vocal folds. Applied and Computational Mechanics 4(2010) 215-224.

On Connectivity of Ad Hoc Network Using Fuzzy Logic

Mohit Jain
Division of Computer Engineering
Netaji Subhas Institute of Technology (NSIT)
University of Delhi
New Delhi, INDIA
Jain.mohit@nsitonline.in

Satish Chand
Division of Computer Engineering
Netaji Subhas Institute of Technology (NSIT)
University of Delhi
New Delhi, INDIA
Satish@nsit.ac.in

Abstract— This paper investigates the use of fuzzy logic approach to understand the level of nodes connectivity in the ad hoc networks that can be employed to generate the list of simple if then rules. It helps to determine the transmission range needed in order to achieve the sure connectivity of the wireless multi hop system for a given node density. In other words, if the transmission range of the nodes is given, we can find out the number of nodes needed to cover a certain area. We validate our results through simulation.

.Keywords—Ad hoc networks, connectivity, topology control, critical transmitting range, node density, energy consumption routing ,Critical points, k-connectivity, Fuzzy Logic.

I. Introduction

Mobile ad hoc networks are decentralized, self-organizing networks capable of forming a network without relying on any fixed infrastructure [1,2]. The advantage of such networks is that the mobile devices communicate with each other in peer to peer fashion, establish a self organizing network without the need of any access point or any pre-existing infrastructure. In short, they can be formed in a spontaneous way, that’s why they are known as ad hoc networks [3-7]. One of the main properties of the MANETs is node mobility, which causes network topology to change ,leading to the problem of network connectivity. To achieve a fully connected ad hoc network, there must be a path between each pair of nodes. The connectivity depends on the number of nodes per unit area and their radio transmission range. Each single mobile node contributes to the connectivity of the entire network. The correct adjustment of radio transmission range is therefore an important system feature. Increasing transmission range of nodes will possibly lead to direct link between each pair of nodes. On the other hand, decreasing the transmission range of nodes may lead to them isolated and hence the network disconnectivity.

It is however not decidable how much the transmission range should be considered in order to achieve almost surely connected network. There have been some studies to understand the network connectivity of ad networks by considering the node density and transmission range .In this paper, we apply fuzzy logic[8-11] to understand the network connectivity by considering the same parameters and simulate the result using NS2[12]. Our approach is much simpler and effective.

The organization of the remaining paper is as follows. Section II discusses the related literature. .Section III presents the proposed approach and section IV discussed the fuzzy simulation and section V discuss the network simulation to verify the proposed approach .Finally the paper is concluded in section VI.

II. Literature Review

The problem of node connectivity was very first time discussed by Cheng and Robertazzi in 1989 [13]. They investigated the influence of node density and transmission range of a node‘s broadcast in a multi hop radio network modeled by spatial Poisson process. They suggested that for optimizing the transmission range, its value should be lower bounded to maintain desired network connectivity. However, they could not implement it in real scenario. The paper [14], an extension of [13], discusses the disconnectedness of Poisson distributed nodes. It provides some insights on critical coverage range vs. critical transmission range of the nodes placed in a square area according to Poisson fixed density. This problem is further discussed in [15] for one dimensional line segment that determines the critical transmitting range for Poisson distributed nodes in square area. However, these works [14, 15] are difficult to apply in real scenarios because in a Poisson process the actual number of deployed nodes is a random variable itself whose only average value can be found. The paper [16] discusses the critical power of nodes in a network for transmission to ensure network connectivity by using the percolation theory [17]. The probabilistic lower and upper bounds for isolated and connected nodes fail to explain about nodes not placed independently in a disc. The paper [18] investigates the connectivity of hybrid ad hoc networks consisting the Poisson distributed nodes using the percolation theory [17]. It reports that for populated regions one dim sparse network is well suited and the pure ad hoc network is useful relatively low density areas. For density critical areas the cellular network can provide the acceptable connectivity. The paper [19] estimates the nearest neighbors when network becomes disconnected. In this work, the same number of nearest neighbors is maintained for each node. If each node is connected to less than 0.074logn nearest neighbors the network is asymptotically disconnected. If each node is connected to more than 5.1774logn nearest neighbors, then it is asymptotically connected. The paper [20] discusses the connectivity augmentation problem and determines a set of edges of minimum weight to be inserted in order to make the resulting graph λ-vertex edge connected. It is reported that the problem is NP hard for $\lambda>1$. In [21], the same work has been discussed by using the concept of minimum geometric disk cover (MGDC) problem, commonly used in wireless networking applications or facility location problems. The paper [22] uses the random graph theory and theory of kolmogorov complexity to establish the network connectivity via building local cluster head connections between nearby cluster heads without considering global network topology. In [23], the radio transmission range problem is analyzed and the probabilistic bounds for isolated nodes and connected nodes with uniform nodes on 1- 2- and 3-dimension are calculated. It reports that the transmitting range of nodes can be reduced substantially from the deterministic requirements if there is high probability of connectedness. The paper [24] extends the work [23] and discusses the asymptotic minimum node degree of a graph on uniform points in d dimension .Some more studies on radio transmission range problem are discussed in [25-30]. The works [25-28] however do not consider inhomogeneous nodes .The issue of k-connectivity with respect to different transmission ranges has been discussed in [26, 27]. The same has been analyzed using the stochastic connectivity properties of the wireless multi hop networks in [29]. The paper [30] discusses the connectivity for inhomogeneous node distributions with random waypoint (RWP) nodes. The paper [31] extends the Bettsetter's work [30] by incorporating the deployment border effects on the range to provide k-connectivity. In [32], the critical transmission power based on Bettsetter [30] is discussed to maintain k-connectivity, ensuring k-neighbors of a node is a necessary condition but not the sufficient condition for k-connectivity. It is because the network graph may have critical points which can cause the network failure and destroys the end-to-end network connectivity. In [33, 34], the critical transmitting range for connectivity in both stationary and mobile ad hoc networks has been analyzed. The paper [33] also discusses the probability for establishing a multi hop path between two Poisson distributed nodes on an infinite line with a given distance. The paper [35] discusses about the node that keeps a multi hop path to a fixed base station with the nodes moving in a straight line away from the base station. The k-connectivity concept has been further extended in [36] that studies the critical number of neighbors needed for k-connectivity. There are critical or weak points that play a major role in destroying the network connectivity. The works [26, 27, 29, 31 36] have not discussed critical points.The paper [37] characterizes the critical transmitting range by using the asymptotic distribution of the longest minimum spanning tree [38, 39]. In [40], the problem of minimizing the maximum of node transmitting ranges while achieving connectedness is discussed. The basic assumption here is that the relative distance of all nodes is considered as input to the centralized topology control algorithm. In [41], a distributed topology control protocol is discussed to minimize the energy required to communicate with a given master node. In this work, every node is equipped with a GPS receiver to provide position information. Initially every node iteratively broadcasts its position to different search regions. When the node is able to calculate a set of nodes, called as its enclosure, based on the position information obtained from neighbors, this process stops. Its major drawback is that the number of iterations to determine the enclosure depends on the definition of initial search region, which affects the energy consumption of the protocol. The same problem has been analyzed in [42] using directional information obtained by using multi-directional antenna. But such setup is not possible in sensor networks because the nodes are very simple and have no centralized communication facility.

The above mentioned works have used graph theory [43], modern graph theory [44],and Probability theory [45], Statistics for spatial data [46], Random Graphs [47], Geometric Random Graphs [48] in order to work on the problem of node connectivity in MANETs. We use fuzzy logic to address the network connectivity problem in ad hoc networks . Our method provides better results, than the existing methods discuss in literature [26-30].

III. Proposed Work

To achieve a fully connected ad hoc network there must be a path between each pair of nodes. The connectivity therefore depends on the number of nodes per unit area (node density) and radio transmission range. Each single mobile node contributes to the connectivity of the entire network. The correct adjustment of radio transmission range is therefore an important system feature. Consider Fig 1in which O is a given node .If some node A is in the inner circle, then the node O has direct link with node A. If the node O has transmission range r and some node B lies in outer circle, then there must be some other node C in the range of O which lies in the communication range of both O as well as node B lying in the outer circle.

Increasing the transmission power of node O will increase transmission range that will help to cover more nodes like B via direct link. On the other hand, decreasing transmission power may lead to its isolation. It however cannot be decided to have almost a surely connected network for the given transmission ranges of the nodes .In this paper we address this problem using fuzzy logic. We estimate the number of nodes and transmission range needed for a given area to ensure network connectivity. We take node density and transmission range as two inputs and one output i.e. network connectivity. Both the input variables take three linguistic values: low, high, very high and the output variable has two linguistic values: surely and poorly. The rule base is given by Table I

ISBN: 978-1-61804-230-9

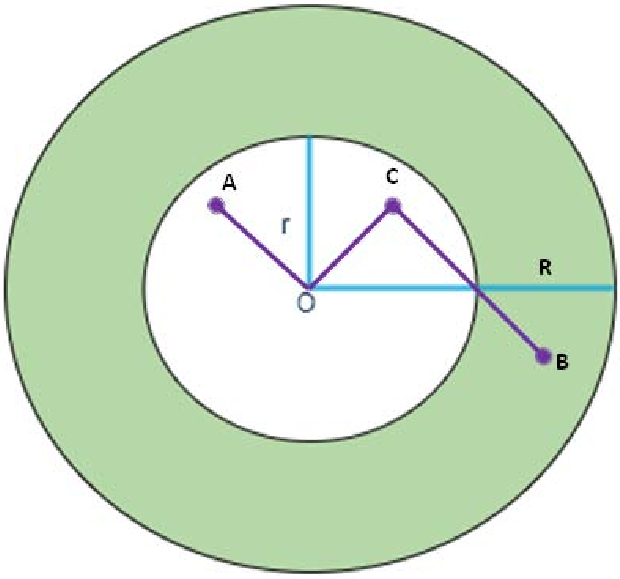


Fig. 1 Effect of transmission range on network connectivity

Table I. Fuzzy Rules for Network Connectivity

Node Density / Transmission Range	Low	High	Very High
Low	Poorly	Poorly	Poorly
High	Surely	Surely	Surely
Very High	Surely	Surely	Surely

The node density parameter assumes values in the set for 500m x 500m {0.000028,0.00004,0.00008,0.00012,0.000016,0.0002,0.00024,0.00028,0.00032,0.00036,0.0004} and for 1000m x 1000m {0.000007,0.00001,0.00002,0.00003,0.00004,0.00005,0.00006,0.00007,0.00008,0.00009,0.0001} and the transmission range assumes values in the set {40,60,80,100,120,150,160,180,200,220,250,300} and the network connectivity assumes values in the set {10,20,30,40,50,60,70,80,90,100}.The membership function for input parameters are triangular as shown in Fig 2 ,Fig 3and Fig4and that for output parameter is Gaussian , as shown in Fig 5

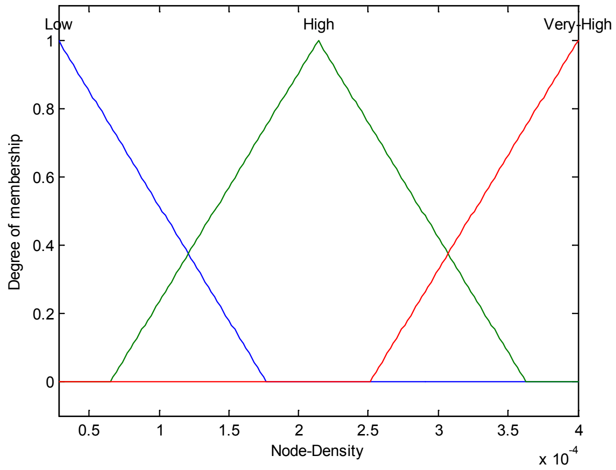


Fig 2: Illustration of Triangular Membership Function of Node Density for 500m x 500m

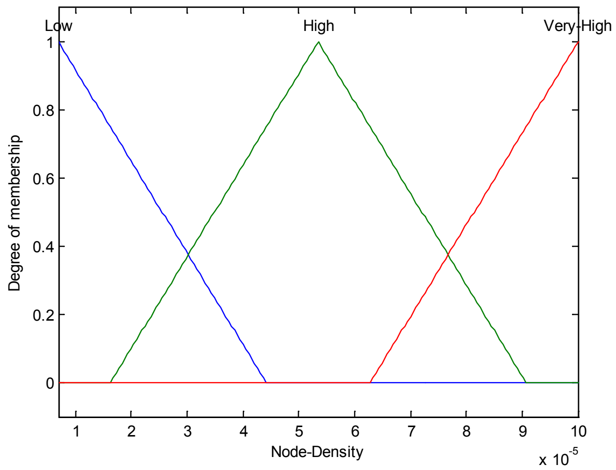


Fig 3: Illustration of Triangular Membership Function of Node Density for 1000m x 1000m

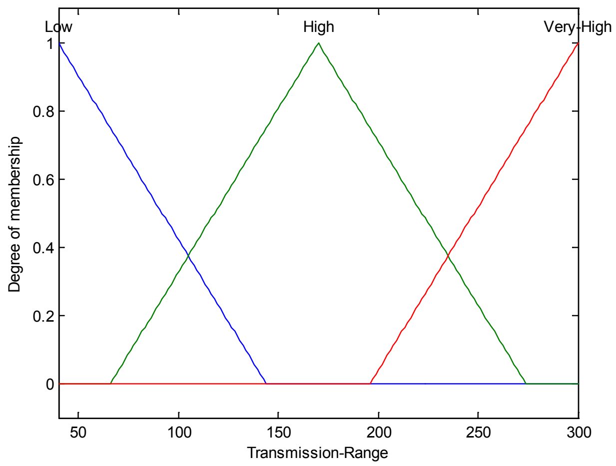


Figure 4: Illustration of Triangular Membership Function of Transmission Range

ISBN: 978-1-61804-230-9

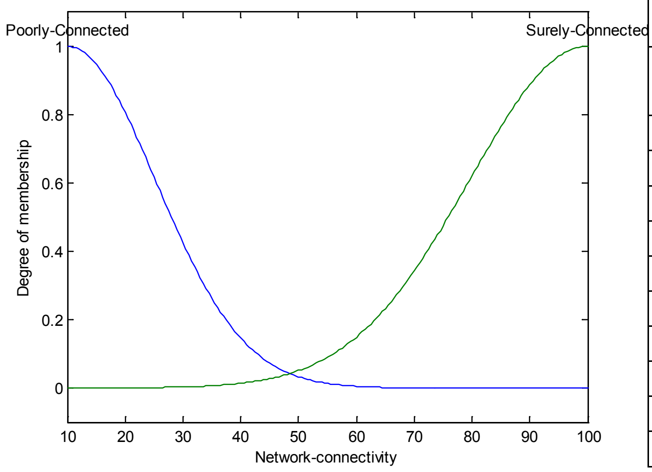


Figure 5: Illustration of Gaussian Membership Curve of Network Connectivity

IV. Fuzzy Simulation

We have taken the normalized value for each parameter because the actual value of the parameters might be different for different networks. The crisp normalized values are converted into fuzzy variable. We have estimated the node density by varying the number of nodes from 7 to 100 for the given area 500x500m^2 and 1000x1000m^2. In fuzzy simulation [49] we have estimated the network connectivity, based on node density and transmission range, as shown in Fig6 and Fig7.

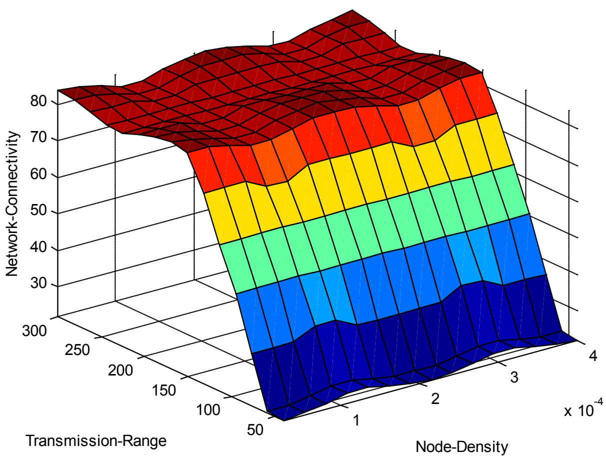


Fig6. Illustration of fuzzy simulation for terrain of size 500mx500m

Fig6 provide the information about the network connectivity which is summaries in tables II with respected to transmission range 80m, 100m, 150m and 250m for terrain of size 500m x 500m.

Table II. Analysis of Fig6 with respect to transmission range 80m, 100m, 150m, 250m

Node	Transmission Range			
Density	80	100	150	250
$0.28x10^{-4}$	38.72	53.30	80.72	81.70
$0.4x10^{-4}$	38.72	53.30	80.72	81.70
$0.8x10^{-4}$	39.70	54.21	81.56	81.92
$1.2x10^{-4}$	39.70	54.21	81.56	81.92
$1.6x10^{-4}$	40.64	55.72	82.60	83.0
$2x10^{-4}$	40.64	55.74	82.60	83.12
$2.4x10^{-4}$	42.45	56.81	83.32	83.54
$2.8x10^{-4}$	42.64	57.20	83.67	83.71
$3.2x10^{-4}$	43.40	64.45	84.20	84.50
$3.6x10^{-4}$	43.56	69.20	84.67	85.0
$4x10^{-4}$	44.26	79.80	85.10	87.34

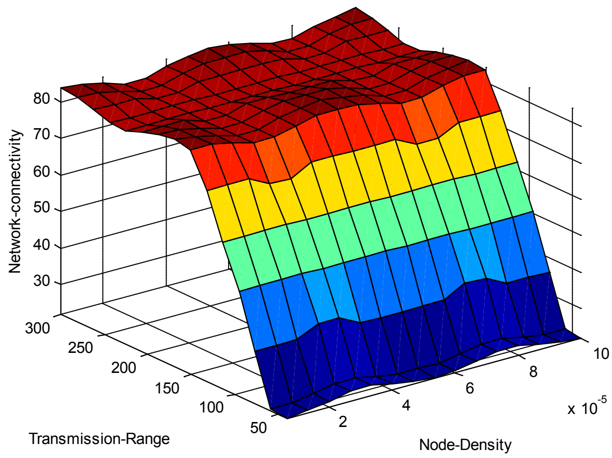


Fig7. Illustration of fuzzy simulation for terrain of size 1000m x 1000m

Fig7 provide the information about the network connectivity which is summaries in tables III with respected to transmission range 80m, 100m, 150m and 250m for terrain of size 1000m x 1000m.

Table III. Analysis of Fig7 with respect to transmission range 80m, 100m, 150m, 250m

Node	Transmission Range			
Density	80	100	150	250
$0.7x10^{-5}$	37.72	51.30	78.72	80.70
$1x10^{-5}$	37.78	51.32	78.74	80.72
$2x10^{-5}$	38.30	52.21	79.56	81.50
$3x10^{-5}$	38.34	52.28	79.65	81.62
$4x10^{-5}$	39.24	53.40	80.27	81.85
$5x10^{-5}$	39.30	53.48	80.35	82.12
$6x10^{-5}$	40.41	54.32	81.30	82.54
$7x10^{-5}$	40.53	54.41	81.57	82.71
$8x10^{-5}$	58.20	65.24	82.20	83.50
$9x10^{-5}$	62.56	67.20	83.67	84.54
$10x10^{-5}$	63.26	74.50	84.10	86.34

ISBN: 978-1-61804-230-9

For example in order to position 100 nodes or more nodes having terrain of 500 x 500 m^2 and 1000x1000 m^2 , we need the transmission range of all nodes to 110 or higher then that in order to make our network surely connected From our fuzzy simulation we can also predict the vice versa i.e. if we know that our nodes are capable of transmitting the range of 100 or less then that, then we need at least 800 nodes or more to make the network surely connected. We will validate these results using NS2

V. Network Simulation

In order to verify our approach to measure the level of network connectivity based on the parameters node density and transmission range, we use NS-2.34[12]. We generate the network scenario with the help of following simulation parameters:

Table IV: Simulations Parameters

Simulation Parameter	Value
Routing Protocol	AODV
MAC Layer	IEEE802.11
No. of Nodes	7
Packet Size	512B
Initial Energy	1.5W
Rx Power Consumption	0.1W
Tx Power Consumption	0.1W
Terrain Size	500m*500m,1000m*1000m
Transmission Range	250m , 150m ,100m
Simulation Time	200sec
Traffic Source	TCP

We simulate the network connectivity by varying the transmission range for number of nodes for two different terrain of size 500m*500m and 1000m*1000m.

*For 500m*500m area and transmission range 250m , 150m , 100m and number of nodes 7 , we measure the following network parametres :*

Table V Performance parameter for transmission range in 500m*500m

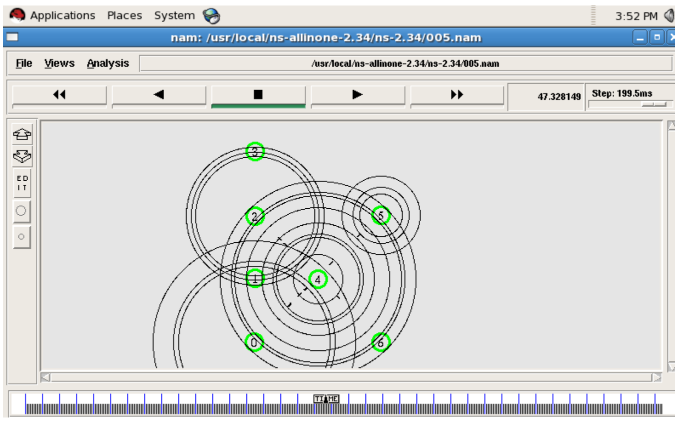
Network Parameters	Trans mission Range		
	250m	150m	100m
Generated Packet	8493	2913	25
Received Packet	8385	2901	0
Forwarded Packet	3457	943	0
Packet Delivery Ratio	98.72	99.58	0
Total Packets Dropped	60	17	0
Average end-to-end delay	80.8323ms	248.839ms	0

*For 1000m*1000m area and transmission range 250m, 150m, 100m and number of nodes 7, we measure the following network parameters*

Table VI. Performance parameter for transmission range in 1000m*1000m

Network Parameters	Trans mission Range		
	250m	150m	100m
Generated Packet	4436	1454	20
Received Packet	4320	1403	0
Forwarded Packet	1477	338	0
Packet Delivery Ratio	97.38	96.49	0
Total Packets Dropped	43	12	0
Average end-to-end delay	76.9442ms	248.62ms	0

From Table V and Table VI, it can be observe that with the transmission range 250m and 150m, there are fewer packets generated and average end to end delay increases while packet delivery ratio drops by 0.8%. This result can be evident from Table II and Table III which shows strong network connectivity for the 7 no. of nodes in area of size 500m x 500m and 1000m x 1000m with respect to transmission range 150m and 250m. For 150m the, network connectivity is 80.72% for terrain of 500m x 500m and 78.72% for terrain of 1000m x 1000m. Also, for 250m, the network connectivity is 81.70% for terrain of 500m x 500m and 80.70% for terrain of 1000m x 1000m. We can also visualize the strong network connectivity with the help of ns simulator, as shown in Fig8.



ISBN: 978-1-61804-230-9

Fig 8. Illustration of Strong Network Connectivity in terrain of size 500m*500m using 7 nodes with respect to transmission range 250m.

For Transmission range 100m, Although Packets are generated in both Table V and Table VI but are not forwarded to the desired destination .This again can be evident from table II and table III which shows network disconnectivity for the 7 no. of nodes in area of size 500m x 500m and 1000m x 1000m with respect to transmission range 100m. The network connectivity for terrain of size 500m x 500m is 53.30% and for 1000m x 1000m it is 51.30%. We can visualize this poor network connectivity with the help of ns simulator as shown in Fig9 and Fig 10. In Fig9 packets are generated but do not forwarded to the desired destination. Also, after some interval the network become completely isolated as shown in Fig10.

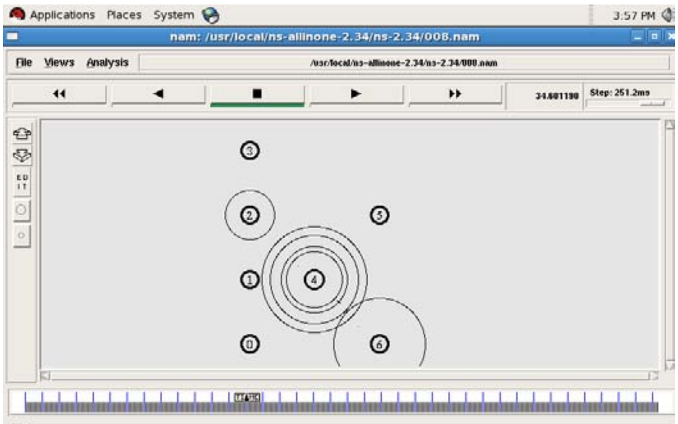


Fig9: Illustration of Network Disconnectivity in terrain of size 500m x 500m using 7 nodes with respect to transmission range 100m

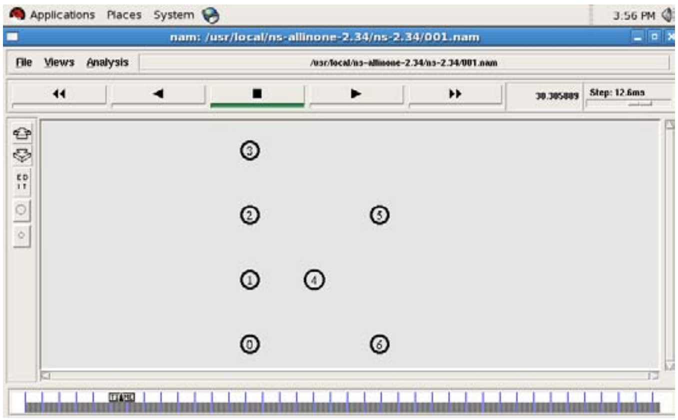


Fig10. Network become isolated

VI. Conclusion

In this paper we have proposed fuzzy logic based solution for the node connectivity problem in moble ad hoc network. For the given node density ,we can decide the transmission range to ensure the network connectivity and vice versa . We have validated the results using NS2 simulator . This study can be helpful in designing the mobile ad hoc network.

References

[1] S.S Basu and A.Chaudhari, "Self Adaptive Topology Management for Mobile Ad Hoc Network," IE(I) Journal-ET vol 84,July 2003

[2] David B.Johnson and David A.Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks,", Computer Science Department , Carneige Mellon University,5000 Forbes Avenue, Pittsburg,PA 15,213-3891.

[3] J.Jublin and J.Tornow,"The DARPA Packet Radio Network Protocols," Proc ,IEEE, vol.75, no.1,1987,pp.21-32.

[4] B.M. Leiner, D.L.Nielson, and F.A.Tobagi,eds.,Proceedings of the IEEE, Special issue on Packet Radio Networks,vol.75, Jan1987

[5] Z.J.Haas,M.Gerla,D.B.Johnson, C.E.Perkins,M.B. Pursley,M.Steenstrup, and C.-K. Toh,eds.,IEEE Journal on Sel. Areas in Comm,.Special issue on Wireless Ad Hoc Networks, vol17, August 1999.

[6] P.Kermani and N.H. Vaidya, eds.,IEEE Personal Communication, Special issue on Advances in Mobile Ad Hoc Networking, Feb 2001.

[7] C.E.Perkins,ed.,Ad Hoc Networking.Addison-Wesley,2001.

[8] L.A.Zadeh,"Fuzzy Sets",Inform. Contr., vol.8,no-3,pp 338-353,June 1965.

[9] L.X.Wang and J.M.Mendel."Generating Fuzzy Rules from Numerical Data With Applications," USC-SIPI Technique Report #169 ,1991

[10] C.C.Lee,"Fuzzy Logic in Control Systems: Fuzzy Logic Controller-Part I,"IEEE Trans. Sys.Man Cybern,vol .SMC-20,no.2,pp 408-418,Mar./April 1990.

[11] C.C.Lee,"Fuzzy Logic in Control Systems: Fuzzy Logic Controller-Part II,"IEEE Trans. Sys.Man Cybern,vol .SMC-20,no.2,pp 419-435,Mar./April 1990

[12] Network Simulator,NS-2,Homepage, http://www.isi.edu/nsnam/ns

[13] Cheng ,Y.-C.and Robertazzi,T.G. ," Critical connectivity phenomena in multihop radio models",IEEE Trans.Communication,Vol 37,1989,pp 770-777

[14] Philips,T.K.,Panwar,S.S and Tantawi,A.N. ,"Connectivity properties of a packet radio network model", IEEE Trans.Inform. Theory,Vol 35, 1989,pp 1044-1047

[15] Piret,P.," On the connectivity of radio networks", IEEE Trans.Inform.Theory,vol 37,1991,pp 1490-1492

[16] Gupta,P. and Kumar, P.R,"Critical power for asymptotic connectivity in wireless networks", In McEneany ,W.M., Optimization and Applications,1998,pp.547-566.Birkhauser

[17] R.Meesterand,R.Roy,Continuum Percolation.Cambridge,Mass.:Cambridge University Press,1996

[18] Dousse,O.,Thiran,P.and Hasler,M."Connectivity in ad hoc and hybrid networks,"In Proc.IEEEInfocom,New York,USA,June 2002, pp.1079-1088

[19] F.Xue, and P.R.Kumar,"The number of neighbors needed for connectivityof wireless networks", Wireless Networks,Vol10(2), 2004, pp.161-181

[20] S.Khuller,"Approximation algorithms for finding highly connected subgraphs",In D.S. Hochbaum, editor,Approximation algorithms for NP hard problems.PWS Publishing Co.,1996

[21] G.Ausiello,and P.Crescenzi, and G.Gambosi, and V.Kann, A.Marchetti-Spaccamela,andM.Protasi,"Complexityandapproximation:Combinatorial optimization problems and their approximability properties",Springer-Verlag,Berlin,1999.

[22] Chlamtac,I.andFargo,A."A new approach to the design and analysis of peer to peermobilenetworks", ACM/Kluwer Wireless Networks,vol5,1999,pp.149-156

[23] Santi,P.,Blough,D.M. and Vainstein,F."A probabilistic analysis for the radio range assignmnet problem in ad hoc networks",In proc ACM mobiHoc,Long Beach,USA,October 2001,pp.212-220.

[24] M.J.B.Appel and R.P.Russo,"The minimum vertex degree of a graph on uniform points in $[0,1]^{d}$", Adv in Applied Probability,vol29,1997,pp.582-594

[25] Krishnamachari,B.,Wicker,S.B.andBejar,R.,"Phase transition phenomenon in wireless ad hoc networks",In proc.IEEE Globecom,San Antonio,USA,November 2001,pp.2921-2925

[26] Bettstetter,C.,"On the minimum node degree and connectivity of wireless multihop network",In Proc .ACM MobiHoc,June 2002,pp.80-91.

[27] Bettsetter,C.,"On the connectivty of wireless multihop networks with homogeneous and inhomogeneous range assignmnet",In Proc.IEEE VTC,Vancouver,Canada,September 2002,pp.1706-1710.

[28] Bettstetter,C. and Zangl,J.,"How to achieve a connected ad hoc network with homogeneous range assignmnet:an analytical study with considerationof border effects",In Proc.IEEE Conf.on Mobile and WirelessCommunicationNetworks(MWCN),Stockholm,Sweden,September 2002,pp.125-129

[29] Bettstetter,C.,"On the connectivity of ad hoc networks,"The Computer journal,Special issue on mobile and pervasive computing,47(4):432-447,July 2004.

[30] Bettstetter,C,"Topology properties of ad hoc networks with random waypoint mobility",In Proc.ACM MobiHoc,Annapolis,USA,June 2003,Short Paper

[31] Q.Ling and Z.Tian,"Minimum node degree and k-connectivity of a wireless multihop network in bounded area,"Proc .of IEEE Globecom,2007.

[32] H.Zang and J.C.Hou,"On the critical total power for asymptotic k-connectivity in wireless networks,"IEEE/ACM Transactions on Networking,April 2008.

[33] Santi,P.and Blough,D.M,"The critical transmitting range for connectivity in sparse wireless ad hoc networks",IEEE Trans.Mobile Computing,vol2,2003,pp.25-39.

[34] Desai,M.and Manjunath,D."On the connectivity in finite ad hoc networks",In Proc IEEE Commun.Lett.,vol6,2002,pp.437-439

[35] Nakano,K.,Shirai,Y.,Sengoku,M.and Shinoda,S."On connectivity and mobility in mobile multi hop wireless networks",In Proc.IEEE VTC,Jeju,Korea,April 2003,pp.89-98

[36] P.J.Wan and C.W.Yi,"Asymptotic critical transmission radius and critical neighbor number for k-connectivityin wireless ad hoc networks",5th ACM Symposium on Mobile Ad Hoc Networking and Computing(MobiHoc),2004.

[37] P.Panchapakesan and D.Manjunath,"On the transmission range in dense ad hoc radio networks",Proc .IEEE SPCOM'01,2001.

[38] M.D.Penrose,"A strong law for the longest edge of the minimal spanning tree,"The Annals of Probability,vol27(1),1999,pp.246-260

[39] M.D.Penrose,"The longest edge of the random minimal spanning tree,"The Annals of Probability,vol15(2),1999,pp.340-361

[40] R.Ramanathan, and R.Rosales-Hain,"Topology Control of Multihop Wireless Networks using Transmit Power Adjustment",Proc IEEE Infocom 2000,pp.404-413

[41] V.Rodolpu, and T.H.Meng,"Minimum energy mobile wireless networks",IEEE J.Selected Areas in Comm.,vol17(8),August 1999,pp.1333-1344.

[42] L.Li,J.H.Halpern,P.Bahl,Y.Wang,andR.Wattenhofer,"Analysis of a cone-based distributed topology control algorithm for wireless multi hop networks".to appear in Proc.ACM Symp. On principles of Distributed Computing(PODC),August2001.

[43] R.Diestel,Graph theory,Springer,2nd ed..2000.

[44] B.Bollobas,Modern Graph Theory,Springer,1998.

[45] W.Feller,An Introduction to Probability theory and its Applications,John Wiley&Sons,New York,1950.

[46] S.Janson,T.Luczak,andA.Rucinski,RandomGraphs,JohnWiley&Sons,New York,2000.

[47] N.A.C.Cressie,Statistics for Spatial Data,John Wiley&Sons,1991.

[48] J.Diaz,M.D.Penrose,J.Petit,and M.Serna,"Convergence Theorems for SomeLayout Measures on Random Lattice and Random Geometric Graphs,"Combinatorics,Probability,andComputing,no.6,2000,pp.489-511.

[49] Maths Works ,Fuzzy Logic ToolBoox User's Guide,Jan,1998

Financial Economic Aspects of Precious Metals

A. Ioana, A. Semenescu, C.F. Preda

Abstract— This article presents specific aspects of financial-economic and management models for precious metals. Thus, we present the main types of financial-economic analysis: financial-economic statistical analysis, financial-economic dynamic analysis, financial-economic static analysis, financial-economic mathematical analysis, financial-economic psychological analysis. Also we present the main object of the financial-economic analysis: the financial-economic technological activity analysis of a company, the financial-economic analysis of the production costs, the financial-economic analysis of equipment, the financial-economic analysis of labor productivity, the financial-economic analysis of the goods flow. Also, this paper presents an algorithmic analysis of the marketing mix for precious metals. It also analyzes the main correlations and their optimizing possibilities through an efficient management. Thus, both the effect and the importance of the marketing mix, for components (the four “P-s”) are analyzed for precious metals, but their correlations as well, with the goal to optimize the specific management. There are briefly presented the main correlations between the four marketing mix components. We also present and analyze in our article new concepts such as: Level of Precious Metals Product Completion (LPMPC), Precious Metals Quality Control Activity (PMQCA), Precious Metals Cost Control Activity (PMCCA), Precious Metals Profit Planning (PMPP).

Keywords— Financial - Economic Analysis, Management Models, Precious Metal Product.

I. Introduction

According (***, 2002) finance notion has three meanings, namely:

- The practice of manipulating and managing money.
- The capital involved in a project, especially the capital that has to be raised to start a new business.
- A loan of money for a particular purpose, especially by a finance house.

A Finance Company (FC) is a company that provides finance, normally in the form of loans. As it tends to finance ventures with a high risk factor, the cost of borrowing is likely to be higher than that made by a clearing bank.

In Figure 1 we present the scheme of a Finance Company (FC) and correlations with risk factors.

Not coincidentally it is said that the bank gives you an umbrella when is a good weather (when you can have arisen only from the sun) that asks you back when is bad weather ("Financial Storm").

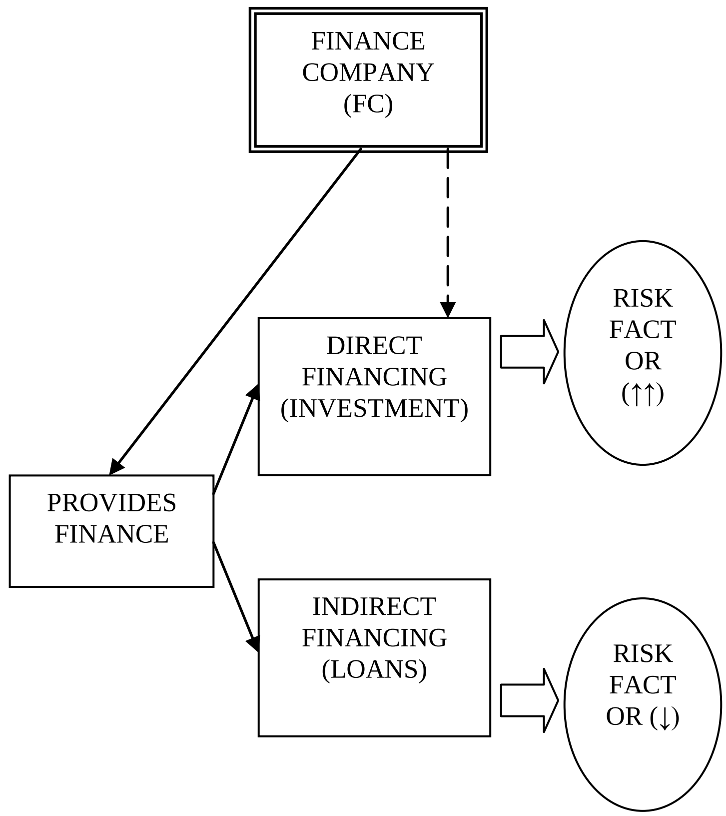


Fig. 1. Scheme of a Finance Company (FC) and correlations with risks factors

The economic analysis is a research method, based on decomposing and dismantling an object or a phenomenon in its components or its basic elements (Ioana, 2009, 2007).

The etymology of the term "economic analysis" is Greek word "analisi" (Ioana et al, 2013,a, *** 2002). The term “analysis” is from Medieval Latin “analysis”, from Ancient Greek: “ἀνάλυσις” (analusis), ἀναλύω (analuō, “I unravel, investigate”), ἀνά (ana, “on, up”) + λύω (luō, “I loosen”).

The economic analysis examines the activities or phenomena from the economic point of view. The essential issue when performing economic analysis is that it observes the structural relationships, including functional relationships and the cause and effect relationships (Kotler, 2003, Patriche, 1994).

The economic analysis may be:

- Financial analysis
- Psychological analysis
- Diagnostic analysis
- Statistical analysis
- Mathematical analysis
- Dynamic analysis

The object of the financial analysis may be:

- The financial - economic activity analysis of a company
- The financial - economic analysis of equipment
- The financial - economic analysis of the production costs
- The financial - economic analysis of labor productivity
- The financial - economic analysis of the goods flow
- The technological activity analysis of a company

The financial - economic synthesis is an important component of the financial - economic analysis, which recomposes the parts or elements of an object or phenomenon as a whole.

Whilst the analysis process involves the decomposition of a phenomenon, the synthesis process examines the elements of an object, studies together as a whole.

II. METHODOLOGY

Adoption of innovation (Ioana, 2004, Ioana, 1998, Scarlat, 2003) has been suggested that those consumers who eventually accept an innovation can fall into five groups shown in figure 2 - Innovation Groups Adoption (IGA).

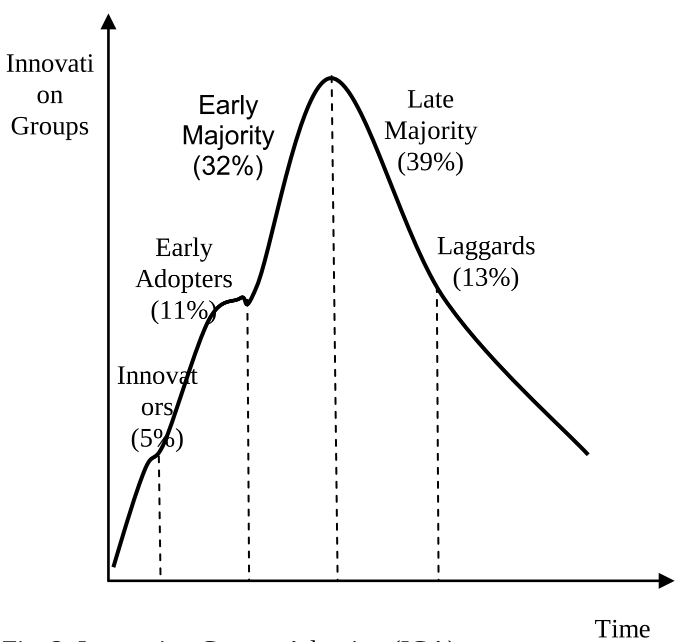


Fig. 2. Innovation Groups Adoption (IGA)

The consumers who eventually accept an innovation can fall into the five groups (figure 2). The main characteristics of the five groups of consumers are:

- ➢ Innovators – the first to buy and use new products.

• The innovators are clearly critical to the process of adoption, including financial – economic adoption.
• They are likely to communicate with and persuade others to try the product, including financial – economic product.
• They put the innovation on show to create the image of being venturesome.
• Without their support an innovation is unlikely to be successful.
• This group is defined as the first 5% to adopt the new product.

- ➢ Early adopters – tend to the opinion leaders and to adopt new ideas early but carefully.

• This group is defined as the next 11% of the adopting consumers.

- ➢ Early majority – people regards as being deliberate in their decisions, who are rarely leaders.

• These form the next 32% of the adopting consumers.

- ➢ Late majority – sceptics who only adopt an innovation after most other people have tried it.

• These form the next 39% of the adopting consumers.

- ➢ Laggards – the most tradition-bound.

• Laggards are suspicious of changes and innovation.
• They tend only to adopt the innovation when it has become widely accepted.
• Laggards are about 13%.

Figure 3 presents a model of financial - economic analysis.

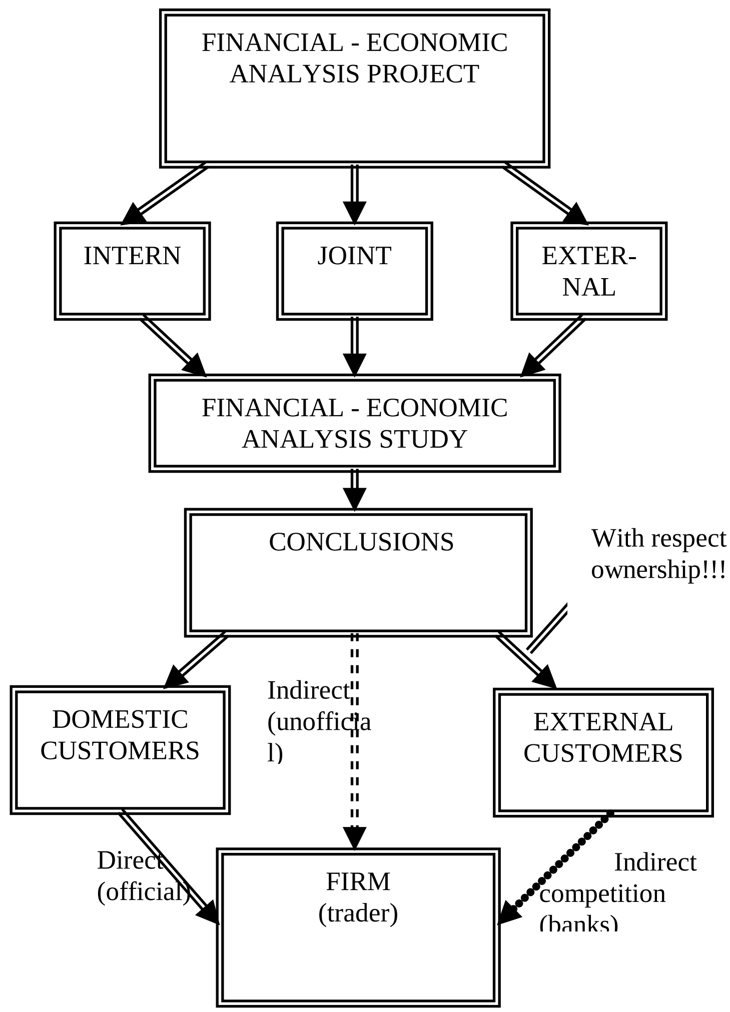


Fig. 3. Model of financial - economic analysis

The model of financial - economic analysis presented has as its starting point the 3 categories of performers (Ioana and Semenescu, 2013, Ioana et al, 2013b):

- ➢ Intern performers

ISBN: 978-1-61804-230-9

- Joint performers
- External performers

Information (including conclusions) of economic analysis conducted can be passed to the three distinct pathways:

- Domestic Customers → Direct (Official) → Company (trader)
- Indirect (unofficial) → Company (trader)
- External Customers (with respect ownership) → Indirect Competition (banks) → Company (trader)

The four components of the marketing mix (the four "P-s"):

- Product (P1);
- Price (P2);
- Promoting (P3);
- Placement-Distribution (P4).

and their correlation is very important for an efficient management in materials' industry.

The analysis of the correlations between the 4 "P-s" (the four components of the marketing mix) and their management in metallurgy are also very important.

III. The Main Correlations Between Marketing Mix Components For Precious Metals

In figure no. 4 there are briefly presented the main correlations between the 4 marketing mix components (the 4 "P-s") for a product within the materials' industry, including precious metals industry.

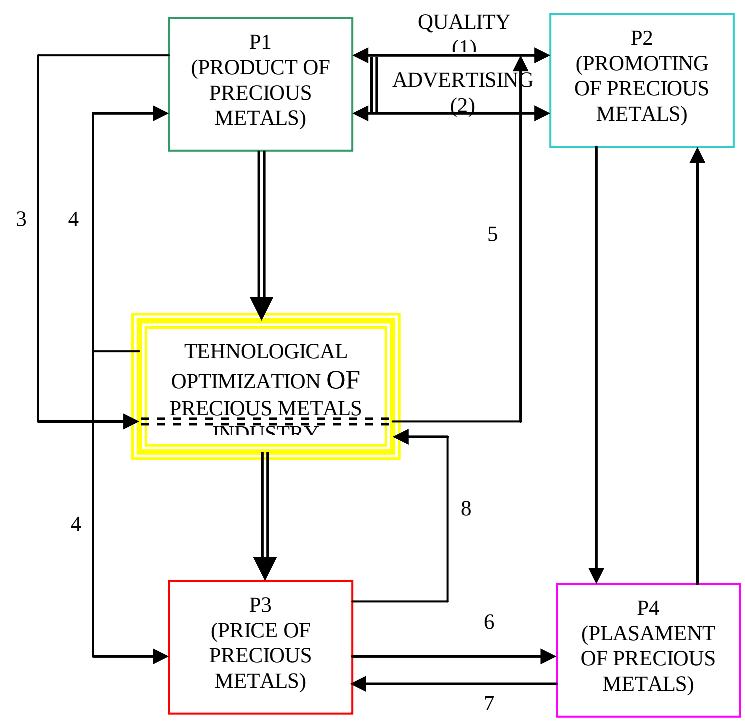


Fig. 4. The main correlations between the four marketing mix components for precious metals product

Aspects regarding management:

(1) – The biunivocal correlation Product (P1) - Promoting (P2) is based on assuring the quality of the product.

(2) – An important role in optimizing the correlation (1) is held by advertising directly correlated with the product' s quality level.

(3) – The Product (P1) needs and determines technological development for assuring the quality technical requirements.

(4) – The biunivocal correlation Product (P1) – and Price (P3) is based on cutting of production costs.

(5) – The level of technological optimization is directly correlated with the product's quality.

(6) – A good (low) price of the product assures good placement condition of it.

(7) – A good placement of the product can lead to a good price (optimal in direct correlation with the sales level).

The lower the price (P3) is (which is facilitated by a high level of technological optimization), the higher the profit (benefit) is, which allows investing it in research-development.

The analysis of the main correlation between the 4 marketing mix components in a case of a product from the materials' industry (presented in figure 3) highlights the importance of management in order to optimize that product.

Figure 5 presents the main correlations between the functional and constructive betterments regarding a product of precious metals.

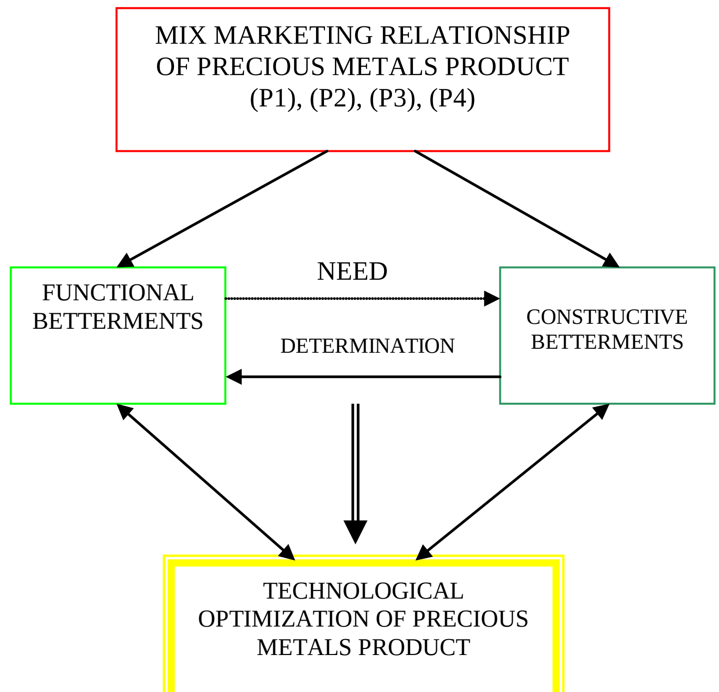


Fig. 5. The main correlation between the functional and constructive betterments in the technology of precious metals product

It is to be noticed that in order to obtain an optimal marketing mix for a product from the materials' industry, the technological optimization management must focus on both functional and constructive betterments.

The functional betterments need constructive betterments and constructive betterments generate functional betterments.

An important component of the marketing mix for precious metals product is the quality and the cost control activity.

The main steps of the quality and cost control activity management for precious metals product are:

- For Level of Precious Metals Product Completion (LPMPC), with the following components:
- Precious Metals Product Planning (PMPP)
- Precious Metals Product Conceiving (PMPC)
- Precious Metals Preparing Production (PMPP)
- Precious Metals Production (PMP)
- Precious Metals Marketing (PMM)
- For Precious Metals Quality Control Activity (PMQCA), with the following components:
- Planning the Overall Precious Metals Product Quality (POPMPQ)
- Planning the Actual Precious Metals Product Quality (PAPMPQ)
- Conceiving the Actual Precious Metals Product Quality (CAPMPQ)
- Prescribing the Precious Metals Quality Control Elements (PPMQCE)
- QC Process Control (QCPC)
- Researches Regarding Quality in the Market (RRQM)
- For Precious Metals Cost Control Activity (PMCCA), with the following components:
- Precious Metals Profit Planning (PMPP)
- Precious Metals Planning Raw Materials (PMPRM)
- Precious Metals Cost Cut of Program (PMCCP)
- Precious Metals Raw Materials Cost Evaluation (PMRMCE)
- Precious Metals Real Cost Evaluated of Cost Components (PMRCECC)
- Activity of Raw Materials Cost Cutting (ARMCC)
- Present Situation Evaluation and Analysis of the Disparity (PSEAD)

The level of the product's completion is to be noticed. This based on the following activities: planning the product, conceiving the product, preparing production and marketing.

Another management and economic analysis model for precious metals product is Turnover Analysis (TA) in conjunction with the Production Capacity (PC) and demand (D). Figure 6 shows the schematic of this model.

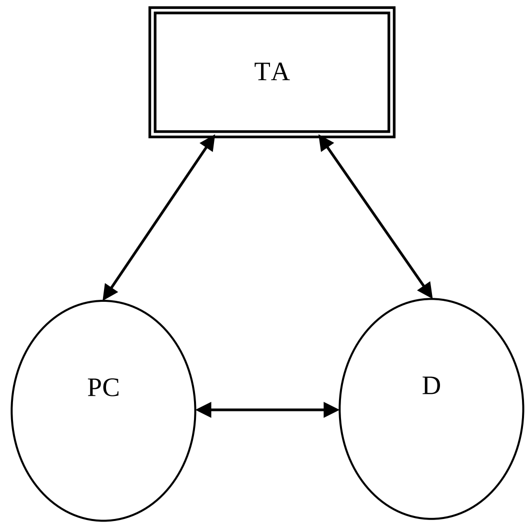


Fig. 6. Model of Turnover Analysis (TA) in conjunction with the Production Capacity (PC) and demand (D)

For turnover analysis (T) in conjunction with the production capacity (PC) and demand (D) I define the following indicators:

- Utilization of Production Capacity (UPC) = {[Production Made (PM)] / [Production Capacity (PC)]} x 100 [%]
- Meet Demand (MD) = {[Turnover from Core Business (TCB)] / [Demand (D)]} x 100 [%]
- Relationship between Turnover and Production Made (T/PM) = Turnover from Core Business (TCB) / Production Made (PM).

IV. Conclusion

Financial - Economic Analysis (FEA) is very important for business activity efficiency, including for precious metals product.

The correct choice of the object of financial - economic analysis (the technological activity analysis of a company, the analysis of equipment, the analysis of the production costs, the analysis of labor productivity, the analysis of the goods flow etc) is also very important.

The synthesis is an important component of the financial - economic analysis.

Analysis of five consumer groups (Innovators, Early Adopters, Early Majority, Late Majority, Laggards) highlighted:

- Innovators put the innovation on show to create the image of being venturesome. They are likely to communicate with and persuade others to try the product, including for precious metals product. The innovators are clearly critical to the process of adoption.
- Early adopters – tend to the opinion leaders and to adopt new ideas early but carefully.
- Early majority – people regards as being deliberate in their decision, who are rarely leaders.
- Late majority – sceptics who only adopt an innovation after most other people have tried it.

ISBN: 978-1-61804-230-9

- Laggards are suspicious of changes and innovation. They tend only to adopt the innovation when it has become widely accepted.

Model of financial - economic analysis presented has as its starting point the three categories of performers: intern performers, joint performers, external performers.

The marketing mix analysis for precious metals product highlights the importance of the technological optimization in order to obtain an optimum in the field.

The technological optimization in the precious metals product is based on functional and constructive betterments. The optimization of the biunivocal correlation between them (need-determination) assures the efficiency of the marketing mix of precious metals product.

The main correlations between the four marketing mix components for precious metals product mainly highlights: an important role in optimizing the correlation Product - Promoting is held by advertising directly correlated with the product' s quality level; the level of technological optimization is directly correlated with the product's quality; the precious metals product needs and determines technological development for assuring the quality technical requirements; a good placement of precious metals product can lead to a good price (optimal in direct correlation with the sales level).

The quality and the cost control activity are very important components of the marketing mix for precious metals product.

We have presented and analyzed new concepts such as: Level of Precious Metals Product Completion (LPPMC), Precious Metals Quality Control Activity (PMQCA), Precious Metals Cost Control Activity (PMCCA).

An important management and economic analysis model is Turnover Analysis (TA) in conjunction with the production capacity (PC) and demand (D).

REFERENCES

[1] Ioana, A. (2009). Managementul activității financiar-contabile și analize economice. Teorie și Aplicații, Editura Politehnica Press, București.

[2] Ioana, A. (2007). Managementul producţiei în industria materialelor metalice. Teorie şi Aplicaţii., Editura PRINTECH Bucureşti.

[3] Ioana, A. (2004). Plan de afaceri pentru înfiinţarea şi dezvoltarea unui Centru de Cercetare, Perfecţionare şi Consultanţă în domeniul Materialelor Speciale şi al Ecometalurgiei "CCPC MAT-ECO", Universitatea "Politehnica" din Bucureşti, Şcoala de Studii Academice Postuniversitare de Management, Bucureşti.

[4] Ioana, A. (1998). Optimizarea performanţelor funcţionale şi tehnologice ale cuptorului cu arc electric cu preîncălzirea încărcăturii şi insuflare de pulberi pentru elaborarea oţelurilor de înaltă calitate, Teză de Doctorat, Universitatea "Politehnica" din Bucureşti.

[5] Ioana, A., Semenescu, A. (2013). Technological, Economic, and Environmental Optimization of Aluminum Recycling, Journal of the Minerals, Metals & Materials Society, JOM: Volume 65, Issue 8, ISSN 1047-4838 (ISI-Web of Science/Science Citation Index Expanded).

[6] Ioana, A., Semenescu, A., Marcu, D., Ghiban, A., Alina Nicoleta Colan, (2013). Managementul Calită□ii. Teorie şi Aplica□ii. Editura Matrix Rom, Bucureşti, 2013.

[7] Ioana, A., Semenescu, A., Preda, C.F. (2013). Knowledge Management Innovation for Sustainable Development in the Context of the Economic Crisis, ISI Proceedings of the 2013 International Conference on Environment, Energy, Ecosystems and Development (EEEAD 2013), ISBN: 978-1-61804-211-8, pp. 21-26, Venice, Italy.

[8] Kotler, Ph. (2003). Managementul marketingului, Editura Teora, Bucureşti.

[9] Nicolescu, O. (2000). Sisteme metode şi tehnici manageriale ale organizaţiei, Editura Economică, Bucureşti.

[10] Patriche, D. (1994). Marketing industrial, Editura Expert, Bucureşti.

[11] Scarlat, C. (2003). Antreprenoriat şi managementul întreprinderilor mici si mijlocii, Editura Printech, Bucureşti.

[12] *** (2002). A Dictionary of Business, Oxford University Press, New York.

Adrian Ioana, birth at Curtea de Arges, Romania, 10.23.1958. Educational background is: 2004, University "Politehnica" of Bucharest, Romania, School of Postgraduate Academic Studies in Management, 1990 - 1998: Material and Equipment for Metallurgical plant Department, Science and Engineering Materials Faculty, University "Politehnica" of Bucharest, Romania, title: doctor engineer (1998), title of the doctoral degree paper: "The electric arc furnaces (EAF) functional and technological performances with the preheating of the load and powder blowing optimization for the high quality steel processing", 1979 - 1984: Polytechnic Institute Bucharest, Romania, Metallurgical Faculty, training Extraction Metallurgy.

He has an extensive work experience: **1998 – present**: University "POLITEHNICA" of Bucharest, 313 Spl. Independenţei Romania, LECTURER DOCTOR ENGINEER, "Material and Equipment for Metallurgical Plant" Department, and "Materials' Making and Eco-Metallurgy" Department, Engineering and Management of Metallic Materials' Making" Department. He has published: 16 speciality books (eq. "Production Management of Metallic Materials Industry. Theory and Applications", PRINTECH Publishing Bucharest, ISBN 978-973-758-1232, 232 pg., 2007, "Precious Metallic Materials' Engineering in the Financial – Banking Circuit", Fair Partners Publishing, Bucharest, 2002, ISBN Code 973-99 985-5-0; "Optimal Managing of Electric Arc Furnaces", Fair Partners Publishing, Bucharest, 2002, ISBN Code 973-8470-04-8), 21 ISI Articles, 32 BDI Articles, over 100 Proceedings Conference Articles. He has published 7 patents (and two patent applications pending assessment) that he has obtained (all designed and implemented in industry) are relevant in this way. **1997-1998**: GRANTMETAL S.A. Bucharest. He has worked as an ENGINEER for the Technical Department, solving burner, furnace and energy saving matters. **1987 - 1996**: METALLURGICAL RESEARCH INSTITUTE, BUCHAREST (ICEM) and POLYTECHNICAL INSTITUTE BUCHAREST. He has working as a MAIN SCIENTIFIC RESEARCHER 2-ND DEGREE within the Laboratory for Economy of Fuel and Energy, developing an activity of design, planning and drawing up of "know-how" for burner and recovery plants of secondary energy resources. **1987 – 1995**: ASSOCIATED PROFESSOR, Material and Equipment for Metallurgical plant Department, Science and Engineering Materials Faculty, University "Politehnica" Bucharest. **1986**: POLYTECHNICAL INSTITUTE BUCHAREST, METALLURGICAL FACULTY. He worked as an ASSISTANT-ENGINEER in the field of "Technological metallurgical outfits and equipment", giving lectures concerning theory and practice (laboratory) on "Metallurgical Equipment" and "Metallurgical Automatizations". **1984 - 1986**: THE SPECIAL STEEL COMPLEX Târgovişte (C.O.S.T). He worked as a METALLURGICAL ENGINEER in the "Electrical Steel Plant" Department, in the sectors: Preparation-burdening of melting stock, steel making (electric arc furnace), casting, primary heat treatment, conveyance-deliver production.

Prof. Ioana was Secretary of "Authorization and Homologation of Burning Installations Board", in the "Metallurgical Research Institute (ICEM)", "Combustion and Energy Economy Laboratory (LECE)". Prof. Ioana is a member of the " Romanian Engineers' General Association" (AGIR) and a honor member of the "Romanian Metallurgy Society" (SRM), WSEAS member, Research Cooperative membre, CrossRef member. Prof. Ioana is an authorized 2nd degree specialist of the "Natural Distribution of Gases" (ROMGAZ), in the fields: designing, making and exploitation of the using installation, exploitation and making of all works from the natural gases distribution system (AUT. No. 867R/1997). The main awards and committees and publications are: **2002**: The Diploma and Gold Medal at "The 51th World Exhibition of Innovation, Research and New Technology-EUREKA 2002", Brussels, Belgium, with the invention "Lance for Post-Combustion".

1999 : “Henry ford European Conservation Awards – Recognising the Outstanding Contribution to the Conservation of the Natural Environment and Cultural Heritage”, for the paper: “Industrial Burning Systems, Ecologically and Energetically Advanced”. **1998**: Second Prize at The International Symposium “Traditions and Perspectives in The Romanian Metallurgical School”, with the paper “Characteristic Features in The Fiability Metallurgical Equipments Determination”. **1996**: “GREAT AWARD” of The Romanian Company of Metallurgy (SRM) with the paper: “Economical, ecological and lasting heating systems of metals and non-metallic materials all over the world using the computer and the regenerative burners”.
1989: “First Prize” of “METALLURGICAL RESEARCH INSTITUTE (ICEM)’ Symposium”, with the paper: “Furnaces that Work with Residual Combustible Gases”.

Paradigm changes in Project Management and their impact on knowledge of project managers

B. Lacko, M. Polčáková, K. Hrazdilová Bočková

Abstract—The contribution describes 8 paradigms employed by mankind for project design and realization. The accent that dominated at each time and affected all other aspects was used for the paradigm characteristics and also as a standard for paradigm allocation. The goal, time-frame, costs, risks and quality of project realization were always important, but these factors had different priorities at different times. Project Management is developing dynamically just as the world in which Project Management is carried out. The article also presents the second aspect to these innovations the paradigm of project management: what these changes had impact on the required knowledge of project management. This paper was done within project denoted as KEGA - 003DTI-4/2014.

Keywords—competence, paradigm, project management

I. Introduction

MANKIND has been involved in the effort to manage extensive complex projects requiring the co-ordination of many people since time immemorial, but only after the year 1900 did the term Project Management come into wide use.

It is not without interest to learn how the paradigm of Project Management has changed in the course of mankind's development on our planet. By the term paradigm we understand a specific approach to a specific solution to a problem in which members of a certain expert community share the same set of opinions, approaches, values and techniques.

The following introduction outlines 8 paradigms employed by mankind for project design and realization. The accent that dominated at each time and affected all other aspects was used for the paradigm characteristics and also as a standard for paradigm allocation. The goal, time-frame, costs, risks and quality of project realization were always important, but these factors had different priorities at different times. Project Management is developing dynamically just as the world in which Project Management is carried out.

The article also presents the second aspect to these innovations the paradigm of project management: what these changes had impact on the required knowledge of project management.

II. Scarcity of Resources - the Beginning of Civilization

Man can use his intellect to manage himself or his community, while able to reach the target of his endeavor as well as the ways to achieve the chosen targets. These targets are very heterogeneous and they are connected to very developed trains of thought in the human brain. In this way, man differs significantly from animals which are only able to attain their biological targets (self-preservation, species preservation, etc.) with their instinctive reflexes.

It is certain that as early as prehistoric times, man began to consider in his own way how to design and manufacture things that would help him fly like the birds, be as powerful as nature's volcanic phenomena, control and exploit natural elements, multiply his abilities and senses, etc.

Man's projects, however, were strongly affected by the scarcity of all imaginable resources. Above all, he did not have enough essential skills. He had no helpful tools and machines at his disposal, thus mankind was not able to co-ordinate major projects as a result of communication problems, scarcity of required raw materials, etc. We could continue citing the reasons for a very long time. Therefore man had no other choice but to transform his unrealizable imagination into myths and legends, and continue hacking his Flintstone to the shape of a fist wedge or building his primitive hovels.

III. The goal - Antiquity and the Middle Ages

Two thousand years ago the Greek writer Antipater from Sidon made a list of the most impressive and beautiful manmade objects [1]. Since then this list has been known as The Seven Wonders of the Ancient World.

- The Pyramids of Egypt at Gíza
- The Lighthouse on Pharos Island of Alexandria Harbour
- The Hanging Gardens of Queen Semiramis of Babylon
- The Tomb of Mausolus at Halikarnas
- The Temple of Artemis at Ephesus
- The Status of Zeus at Olympia
- The Colossus of Rhodes at the entry to Lindos Harbour

Naturally he could not include the awesome Great Wall of China or imposing monuments made by the Incas, Aztecs,

ISBN: 978-1-61804-230-9

Mayas and others in the Americas, because at that time there were no existing contacts with the remote civilizations.

Of the above mentioned Seven Wonders, only The Pyramids of Egypt survive. We can however still admire the beautiful cathedrals and castles of Antiquity and the Middle Ages.

All of these magnificent monuments demonstrate that as early as ancient time's people were able to realize large projects, though they did not call them such. In Ancient China there existed a "Book of Changes" (I-ťing) that basically postulated (among other things) the successful Project Management of that time.

The goal was active power and everything was subordinated to it. The will of a certain community or the decision of a sovereign who was able to realize his ideas with enlightened wisdom or tyrannical power was concentrated on the goal. The fulfilment of the goal sometimes lasted for decades and the exact date of completion was not even planned. They took steps slowly and circumspectly. Nobody could estimate the required amount of money and in fact it was not necessary. As much money could be collected, was provided for construction to cover the building requirements. With other resources the story was similar, especially with the human ones. Today, nobody knows how much the pyramids cost in terms of human lives, for instance. The goal was an absolute priority and everybody had to contribute to its realization in accordance with their abilities, even if they could lose their lives in the process (they would certainly have lost their lives anyway if they had defied the goal or refused to participate on its realization). What if further resources were required? War would be waged, further slaves acquired, gold and other treasures brought from defeated countries, etc.

When we realize how primitive the means available for mankind at the time in question were, we have to admire their human will-power and perseverance.

IV. COSTS - THE MAGIC FORMULA OF MODERN CAPITALIST TIMES

Capitalism introduced a new force into the world - the power of money. Not, that money did not game a major role till then. Before capitalism thought, there were other factors more crucial, for instance, religion (Christian or other), the absolute power of an individual, etc., and money had to be subordinate to them. But now money ruled the world. The man who wanted to begin a project needed enough financial means and, in particular, had to promise that the realization of the project would bring in even more money upon completion.

At the very beginning the question "How much it will cost?" logically emerged. It was not enough to declare the goal. Whoever didn't have enough money had to table the idea no matter how good or ambitious it was. If in the course of the project realization a shortage of money occurred, it was evident that the project would be stopped. For that reason much attention was paid to the planning of the expenditure outlay from beginning to end. The same attention was paid to finance monitoring (unless it was a fraud from the very beginning). A great number book-keepers, bank clerks, confidential clerks, controllers and other workers took care of these financial problems (we may note that they worked very methodically and pedantically and often achieved better results than people do today using computers). The goal was reached relatively quickly if there was enough money and if not, the whole project dragged on. But how speed the process up without sufficient financial means? It was difficult to find an answer at the time, since everybody relied only on the power of money.

V. TIME - THE DETERMINANT OF THE MODERN ERA

The First World War and the Second World War, as well as so-called The Cold War, caused paradigm changed in project management. Wars, especially the aggressive ones, were usually waged by countries that had sufficient financial means. Otherwise they would not have been able to equip their armies with weapons in the required quantities. In reality it turned out that there was enough money for armament during the preparation for war as well as during war. Very often both adversaries managed to mobilize the required finances and other resources. Time became a critical factor in such cases. The ability to develop a new weapon more quickly, the ability to equip the army with weapons more quickly, the ability to mobilize more quickly, etc. - that was crucial for military superiority and ultimately for victory. Thus the pursuit of the shortest possible time frames dominated all else. During the First World War Gantt´s diagrams came into existence, when arose a need to shorten the process of the ship manufacturing necessary for the transportation of American soldiers to Europe. The network analysis methods came into existence thanks to the arms race during the Cold War and Iron Curtain days, when American industry was expected to manufacture the POLARIS rocket sooner than the Soviet Union could manufacture an intercontinental rocket. No resources or finances were spared in matters concerning national defense or war. The countries did not even mind the loss of the lives of their nationals and foreign citizens, nor the financial means necessary to surpass the adversary in time.

This paradigm was quickly transferred from the armaments industry field to economic competition. To succeed in business, the businessman had to introduce a new product in the market before his competitor, be first in cost reduction and thus gain a competitive advantage. The time aspect also showed up in normal national-economic overviews when it was learnt that not only the amount of capital mattered, but also its effectiveness and productivity in terms of time.

VI. THE RISKS OF ADVERSE EFFECTS - THE SECOND HALF OF THE TWENTIES

The second half of the twenties found mankind at the threshold of gigantic projects and project managers (a brand new profession now), were charged with realization of these projects. We may at least mention the most important projects: neutron bombs, space rockets, man's orbits round the Earth and the landing on the moon’s surface, nuclear power-stations, huge dams, the realization of supercomputers and supersonic airplanes.

Project management had its own methods (PERT, CPM, MPM) used at that time, its calculation being performed by means of computers, and it employed many skilled experts.

Certain projects were part of the world powers´ state politics and their aim was to acquire an international reputation. Considerable financial means were earmarked for these projects, top experts were engaged in them and they often were carried out within a relatively short time limit.

The reduction of the risks adverse effects for these projects became a fundamental issue. The deaths of many cosmonauts, collapses of dams, disasters in nuclear power-stations, plane crashes, and the negative impacts of many projects on the environment - all these warning signals attracted the attention of the public as well as experts to the question of how to analyze the possible risks of the negative impacts ahead of the realization of these projects. The major financial means and huge resources mobilized in order to realize these projects made a precise analysis of the threats involved in each particular project necessary. The project could have been jeopardized by many typical things such as the death of a project leader (for instance the situation in the Soviet Union’s space program after the death of S. Koroljov).

And what about the risk from domination of intelligent robots or supercomputers! [7]

Western businessmen had a sufficient quantity of financial means and did not lack for potential and promising projects. The problem was, however, choosing

a project that would promise the best success with reasonable risks, and whose design guaranteed the optimal risk management. Risk engineering had been used only in the insurance and bank field until then, but now it began to be used for solutions for technical equipment, and during project design and realization.

VII. The Phenomenon of Changes - the Nineties

In the nineties of the 20th century, projects were under the pressure of far more changes than ever before. This turbulent era - a time with a high frequency of big and small changes that often worked against each other - affected a whole range of projects. The sudden changes affected (and still affect) the definition of goals, costs and available resources, made it impossible to use successful statistical analysis for prognosis and estimation, and destabilized project teams and working environments in companies where these projects were carried out.

Change management had to be fully implemented in project design and management in order to have at least some chance of success. Changes within the project or changes relating to the project became a threat to the projects and amplified the risks of project failure. The Philosophy of short project planning became a defense against the consequences of the instability of the environment where these projects were carried out. This philosophy was in conflict with a wide range of problems that were supposed to be solved. Change management has thus become an essential factor in project management. We may note that publication of the ISO 10 007 standard was initiated by an effort to introduce at least systematic filing and evaluation of changes [6]. This filing was very laborious and required efficient utilization of computers not only to reduce administrative demands, but also to guarantee the rapid and extensive knowledge of changes.

We may note that our project teams often don’t cope with change management very well and go from one extreme to another. They either deny and ignore the changes, or are ruled by these changes.

VIII. Project Limitations - the Situation at the Turn of the Millennium

A contemporary problem affecting projects being designed and carried out is the fact that a project faces a whole range of limitations. It must be brought to completion in the shortest possible time, it is required that costs be kept to a minimum, only limited resources are available for the realization of the project, any foreseeable changes limit the choice of the solutions which would be available were there no changes, etc. Questions such as these very often are not posed: How much will it cost? How long will it take? What will be required? All such matters are often determined upon assignment of the project. In brief: This and that must be done by a certain time, while a predetermined sum and only certain precisely stipulated resources have been set aside for the completion of the project. Thus the fundamental problem becomes whether to decide for or against the realization of the project formulated this way. Often no allowances are made for such a discussion, however, and it is simply required that the project be carried through with the given limitations.

The above-mentioned problems are not specific to current projects and project management. They have a broader context, as described in the publication of the theory of constraints. - TOC.

This situation calls for new approaches to the drafting and realization of projects. One method which captured people’s attention was Goldratt´s Critical Chain [2], which brilliantly integrates classical a priori planning approaches with risk management and flexible follow-up intervention for changes in the project, utilizing multi-level time and funding reserves. .

Questions regarding effective teamwork are currently at the forefront, especially in regard to rapid communication between all participants in a project by means of a computer network. The programming products that are employed in support of project management must ensure the speedy and complex reassessment of the original project design in view of the current state, as well as the original and more recent limitations.

IX. Project Quality at the Dawn of the New Millennium

The publication of the ISO 10 006 standards in 1997 marked a change in the paradigm for the third millennium.

The influence of the philosophy of Total Quality Management (TQM.) is clearly evident - namely, doing the correct things correctly. [5] The fact that even project management was covered by independent standards for quality norms has had great influence on both the aspect of quality as well as the field of project management.

This is partly in reaction to several unsuccessful attempts to establish quality with the TQM, philosophy in a range of companies. Those companies saw the establishment of a quality control system according to the ISO 9 000 standards not as a start-up of fundamental and permanent change in the approaches of management and employees towards quality, but merely as a one-time formal administrative measure for the company. As a rule, such approaches ended unsuccessfully. Now is the time to establish a system of quality management that involves obtaining a certificate, or improving quality, that is always conceptualized as a quality project.

For the field of Project Management the ISO 10 006 standard represents a clear definition of a project as a process that brings a planned change, while this process, and the project's success, is dependent upon a defined multiplicity of processes, each of which must be correctly carried out.

As well, the use of Project Management in connection with quality came right at the time that efforts to procedurally integrate the technical quality, safety and environmental awareness of processes were peaking. Thus the expected, axiomatic transition to a philosophy of global quality has been fully borne out, and Project Management, in those cases, serves as an effective instrument towards the realization of such a philosophy.

The need for improvements in project quality was demonstrated by a survey of the participants at the Project Leadership 99 Conference in Chicago. It revealed that although many American companies had instituted project management, problems persisted with regard to the quality of specific projects.

The ISO 10 006 standard actually came out in the year 1997, but its quality requirements which outline processes are on the one hand difficult to fulfil, and on the other hand, the effects of the standard are not manifested immediately upon its publication. Therefore it is expected to perform its unique function and profoundly influence the field of project management only after the year 2000.

X. Next Trends

How will Project Management develop in the course of the third millennium?

We may assume that in the third millennium, Project Management will to a wide extent use expert systems founded on the knowledge and experience gained from completed projects in order to improve universal knowledge. In particular, this knowledge will be used to optimize the design and management of projects through the use of sophisticated programs based on artificial intelligence (neuron networks, genetic algorithms, etc.) The globalization of projects, and the fact that the managers and participants in projects are often stretched to their limits, calls for the creation of virtual project teams working through computer networks to run projects. Computerized models and simulations will be used during the designing of projects to better predict outcomes. All of this is obviously intended to optimally design and carry out projects. Thus, perhaps the most appropriate title for the paradigm of the future should be optimization.

Conclusion

The above-mentioned paradigms followed in the wake of various principal approaches to projects in different periods of time. The dynamics inherent in their changes reflect the dynamics of the development of Project Management. We could certainly plot out even more eras on a time axis were we to use different criteria: for instance, according to the working methods employed in the design and realization of the project (intuitive approach, heuristic methods, exact methods and methods employing computers), according to the tools used (whether none are used, or organizational aids, or computers) or according to the influence of other disciplines (mathematical statistics, operating research, system theory, management theories, risk engineering), etc. A complete taxonomy, however, was not the aim of this contribution.

It is understandable that changes of Project Management paradigm had necessarily an impact on the required knowledge of the people, who were brought to the head of projects in that time, that led them to a successful end.

To the beginning of the Second World War it was the knowledge that can get only experience. Project managers cannot rely on any of the specialized methods that could be used to successfully design and to successful project management.

Only since the 1970s have been required from the knowledge of project manager's methods such as CPM, PERT, EVM, and many others, including such techniques for example the technique of Gantt charts.

Today certification is required and verified knowledge of a wide range of competencies (see e.g. publications [8], which declare the competence of 20 technical, 20 behavioral and 13 contextual competencies. Together there are 48.

In closing, it must be emphasized that a good knowledge of the currently used paradigms allows project managers to correctly understand the demands made on projects, project management, and finally on project managers themselves.

This paper was done within project denoted as KEGA - 003DTI-4/2014.

References

[1] *Caselli, G.: Wonders of the world. Dorling Kondersley Ltd., London*

[2] *Goldratt, E.M.: Critical Chain. Great Barrington, North River Press 1997*

[3] *ISO 10 006 Quality management – Guidelines to quality in project management*

[4] *Lacko, B.: New Role of Project Management. In: Proceedings of the International Seminar „Project Management in the New Europe". Internet Czechoslovakia Prague 1993, page 127-130.*

[5] Ishikawa, K.: What is total quality control? (The Japanese way). Prentice Hall 1991
[6] ISO 10 007 Quality Management – Guidelines for configuration Management
[7] Čapek, K: R.U.R.(Rossum´s Univeral Robots).
[8] ICB - IPMA Competence Baseline (Version 3.0, June 2006 IPMA). Nijkerk (The Netherlands) 2006.

Authors Index

ISBN: 978-1-61804-230-9