



---

# The 15<sup>th</sup> World Multi-Conference on Systemics, Cybernetics and Informatics

July 19<sup>th</sup> - July 22<sup>nd</sup>, 2011 – Orlando, Florida, USA

## PROCEEDINGS Volume I

Edited by:

Jorge Baralt  
Nagib Callaos  
Chan Ham  
Natalja Lace  
William Lesso  
C. Dale Zinn



**Organized by**  
**International Institute of Informatics and Systemics**  
Member of the International Federation for Systems Research (IFSR)



## **COPYRIGHT**

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy for private use. Instructors are permitted to photocopy, for private use, isolated articles for non-commercial classroom use without fee. For other copies, reprint, or republication permission, write to IIS Copyright Manager, 13750 West Colonial Dr Suite 350 – 408, Winter Garden, Florida 34787, U.S.A. All rights reserved. Copyright 2011. © by the International Institute of Informatics and Systemics.

The papers of this book comprise the proceedings of the conference mentioned on the title and the cover page. They reflect the authors' opinions and, with the purpose of timely disseminations, are published as presented and without change. Their inclusion in these proceedings does not necessarily constitute endorsement by the editors.

ISBN- 978-1-936338-28-3 (Collection)

ISBN-978-1-936338-29-0 (Volume I)



## PROGRAM COMMITTEE

**Chairs:** C. Dale Zinn (USA)  
Jorge Baralt (Venezuela)

Abusitta, Adel	Ajman University of Science and Technology	UAE
Acharya, Sushil	Robert Morris University	USA
Adamopoulou, Evgenia	National Technical University of Athens	Greece
Affenzeller, Michael	Upper Austrian University of Applied Sciences	Austria
Aguirre-Muñoz, Zenaida	Texas Tech University	USA
Aksoy, M. S.	King Saud University	Saudi Arabia
Al Obaidy, Mohaned	Gulf College	Oman
Alexandrov, Natalia	National Aeronautics and Space Administration	USA
Alhamouz, Sadeq	Umm Al-Qura University	Saudi Arabia
Als, Adrian	University of the West Indies	Barbados
Alvarado Moore, Karla	University of Central Florida	USA
Alzamil, Zakarya	Riyadh College of Technology	Saudi Arabia
Anantharaj, Valentine	Mississippi State University	USA
Andina, Diego	Technical University of Madrid	Spain
Anton, José M.	Technical University of Madrid	Spain
Anunciação, Pedro	Instituto Politécnico de Setúbal	Portugal
Aoki, Toru	Shizuoka University	Japan
Aruga, Masahiro	Tokai University	Japan
Assaf, Mansour	The University of Trinidad and Tobago	Trinidad and Tobago
Astakhov, Vadim	University of California San Diego	USA
Aukstakalnis, Nerijus	Kaunas University of Technology	Lithuania
Bangert, Patrick	Algorithmica Technologies	USA
Barkana, Atalay	Anadolu University	Turkey
Barkstrom, Bruce	George Washington University	USA
Barros-Justo, José Luis	Universidad de Vigo	Spain
Baruah, Debendra C.	Tezpur University	India
Basso, Giuliano	Institute of Electrical and Electronics Engineers	Belgium
Belcher, E. Christina	Trinity Western University	Canada
Benbouziane, Mohamed	Tlemcen University	Algeria
Benedicenti, Luigi	University of Regina	Canada
Bennett, Leslie	University of Louisville	USA
Bermeo, José	Universidad de Los Andes	Colombia
Bernardino, Jorge	Instituto Superior de Engenharia de Coimbra	Portugal
Bezuglov, Anton	University of South Carolina	USA
Bhat, Talapady N.	National Institute of Standards and Technology	USA
Bhattacharyya, S.	University of Kentucky	USA
Bidarra, José	Universidade Aberta	Portugal
Blair, Madelyn	Pelerei, Inc.	USA
Boguslavsky, Andrey A.	Keldysh Institute for Applied Mathematics	Russian Federation
Bolboaca, Sorana D.	University of Medicine and Pharmacy	Romania

Bönke, Dietmar	Reutlingen University	Germany
Borchers, Carsten H. J.	University of Hanover	Germany
Botto, Todd	Quinnipiac University	USA
Boukachour, J.	Le Havre University	France
Bradl, Peter	University of Applied Sciences Wuerzburg	Germany
Branski, L. K.	University of Texas Medical Branch	USA
Broussard, Randy P.	United States Naval Academy	USA
Bueno, Newton Paulo	Federal University of Viçosa	Brazil
Burke, David	Robert Morris University	USA
Burnett, Andrea	University of the West Indies	Barbados
Butrous, Nasir	Australian Catholic University	Australia
Buzzi, Maria Claudia	Institute for Informatics and Telematics	Italy
Byun, Juman	The George Washington University	USA
Caldera, Lizeth	Florida International University	USA
Calenzo, Patrick	Institut Matériaux Microélectronique Nanosciences de Provence	France
Calvo, Andres	The University of Dayton	USA
Cano, Julio	Universidad Carlos III de Madrid	Spain
Cao, Hong	Southern Medical University	China
Cárdenas, Henry E.	Louisiana Tech University	USA
Cardoso, Eduardo	Tecnológico de Monterrey	Mexico
Carvalho, Marco	Institute for Human and Machine Cognition	USA
Cázares-Rangel, Víctor	Tecnológico de Monterrey	Mexico
Cerny, Václav	Veřejní Pøístupné Informace	Czech Republic
Cha, Seung Tae	Korea Electronic Power Research Institute	South Korea
Cha, Sung-Hyuk	Pace University	USA
Chandra, Vigyan	Eastern Kentucky University	USA
Chen, Chuen-Yih	Chang Jung Christian University	Taiwan
Chen, Huei-Huang	Tatung University	Taiwan
Chen, Lisa Y.	I-Shou University	Taiwan
Chen, Shih-Chih	National Taiwan University	Taiwan
Chen, Yil	Huafan University	Taiwan
Chen, Yuhua	University of Houston	USA
Cherinka, R.	The MITRE Corporation	USA
Cherry, Barbara	Indiana University	USA
Chien, Steven	New Jersey Institute of Technology	USA
Chiou, Richard	Drexel University	USA
Chiou, Yin-Wah	Nanhua University	Taiwan
Cho, Kyoung-Rok	Chungbuk National University	South Korea
Cho, Tae Won	Chungbuk National University	South Korea
Cho, Vincent	The Hong Kong Polytechnic University	Hong Kong
Choi, Seung-Seok	Pace University	USA
Choi, Yu-Lee	Chungbuk National University	South Korea
Choo, Jinboo	Korea Electric Power Research Institute	South Korea
Chou, Andrew	Kainan University	Taiwan
Chowdhury, Masud H.	University of Illinois at Chicago	USA
Cipolla F., Francisco	F&F Multimedia Communications Corp.	Italy
Cirella, Jonathan	RTI International	USA
Clarke, Tim	University of Wales	UK
Clegg, Warwick	Victoria University of Wellington	New Zealand
Coffman, Michael G.	Southern Illinois University Carbondale	USA

Cohen, Bernard	City University London	UK
Contreras, Sebastián	Universidad de Los Andes	Colombia
Cote, Paul	Benet Laboratories	USA
Cowan, Jimmy	Florida Institute of Technology	USA
Cripe, Billy	Oracle Corporation	USA
Curran, Kevin	University of Ulster	UK
Dawoud, Dawoud	University of KwaZulu Natal	South Africa
De Volder, Dennis	Western Illinois University	USA
Demestichas, K.	National Technical University of Athens	Greece
Desa, Shakinaz	Universiti Pendidikan Sultan Idris	Malaysia
Diallo, Saikou Y.	Old Dominion University	USA
Dierneder, Stefan	Linz Center of Mechatronics GmbH	Austria
Djukom, C. D.	University of Texas Medical Branch at Galveston	USA
Dolan, Dan	South Dakota School of Mines & Technology	USA
Dosi, Vasiliki	University of Ioannina	Greece
Duman, Hakan	University of Essex	UK
Dunning, Jeremy	Indiana University	USA
Dziong, Zbigniew	University of Québec	Canada
Edwards, Stephen H.	Virginia Tech	USA
El-Halafawy, Farag Z.	Menoufiya University	Egypt
Elmahboub, W. M.	Hampton University	USA
Emdad, F.	University of Texas Medical Branch	USA
Erickson, Sarah	Florida International University	USA
Erkollar, Alptekin	University of Applied Sciences	Austria
Eshraghian, Kamran	Chungbuk National University	South Korea
Estévez, Leonardo	Texas Instruments	USA
Eye, John	Southern Utah University	USA
Fang, Rong Jyue	Southern Taiwan University of Technology	Taiwan
Fisher, Wendy	The Open University	UK
Fox, Kelly	Texas Tech University	USA
Freund, Rodolf	Vienna University of Technology	Austria
Fu, Shih-Lung	National Taiwan University	Taiwan
Fu, Xiuju	Institute of High Performance Computing	Singapore
Fu, Yonggang	Shanghai Jiao Tong University	China
Fuhrer, Patrik	University of Fribourg	Switzerland
Fujikawa, Takemi	University of Western Sydney	Australia
Fujita, Naoyuki	Japan Aerospace Exploration Agency	Japan
Fumizawa, Motoo	Shonan Institute of Technology	Japan
Gagnon, Francois	École de Technologie Supérieure	Canada
Ganapathi, Nanthini	Research Triangle Institute	USA
Ganapathy, Velappa	Monash University	Malaysia
Ganchev, Ivan	University of Limerick	Ireland
García-Atanacio, César	Instituto Nacional de Investigaciones Nucleares	Mexico
Gardezi, A. K.	Colegio de Postgraduados	Mexico
Garibaldi, Jonathan M.	University of Nottingham	UK
Georgescu, Vasile	University of Craiova	Romania
Gesekus, Tim	DFS Deutsche Flugsicherung GmbH	Germany
Glotzbach, Ronald J.	Purdue University	USA
Godavarty, Anuradha	Florida International University	USA
González, Jean	Florida International University	USA
Gordon, Richard	University of KwaZulu Natal	South Africa

Goriachkin, Oleg	Technical and Engineering Universities of Russia	Russian Federation
Gosselin, Clément	Universite Laval	Canada
Goulding, Tom	Daniel Webster College	USA
Gregory, Mark A.	RMIT University	Australia
Guadarrama, Javier de J.	Instituto Tecnológico de Toluca	Mexico
Guevara López, Miguel	University of Porto	Portugal
Guiasu, Radu Cornel	York University	Canada
Guiasu, Silviu	York University	Canada
Gulez, Kayhan	Yıldız Teknik Üniversitesi	Turkey
Gustavsson, Rune	Blekinge Institute of Technology	Sweden
Gyires, Tibor	Illinois State University	USA
Haba, C. G.	Technical University of Iasi	Romania
Hallot, Frédéric	Royal Military Academy	Belgium
Ham, Chan	Southern Polytechnic State University	USA
Hamanaka, Masatoshi	University of Tsukuba	Japan
Hammond, Bruce R.	Saint Leo University	USA
Hangai, Seiichiro	Tokyo University of Science	Japan
Hao, Tianyong	City University of Hong Kong	China
Hardy, Frank	University of South Carolina Upstate	USA
Hardy, Leon C.	Embry-Riddle Aeronautical University	USA
Hashimoto, Shigehiro	Osaka Institute of Technology	Japan
Heiserich, Gerd	Institute of Transport and Automation Technology	Germany
Hemmelman, Brian	The South Dakota School of Mines and Technology	USA
Hendel, Russell Jay	Towson University	USA
Henninger, Michael	PH Weingarten	Germany
Herget, Josef	University of Applied Sciences	Switzerland
Herndon, D. N.	University of Texas Medical Branch at Galveston	USA
Herr, Stephan	DFS Deutsche Flugsicherung GmbH	Germany
Hetzer, Dirk	Media Broadcast	Germany
Higashiyama, Yoichi	Ehime University	Japan
Hochin, Teruhisa	Osaka Prefecture University	Japan
Hodge, Diane M.	Radford University	USA
Hofmeister, Helge	BASF IT Services	Belgium
Hong, Yun-Ki	Chungbuk National University	South Korea
Horne, Jeremy	Maricopa Community College	USA
Hsu, Shu-Mei	Tatung University	Taiwan
Huang, Pin Chia	I-Shou University Kaohsiung	Taiwan
Huang, Ruhua	Wuhan University	China
Huang, Sheng He	University of Southern California	USA
Hvass, Michael	IBM Denmark	Denmark
Iembo, Rosanna	University of Calabria	Italy
Imai, Michita	Keio University	Japan
Imamura, Nobuaki	Kobe City College of Technology	Japan
Ishikawa, Hiroshi	NUIS	Japan
Ito, Akinori	Tohoku University	Japan
Jäntschi, Lorentz	Academic Direct Organization	Romania
Jiménez R., Lourdes	University of Alcala	Spain
Johnson, Mark	Benet Laboratories	USA
Jones, Paul	University of Cincinnati	USA
Joordens, Steve	University of Toronto Scarborough	Canada
Jung, Kyung Im	Samsung Electronics	South Korea

Jung, Kyung Kwon	Dongguk University	South Korea
Jung, Seul	Chungnam National University	South Korea
Kadoch, Michel	University of Québec	Canada
Kamejima, Kohji	Osaka Institute of Technology	Japan
Kaneko, Takashi	Niigata University	Japan
Kao, Chih-Yang	Ming-Chuan University	Taiwan
Karamat, Parwaiz	The Open Polytechnic of New Zealand	New Zealand
Kasapoglu, Ercin	Hacettepe Üniversitesi	Turkey
Kaszubiak, J.	University Magdeburg	Germany
Kawaguchi, Masashi	Suzuka National College of Technology	Japan
Kehtarnavaz, Nasser	University of Texas at Dallas	USA
Khaled, Pervez	University of Illinois at Chicago	USA
Khan, Faisal	Khalifa University of Science	UAE
Khatri, Anil	Bowie State University	USA
Kim, Jeongdae	Information and Communications University	South Korea
Kim, Kyungwoo	University of Florida	USA
Kim, Seok-Man	Chungbuk National University	South Korea
Kim, Yeo Jin	Samsung Electronics	South Korea
Kim, Yeon-Ho	Chungbuk National University	South Korea
Kim, Yong Hak	Korea Electric Power Research Institute	South Korea
Kincaid, Rex K.	The College of William and Mary	USA
Kira, Dennis S.	Concordia University	Canada
Kizirian, Robin	Drexel University	USA
Klapp, Jaime	Instituto Nacional de Investigaciones Nucleares	Mexico
Kobal, Damjan	University of Ljubljana	Slovenia
Koczkodaj, Waldemar	Laurentian University	Canada
Komatsu, Takanori	Shinshu University	Japan
Kouki, A.	École de Technologie Supérieure	Canada
Kozma-Bognár, V.	University of Pannonia	Hungary
Krakovska, Monika	Jagiellonian University	Poland
Kreisler, Alain	Université Pierre et Marie Curie	France
Kromrey, Jeffrey D.	University of South Florida	USA
Kronreif, Gernot	Austrian Research Centers	Austria
Krothapalli, Sreenivasa	Indian Institute of Technology Kharagpur	India
Kuftin, Felix A.	KTTI	Russian Federation
Kung, C. M.	Shih Chien University	Taiwan
Kuragano, Tetsuzo	The American Society of Mechanical Engineers	Japan
Kutter, Anna K.	PH Weingarten	Germany
Kwon, Yongjin (James)	Ajou University	South Korea
Lahlouhi, Ammar	University of Biskra	Algeria
Lai, James C. K.	Idaho State University College of Pharmacy	USA
Latawiec, Krzysztof J.	Opole University of Technology	Poland
Latchman, Haniph A.	University of Florida	USA
Lee, Hwajung	Radford University	USA
Lee, Jae-Suk	Gwangju Institute of Science and Technology	South Korea
Lee, Kangsun	MyongJi University	South Korea
Lee, Nam Ho	Korea Electric Power Research Institute	South Korea
Lee, Sang-Jin	Chungbuk National University	South Korea
Lee, Suk Bong	Samsung Electronics	South Korea
Lee, Yih-Jiun	Chien Kuo Technology University	Taiwan
Lee, Yusin	National Cheng Kung University	Taiwan

Letellier, T.	Université Victor Segalen	France
Li, Lihong	City University of New York	USA
Lin, Shu-Chiung	Tatung University	Taiwan
Linares, Oscar	Universidad Distrital Francisco José de Caldas	Colombia
Lipikorn, Rajalida	Chulalongkorn University	Thailand
Lipinski, Piotr	Technical University of Lodz	Poland
Litvin, Vladimir	California Institute of Technology	USA
Liu, Jun	University of Ulster	UK
Liu, Kuo-Shean	Tatung University	Taiwan
Liu, Shih-Chi	Tatung University	Taiwan
Livne, Nava L.	University of Utah	USA
Livne, Oren E.	University of Utah	USA
Long, Changjiang	Huazhong University	China
López Román, Leobardo	University of Sonora	Mexico
Lou, Shi Jer	National Pingtung University of Science and Technology	Taiwan
Loutfi, Mohamed	University of Sunderland	UK
Love C., Gloria	Dillard University	USA
Lowe, John	University of Bath	UK
Lowry, Pam	Lawrence Technological University	USA
Luh, Guan Chun	Tatung University	Taiwan
Lui, Wen Lik Dennis	Monash University	Malaysia
Lyell, Margaret	Intelligent Automation, Inc.	USA
Ma, Yongqing	Victoria University of Wellington	New Zealand
Machotka, Jan	University of South Australia	Australia
Mahendran, Francis	Motorola Software Group Singapore	Singapore
Mahgoub, Ahmed G.	Alexandria University	Egypt
Manley, Denis	Dublin Institute of Technology	Ireland
Mansikkamäki, Pauliina	Tampere University of Technology	Finland
Marino, Mark	Erie County Community College	USA
Martínez, Pablo	RIKEN	Japan
Martínez, Sergio	Florida International University	USA
Martínez M., Natividad	Universidad Carlos III de Madrid	Spain
Marwala, Tshilidzi	University of Johannesburg	South Africa
Mascari, Jean-François	Istituto per le Applicazioni del Calcolo	Italy
Masoumi, Nasser	University of Tehran	Iran
Mathews, Brian	University of Bedfordshire	UK
Matsumoto, Kazunori	KDDI R&D Laboratories Inc.	Japan
Matsuno, Akira	Teikyo University	Japan
Mayer, Daniel	University of West Bohemia	Czech Republic
Mbobi, Aime Mokhoo	Centennial College	Canada
Medina, Omar	Universidad de Guanajuato	Mexico
Mehrabian, Ali	University of Central Florida	USA
Mellouli, Sehl	Université Laval	Canada
Meyer, Heiko	University of Applied Sciences	Germany
Michaelis, B.	University Magdeburg	Germany
Migliarese, Piero	Università Della Calabria	Italy
Miller, R.	The MITRE Corporation	USA
Minnaert, Eric	South Dakota School of Mines & Technology	USA
Mirza Sebzali, Yussef	Kuwait Institute for Scientific Research	Kuwait
Mistry, Jaisheel	University of the Witwatersrand	South Africa
Moin, Lubna	Pakistan Naval Engineering College	Pakistan

Moreau, Alban	University of Brest	France
Moreno, Carlos M.	Universidad Central de Venezuela	Venezuela
Moreno Sánchez, Ana	Universidad Politécnica de Madrid	Spain
Morii, Hisashi	Shizuoka University	Japan
Mostafaeipour, Ali	Yazd University	Iran
Muknahallipatna, Suresh	University of Wyoming	USA
Muraleedharan, Rajani	Syracuse University	USA
Naddeo, Alessandro	Università di Salerno	Italy
Nadworny, Margaret	Global Software Group, Motorola	USA
Nagai, Yasuo	Tokyo University of Information Sciences	Japan
Nagalakshmi, Vadlamani	Gandhi Institute of Technology and Management	India
Nagaoka, Tomoyuki	Tokyo Metropolitan University	Japan
Nahmens, Isabelina	University of South Florida	USA
Naillon, Martina	Co-Decision Technology	France
Nair, V. S. Sukumaran	Southern Methodist University	USA
Nakashima, Takuya	Shizuoka University	Japan
Nam, Su Chul	Korea Electronic Power Research Institute	South Korea
Nave, Felecia M.	Prairie View A&M University	USA
Nawawi, S. W.	Universiti Teknologi Malaysia	Malaysia
Nazmy, Taymoor M.	Ain Shams University	Egypt
Nedic, Zorica	University of South Australia	Australia
Nelwamondo, Fulufhelo	University of the Witwatersrand	South Africa
Neo, Yoichiro	Shizuoka University	Japan
Nguyen, Mai	RTI International	USA
Nguyen, Patricia	RTI International	USA
Ni, Xingliang	University of Science and Technology of China	China
Noh, Bong Nam	Chonnam National University	South Korea
Nonaka, Hidetoshi	Hokkaido University	Japan
Novikov, Oleg	Tomko Inc.	Russian Federation
O'Brien, Ann	Loughborough University	UK
Obrebski, Jan B.	Politechnika Warszawska	Poland
Ocelka, Tomas	Institute of Public Health	Czech Republic
Oh, Yun Sang	Samsung Electronics	South Korea
Ojala, Pasi	University of Oulu	Finland
Olla, Phillip	Madonna University	USA
Olson, Patrick C.	Aware Consulting Group	USA
Ong, Vincent Koon	University of Bedfordshire	UK
Osadciw, Lisa Ann	Syracuse University	USA
Ostergaard, Soren Duus	IBM Europe	Denmark
Overmeyer, Ludger	Institute of Transport and Automation Technology	Germany
Oya, Hidetoshi	Shonan Institute of Technology	Japan
Ozdemir, Ahmet S.	Marmara University	Turkey
Palesi, Maurizio	University of Catania	Italy
Paré, Dwayne E.	University of Toronto Scarborough	Canada
Park, Kyung-Chang	Chungbuk National University	South Korea
Park, Youngc	Baekseok University	South Korea
Peng, Jian	University of Saskatchewan	Canada
Petkov, Emil	Northumbria University	UK
Pfeifer, Michael	Technical University of Dortmund	Germany
Phillips, C. Dianne	NorthWest Arkansas Community College	USA
Phillips, James	Louisiana Tech University	USA

Pierce, Ryan M.	Arkansas State University	USA
Pitzer, Erik	University of Applied Sciences Hagenberg	Austria
Platt, Glenn	CSIRO	Australia
Polanski, Andrzej	Silesian University of Technology	Poland
Poon, Gilbert	University of Waterloo	Canada
Popentiu, Florin	University of Oradea	Romania
Postolache, Octavian	Instituto de Telecomunicações	Portugal
Poulin, Régis	Universite Laval	Canada
Pulcher, Karen L.	University of Central Missouri	USA
Pun, Daniel	Central Queensland University	Australia
Putnam, Janice	University of Central Missouri	USA
Quan, Xiaojun	University of Science and Technology of China	China
Quintyne, Vasco	University of the West Indies	Barbados
Rachev, Boris	Technical University of Varna	USA
Rahman, Anuar Abdul	Pusat Tenaga Malaysia	Malaysia
Rahman, Mohammad	University of Texas at Dallas	USA
Ramachandran, S.	Indian Institute of Technology Madras	India
Ramirez C., Guillermo	Soka University	Japan
Ratkovic K., Nada	University of Belgrade	Serbia
Ren, Jianfeng	Qualcomm Inc.	USA
Revetria, Roberto	Università degli Studi di Genova	Italy
Riaz M., Mohammad	University College of Engineering and Technology	Pakistan
Rodríguez, María D.	Universidad de Alcalá	Spain
Ropella, Glen E. P.	Tempus Dictum Inc.	USA
Rossignol, R.	Université Victor Segalen	France
Rossmann, Jürgen	RWTH Aachen University	Germany
Ruan, Tongjun	New Mexico Tech	USA
Rutkauskas, Aleksandras	Vilnius Gediminas Technical University	Lithuania
Saadane, Rachid	Institut Eurecom	France
Sahara, Tomohiro	Osaka Institute of Technology	Japan
Salazar, Dora	Texas Tech University	USA
Saleh, Magda M.	Alexandria University	Egypt
Sanna, Andrea	Politecnico di Torino	Italy
Šaruckij, Mark	Cracow University of Economics	Poland
Sateesh Reddy, J.	Indian Institute of Technology Madras	India
Sato, Tomoaki	Hirosaki University	Japan
Sax, Eric	Mercedes-Benz Technology Gmbh	Germany
Schaeffer, Donna M.	Marymount University	USA
Schlette, Christian	RWTH Aachen University	Germany
Schluse, M.	RWTH Aachen University	Germany
Schrader, P. G.	University of Nevada	USA
Schulz, Lennart	Institute of Transport and Automation Technology	Germany
Schumacher, Jens	University of Applied Sciences Vorarlberg	Austria
Seepold, Ralf	Universidad Carlos III de Madrid	Spain
Segall, Richard S.	Arkansas State University	USA
Seitzer, Jennifer	The University of Dayton	USA
Sert, Yasemin	University of South Florida	USA
Shaw, Jill	The Open University	UK
Shayeghi, Hossien	Iran University of Science	Iran
Shim, Eung Bo	Korea Electric Power Research Institute	South Korea
Shin, Jeong Hoon	Korea Electronic Power Research Institute	South Korea

Shin, Jungpil	University of Aizu	Japan
Shiraishi, Yoshiaki	Nagoya Institute of Technology	Japan
Shklyar, Benzion	Holon Academic Institute	Israel
Sim, Sang Gyoo	Samsung Electronics, R. O. Korea	South Korea
Sinkunas, Stasys	Kaunas University of Technology	Lithuania
Sleit, Azzam Talal	University of Jordan	Jordan
Soeiro, Alfredo	University Porto	Portugal
Song, Hong Jun	University of Sydney	Australia
Soundararajan, Ezekiel	Indiana University of Pennsylvania	USA
Srisawangrat, Songsiri	University of Regina	Canada
Starosolski, Zbigniew	Silesian University of Technology	Poland
Stasytyte, Viktorija	Vilnius Gediminas Technical University	Lithuania
Stößlein, Martin	Universität Erlangen-Nürnberg	Germany
Stubberud, Stephen	The Boeing Company	USA
Su, J. L.	Shanghai University	China
Su, Wei	National University of Defense Technology	USA
Sun, Baolin	Wuhan University	China
Suzuki, Motoyuki	Tohoku University	Japan
Swart, William	East Carolina University	USA
Szygenda, Stephen A.	Southern Methodist University	USA
Takemi, Fujikawa	University of Western Sydney	Australia
Tam, Wing K.	Swinburne University of Technology	Australia
Tappert, Charles C.	Pace University	USA
Taylor, Stephen	Sussex University	UK
Tchier, Fairouz	King Saud University	USA
Teng, Chia-Chi	Brigham Young University	USA
Teshigawara, Yoshimi	Soka University	Japan
Thakur, Amrit' Anshu	Mississippi State University	USA
Theologou, Michael	National Technical University of Athens	Greece
Thornton, Mitchell A.	Southern Methodist University	USA
Till, Robert	John Jay College	USA
Trahan, Shane	RTI International	USA
Traum, Maria	Johannes Kepler University	Austria
Trimble, Robert	Indiana University of Pennsylvania	USA
Trzaska, Mariusz	Polish Japanese Institute of Information Technology	Poland
Tsai, Li-Hung	Tatung University	Taiwan
Tsaur, Woei-Jiunn	Da-Yeh University	Taiwan
Tseng, Ko Ying	Tatung University	Taiwan
Tseng, Tzu-Liang (Bill)	The University of Texas at El Paso	USA
Tsiligaridis, John	The State University of New York at Buffalo	USA
Tsubaki, Michiko	The University of Electro-Communications	Japan
Turnitsa, Charles D.	Old Dominion University	USA
Valakevicius, Eimutis	Kaunas University of Technology	Lithuania
Van Delden, Sebastian	University of South Carolina Upstate	USA
Vargoz, Erik P.	The College of William and Mary	USA
Vasinek, Vladimir	Technical University of Ostrava	Czech Republic
Venkataraman, S.	All India Association for Micro Enterprise Development	India
Verlinde, Patrick	Royal Military Academy	Belgium
Viloria, Orlando H.	Universidad Simón Bolívar	Venezuela
Vitral, Renan	BITSYS	USA
Volkov-Husovic, T.	University of Belgrade	Yugoslavia

Voss, Andreas	Dortmund University of Technology	Germany
Wagner, Stefan	Upper Austrian University of Applied Sciences	Austria
Walter J., Ammann	Institute for Snow and Avalanche Research	Switzerland
Wang, Lei	University of Houston	USA
Warwick, Jon	London South Bank University	UK
Wei, Xinzhou	New York City College of Technology	USA
Wells, Harvey	King's College London	UK
Whitbrook, Amanda M.	University of Nottingham	UK
Wirtz, Guido	Universität Bamberg	Germany
Woodhead, Steve	University of Greenwich	UK
Woodthorpe, John	The Open University	UK
Wu, Chun Yin	Tatung University	Taiwan
Yamada, Seiji	National Institute of Informatic	Japan
Yamaguchi, Akira	Meisei University	Japan
Yan, Kuo Qin	Chaoyang University of Technology	Taiwan
Yan, Mu Tian	Huafan University	Taiwan
Yang, Huiqing	Virginia State University	USA
Yang, Hung Jen	National Kaohsiung Normal University	Taiwan
Yang, Ming	Jacksonville State University	USA
Yang, Sung Un	Syracuse University	USA
Yang, Yueh-Ting	Drexel University	USA
Yasser, Muhammad	BHK	Japan
Yatsymirskyy, Mykhaylo	Technical University of Lodz	Poland
Yazawa, Toru	Tokyo Metropolitan University	Japan
Yilmaz, Levent	Auburn University	USA
Yoon, Changwoo	Electronics and Telecommunications Research Institute	South Korea
Yoshida, Eri	Toyohashi University of Technology	Japan
Yoshida, Takahiro	Tokyo University of Science	Japan
You, Younggap	Chungbuk National University	South Korea
Yu, Xin	University of Bath	UK
Zadeh, Jeff	Virginia State University	USA
Zarama, Roberto	Universidad de los Andes	Colombia
Zaretsky, Esther	Givat Washington College of Education	Israel
Zelinka, Tomas	Czech Technical University in Prague	Czech Republic
Zeller, Andrew J.	Norisbank	Germany
Zhang, Jinfang	University of Waterloo	Canada
Zhang, Linda L.	University of Groningen	Netherlands
Zhang, Qingyu	Arkansas State University	USA
Zhang, Xiaozheng Jane	California Polytechnic State University	USA
Zhonghua, Fang	Shanghai Institute of Technical Physics	USA
Zhou, Dan	University of California San Diego	USA
Zhu, Hui	Soochow University	China
Zobaa, Ahmed	Cairo University	Egypt



## ADDITIONAL REVIEWERS

Abbas-Turki, Abdeljalil	University of Technology of Belfort-Montbéliard	France
Abdel Hafez, Hoda	Suez Canal University	Egypt
Abdul, Hamed	University of Management and Technology	Pakistan
Aboalsamh, Hatim	King Saud University	Saudi Arabia
Abu Bakar, Ahmad B.	International Islamic University of Malaysia	Malaysia
Adela, Ionescu	University of Craiova	Romania
Agulhari, Cristiano M.	State University of Campinas	Brazil
Akbar Hussain, D. M.	Aalborg University Esbjerg	Denmark
Aleksic-Maslac, Karmela	Zagreb School of Economics and Management	Croatia
Alexik, Mikulas	University of Zilina	Slovakia
Almulla, Mohammed	Kuwait University	South Korea
Alzaid, Hani	Queensland University of Technology	Australia
Andreae, John H.	University of Canterbury	New Zealand
Anido, Cecilia	Far Eastern University	Philippines
Apolloni, Bruno	University of Milan	Italy
Araújo, Armando	University of Porto	Portugal
Arora, Rajan	Cadence Design Systems	India
Arunyawat, Uraiwan	Kasetsart University	Thailand
Asatryan, David	Institute for Informatics and Automation Problems of NAS	Armenia
Atmaca, Tulin	Telecom SudParis	France
Avny, Amos	Omnidev International Consultants	Israel
Bacha, Faouzi	University of Tunis	Tunisia
Bagnoli, Franco	University of Florence	Italy
Balas, Valentina	Aurel Vlaicu University of Arad	Romania
Bamidis, Panos	Aristotle University of Thessaloniki	Greece
Bardine, Alessandro	University of Pisa	Italy
Bargiotas, Dimitrios	Technological Education Institution of Chalkida	Greece
Bartolini, Sandro	University of Siena	Italy
Batos, Vedran	University of Dubrovnik	Croatia
Behzadan, Amir	University of Central Florida	USA
Bessmertny, Igor	University of Information Technologies	Russian Federation
Bettaz, Mohamed	Philadelphia University	Jordan
Bisceglia, Bruno	University of Salerno	Italy
Boonjing, Veera	King Mongkut's Institute of Technology Ladkrabang	Thailand
Booth, Marcia	West Virginia State University	USA
Borges, Carlos C. H.	National Laboratory for Scientific Computation	Brazil
Borworn, Papasratorn	King Mongkut's University of Technology	Thailand
Boudaba, Madjid	Electronic Solutions GmbH	Germany
Buglione, Luigi	Engineering IT	Italy
Burita, Ladislav	Military Academy in Brno	Czech Republic

Burton, Peter G.	Australian Catholic University	Australia
Camilleri, Mario	University of Malta	Malta
Canbolat, Huseyin	Mersin University	Turkey
Carnes, Patrick	Kirtland Air Force Base	USA
Castellini, Horacio	National University of Rosario	Argentina
Cechova, Ivana	University of Defence	Czech Republic
Chang, Ching Pou	Hsiuping Institute of Technology	Taiwan
Chang, Hsung Pin	National Chung Hsing University	Taiwan
Chang, Weng-long	National Kaohsiung University of Applied Sciences	Taiwan
Chang, Yu-Ju	National Chi Nan University	Taiwan
Chen, Chin-Chou	Chunghwa Telecom	Taiwan
Chen, Jingchao	Donghua University	China
Chen, Kuan-Hung	Feng-Chia University	Taiwan
Chen, Liang-Bi	National Sun Yat-Sen University	Taiwan
Chengalvarayan, R.	AT&T Labs Research	USA
Chiu, Shao-I	Taipei College of Maritime Technology	Taiwan
Clément, Alain	Université d'Angers	France
Corcuera, Pedro	University of Cantabria	Spain
Corsano, Gabriela	Instituto de Desarrollo y Diseño	Argentina
Costa, João	Federal University of Pará	Brazil
Crampes, Michel	Ecole des Mines d'Ales France	France
Cristóbal-Salas, Alfredo	University of Veracruz	Mexico
Cuvillo, Carlos del	Technical University of Madrid	Spain
D`Silva, Icy	University of Guelph	Canada
Dahlan, Jaslin	Mara University of Technology	Malaysia
Daley, Wayne	Georgia Tech Research Institute	USA
Datcu, Dragos	Delft University of Technology	Netherlands
De Diego, José Ramón	Universidad de Oviedo	Spain
De la Rosa, Guadalupe	Universidad de Guanajuato	Mexico
Di Rico, Gianluca	Osservatorio Astronomico di Teramo	Italy
Dimililer, Kamil	Near East University	Cyprus
Dizdaroglu, Bekir	Karadeniz Technical University	Turkey
Dobritoiu, Maria	University of Petrosani	Romania
Dvornik, Josko	University of Split	Croatia
Dzhusupova, Zamira	United Nations University	Macau
Eaglestone, Barry	University of Sheffield	UK
Ediberidze, Alexeender	Technical University of Georgia	Georgia
El Bakkali, Hanan	ENSIAS	Morocco
El Manouni, Said	Al-Imam University	Saudi Arabia
Elgamal, Mohamed	University of Victoria	Canada
El-Nadi, Khairia	Alexandria University	Egypt
Erol, Cemil B.	TUBITAK	Turkey
Farooqui, Aamir A.	King Fahd University of Petroleum and Minerals	USA
Faulín, Javier	Public University of Navarre	Spain
Félez, Jesús	Technical University of Madrid	Spain
Fillion, Gerard	University of Moncton	Canada
Fine, Terri	University of Central Florida	USA
Fiorini, Rodolfo	Polytechnic of Milan	Italy
Florea, Adrian	University of Sibiu	Romania
Florescu, Gabriela	National Institute for Research and Development in Informatics	Romania
Floyd, Raymond	Innovative Insights, Inc.	USA

Fox, Richard	Northern Kentucky University	USA
Fukase, Masa-Aki	Hirosaki University	Japan
Fúster-Sabater, Amparo	Consejo Superior de Investigaciones Científicas	Spain
García Marco, Francisco	University of Zaragoza	Spain
Genser, Robert	Institute for Handling Devices and Robotics	Austria
Giampapa, Joseph	Carnegie Mellon University	USA
Gini, Giuseppina	Polytechnic of Milan	Italy
Golayoglu F., Afet	Baskent University	Turkey
Gonçalves, Ricardo	New University of Lisbon	Portugal
Grygierzec, Wieslaw	University of Agriculture in Krakow	Poland
Guang, Liang	University of Turku	Finland
Haack, Bertil	Technical University of Applied Sciences Wildau	Germany
Haddad, Hisham	Kennesaw State University	USA
Hameed, Shihab A.	International Islamic University Malaysia	Malaysia
Harib, Khalifa	United Arab Emirates University	UAE
Hartley, Roger	University of Leeds	UK
Hu, Hong	Hampton University	USA
Huang, Yo-Ping	National Taipei University of Technology	Taiwan
Hussain, Aini	National University of Malaysia	Malaysia
Ianigro, Massimo	Italian National Research Council	Italy
Ibrahim, Azzam	Hamad Medical Corporation	USA
Ilkova, Tatiana	Institute of Biophysics and Biomedical Engineering	Bulgaria
Inci, A. Can	Bryant University	USA
Ingber, Lester	Lester Ingber Research	USA
Intakosum, Sarun	King Mongkut`s Institute of Technology Ladkrabang	Thailand
Iovan, Stefan	Romanian Railway IT Company	Romania
Ishikawa, Yasushige	Kyoto University of Foreign Studies	Japan
Ivlev, Oleg	University of Bremen	Germany
Janota, Ales	University of Zilina	Slovakia
Jaoua, Ali	University of Qatar	Qatar
Job, Remo	University of Trento	Italy
Jwo, Dah-Jing	National Taiwan Ocean University	Taiwan
Kalpic, Damir	University of Zagreb	Croatia
Karigar, Chandrakant	Bangalore University	India
Kawarazaki, Noriyuki	Kanagawa Institute of Technology	Japan
Kim, Do-Hoon	Kyung Hee University	South Korea
Kim, Hyung Nam	Virginia Polytechnic Institute and State University	USA
Kim, Hyunju	Jackson State University	USA
Kim, Mooyoung	Sejong University	South Korea
Kiriazov, Petko	Bulgarian Academy of Sciences	Bulgaria
Klamka, Jerzy	Silesian University of Technology	Poland
Kong, Hyung-Yun	University of Ulsan	South Korea
Krishnam, Manjunath	Intellon Corporation	USA
Kroumov, Valeri	Okayama University of Science	Japan
Ksenofontov, Alexandre	Moscow Engineering Physics Institute	Russian Federation
Kurtulus, Kemal	Istanbul University	Turkey
Kurubacak, Gulsun	Anadolu University	Turkey
Lai, Cristian	Center for Advanced Studies, Research and	Italy
Lappas, Georgios	Technological Educational Institution of Western Macedonia	Greece
Lara, Teodoro	University of the Andes	Venezuela
Las-Heras, Fernando	University of Oviedo	Spain

Law, Rob	Hong Kong Polytechnic University	Hong Kong
Ledesma Orozco, Sergio E.	Guanajuato University	Mexico
Lee, DoHoon	Pusan National University	South Korea
Lee, Jong Hyup	Inje University	South Korea
Lee, Jong Kun	Changwon National University	South Korea
Lee, Richard C. H.	University of Hong Kong	Hong Kong
Lehman, Mario	Sofilab S.A. de C.V	Mexico
Leporati, Francesco	Pavia University	Italy
Li, Di	South-China University of Technology	China
Li, Weigang	University of Brasilia	Brazil
Li, Zheng	Alcatel-Lucent	USA
Liang, Lou Y.	Leader University	Taiwan
Liao, Chu-Hsien	Food Industry Research & Development Institute	Taiwan
Lin, Hong	University of Houston Downtown	USA
Liu, Hong	Embry-Riddle Aeronautical University	USA
Liu, HsinChih Frank	Shih-Chien University	Taiwan
Liu, Yonghuai	University of Wales Aberystwyth	UK
Lo, Yu-Kang	National Taiwan University of Science and Technology	Taiwan
Lopes da Silva, Paulo A.	Military Engineering Institute	Brazil
Luna Rodríguez, Juan J.	Cordoba University	Spain
Machado, José	Institute of Engineering of Porto	Portugal
Mainguenaud, Michel	Institut National des Sciences Appliquées	France
Malan, Danie	University of Pretoria	South Africa
Mansour, Essam	International University in Germany	Germany
Mansour, Samah	Grand Valley State University	USA
Martín, José L.	University of the Basque Country	Spain
Martínez-Carreras, M. A.	University of Murcia	Spain
Mayer, Bernd	Emergentec Biodevelopment GmbH	Austria
Mehata, K. M.	Anna University	India
Memon, Qurban	Hamdard University	Pakistan
Meyer, Heiko	Gefasoft AG	Germany
Milani, Mattia	Intercom R. R. L.	Italy
Moharrum, Mohammed	Old Dominion University	USA
Mohsen, Ghribi	University of Moncton	Canada
Montisci, Augusto	University of Cagliari	Italy
Moody, Scott	The Boeing Company	USA
Moreno, María N.	University of Salamanca	Spain
Morgan, Theresa	Wudang Research Association	USA
Mozar, Stefan	Electrical Testing Services Pty Ltd.	Australia
Mtika, Chatonda	Algis Autograding, Inc.	Canada
Muni, Bishnu Prasad	Bhel Corporate R&D	India
Musio, Carlo	Institute of Cybernetics Eduardo Caianiello	Italy
Nagar, Atulya Kumar	Liverpool Hope University	UK
Nagy, Endre L.	Society of Instrument and Control Engineers	Japan
Nahm, In Hyun	Sun Moon University	South Korea
Narasimhan, Lakshmi	University of Newcastle	USA
Nemec, Juraj	Matej Bel University	Slovakia
Ng, Kia	University of Leeds	UK
Nieuwenhuis, Lambert	University of Twente	Netherlands
Occelli, Sylvie	Institute for Economic & Social Research of Piedmont	Italy
Odetayo, Michael	Coventry University	UK

Ogorodnikov, Dmitri	Mount Sinai School of Medicine	USA
Okatan, Ali	Halic University	Turkey
Ortiz, Andrés	University of Malaga	Spain
Osborne, King	University of Central Florida	USA
Otamendi, Javier	Rey Juan Carlos University	Spain
Pandey, Santosh K.	Institute of Chartered Accountants of India	India
Pantelelis, Nikos G.	National Technical University of Athens	Greece
Park, Se Hyun	Chung-Ang University	South Korea
Parrilla Roure, Luís	University of Granada	Spain
Passos, Nelson Luiz	Midwestern State University	USA
Patel, Susmit H.	MITRE Corporation	USA
Pavone, Mario	University of Catania	Italy
Pisarchik, Alexander	Center for Research in Optics	Mexico
Plakitsi, Katerina	University of Ioannina	Greece
Poh, Elsa	Eastern Michigan University	USA
Pokrywka, Rafal	IBM SWG Laboratory	Poland
Prakash, Naveen	Jaypee University of Information Technology	India
Priest, John	University of Texas Arlington	USA
Pruski, Alain	Paul Verlaine University – Metz	France
Rahnama, Ali	Amirkabir University of Technology	Iran
Resta, Marina	University of Genova	Italy
Rizki, Mateen	Wright State University	USA
Romanov, Sergey	Russian Academy of Sciences	Russian Federation
Rot, Artur	Wroclaw University of Economics	Poland
Rotondi, Guido	Italian National Statistical Institute	Italy
Rusconi, Andrea	Selex Galileo	Italy
Saitta, Francesco	University of Palermo	Italy
Saleemi, Muhammad M.	Åbo Akademi University	Finland
Sarukhanyan, Hakob G.	Institute for Informatics and Automation Problems of NAS	Armenia
Schreurs, Jeanne	Hasselt University	Belgium
Segovia, Hugo	Central University of Venezuela	Venezuela
Selcuk, Artut	Sabancı University	Turkey
Shareef, Ali	University of Maine	USA
Shing, Chen-Chi	Radford University	USA
Silva, Geraldo	Estadual Paulista University	Brazil
Simeonov, Stanislav	Burgas Free University	Bulgaria
Simurina, Jurica	University of Zagreb	Croatia
Singh, Vijander	Netaji Subhas Institute of Technology	India
Smith Nash, Susan	University of Oklahoma	USA
Snow, Richard	Embry-Riddle Aeronautical University	USA
Sokolov, Sergey	Keldysh Institute for Applied Mathematics	Russian Federation
Somkantha, Krit	Rajabhat University Udonthani Thailand	Thailand
Sousa, António	Institute of Biomedical Engineering	Portugal
Stanchev, Peter	Kettering University	USA
Stasytyti, Viktorija	Vilnius Gediminas Technical University	Lithuania
Štork, Milan	University of West Bohemia	Czech Republic
Sutherland, Trudy	Vaal University of Technology	South Africa
Szabo, Raisa	Nova Southeastern University	USA
Talla, Malleswara	Concordia University	Canada
Talukdar, Fazal Ahmed	National Institute of Technology Silchar	India
Tanaka, Hirokazu	Toshiba Corporation	Japan

Tenreiro Machado, J. A.	Institute of Engineering of Porto	Portugal
Thissen, M. Rita	RTI International	USA
Thurasamy, Ramayah	Science University of Malaysia	Malaysia
Tickle, Andrew	Coventry University	UK
Tico, Marius	Nokia Research Center	USA
Trobec, Roman	Jozef Stefan Institute	Slovenia
Tsai, Ping-Sing	The University of Texas - Pan American	USA
Uddin, Vali	Pakistan Naval Engineering College	Pakistan
Vaganova, Natalia A.	Russian Academy of Sciences	Russian Federation
Varriale, Luisa	Parthenope University of Naples	Italy
Vasilache, Simona	University of Tsukuba	Japan
Vázquez, Ernesto	Autonomous University of Nuevo Leon	Mexico
Vdović, Roberto	University of Zagreb	Croatia
Verber, Domen	University of Maribor	Solomon Islands
Vint, Larry	Griffith University	Australia
Volf, Jaromir	Czech Technical University in Prague	Czech Republic
Wada, Shigeo	Tokyo Denki University	Japan
Walters, Sharon	Southern Illinois University	USA
Wang, Chian	National Changhua University of Education	Taiwan
Wang, Fuchung	National Chengchi University	Taiwan
Wang, Xiao Hui	University of Pittsburgh	USA
Wang, Xihuai	Shanghai Maritime University	China
Wang, Yi-Fan	National Taipei College of Business	Taiwan
Wen, Feng	Southwest Jiaotong University	China
Wesley, Joan	Jackson State University	USA
Wicker, Dorothy	Walden University	USA
Wilson, Hentie	University of South Africa	South Africa
Wiriyasuttiwong, W.	Srinakharinwirot University	Thailand
Xie, Xianghui	Chinese Academy of Sciences	China
Yahiaoui, Azzedine	Eindhoven University of Technology	Netherlands
Yang, Jingyun	Harvard Medical School	USA
Yildirim, Yakup	Bilkent University	Turkey
Yin, Jianwen	Dell Computer Corporation	USA
Yossakda, Nipa	Northwestern Polytechnic University	USA
Yourguelenas, Youri V.	General Physics Institute Ras	Russian Federation
Zaman, Saiful	Sultan Qaboos University	Oman
Zhang, Xiao-Dong	Xi'an Jiaotong University	China
Zhang, Yanlong	Manchester Metropolitan University	UK
Zhong, Cheng	Guangxi University	China
Zolotova, Iveta	Technical University Kosice	Slovakia
Zorzo, Sérgio Donizetti	Federal University of São Carlos	Brazil
Zuzeviciute, Vaiva	Vytautas Magnus University	Lithuania



## ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

Abdulah Zadeh, Abdulah	Shahid Beheshti University	Iran
Alexik, Mikulas	University of Zilina	Slovakia
Alsmadi, Izzat	Yarmouk University	Jordan
Álvarez González, Ricardo	Benemérita Universidad Autónoma de Puebla	Mexico
Al-Yasiri, Adil	University of Salford	UK
Anderson, Janice	University of North Carolina at Chapel Hill	USA
Anita, Mary	Anna University	India
Antidze, Jemal	I.Vekua Scientific Institute of applied mathematics	Georgia
Arakaki, Reinaldo Gen I.	Faculdade de Tecnologia de São José dos Campos	Brazil
Ashoori, Maryam	University of Ottawa	Canada
Avgerou, Chrisanthi	London School of Economics and Political Science	UK
Barb, Adrian	Penn State University	USA
Basu, Aryabrata	University of Georgia	USA
Becker, Matthias	University of Flensburg	Germany
Berghaus, Martin	TU Dortmund	Germany
Berleant, Daniel	University of Arkansas at Little Rock	USA
Bethke, Albert	RTI International	USA
Bi, Zhuming	Indiana University	USA
Bönke, Dietmar	Reutlingen University	Germany
Bubnov, Alexey	Academy of Sciences of the Czech Republic	Czech Republic
Cardenas A., Diana M.	Universidad Nacional de Colombia	Colombia
Chae, Yeon Sik	MINT	South Korea
Chandra, Dr Pravin	GGs IPU Delhi	India
Chang, Chih-hung	Hsiuping Institute of Technology	Taiwan
Chang, Weng-long	National Kaohsiung University of Applied Sciences	Taiwan
Chemak, Chokri	University for Information Science and Technology	France
Chen, Chun-hsien	Nanyang Technological University	Singapore
Chen, Ke	School of Computer Science- The University of Manchester	UK
Chen, Mo	State University of New York at Binghamton	USA
Chen, Thomas	Swansea University	USA
Choi, Ho Yong	Chungbuk National University	South Korea
Choi, Tae-yong	Korea Institute of Machinery and Materials	South Korea
Chu, Chih-Ping	National Cheng Kung University	Taiwan
Chu, Xiumin	Intelligent Transport System Research Center	China
Contreras, Leonardo	Universidad Simón Bolívar	Venezuela
Dacey, Simon	Unitec New Zealand	New Zealand
Daneshmand M., Amin	Malayer Azad University	Iran
De, Dr. Sadhan K.	Indian Institute of Technology Kharagpur	India
De Groot, Peter	University of Southampton	UK
Deshmukh, S.G.	ABV Indian Institute of Information Technology and	India

Durai Raj, K. Antony A.	Management Gwalior	India
Dvorak, Zdenek	Infosys Technologies Limited	Slovakia
Dvornik, Josko	Univerzity of Zilina	Croatia
Ekstrom, Joseph	University of Split	USA
El Nasan, Adnan	Brigham Young University	USA
El Oualkadi, Ahmed	American University in Dubai	UAE
Elmahboub, Widad	Abdelmalek Essaadi University	Morocco
Erdogan, Eylem	Hampton University	USA
Essam A., Ahmed	Kadir Has University	Turkey
Estrela, Vania	Sohag University	Egypt
Ezekiel, Soundararjan	State University of Rio de Janeiro	Brazil
Faruquzzaman B., A.	Indiana University of Pennsylvania	USA
Ferreira, Andrea	Bangladesh University of Engineering and Technology	Bangladesh
Figueiredo, Josiel	Federal University of Ceara	Brazil
Forcolin, Margherita	UFMT	Brazil
Franz, Heike	Insiel S.P.A.	Italy
Garrett, Aaron	Electronic Data Systems Corporation	Germany
Gerber, Mariana	Jacksonville State University	USA
Gini, Giuseppina	Nelson Mandela Metropolitan University	South Africa
Goron, Juan	Polytechnic of Milan	Italy
Gupta, M.P	California School of Mines	USA
Hagemeyer-Klose, Maria	Indian Institute of Technology Delhi	India
Ham, Chan	Technical University of Munich	Germany
Hamie, Ali	Southern Polytechnic State University	USA
Haro Aguilar, Gabriel	University of Brighton	UK
Hrubes, Pavel	Colegio de Postgraduados en Ciencias Agrícolas	Mexico
Hu, Dali	Czech Technical University Praha	Czech Republic
Hu, Vincent	Jiangxi University of Finance and Economics	China
Huang, Shih-kun	National Institute of Standards and Technology	USA
Huang, Yi-hung	National Chiao Tung University	Taiwan
Hussein Ali, Samaher	National Taichung University	Taiwan
Hwang, Kofo	Babylon University	Iraq
Ibrahim, Azzam	Overseas Chinese University	Taiwan
Ilkova, Tatiana	Hamad Medical Corporation	USA
Imada, Akira	Bulgarian Academy of Sciences	Bulgaria
Janota, Ales	Brest State Technical University	Belarus
Jiang, Yuantao	University of Zilina	Slovakia
Johannesson, Paul	Shanghai Maritime University	China
Jouis, Christophe	SYSLAB	Sweden
Juan, Ángel A.	University Paris 3	France
Judd, Dave	Open University of Catalonia	Spain
Jung, Kyung Kwon	MDE/OEAA	USA
Karatoy, Hilal	Dongguk University	South Korea
Katar Srinivas	Technische Universiteit Eindhoven	Netherlands
Kazi, Hameedullah	Atheros Communications- Ocala- Florida	USA
Keesler, Venessa	Isra University	Pakistan
Kelemen, Miroslav	MDE/OEAA	USA
Kim, Hyungseok	Armed Forces Academy	Czech Republic
Kritzinger, Elmarie	Chonbuk National University	South Korea
Kumamaru, Sigehiro	University of South Africa	South Africa
	Hyogo University	Japan

Kumar, Mukesh	Delhi Public School - R.K Puram	India
Kurubacak, Gulsun	Anadolu University	Turkey
Kwon, Heyun-young	Kwangwon University	South Korea
Lee, Cheng-chi	Library and Information Science- Fu Jen Catholic University	Taiwan
Levesque, Mario	Memorial University in Newfoundland	Canada
Levian, Dina	D. Yellin College of Education	Israel
Li, Guoqiang	University of Missouri-St. Louis	USA
Liang, Jiaqi	Georgia Institute of Technology	USA
Libati, Hasting	Copperbelt University	Zambia
Lin, Yu-Ju	University of Florida	USA
Lisboa, Mauricio	CEPEL	Brazil
Liu, C. C.	Ching Yun University	Taiwan
Liu, Chao-liang	Asia University	Taiwan
Lodderman, Frederick	California School of Mines	USA
López Trujillo, Marcelo	Universidad de Caldas	Colombia
Luh, Guan Chun	Tatung University	Taiwan
Luhanga, Pearson	University of Botswana	Botswana
Lunt, Barry	Brigham Young University	USA
Lv, Nengchao	Wuhan University of Technology	China
Malzahn, Don	Wichita State University	USA
Mansoor, Wathiq	American University in Dubai	UAE
Mastacan, Lucian	Gheorghe Asachi Technical University	Romania
Matellan, Vicente	Universidad de Leon	Spain
Meladze, Hamlet	St. Andrew Georgian University	Georgia
Méndez Páez, Rafael	Pontificia Universidad Javeriana	Colombia
Mihci, Gurkan	Izmir Economy University	Turkey
Miro, Kraetzl	DSTO	Australia
Moeller, Knut	Furtwangen university	Germany
Morgan, Theresa	Wudang Research Association	USA
Morshed, Ahmed Hesham	Ain Shams University	Egypt
Mostafazadeh, Mehrdad	IIEES	Iran
Muis, Abdul	University of Indonesia	Indonesia
Muller, Logan	AIT	Thailand
Murch, Ron	University of Calgary	Canada
Murugesan, Karuppasamy	Atilin University	India
Myrick, Stephanie	Naval Research Laboratory	USA
Okatan, Ali	Halic University	Turkey
Oliveira, Hueder Paulo M.	Universidade Camilo Castelo Branco	Brazil
Olson, Patrick	National University	USA
Oreku, George	Tanzania Industrial Research	Tanzania
Otahal, Stanislav	Charles University	Czech Republic
Pandey, Abhishek	IIT Allahabad	India
Park, Cheol Hoon	Korea Institute of Machinery and Materials	South Korea
Pascu, Anca	Universite de Brest	France
Passos, Nelson Luiz	Midwestern State University	USA
Paterson, Matthew	University of Ottawa	Canada
Peng, Ya-Fu	Ching Yun University	Taiwan
Perunicic, Mihailo	Faculty of Technology	Yugoslavia
Petrella Toffolo, Carlos	Universidad Católica	Uruguay
Picariello, Antonio	Universita` di Napoli Federico II	Italy
Pietsch, Joachim	IADT	Ireland

Plotnick, Linda	Jacksonville State University	USA
Pugliese, Andrea	University of Calabria	Italy
Puigt, Matthieu	Institute of Computer Science	Greece
Puthooran, Emjee	IIT Roorkee	India
Quirolgico, Stephen	NIST	USA
Rao, K Sreenivasa	IIT Kharagpur	India
Ren, Jianfeng	Qualcomm Inc.	USA
Rigueira, Glaucio	CEPEL	Brazil
Rike, Robert	Environmental Systems Research Institute - Inc.	USA
Rodionov, Alexey	Institute of Computational Mathematics and Mathematical Geophysics	Russian Federation
Rodríguez de la O, Jose L.	Universidad Autonoma Chapingo	Mexico
Roe, Marvin	Naval Research Laboratory	USA
Ruiz-Vanoye, Jorge A.	National Center for Research and Technological Development	Mexico
Rusconi, Andrea	Selex Galileo	Italy
Schaeffer, Donna M.	Marymount University	USA
Sebastian, van Delden	University South Carolina Upstate	USA
Shafir, Abraham I.	Giv'at Washington Academic College of Education	Israel
Shaya, Fadia	University of Maryland School of Pharmacy	USA
Shen, Junan	Georgia Southern University	USA
Shi, Lei	University of Maryland, Baltimore County	USA
Shufeng, Wang	Shandong University of Technology	China
Singhal, Anoop	NIST	USA
Smith, David	Indiana University of Pennsylvania	USA
Sokolov, Sergey	Keldysh Institute for Applied Mathematics	Russian Federation
Song, Chul Gyu	Chonbuk National University	South Korea
Starek, Tomas	Telematix Services S. A.	Czech Republic
Sushil, Sushil	IIT Delhi	India
Tarcsi, Ádám	Eotvos Lorand University	Hungary
Thiel, Klaus	MES Consult	Germany
Trifas, Monica	Jacksonville State University	USA
Tsai, Ming-Tsung	Southern Taiwan University	Taiwan
Tsiligaridis, John	The State University of New York at Buffalo	USA
Varma, Shirshu	Indian Institute of Information Technology Allahabad	India
Vollmer, Thomas	University of Hamburg	Germany
Wang, John	Montclair State University	USA
Wang, Qun	University of Geoscience	China
Wilson, James	James Madison University	USA
Xiao, Yan	Baylor Health	USA
Yang, Ming	Montclair State University	USA
Yeon, Seung Jun	Electronics and Telecommunications Research Institute	South Korea
Yili, Fu	Harbin Institute of Technology	China
Yngstrom, Louise	Stockholm University	Sweden
Yoshikawa, Takeshi	Hokkaido University	Japan
Yuan, Xiang	Shanghai Maritime University	China
Zalama, Eduardo	Universidad de Valladolid	Spain
Zhan, Jing	Chinese University of Hong Kong	Hong Kong
Zhu, Zhenming	Chinese Academy of Science	China
Zhuming, Bi	Indiana University / Purdue University	USA
Zinner Henriksen, Helle	Copenhagen Business School	Denmark



**HONORARY PRESIDENTS OF PAST CONFERENCES**

Bela Banathy  
Stafford Beer  
George Klir  
Karl Pribram  
Paul A. Jensen  
Gheorghe Benga

**HONORARY CHAIR**

William Lesso

**PROGRAM COMMITTEE CHAIRS**

C. Dale Zinn  
José Ferrer

**GENERAL CHAIR**

Nagib Callaos

**ORGANIZING COMMITTEE CHAIRS**

Belkis Sánchez  
Andrés Tremante

**CONFERENCE PROGRAM MANAGER /  
HARDCOPY PROCEEDINGS PRODUCTION CHAIR**

María Sánchez

**TECHNICAL CONSULTANT ON COMPUTING SYSTEM /  
CD PROCEEDINGS PRODUCTION CHAIR**

Juan Manuel Pineda

**META REVIEWERS SUPPORT**

Maria Sánchez  
Dalia Sánchez

**SYSTEM DEVELOPMENT, MAINTENANCE AND DEPLOYMENT**

Dalia Sánchez  
Keyla Guédez  
Nidimar Diaz

**OPERATIONAL ASSISTANTS**

Marcela Briceño  
Cindi Padilla

**HELP DESK**

Louis Barnes  
Sean Barnes  
Marisela Jiménez



**Number of Papers Included in these Proceedings per Country**  
(The country of the first author was the one taken into account for these statistics)

<b>Country</b>	<b># Papers</b>	<b>%</b>
<b>TOTAL</b>	<b>193</b>	<b>100.00</b>
United States	37	19.17
Japan	11	5.70
South Korea	11	5.70
Latvia	10	5.18
Germany	9	4.66
China	8	4.15
Taiwan	8	4.15
Czech Republic	7	3.63
India	6	3.11
Mexico	6	3.11
Turkey	6	3.11
Australia	5	2.59
Brazil	5	2.59
Canada	5	2.59
Italy	5	2.59
Spain	5	2.59
United Kingdom	5	2.59
France	4	2.07
Lithuania	4	2.07
Croatia	3	1.55
Denmark	3	1.55
Egypt	3	1.55
Israel	3	1.55
Poland	3	1.55
Finland	2	1.04
Greece	2	1.04
Iran	2	1.04
Malaysia	2	1.04
Slovakia	2	1.04
Austria	1	0.52
Belgium	1	0.52
Colombia	1	0.52
Cyprus	1	0.52
Indonesia	1	0.52
Netherlands	1	0.52
Portugal	1	0.52
Russian Federation	1	0.52
South Africa	1	0.52
Switzerland	1	0.52
Vietnam	1	0.52



## Foreword

Our purpose in the 15<sup>th</sup> World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2011) is to provide, in these increasingly related areas, a multi-disciplinary forum, to foster interdisciplinary communication among the participants, and to support the sharing process of diverse perspectives of the same transdisciplinary concepts and principles.

Systemics, Cybernetics and Informatics (SCI) are being increasingly related to each other in almost every scientific discipline and human activity. Their common transdisciplinarity characterizes and communicates them, generating strong relations among them and with other disciplines. They work together to create a whole new way of thinking and practice. This phenomenon persuaded the Organizing Committee to structure WMSCI 2011 as a multi-conference where participants may focus on one area, or on one discipline, while allowing them the possibility of attending conferences from other areas or disciplines. This systemic approach stimulates cross-fertilization among different disciplines, inspiring scholars, originating new hypothesis, supporting production of innovations and generating analogies; which is, after all, one of the very basic principles of the systems' movement and a fundamental aim in cybernetics.

WMSCI 2011 was organized and sponsored by the International Institute of Informatics and Systemics (IIS), member of the International Federation for Systems Research (IFSR). IIS is an organization dedicated to contribute to the development of the Systems Approach, Cybernetics, and Informatics potential, using both: knowledge and experience, thinking and action, for the:

- a) identification of synergetic relationships among Systemics, Cybernetics and Informatics, and between them and society;
- b) promotion of contacts among the different academic areas, through the transdisciplinarity of the systems approach;
- c) identification and implementation of communication channels among the different professions;
- d) supply of communication links between the academic and professional worlds, as well as between them and the business world, both public and private, political and cultural;
- e) stimulus for the creation of integrative arrangements at different levels of society, as well as at the family and personal levels;
- f) promotion of transdisciplinary research, both on theoretical issues and on applications to concrete problems.

These IIS objectives have directed the organizational efforts of yearly WMSCI/ISAS conferences since 1995.

On behalf of the Organizing Committee, I extend our heartfelt thanks to:

1. the 496 members of the Program Committee from 69 countries;
2. the 821 additional reviewers, from 86 countries, for their **double-blind peer reviews**;
3. the 529 reviewers, from 68 countries, for their efforts in making the **non-blind peer reviews**. (Some reviewers supported both: non-blind and double-blind reviewing for different submissions).

A total of 2461 reviews made by 1350 reviewers (who made at least one review) contributed to the quality achieved in WMSCI 2011. This means an average of 6.29 reviews per submission (391 submissions were received). *Each registered author had access, via the conference web site, to the reviews that recommended the acceptance of their respective submissions.* Each registered author could also get information about: 1) the average of the reviewers evaluations according to 8 criteria, and the average of a global evaluation of his/her submission; and 2) the comments and the constructive feedback made by the reviewers, who recommended the acceptance of his/her submission, so the author would be able to improve the final version of the paper.

In the organizational process of WMSCI 2011, about 391 papers/abstracts were submitted. These pre-conference proceedings include about 193 papers that were accepted for presentation from 40 countries. I extend our thanks to the invited sessions' organizers for collecting, reviewing, and selecting the papers that will be presented in their respective sessions. The submissions were reviewed as carefully as time permitted; it is expected that most of them will appear in a more polished and complete form in scientific journals.

This information about WMSCI 2011 is summarized in the following table, along with the other collocated conferences:

Conference	# of submissions received	# of reviewers that made at least one review	# of reviews made	Average of reviews per reviewer	Average of reviews per submission	# of papers included in the proceedings	% of submissions included in the proceedings
<b>WMSCI 2011</b>	<b>391</b>	<b>1350</b>	<b>2461</b>	<b>1.82</b>	<b>6.29</b>	<b>193</b>	<b>49.36%</b>
<b>IMETI 2011</b>	<b>212</b>	<b>679</b>	<b>1431</b>	<b>2.11</b>	<b>6.75</b>	<b>88</b>	<b>41.51%</b>
<b>IMSCI 2011</b>	<b>276</b>	<b>856</b>	<b>2104</b>	<b>2.46</b>	<b>7.62</b>	<b>124</b>	<b>44.93%</b>
<b>CISCI 2011</b>	<b>388</b>	<b>973</b>	<b>2359</b>	<b>2.42</b>	<b>6.08</b>	<b>173</b>	<b>44.59%</b>
<b>TOTAL</b>	<b>1267</b>	<b>3858</b>	<b>8355</b>	<b>2.17</b>	<b>6.59</b>	<b>578</b>	<b>45.62%</b>

We also extend our gratitude to the focus symposia organizers, as well as to the co-editors of these proceedings, for the hard work, energy and eagerness they displayed preparing their respective sessions. We express our intense gratitude to Professor William Lesso for his wise and opportune tutoring, for his eternal energy, integrity, and continuous support and advice, as the Program Committee Chair of past conferences, and as Honorary President of WMSCI 2011, as well as for being a very caring old friend and intellectual father to many of us. We also extend our gratitude to Professor Belkis Sanchez, who brilliantly managed the organizing process.

Our gratitude to Professors Bela H. Banathy, Stafford Beer, George Klir, Karl Pribram, Paul A. Jensen, and Gheorghe Benga who dignified our past WMSCI conferences by being their Honorary Presidents. Special thanks to Dr. C. Dale Zinn and Professor Jorge Baralt for co-chairing WMSCI 2011 Program Committee and to professors Andrés Tremante and Belkis Sánchez for co-chairing the Organizing Committee. We also extend our gratitude to Drs., Louis H. Kauffman, Leonid Perlovsky, Stuart A. Umpleby, Thomas Marlowe, Ranulph Glanville, Karl H. Müller, Shigehiro Hashimoto, T. Grandon Gill, Alec Yasinsac, Marta White Szabo, Jeremy Horne, Mario Norbis, Ham Chan, Felix Soto-Toro, Susu Nousala, and Dipl.-Math Norbert Jastroch, for accepting to address the audience of the General Joint Plenary Sessions with keynote conferences.

Many thanks to Drs. Dale Zinn, Jorge Baralt, Hsing-Wei Chu, Andrés Tremante, Friedrich Welsch, Thierry Lefevre, José Vicente Carrasquero, Angel Oropeza, and Freddy Malpica for chairing and supporting the organization of the focus symposia and conferences in the context of, or collocated with, WMSCI 2011. We also wish to thank all the authors for the quality of their papers.

We extend our gratitude as well to Maria Sanchez, Juan Manuel Pineda, Leonisol Callaos, Dalia Sánchez, Keyla Guedez, Nidimar Díaz, Marcela Briceño, Cindi Padilla Louis Barnes, Sean Barnes, Marisela Jiménez, Noraima Castellano, Abrahan Marin, and Freddy Callaos for their knowledgeable effort in supporting the organizational process producing the hard copy and CD versions of the proceedings, developing and maintaining the software supporting the interactions of the authors with the reviewing process and the Organizing Committee, as well as for their support in the help desk and in the promotional process.

Professor Nagib C. Callaos  
*WMSCI 2011 General Chair*



**VOLUME I**

**CONTENTS**

<b>Contents</b>	i
<b>Computational Models of Applied Mathematics and Theoretical Physics - Invited Session Organizer: Salah Doma (Egypt)</b>	
Ahmed, Shaimaa B. (Egypt): "Relationship Quality as Predictor of B2B Customer Loyalty"	1
Doma, S. B. (Egypt): "Single-Particle Schroedinger Fluid and the Axially Deformed Even-Even Nuclei"	7
Saleh, M. M.; Elgarawany, H. (Egypt): "Sensory Activities Proposed for the Integration of Learning for Kindergartners Cerebral"	13
<b>Cooperative Analysis and Design of Multidisciplinary Systems - Invited Session Organizer: Chan Ham (USA)</b>	
Ham, Chan *; Burton, Matthew *; Lin, Kurt *; Joo, Young Hoon ** (* USA, ** South Korea): "Development of a Passive Magnetic Bearing System for a Flywheel Energy Storage"	19
Ko, Won *; Lee, Jong-Hyun *; Ham, Chan ** (* South Korea, ** USA): "A Grid-Connected Desalination Plant Operation"	24
Lin, Kuo-Chi *; Joo, Young Hoon **; Helkin, Steven *; Ham, Chan * (* USA, ** South Korea): "A Novel Design of Wave Energy Harvest Device with Flywheel Energy Storage System"	28
Park, Cheol Hoon; Choi, Sang-Kyu; Ham, Sang Yong (South Korea): "Rotor Dynamic Analysis and Experiment of 5kWh Class Flywheel Energy Storage System"	34
Park, Cheol Hoon; Son, Young Su; Ham, Sang Yong; Lee, Sung Whee; Lim, Hyung Chul (South Korea): "Controller Design and Simulation for Tracking Mount of Movable SLR, ARGO-M"	40
Pfisterer, Michael; Schillo, Kevin; Valle, Christopher; Lin, Kuo-Chi; Ham, Chan (USA): "The Development of a Propellantless Space Debris Mitigation Drag Sail for LEO Satellites"	44
Soto Toro, Felix A.; Vu, Bruce T.; Hamilton, Mark S. (USA): "Qualification of Electrical Ground Support Equipment for New Space Programs"	49
Vázquez Ramos, Gabriel; Yuan, Jiann-Shiun (USA): "Development of a Novel Wireless Electric Power Transfer System for Space Applications"	53

## **Towards Business Sophistication and Sustainability - Invited Session**

**Organizer: Natalja Lace and Elina Gaile-Sarkane (Latvia)**

Bistrova, Julia; Lace, Natalja (Latvia): "Evaluation of Corporate Governance Influence on Stock Performance of CEE Companies"	59
Ciemleja, Guna; Lace, Natalja (Latvia): "The Model of Sustainable Performing of SMEs in Context of Company's Life Cycle"	65
Eriņš, Ingars; Taraškevičs, Ronalds (Latvia): "Dynamic Financing Models of Tertiary Education According to Market Needs"	71
Gaile-Sarkane, Elīna; Ščeulovs, Deniss (Latvia): "Information Technologies for Sustainable Business Development: Case of Latvia"	76
Gaile-Sarkane, Elina; Andersone, Ieva (Latvia): "Modeling of Consumer Behavior for Business Sophistication"	82
Ginevičius, Romualdas; Podvezko, Valentinas; Ginevičius, Adomas (Lithuania): "Determining the Significance of the Criteria Describing Enterprise Marketing"	88
Kocmanova, Alena; Docekalova, Marie; Nemecek, Petr; Simberova, Iveta (Czech Republic): "Sustainability: Environmental, Social and Corporate Governance Performance in Czech SMEs"	94
Kozlovskis, Konstantins; Lacis, Jurijs (Latvia): "Portfolio Management Modified by GARCH-Type Models and Moving Average Correlation in the Global Currency Market"	100
Lapina, Inga; Aramina, Daira (Latvia): "Changing of Topicality of Human Competencies within Companies' Life Cycle"	106
Oganisjana, Karine; Koke, Tatjana (Latvia): "Nurturing Opportunity Identification for Business Sophistication in a Cross-Disciplinary Study Environment"	112
Rivza, Sandija; Pilvere, Irina; Rivza, Peteris; Rivza, Baiba (Latvia): "Conceptual Models of Risk Assessment in Renewable Energy Production in Latvia"	118
Rutkauskas, Aleksandras Vytautas; Lapinskaitė-Vvohlfahrt, Indrė; Stasytytė, Viktorija; Uturytė-Vrubliauskienė, Laura; Michnevič, Edvard (Lithuania): "Marketing Portfolio as a Media for Marketing Assets Interaction to Optimize Marketing Efficiency Measured with Adequate Metric"	123
Rutkauskas, Aleksandras Vytautas; Stasytytė, Viktorija (Lithuania): "Markowitz Random Field as a Stand for Investment Analysis and Decision Making"	131
Stankeviciene, Jelena (Lithuania): "Multi-Attribute Value Measurement Approach to Assess Business Sustainability"	137
Titko, Jelena; Kozlovskis, Konstantins (Latvia): "Development of Bank Value Model"	143

## **Communication and Network Systems, Technologies and Applications**

Atac, Goktug; Dag, Tamer (Turkey): "An Energy Efficient Routing Algorithm (X-Centric Routing) for Sensor Networks" 149

Kupwade Patil, Harsh; Camp, Joseph; Szygenda, Stephen A. (USA): "Identity Based Authentication Using a Cross Layer Design Approach in Wireless Sensor Networks" 154

Zelinka, Tomas; Lokaj, Zdenek (Czech Republic): "Data Security in Transportation Solutions" 160

## **Computer Science and Engineering**

Chandramouli, Ramaswamy (USA): "Cloud Service Feature Driven Security Policies for Virtualized Infrastructures" 166

Fries, Terrence P. (USA): "Network Intrusion Detection Using an Evolutionary Fuzzy Rule-Based System" 172

Higashiyama, Yoichi \*; Rumchev, Ventsi \*\* (\* Japan, \*\* Australia): "Number of Product Terms in Reliability Formula of Weighted-k-out-of-n System by SDP Method" 178

Horne, Jeremy (Mexico): "A New Three Dimensional Bivalent Hypercube Description, Analysis, and Prospects for Research (Invited Paper)" 184

Park, Yong Kuk; Lee, Min Goo; Jung, Kyung Kwon; Yoo, June Jae; Kim, Hyeong-Seok (South Korea): "Optimum Node Deployment of the Wireless Sensor Network System" 196

Razavi, Seyed Naser \*; Gaud, Nicolas \*\*; Koukam, Abderrafiâa \*\*; Mozayani, Naser \* (\* Iran, \*\* France): "Using Motion Levels of Detail in the Fast Multipole Method for Simulation of Large Particle Systems" 200

## **Concepts, Principles, Methodologies and Applications of Cybernetics**

Lu, Shen; Segall, Richard S. (USA): "Statistical Quality Control of Microarray Gene Expression Data" 206

## **Control Systems, Technologies and Applications**

Kamenicky, Jan; Cermakova, Hana; Soušek, Radovan; Němec, Vladimír (Czech Republic): "Multiple Measurement of Physical Quantity from the Risk Evaluation Point of View" 212

López, Joaquín; Pérez, Diego; Gayoso, Miguel; Paz, Enrique (Spain): "Monitoring and Debugging Distributed Autonomous Systems Using Petri Nets" 216

Wu, Qing; Ma, Yulin (China): "Robust Following Maneuver Design in a Simulated Autonomous Platooning Control System" 222

Zhang, Yong-De; Chen, Hao; Zhao, Yan-Jiang; Wang, Jin-You (China): "Control and Experimental Study of Flexible Needle Insertion Mechanism" 228

### **Ethics, Cybernetics and Informatics**

Srinivasan, Satish Mahadevan \*; Tripathi, Abhishek \*; Srinivasan, Sambasivam \*\* (\* USA, \*\* India): "Should Internet be Governed or Not: Case of Google in China" 233

### **Human Information Systems**

Biskri, Ismaïl; Hillali, Hassane; Descoteaux, Steve; Rompré, Louis; Amar Bensaber, Boucif (Canada): "Assistance in Similarity Classes Interpretation" 238

Chase, Phillip H. (USA): "Visual Assessment Design System" 243

### **Image and Multidimensional Signal Processing**

Chen, Te-Yu; Ou, Hsia-Hung; Yang, Yong-Jay; Tseng, Jing-Hang (Taiwan): "An Information Hiding Technique Based on Visual Cryptography" 249

Ham, Woonchul; Kwon, Hyeokjae; Enkhbaatar, T. (South Korea): "Implementation of 3D Stereo Image Capture System Based on Multi-Segmented Method" 254

Kamejima, Kohji (Japan): "Simulated Saccade for Saliency Sensitive Pattern Capturing in Naturally Complex Scenes" 260

Nachtegael, Mike (Belgium): "A Fuzzy Multistep Filter for Video Noise Removal" 266

Trifas, Monica; Straub, Jeremy (USA): "Super Resolution: A Database Driven Inference Approach" 272

Volf, Jaromír; Oliva, Jan; Vlcek, Josef (Czech Republic): "Sensor for the Tactile Information" 277

Zhu, Guopu \*; Zhang, Shuqun \*\* (\* China, \*\* USA): "Bidirectional Evolution of Morphological Level Set for Fast Image Segmentation" 281

### **Image, Acoustic, Speech and Signal Processing**

Artut, Selcuk; Güzererler, Ahmet (Turkey): "SonicFields: An Immersed Sonic Experience (Sonicfields.net)" 285

Chen, Meng; Luo, Dean; Wang, Lan (China): "Automatic Scoring in a Task of Retelling Stories for Language Learners" 289

Trenchard, Michael; Lohrenz, Maura; Gray, Tom (USA): "Color Compression Optimization Techniques for Localized Aeronautical Charts" 293

### **Information and Communication Technologies**

Balch, Robert; Parker, Andrew (USA): "An Oil and Gas Regulatory Data Mapping Portal" 299

Fathy El Dessouky, Naglaa (Canada): "The Virtual Workplace: A New Alternative Merging Technology and Policy for more Environmental, Economic and Social Sustainability" 303

Fernández, Rocío; Gil, Ignacio; Palacios, Daniel; Devece, Carlos (Spain): "Technology Platforms in Distance Learning: Functions, Characteristics and Selection Criteria for Use in higher Education" 309

Narasimhan, Sreenivasan; Latchman, Haniph (USA): "Introduction to the Border Gateway Protocol – Case Study Using GNS3" 315

Spöttl, Georg; Schulte, Sven (Germany): "Work Process Oriented Learning via Mobile Devices – Theoretical Basics and Examples for a (New) Didactical Approach" 321

### **Information Systems Development**

Nonaka, Hidetoshi (Japan): "Virtual 3D Manipulation Using Cutting Plane Lines" 327

Ostergaard, Soren Duus (Denmark): "Arctic Challenges – Developing the ICT Strategy for Greenland – a Case Study" 331

Zaretsky, Esther (Israel): "Professional Development of Teacher Educators through Developing their Computer Skills" 337

### **Information Systems, Technologies and Applications**

Huang, Youfang; Yu, Siqin; Huang, Shunquan (China): "On Specialized Third-Party Shipbuilding Logistics Companies in Shanghai" 343

Yeh, Chung-Wei; Wu, Kee-Rong (Taiwan): "Design of Customized Service Using Data Mining - Electronic Coupons" 349

Zhou, Lixia (China): "Study on Scheme for University Information System Integration Based on SOA" 355

### **Optical Systems, Technologies and Applications**

Fumizawa, Motoo; Ohkawa, Syuhei (Japan): "Exchange Flow Rate Measurement Technique in Density Different Gases" 359

### **Robotics**

de Faria, Elmo B.; Pereira, Mauricio F. L.; Bueno, Jesús F.; Silva, Carla M.; Carvalho, Sidney R. D.; Joazeiro, Sayuri A.; Santos, Wellington S. E. (Brazil): "Hardware System for Robot and Proprioceptive Algorithms" 365

Do, Hyun Min; Park, Chan Hun; Park, Dong Il; Kim, Byung In; Kyung, Jin Ho (South Korea): "Tracking Performance Improvement for a Parallel Robot Based on Synchronized Control" 371

Donado, A.; Castañeda-Camacho, J.; Reyes, F.; Mino, G.; Muñoz-Hernández, G. (Mexico): "On the Design of a Bluetooth Data Acquisition Card for the Control of Manipulator Robots" 376

## **Systemics**

Corea, Stephen (UK): "A Systems Approach for Disclosing the Dialectics of Emergent Technology-Based Organisational Performance along Axiological Dimensions" 381

Horne, Jeremy (Mexico): "Is Reality Digital or Analog? (Invited Plenary Keynote Address)" 385

**Authors Index** 395

# **Relationship Quality as Predictor of B2B Customer Loyalty**

**Shaimaa B. Ahmed**

Faculty of commerce, Business Administration Department, Alexandria University  
Email: Shaimaa\_ahmed24@yahoo.com

## **Abstract**

Relationship marketing has become extremely important recently due to the fierce competition in today's marketplace. Companies are required to build long-term profitable relationship with customers and to achieve customer loyalty. Also, switching behaviors frequently occur among most of targeted customers. Fewer studies, however, discuss the effects of relationship quality efforts on customer loyalty. Therefore, this study is aimed to investigate the impact of relationship quality on customer loyalty in B2B context in the Egyptian shipping services sector. Building on prior research, we propose relationship quality as a higher construct comprising trust, commitment and satisfaction. An analytical model is developed as a guideline to test the relationships between relationship quality dimensions and customer loyalty.

## **1. Introduction**

Today's marketplace is characterized by highly saturated markets, increased competition and a greater need to pay attention to customer needs. Matching the growing complexity of the business environment has led to an ever-more diversified and demanding customer base. Customer can easily defect to competitors who promise better offerings at lower prices [1].

The key success factor to survive in mature markets relies on sustaining long-term relationships with stakeholders. The challenge all marketers face today is in finding ways of increasing customer loyalty. Transforming indifferent customers into loyal ones and establishing a long term relationship with customers is critical for organizational success. [1].

Marketing attention shifted gradually but definitely from mutually independent transactions to loyalty based repeat purchases and cross-sell opportunities. Relationship quality is perceived as a tool from which customer loyalty can be secured and, as a result, the attainment of higher competitiveness and enhanced customer satisfaction can be achieved. Numerous factors that influence relationship marketing success, three factors consistently identified as important: trust [2], relationship commitment [3] and satisfaction [4].

There is little agreement among researchers as to which individual or composite relational effect best captures the key aspects of a relationship that most affect customer loyalty or performance [5]. For this reason, this study aims to empirically investigate the impact of relationship quality dimensions on customer loyalty.

## **2. Literature Review**

### **2.1 Relationship Quality**

Over the last two decades, Relationship Quality (RQ) has become one of the pillars of Relationship Marketing (RM) and gained in importance as a measure of successful business relationships [5].

There is no consensus on a definition of RQ. However, there is some degree of consensus that RQ is a higher-order construct comprising several different, though related dimensions [6].

Holmlund's definition is considered the one that best reflects the nature and domain of Relationship Quality: "Relationship quality is the cognitive evaluation of business interactions by key individuals in the dyad, comparatively with potential alternative interactions" [7].

Trust, commitment and satisfaction should be included as dimensions of relationship quality. Indeed, these three constructs, also referred to as the building blocks of RQ, are commonly

viewed as important results of good quality relationships.

**2.1.1 Trust**

Trust is defined as the ability and willingness to rely on the relationship manager’s integrity and behavior (i.e. trustworthiness) so that the long-term expectations of the buyer will be met [2].

Many researches have suggested that customers’ trust is a significant role in building long-term relationship and achieving customer loyalty.

**2.1.2 Commitment**

It is defined as an enduring desire to maintain a valued relationship”. In marketing-practice and research it is agreed that mutual commitment among partners in business relationships produces significant benefits for companies. Commitment is important to the study of relationships because it not only signals enduring stability at the conceptual level, but also serves as a reliable surrogate measure of long-term relationships at the operational and empirical level [3].

**2.1.3 Satisfaction**

It is defined as the assurance, perceived by the buyer, regarding the relationship manager’s future performance, given that past performance has been consistently satisfactory [8].

In relationship marketing literatures, customer satisfaction has also been thought to be a key performance indicator for evaluating the quality of a relationship between service provider and customers. In business-to-business research, several authors show that a link between satisfaction and loyalty exists.

**2.2 Customer Loyalty**

Customer loyalty is the most important goal of implementing relationship marketing activities. It is defined as a “deeply held commitment to rebury a preferred product or service consistently in the future, thereby causing repetitive same-brand or same

brand-set purchasing, despite situational influences and marketing efforts having the potential to cause switching behavior”. Loyal customers are less likely to switch to a competitor solely because of price, and they even make more purchases than non-loyal customers. Loyal customers are also considered to be the most important assets of a company. It is thus essential for vendors to keep loyal customers who will contribute long-term profit to the business organizations [9]. In general, customer loyalty is the final purpose that firms implement relationship marketing.

**3. Research Hypotheses**

Based on the evidence of past research on loyalty and relationship quality, it is possible to lay out the following research questions: Does relationship quality influence customer loyalty? Does the suggested relationship quality dimensions influence customer loyalty? In order to examine these relationships we developed the following research hypotheses:

H1: Relationship Quality has significant positive effect on customer loyalty.

This effect must be evidence through the following hypotheses:

H1 a) Trust is positively related to customer loyalty.

H1 b) Commitment is positively related to customer loyalty.

H1 c) Satisfaction is positively related to customer loyalty.

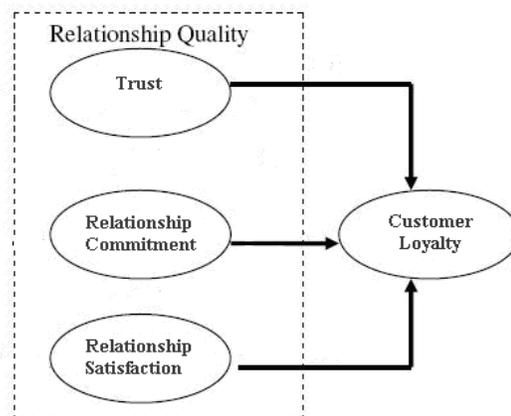


Figure 1 Conceptual Framework

## **4. Research Methodology**

### **4.1 Selection of sample**

This study is empirical and conducted in a business-to-business setting of the shipping and freight delivery service industry in Egypt. The survey is targeted at business customers; the study was carried out on a convenience sample of 500 respondents, through distribution of structured questionnaires to customers within Egypt. A total of 310 usable questionnaires were collected.

### **4.2 Data Analysis**

The statistical package SPSS (version 17.0) was used for data analysis. A two-step detailed statistical analysis of data was involved. First, Factor analysis was performed to extract the underlying factor of relationship quality; second, multiple regression analysis was performed to understand the relationship quality-customer loyalty relationship.

### **4.3 Scale and Measurement**

As discussed in the above section, we have identified relationship quality to be multidimensional construct consisting of three components: trust; commitment and satisfaction.

Each of the components should be measured reliably with a multi-item scale. To measure relationship quality and customer loyalty we used a five point likert scale (1 = strongly disagree” and 5 = strongly agree”) for the 14 statement/ items of the three dimensions of the scale.

We used existing measures whenever possible to develop items for each dimension of the relationship quality and customer loyalty. We developed initial items based on the work of [2].

We developed 14 items for relationship quality dimensions. Items were pre tested for relevance, face validity, interpretation and readability with marketing academics and company managers and company customer. Therefore some modifications and slight changes in wording were

required to fit the shipping services context.

### **4.4 Measure Reliability**

The first stage focused on the specification of dimensions for the verification of the 14 items scale, its reliability and validity, analyzing the elements and dimensions which determine the relationship quality construct in shipping services. Two-step procedure was followed. First, different analyses of the correlations between the initial scale items were carried out for this purpose, as well as examinations of scale reliability. The correlation matrix contained many high correlations.

The next step was an exploratory factor analysis to initially assess the psychometric properties of the scale. Factor analysis was used to allow the grouping of variables and therefore, simplify data for possible interpretation. The statistical purpose of factor analysis was to determine whether there were linear combinations of variables that will assist in summarizing the data and identifying underlying relationships. The scale items were purified through an iterative process. Items that did not load heavily on the primary factor and items that had significant cross loadings were removed. This resulted in a removal of three more items. The remaining fourteen items were factor analyzed again. The analysis produced a three-factor solution, which accounts for 59.25% of the variation in the data, according to the criteria developed by (Kaiser, 1958). As can be seen in Table 1, these factors are: trust, satisfaction and commitment. Table 1 provides the descriptive output for each factor.

A reliability analysis was conducted on the summated scale of these constructs using reliability coefficients. The internal consistency was assessed by means of the Cronbach’s alpha coefficient. Tables 1 report the reliability of the multi-item scale which using Cronbach’s coefficient alpha. The overall coefficient alpha for

the scale is 0.894 which is greater than the recommended cut-off level of 0.70. With regard to individual subscales, the reliability coefficient of all the three components: trust (0.791), satisfaction (0.770) and commitment (0.658) met the standard. Therefore, these indicated high reliability estimates since reliability figures less than 0.60 are generally considered to be poor, those in the range of 0.70 to be acceptable, while those above 0.80 to be good.

#### **4.5 Construct validity**

According to Cronbach (1970), evidence of construct validity exists when the pattern of correlation among variables conforms to what is predicted by theory. Therefore convergent and nomological validities are examined.

### **5. Results**

The data analysis was conducted in two steps. First, an exploratory factor analysis with varimax rotation was performed to determine the underlying dimensions of Relationship quality construct and testing of the measurement model for relationship quality construct in order to determine if the extracted dimensions in step 1 offered a good fit to the data. Finally, we examined the hypothesized relationships. These steps are discussed in the following subsections.

#### **5.1 Factor Analysis**

The three factors shown in Table 1 relate to elements of relationship quality in existing literature and collectively, these retained factors account for 59.25 percent of the total variance in the 14 variables. Based on the items loading on each factor, the factors were, respectively, labeled as trust, commitment and satisfaction.

##### **5.1.1 Factor 1 Trust**

The respondents felt that trust is a very important aspect of relationship marketing. This factor accounted for 19.7 % of variance and its reliability was 0.791.

##### **5.1.2 Factor 2 Commitment**

All the three items loading on this factor relate to different aspects of commitment dimension of relationship quality. The respondents felt that if an organization wanted build and maintain long-term relationship with the customers, it should cultivate mutual commitment. This factor accounted for 19.7 percent of variance and its reliability was 0.658.

##### **5.1.3 Factor 3 Satisfaction**

The respondents felt that if an organization wanted to build and maintain long-term relationship with the customers, it should cultivate customer satisfaction. This factor accounted for 19.9 % of variance and its reliability was 0.770.

### **5.2 Test of Hypotheses**

Hypothesis 1 postulates that relationship quality has significant positive effect on customer loyalty. In order to test this hypothesis, we conducted a multiple regression analysis using customer loyalty as the dependent variable, and the various components of relationship quality (trust, satisfaction and commitment) as the predicting variables. Thus, we can identify the relative importance of relationship quality components in explaining the variation in the customer loyalty levels. However, since the three components of relationship quality are closely related, multi-collinearity may exist. Therefore, to remove this impact, a two -stage analysis was conducted. First, the 14 items of the relationship marketing orientation were factor analyzed into three factors. Then the factor scores of the three dimensions were used as independent variables for the regression analysis. Table-2 presents the regression results of the variables; it shows that relationship quality has a significant positive association with customer loyalty where customer loyalty is the dependent variable. As can be seen from the table, many interesting findings emerge. First, the

regression model achieve a high degree of fit, as reflected by an  $R^2$  of 0.56 ( $F = 101.9$ ;  $p < 0.001$ ). Second, trust and satisfaction demonstrate a positive and significant association with customer loyalty. These are generally expected results. Third, commitment, a key construct identified by previous relationship marketing research is found to have significant effect on customer loyalty. Moreover, as can be seen from Table 2, the positive sign of all the three estimates (beta coefficients) shows that the greater the extent of these relationship quality variables, the higher the level of customer loyalty. Therefore, on the basis of the analysis made, it can be said that successful relationship marketing efforts of the Egyptian shipping companies can improve customer loyalty. Finally, the first hypothesis that there is a significant positive relationship between relationship quality variables and customer loyalty has been accepted.

## 6. Discussion

This study provides initial empirical evidence of the relationship quality on customer loyalty. The results of the regression analyses provide support for the hypothesized relationships. Consistent with  $H_1$ , that relationship quality has significant positive effect on customer loyalty.

There are several factors that management of shipping service providers can use to improve their specific service delivery process and to establish long-lasting relationships with their customers. Relationship quality is one of key strategic issues for managers of shipping companies for establishing and maintaining long-term relationships with their clients.

(1) To remain competitive and obtain competitive advantages, the company managers can try to increase relationship and thus customer loyalty by managing each dimension of relationship quality in the context with the Egyptian shipping companies' customers.

(2) The three dimensions of the relationship quality scale may serve training needs for company staff to develop appropriate training programs that can help to improve their understanding of the activities involved in developing relationship quality.

## References

- [1] Bhardwaj, Deepali. (2007). Relationship Marketing in Context to the IT Industry. *VISION-The Journal of Business Perspective*, 11, (2) (April – June), 57- 66.
- [2] Morgan, R. and Hunt, S. (1994), “The Commitment-Trust Theory of Relationship Marketing”, *Journal of Marketing*, 58, (3), 20-38.
- [3] Moorman, C., Zaltman, G. and Deshpande, R. (1992), “Relationships Between Providers and Users of Market Research: The Dynamics of Trust Within and Between Organizations”, *Journal of Marketing Research*, 29, (3), (August), pp. 314-28.
- [4] Sin, Y.M., C.B. Tse, H.M. Yau, R. Chow and S.Y. Lee. (2002). The Effect of Relationship Marketing Orientation on Business Performance in a Service-Oriented Economy. *Journal of Services Marketing*, 16 (7), 656–76.
- [5] Palmatier, Robert W., Dant, Rajiv P., Grewal, Dhruv & Evans, Kenneth R. (2006). Factors Influencing the Effectiveness of Relationship Marketing: A Meta-Analysis. *Journal of Marketing*, 70 (4) (October), 136–53.
- [6] Woo, K. and Ennew, C. T. (2004), “Business-to-Business Relationship Quality: An IMP Interaction-Based Conceptualisation and Measurement”, *European Journal of Marketing*, 38, (9/10), 1252-1271.
- [7] Holmlund, M. (2001), “The D&D Model: Dimensions and Domains of Relationship Quality Perceptions”, *The Service Industries Journal*, 1, (3), 13-36.
- [8] Parsons, A. (2002), “What Determines Buyer-Seller Relationship Quality? An Investigation from the

Buyer's Perspective", The Journal of Supply Chain Management, 38, (2), (Spring), 4-12.  
 [9] Tseng, Yi Ming (2007). The Impacts of Relationship Marketing

Tactics on Relationship Quality in Service Industry, The Business Review, Cambridge; Summer 2007, 7 (2), 310-314.

Table1: Result of exploratory factor analysis of relationship quality (Scale Reliability):

Construct	Factor Loadings	Variance explained	Reliability Cronbach Alpha
Factor1: Trust 1.Company has higher integrity 2.Work in close cooperation 3. Company has good will reputation 4.Company is trust worthy 5.Consistent in providing quality services 6. Company is reliable in providing unique services.	0.590 0.649 0.620 0.769 0.809 0.498	19.698	0.791
Factor 2: Commitment 7. Remain a client to enjoy the relationship. 8. Feel company is looking out for my interest. 9. I have strong sense of loyalty toward company.	0.800 0.689 0.601	19.650	0.658
Factor 3: Satisfaction 10. Company completely meets my expectations. 11. Company shows a sincere interest in solving complaints. 12. Employees work in a tidy professional way. 13. Satisfied with all services offered. 14. Company tries very hard to establish long term relationship.	0.598 0.560 0.792 0.698 0.602	19.902	0.770
Relationship Quality (Scale reliability)		59.25	0.894

Table 2: The effect of relationship quality (Trust, Commitment and satisfaction) on customer loyalty: estimated standardized regression coefficients

Independent Variable	Standardized B	t	Sig.
Trust	0.016	0.190	0.849
Commitment	0.007	0.177	0.792
Satisfaction	0.018	0.186	0.860

# Single-Particle Schroedinger Fluid and the Axially Deformed Even-Even Nuclei

S. B. Doma

Faculty of Information Technology and Computer Sciences, Sinai University, El-Arish, North Sinai, Egypt. E-mail address: sbdoma@su.edu.eg

## Abstract

The study of the axially-symmetric deformed nuclei is well explained in frame work of the single-particle Schrödinger fluid. The calculated values of the moments of inertia of such nuclei by using this concept are in good agreement with the corresponding experimental values. In this paper we apply the concept of the single-particle Schrödinger fluid to calculate the cranking model- and the rigid body model-moments of inertia of some even-even deformed nuclei in the sd-shell, assuming that they are axially symmetric: namely the nuclei  $^{20}\text{Ne}$ ,  $^{24}\text{Mg}$  and  $^{28}\text{Si}$ . The variations of these moments with respect to the deformation parameter  $\beta$  are also given. The obtained results are in good agreement with the corresponding experimental values.

## 1. Introduction

To explain the properties of the deformed nuclei, it is necessary not only to study the collective motions in them but also to consider one-particle states of the nucleons moving in non-spherical potential. In the adiabatic approximation, the one-particle and the collective motions in the nucleus can be separated, so that the total wave function of the nucleus can be represented in the form of a product of two functions, one of which describes the motion of the nucleus as a whole (that is the rotation of the nucleus and the vibrations of its surface) and the other describes the internal motion in a non-spherical nucleus [1].

In treating the internal motion in the nucleus, it is assumed that the individual nucleons move independently in a certain fixed non-spherical field of the nucleus. The Hamiltonian of the internal motion can then be represented, as in the ordinary model, in the form of a sum of one-particle Hamiltonians. Thus, the problem reduces to solving Schrödinger equation for one-particle potential which must be chosen in such a way that it reflects principal features of the motion of the nucleus in non-spherical

nuclei while having at the same time the simplest possible form. The energy levels of a non-spherical nucleus consist of two parts: the levels of the fixed nucleus and the energy of its rotation as a whole.

The single particle Schrödinger fluid [2,3,4] is a concept which is used to describe the motion of a single nucleon in an axially deformed potential of the nucleus. This concept is carried out by a suitable choice of the time- dependent part of the nucleon wave function in the time-dependent Schrödinger equation. This concept can be applied to study the rotational motion of a deformed nucleus.

In this paper we carry out the derivations of this concept and accordingly clarify how the moment of inertia of an axially deformed nucleus can be obtained in frame-work of this concept. As examples for the application of this concept to the calculations of the nuclear moments of inertia we have calculated the cranking-model and the rigid body-model moments of inertia of three of the even-even deformed nuclei in the s-d shell, by assuming that they are axially symmetrical; namely the nuclei  $^{20}\text{Ne}$ ,  $^{24}\text{Mg}$  and  $^{28}\text{Si}$ . The variations of the reciprocal moments of inertia of these nuclei

with respect to the deformation parameter  $\beta$  have been also given in this paper.

## 2. The Fluid Dynamical Equations

We assume that each nucleon (proton or neutron), with mass  $m$  in a nucleus, consisting of  $A$  nucleons, is moving in a single-particle potential  $V(\mathbf{r}, \alpha(t))$ , which is deformed with time  $t$ , through its parametric dependence on a classical shape variable  $\alpha(t)$ . Here,  $\alpha(t)$  is assumed to be an externally prescribed function of  $t$ . Thus, the Hamiltonian for the present problem is given by [2,3]

$$H(\mathbf{r}, \mathbf{v}, \alpha(t)) = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}, \alpha(t)). \quad (2.1)$$

The operator  $\nabla$  in equation (2.1) appeared due to the fact that in, Quantum Mechanics, the operator associated with the particle momentum  $m\mathbf{v}$  is given by  $-i\hbar\nabla$ , where  $\mathbf{v}$  is the particle velocity. The single-particle time-dependent wave function  $\Psi(\mathbf{r}, \alpha(t), t)$  which satisfies the time-dependent Schrödinger wave equation, that describes the motion of a nucleon, is defined as

$$H(\mathbf{r}, \mathbf{v}, \alpha(t))\Psi(\mathbf{r}, \alpha(t), t) = i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}, \alpha(t), t). \quad (2.2)$$

To obtain a fluid dynamical description of the wave function  $\Psi(\mathbf{r}, \alpha(t), t)$ , we use the polar form of the wave function. We first isolate the explicit time dependence in the form

$$\Psi(\mathbf{r}, \alpha(t), t) = \psi(\mathbf{r}, \alpha(t)) \exp \left\{ -\frac{i}{\hbar} \int_0^t \epsilon(\alpha(t')) dt' \right\}, \quad (2.3)$$

where  $\epsilon$  is the energy density which depends on the time through the parameter  $\alpha(t)$ . Then, we write the complex wave function  $\psi(\mathbf{r}, \alpha(t))$  in the following polar form

$$\psi(\mathbf{r}, \alpha(t)) = \Phi(\mathbf{r}, \alpha(t)) \exp \left\{ -\frac{iM}{\hbar} S(\mathbf{r}, \alpha(t)) \right\}, \quad (2.4)$$

where  $\Phi(\mathbf{r}, \alpha(t))$  and  $S(\mathbf{r}, \alpha(t))$  are assumed to be real functions of  $\mathbf{r}$  and  $\alpha(t)$ . Finally, we assume that the function  $\Phi(\mathbf{r}, \alpha(t))$  is positive definite. In the case of rotation, the parameter  $\alpha(t)$  becomes the angle of rotation,  $\theta = \Omega t$ , where  $\Omega$  is the angular velocity.

Substituting equations (2.1), (2.3) and (2.4) into (2.2) and carrying the necessary calculations we get

$$\left[ H - M \left( \frac{\partial S}{\partial t} - \frac{1}{2} \nabla S \cdot \nabla S \right) \right] \Phi = \epsilon \Phi, \quad (2.5)$$

$$\frac{1}{2} \Phi (\nabla^2 S) + (\nabla \Phi) \cdot (\nabla S) = \frac{\partial \Phi}{\partial t}. \quad (2.6)$$

We may call equation (2.5) modified Schrödinger equation because it differs from the usual time-independent Schrödinger equation  $H\Phi = \epsilon\Phi$  by an added term which we refer to as the “dynamical modification potential”

$$V_{dyn} = -M \left[ \frac{\partial S}{\partial t} - \frac{1}{2} (\nabla S) \cdot (\nabla S) \right]. \quad (2.7)$$

## 3. Interpretation of the Probability as a Fluid Continuity Equation

When we identify the probability density of the single-particle as the square of the amplitude  $|\Phi|^2$  and recognize that equation (2.6), when multiplied by  $2\Phi$ , gives

$$\Phi^2 \nabla^2 S + \nabla \Phi^2 \cdot \nabla S = \frac{\partial \Phi^2}{\partial t}, \quad (3.1)$$

then, we will obtain two equations, the first is

$$\rho \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla \rho = -\frac{\partial \rho}{\partial t}, \quad (3.2)$$

where  $\mathbf{v}$  is the irrotational velocity and  $\rho$  is the density. It is the well-known equation of continuity in fluid mechanics. It can be rewritten in the form:

$$\nabla \cdot (\rho \mathbf{v}) = -\frac{\partial \rho}{\partial t}, \quad (3.3)$$

where  $\rho = \Phi^2$  and  $\mathbf{v} = -\nabla S$ .

The second equation is

$$(H + V_{dyn})\Phi = \epsilon\Phi, \quad (3.4)$$

which is a modified Schrödinger equation with

$$V_{dyn} = -M \left( \frac{\partial S}{\partial t} - \frac{1}{2} \mathbf{v}^2 \right). \quad (3.5)$$

#### 4. The Collective Kinetic Energy for the Entire Nucleus

In the adiabatic approximation, where  $\frac{\partial \alpha}{\partial t} \rightarrow 0$ , the collective kinetic energy of a nucleon in the nucleus is given by [2,3]

$$T_K = \frac{1}{2} \int \rho \mathbf{v}_K \cdot (\boldsymbol{\Omega} \times \mathbf{r}) d\mathbf{r}. \quad (4.1)$$

and the collective kinetic energy  $T$  of the nucleus is given by

$$T = \frac{1}{2} M \int \rho_T \mathbf{v}_T \cdot (\boldsymbol{\Omega} \times \mathbf{r}) d\mathbf{r}, \quad (4.2)$$

where  $\rho_T$  is the total density distribution of the nucleus and  $\mathbf{v}_T$  is the total velocity field,

$$\mathbf{v}_T = \frac{\sum_{occ} \rho_K \mathbf{v}_K}{\sum_{occ} \rho_K}. \quad (4.3)$$

#### 5. Cranking Moments and Rigid-body Moments of Inertia

The following-expressions for the cranking moment of inertia, and the rigid moment of inertia  $\mathfrak{I}_{rig}$  as a function of  $q$  are given [3]:

$$\mathfrak{I}_{cr} = \frac{E}{\omega_0^2} \left( \frac{1}{6+2\sigma} \right) \left( \frac{1+\sigma}{1-\sigma} \right)^{\frac{1}{3}} \left[ \sigma^2 (1+q) + \frac{1}{\sigma} (1-q) \right], \quad (5.1)$$

$$\mathfrak{I}_{rig} = \frac{E}{\omega_0^2} \left( \frac{1}{6+2\sigma} \right) \left( \frac{1+\sigma}{1-\sigma} \right)^{\frac{1}{3}} \left[ (1+q) + \sigma(1-q) \right], \quad (5.2)$$

where  $E$  is the total nuclear ground-state energy

$$E = \sum_{occ} [\hbar \omega_x (n_x + n_y + 1) + \hbar \omega_z (n_z + 1)], \quad (5.3)$$

and  $q$  is the ratio of the summed single particle quanta in the y- and z-directions

$$q = \frac{\sum_{occ} (n_y + 1)}{\sum_{occ} (n_z + 1)}. \quad (5.4)$$

$q$  is known as the anisotropy of the configuration. In equations (5.1) and (5.2) the deformation of the potential,  $\sigma$  is defined by [2]

$$\sigma = \frac{(\omega_y - \omega_z)}{(\omega_y + \omega_z)}. \quad (5.5)$$

is a measure of the deformation of the potential.

The oscillator frequencies  $\omega_x$ ,  $\omega_y$  and  $\omega_z$  are expressed in terms of the deformation parameter  $\delta$  as follows [5]

$$\omega_z^2 = \omega_0^2 \left( 1 - \frac{4}{3} \delta \right), \quad (5.6)$$

$$\omega_x^2 = \omega_y^2 = \omega_0^2 \left( 1 + \frac{2}{3} \delta \right). \quad (5.7)$$

The condition of constant volume of the nucleus leads to

$$\omega_x \omega_y \omega_z = \text{const}. \quad (5.8)$$

Keeping this condition in the general case together with (5.6) and (5.7),  $\omega_0$  has to depend on  $\delta$  in the following way [5]

$$\omega_0(\delta) = \omega_0^0 \left\{ 1 - \frac{12}{9} \delta^2 - \frac{16}{27} \delta^3 \right\}^{-\frac{1}{6}}, \quad (5.9)$$

where  $\omega_0^0$  is the value of  $\omega_0(\delta)$  for  $\delta = 0$ . It turns out that  $\delta$  is related to the quantity  $\beta$  by

$$\delta = \frac{3}{2} \sqrt{\frac{5}{4\pi}} \beta. \quad (5.10)$$

The parameter  $\beta$  is allowed to vary in the range  $-0.5 \leq \beta \leq 0.5$ .

Further details concerning the single-particle Schrödinger fluid can found in [7,8].

The non deformed oscillator parameter  $\hbar\omega_0^0$  is expressed in terms of the number of protons  $Z$ , the number of neutrons  $N$  and the mass number  $A$  as follows [6]

$$\hbar\omega_0^0 = \frac{38.6A^{-\frac{1}{3}}}{\left[1 + \frac{1.646}{A} - \frac{0.191(N-Z)}{A}\right]^2}$$

## 6. Results and Conclusions

We have constructed the ground states of the three nuclei  $^{20}\text{Ne}$ ,  $^{24}\text{Mg}$  and  $^{28}\text{Si}$ , which are all even-even (number of protons  $Z$  is even and number of neutrons  $N$  is also even) deformed nuclei. Accordingly, the single particle states in each nucleus are filled with the corresponding wave functions. As a result,  $\mathfrak{I}_{cr}$  and  $\mathfrak{I}_{rig}$  are calculated for each nucleus. Finally the corresponding reciprocal moments  $\frac{\hbar^2}{2\mathfrak{I}_{crank}}$  and  $\frac{\hbar^2}{2\mathfrak{I}_{rigid}}$  are calculated. In Figures-1, -2 and -3 we present the calculated values of the reciprocal moments of inertia according to the cranking model and the rigid-body model (in KeV) as functions of the deformation parameter  $\beta$  for the nuclei  $^{20}\text{Ne}$ ,  $^{24}\text{Mg}$  and  $^{28}\text{Si}$ , respectively.

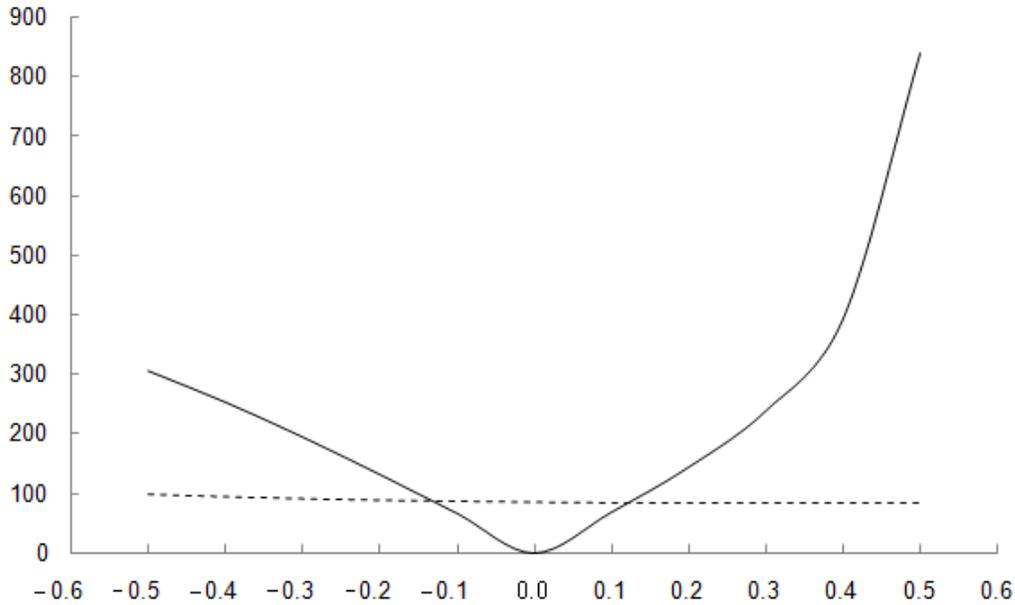


Figure-1 Reciprocal moment of inertia (in KeV) of the nucleus  $^{20}\text{Ne}$  as function of the deformation parameter  $\beta$ . The solid line corresponds to the cranking moment and the dashed line corresponds to the rigid-body moment.

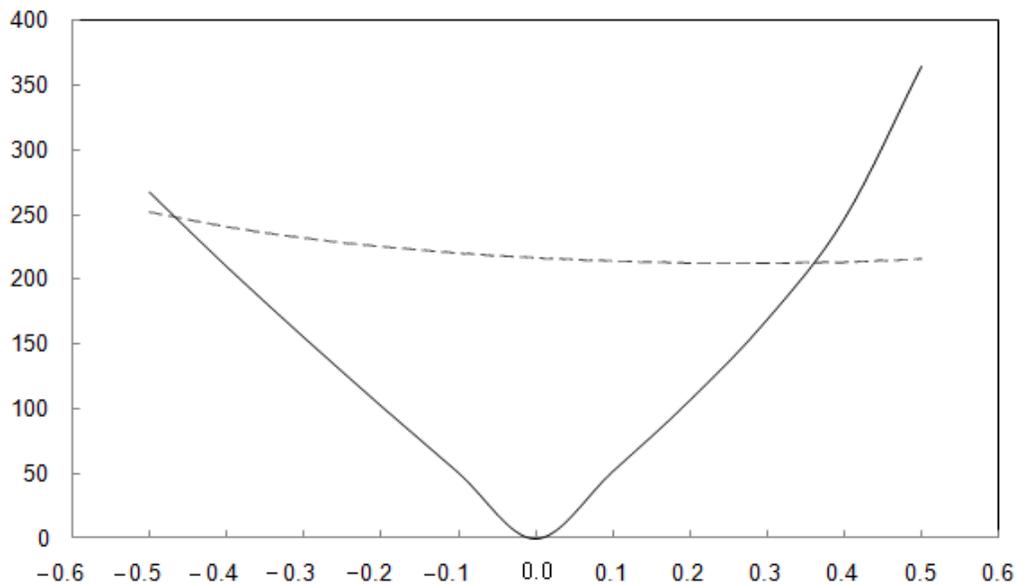


Figure-2 Reciprocal moment of inertia (in KeV) of the nucleus  $^{24}\text{Mg}$  as function of the deformation parameter  $\beta$ . The solid line corresponds to the cranking moment and the dashed line corresponds to the rigid-body moment.

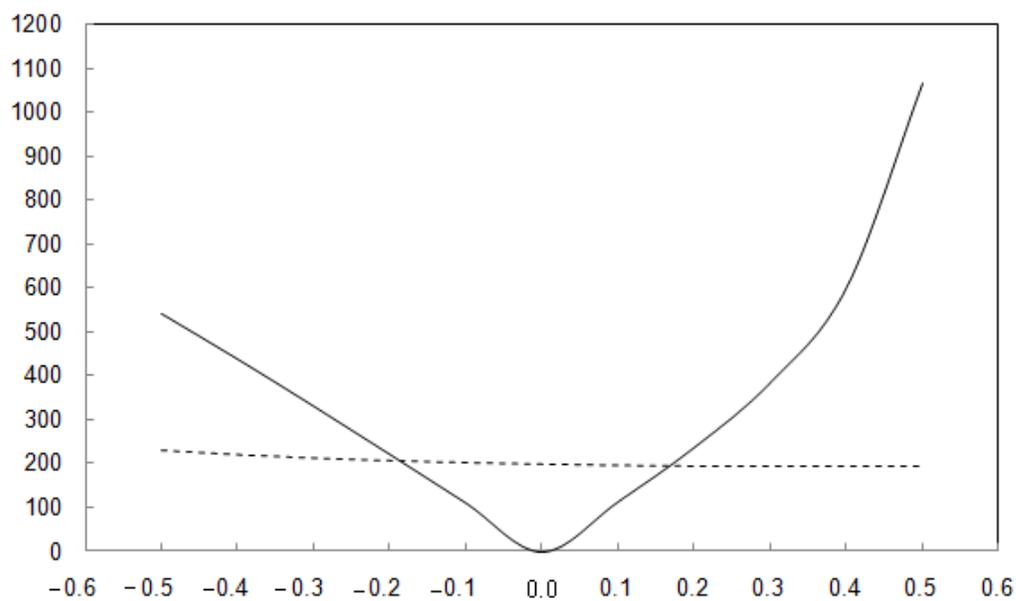


Figure-3 Reciprocal moment of inertia (in KeV) of the nucleus  $^{28}\text{Si}$  as function of the deformation parameter  $\beta$ . The solid line corresponds to the cranking moment and the dashed line corresponds to the rigid-body moment.

The corresponding experimental values [9] of the reciprocal moments of inertia of the three nuclei  $^{20}\text{Ne}$ ,  $^{24}\text{Mg}$  and  $^{28}\text{Si}$  are 279.90, 237.90 and 324.60 (KeV), respectively. The values of the deformation parameter  $\beta$  which produce the best values of the reciprocal cranking-model moments of inertia for the three nuclei are 0.22, 0.39 and 0.26, respectively. From the obtained results it is shown that the assumption that the three mentioned nuclei are axially symmetric is acceptable one and that the concept of the single-particle Schrödinger fluid is a successful model for the calculations of the nuclear moment of inertia. The calculated values of the reciprocal moments of inertia according to the rigid-body model for the two nuclei  $^{20}\text{Ne}$  and  $^{28}\text{Si}$  are not in good agreement with the corresponding experimental values, a fact which is not surprising since the pairing correlation is not taken into account in these calculations.

## References

- [1] A. Bohr and B. Mottelson, Nuclear Structure, Vol. **II**, Benjamin, New York (1975).
- [2] K. K. Kan and J. J. Griffin, Phys. Rev. C **15**; 1126 (1977).
- [3] K. K. Kan and J. J. Griffin, Nucl. Phys. A **301**: 258 (1978).
- [4] S. B. Doma, High Energy Physics and Nuclear Physics, **26(8)**: 836 (2002).
- [5] S. G. Nilsson, Kgl. Danske. Videnskab. Mat. Fys. Medd. **29**: 1 (1955).
- [6] G. A. Lalazissis and C. P. Panos, Phys. Rev. C **51(3)** (1995).
- [7] S. B. Doma and M. M. Amin, International Journal of Modern Physics E, **11(5)**: 455 (2002).
- [8] S. B. Doma and M. M. Amin, The Open Applied Mathematics Journal, **3**: 1 (2009).
- [9] C. L. Dunford and R. R. Kinsey, Nu. Dat. System for Access to Nuclear Data, IAEA-NDS-205, (BNL-NCS-65687)IAEA, Vienna, Austria (1998).

## **Sensory Activities Proposed for the Integration of Learning for Kindergartners Cerebral**

**M. M. Saleh and H. Elgarawany**

**Faculty of Kindergarten, Alexandria University, Sidi Gaber, Alexandria, Egypt**

### **Abstract**

This study aims to discover the importance of Integrative strategic teaching to kindergartner children. Therefore, the study sample consisting of 30 boys and girls from preschool was submitted to the test of brain learning and sensory activities of mathematical skills. Study results have shown that the proposed activities relative to the effectiveness of learning based on brain integration have significant impact on the superiority of the experimental group over the control group. Moreover, the educational need in the Arab world should reconsider the various elements to develop and contribute to the constructive change in the face of the third millennium challenges.

### **1. Introduction**

Brain theory during the twentieth century showed the need to increase reliance on the right half of the brain in learning, making educators develop programs to support the impact of brain in the activities of learners [6]. The results of brain research, found that the theory is based on brain learning, which emphasizes that each individual exercised the learning process "Does Learn", and that there is a need to create a learning environment and work to spend a learner in the educational experience and the freeing of the learner's anxiety and allowing him to active treatment [4].

The cranium is the largest part of the forebrain which is responsible for the performance of mental activities, and is divided into two halves, left and right, controlling the right side of the movement of the body and vice versa. Also, right hemisphere is associated with music, art, and emotional responses, and images. The left part is associated with the speech logic, numeracy, time, details and mathematics; etc. .The right side deals with colleges, while the left deals with the molecules, and work together in harmony and consistency.

In the twenty-first century learning has been focused on hemispheres of

the brain, "the left and right," the learners during the learning process and that, as they interact in any position of learning as an integrated whole and not just one side of the brain, "the left".

The perception of visual mathematics education depend on academic subjects, including mathematics and calculations belong to left lobe of the brain with logical thinking, and other activities while perception visual imagination and visual memory depend on the right.

Mathematics has focused on the basis of mathematical reasoning teacher requirements of the twenty-first century and lead to improved performance, thus improving the level of the understanding, which leads to academic excellence and skills of teaching and learning.

The numerical and spatial experience required in early childhood, and the acquisition of concepts, abstractions, for example numbers are not felt by the child, but can be a sense out of sensory experiences are produced through the opportunity to understand the concepts related to those experiences.

Conducted study of Bransford (2004) and Brewer (2006) are done on the enrichment of uses strategies based

learning on the brain and in raising the performance to remember and absorb the vocabulary of mathematics for the children of the KG2 in one of the schools of California and used in its context a process which is based on the use of the right hemisphere of the brain in the teaching and learning of these children, and that is added to the use of past experience.

The above discussion has already made the idea of this paper in order to pay attention to the children in this age, a study which is based on the authors experiences on the use of mathematics as an entry point for the development of learning based on the brain, both through the sensory contribution to support their learning and develop the skills demanding their contribution to facilitate and deepen their experience of teaching academic all. We made use on the above experience and the importance of mathematics in the development of different concepts in children and in the absence of previous studies in the field phase. The study in the present paper is an attempt for the preparation and application of mathematics sensory activities consistent with the levels of children's growth. The study relies on sensory learning, discovery and imagination, memory and attention as a backbone for the use of a full child's brain to reach the maximum development of the concepts, the child has to get to the great knowledge as an end of the educational process.

## **2. The Problem of the Study**

Away from traditional methods prevailing in the upbringing of children and kindergarten in pursuit of learning more effective, and in keeping with the times in which we live, we can formulate the problem of the current research in the following question:

1 - "What is the effectiveness of proposed activities based on the

mathematical skills to integrate the learning brain dimensions?"

### **Study Sample**

It consists of (30) boys and girls of the experimental group and the same for the control group from attending the second level of kindergarten children of Alexandria Governorate.

### **Hypotheses**

Research seeks to verify the following hypotheses:

- 1 - No statistically significant differences between the mean scores of the experimental group and the mean scores of children in the control group, so for the integration of learning brain to kindergarten children.
- 2 - The activities of the program are efficient.

## **3. Methods of the Study**

Prepare sensory activities proposed in Mathematics for kindergartners which are based on learning activation. Based on learning building on the activity of the learner, and also on the nature of children in the age of the group - the subject of the research was to propose activities which are also to suit them. So, it depend on the transfer of optical image lobe right to determine the sites of spatial concepts, and then translated in the form of output lobe the left and is done through a thread billions of nerve cells caught between the brain which leads to integration of learning (Jensen 2004).

To make learning the best possible in the proposed activities that had been prepared for the passage of the sample experiences, it has been subject to the following:

- Processing of the learning environment in all the diverse activities to give them greater opportunities to support the teacher.
- Taking into account individual differences.
- Provide an opportunity for children to discover patterns with self-

involvement of each other in their discoveries.

- Promote social solidarity.
- Diversification of evaluation methods to improve performance.

#### **Content of the suggested activities**

- Included in its final digit (76) active suitable for application on a target sample.

**Sample of the study:** The research sample consisted of (60) boys and girls divided into two equal groups. The independent variable is applied to the experimental group. The control group is taught in the usual way.

#### **Construction of achievement test**

The dimensions of the test: was committed by three levels: (a visual imagination, and holistic thinking and technical skills) in their interaction with Riziat skills that have been committed within the boundaries of this research.

#### **Drafting vocabulary test**

- Vocabulary test was formulated, in line with the conditions of the final sample and within the limits of their potential and their previous experience, within the framework to examine some of the studies that are concerned with the field of study such as: (Pinketoon, 2005); (Bransfoed, 2004) and (Lee, 2000).

- A review of vocabulary choice to offer some of the arbitrators specialists in the field of childhood, and in the field of mathematics education, was to make appropriate adjustments to achieve a true test and takes his final image.

In Table-1 we present the following objectives of the test and the corresponding vocabulary:

With regard to consistently, a test has been achieved to calculate the coefficient of stability in a re-test on the class from outside the research

sample, twice in a row with a time lag of three weeks, was to reach a value of "0.91", a statistically significant level (0.01), which refers to the stability of the test.

## **4. Results and Discussion**

### **First Search Results:**

1 - The first hypothesis states that "No statistically significant differences between the mean scores of the children of the experimental group and the mean scores of the children in the control group for learning to integrate the brain of the kindergartners"

The results shown in Tables-2, 3 and 4 state that there was no statistically significant differences at 0.5 between the average rate for children grades for each of the experimental and control groups to test the integration of learning brain to kindergarten children.

As is evident from the table low grades, which is expected for members of the sample, before passing through the experience of research, Valtalm traditional position on the left hemisphere is dominant in most educational institutions.

The following Tables show the statistical treatment of experimental research and the control group after the experiment.

1 - The results of Tables-5, 6 and 7 to the existence of statistically significant differences at level of 0.01, 0.02, 0.05, respectively, in the average rate of the degree of each of the experimental and control groups to test the integration of learning brain to debate with the three subject

2 - The second hypothesis states that: "The activities of the program are efficient".

Table-8 shows that the Black's gain ratio is 1.20, which means that all the dimensions of the test are efficient.

Also, it means that children exceeded 70% of the expected gain, a

fact which indicates that the efficiency is high due to the suggested activities.

The authors proved that the change is an integration of both parts of the brain.

The study of the present paper develops the mathematics concepts of

children by sensory activities. The results agree with: (Caine, 2001), (Brower, 2006), (Bransford, 2004), (Costa, 2005), (Lee, 2002), (Del, 2004), (Weidman, 2009), (Sprenger, 2008) and (Penkerton, 2005).

Table-1 Characteristics of brain integration learning test

Serial	The area of vocabulary associated with the left half of the brain, respectively	The number of vocabulary	Numbers Vocabulary As contained in the test	Dimensions of plotted to the test half-right brain
1	The number spatial relations the conclusion classification Conclusion spatial relations	11	7-5-3-1 25-23-18-16 33-29-27	Holistic thinking
2	The numbers spatial relations, classification conclusion, spatial relations	11	8-6-4-2 -15-14-11-10 31-30-21	Imagine Bosra
3	Patterns, Patterns classification, models Conclusion number Corresponds	11	17-13-11-9 24-22-20-19 32-28-26	Technical skills

Table-2 Mean, standard deviation and the value of "T" of two groups as regard the holistic thinking

Group	Average	Standard deviation	Standard error	Value of "T"	Function
Experimental	4.075	4.27	0.98	0.42	Not signifying
Control	4.55	3.92			

Table-3 Mean, standard deviation and the value of (T) of the two groups as regard the visual imagination

Group	Average	Standard deviation	Standard error	Value of "T"	Function
Experimental	3.94	4.99	1.21	0.26	Not signifying
Control	4.25	5.15			

Table-4 Mean, standard deviation and the value of (T) of two groups as regard the technical skills

Group	Average	Standard deviation	Standard error	Value of "T"	Function
Experimental	4.44	4.75	1.19	0.51	Not Significant
Control	3.83	5.16			

Table-5 Mean, standard deviation and the value of (T) The significance of the post test to measure the two sets of search in the center of holistic thinking

Group	Average	Standard deviation	Standard error	Value of "T"	Function
Experimental	9.55	7.9	1.44	3.2	Significant at the level of 0.01
Control	6.15	4.92			

Table-6 Mean, standard deviation and the value of (T) The significance of the post test to measure the two sets of search in the center of visual imagination

Group	Average	Standard deviation	Standard error	Value of "T"	Function
Experimental	9.15	7.72	1.48	2.51	Significant at the level of 0.02
Control	5.99	4.94			

Table-7 Mean, standard deviation and the value of (T) The significance of the post test to measure the two sets of search in the center of technical skills

Group	Average	Standard deviation	Standard error	Value of "T"	Function
Experimental	10.1	7.82	1.48	2.51	Significant at the level of 0.05
Control	6.1	7.3			

Table-8 Averages for the experimental group

Adjusted ratio of the gain	Average	Application Type	Axis
1.24	4.75 9.55	Before After	Holistic thinking
1.24	3.94 9.15	Before After	Visual imagination
1.27	4.44 10.1	Before After	Art skills

### References

1. Bransford, J. and Cocking, R. (2004). How Children Learn Mathematics: Brain, Mine, Experience, and Schools. 4th. Edition. New Jersey, Prentice – Hall, Inc, 122 – 32.
2. Brewer, J. (2006). Brain – Based Learning: The New Learning Model. Retrieved from <http://www.2.educ/McGarthD/Brewer.html>.
3. Cain, L. (2001). Transforming Education: Understanding Why Education Must Change. From <http://www.rewhouseons.org>.
4. Costa, A. (2005). A well Designed School. Environment Facilitate as Brain Learning. educational Research 35,3. 22-31
5. Del, s. (2005). Developing minds. Educational Digest, 66,8, 26-49
6. Hannaford, C. (2004). Smart Moves: why Learning is Not all in Your Head. Retrieved from <http://www.Brainconnection.com>
7. Jensen, E. (2002). Brain – Based Learning. Retrieved from: [Http://www.jensen.Leaningcenter.com](http://www.jensen.Leaningcenter.com)
8. Lee, S. (2002). A Framework for Literacy in a brain-Compatible Environment ERIC, ED432000.
9. Parpara, L. (2002). Learning: What we Have Learned Retrieved from <http://www.users.starget.net/astechrlearning.html>
10. Pinkirton, K. (2005). Using Brain – Based Learning Techniques in Elementary School Mathematics. Teaching and Learning Mathematics 43, 45-49.
11. Sprenger, M. (2008). Brain – Based Learning Highlights. ERIC No. Ej514356
12. Weiman, H. (2009). Gender Differences in cognitive Functioning. <http://homepage.Luc.edu/hetman/GenderDiffs.html>

# Development of a Passive Magnetic Bearing System for a Flywheel Energy Storage

Chan Ham

Mechatronics Engineering, Southern Polytechnic State University, Marietta, Georgia 30060, USA

Matthew Burton and Kurt Lin

Dept. of Mechanical, Materials, and Aerospace, University of Central Florida, Orlando, Florida 32816, USA

Young Hoon Joo

Corresponding author, Kunsan Nat'l University, Kunsan, Jeonbuk, Korea

*Abstract:* Magnetic bearings are an attractive alternative to mechanical bearings in flywheel energy storage systems since they greatly reduce friction and wear. However, new problems are introduced in terms of stabilization; particularly the magnetic bearing must provide enough damping to reduce excessive vibrations. Two novel configurations of passive magnetic bearings are introduced to meet system requirements. The first configuration, utilizing two Halbach arrays, is proposed to introduce a levitating force as well as lateral stability. The second configuration uses a null-flux coil passive magnet array to add lateral stability. This paper will focus on lateral stability. Displacements in the lateral direction will result in a correction of rotor motion based on its velocity. In this work, firstly the configuration of the Halbach and null-flux coil arrays is presented. Finite element models are then developed for both cases to investigate the resulting magnetic field, its reaction to rotor velocity, and its effects on the system. As a result, it shows the effectiveness of the proposed configuration to stabilize the lateral dynamics of a flywheel energy storage system.

## 1. Introduction

It has been proposed that a flywheel energy storage system could be implemented to capture and store the energy captured from diverse energy production systems until it is needed. Later it can be converted into electrical energy. However, a major issue with this approach is the friction caused by traditional ball bearings as well as any potential support system for the flywheel. This friction will remove energy from the system, and in most cases will need to be greatly reduced for the system to function efficiently. To this end, magnetic bearings have been proposed as an alternative method of stabilizing and supporting the flywheel system. Magnetic bearings are placed either on the rotor or the stator with matching permanent magnets to produce a force.

Due to the dynamics of the system, the magnetic field being produced will experience motion relative to the permanent magnet. This results in eddy currents being created and dissipated; the force produced is a damping force proportional to the velocity of the change in magnetic flux. Hence, the resultant force will depend on both the position and velocity of the rotor. (Sze Kwan Cheah, 2008)

This paper will focus on the use of passive magnetic bearings (PMB) in this application. Though

rotor motion cannot be corrected as accurately, PMB do not have to be controlled; thus, they consume less power.

In this paper, two magnetic bearing configurations are presented for use in a flywheel energy storage system. The first PMB configuration is composed of two Halbach arrays. These arrays can be used for stabilization as well as levitation of a vertical flywheel with coils placed on the perimeter and underneath the PMB. It has been shown in previous research that the magnetic field is concentrated on one side of the array and canceled on the other (Thompson W. K., 2006). This leads to the attractive motion correcting and levitation forces required in this system. In this paper, only the forces contributing to torsional damping and lateral motion correction will be discussed.

The second configuration proposed is a null-flux coil array. A null-flux coil uses two juxtaposed magnets with opposite coil windings to produce a magnetic field. Due to eddy current effects a small permanent magnet moving over the PMB will experience a correction in its motion to the center of the two magnets. This will result in the lateral correction of motion in a flywheel rotor. Damping

forces are also produced in the torsional direction, but these will not be investigated.

Once the configuration is fully described and the mathematics is rigorously defined the magnetic fields resulting from these PMB configurations will be modeled. For the Halbach configuration, the damping due to torsional velocity will be discussed as well as forces correcting its lateral motion. These results will be compared for various rotor velocities. For the null-flux coil array, only lateral correcting forces will be discussed.

Some assumptions must be made about the system to ensure accurate results. Neither torsional stability nor angular displacement will be considered in this paper to simplify analysis. The magnetic field results will also be limited to a 2-D plane. With more advanced work in this field these assumptions may be reconsidered and hence more in depth analysis could be performed to investigate the effects of Halbach and null-flux coil arrays in flywheel suspension.

## 2. System and Configuration

### Halbach Array

For the system under consideration, a flywheel rotor will be positioned vertically. As it spins along its axis, several stability issues need to be explored. The rotor will need to be suspended against the force of gravity acting downward. This issue hopes to be solved by the use of the Halbach array; it will be attached to the bottom of the rotor. The magnetic field produced by the Halbach array will act against coils wound vertically. According to the following equation, this should result in a force counteracting gravity. More specifically, this force is given by a pressure over the surface area of the rotor. In the equation below,  $\mu_0$  denotes the permeability of a vacuum, which is a constant.  $P_{mag}$  denotes the magnetic levitation pressure.  $B$  is the value of the magnetic field. It is this value that is to be investigated and determined in this paper.

$$P_{mag} = \frac{B^2}{2\mu_0} \quad (1)$$

It should be noted that this repelling force is due to the most attractive feature of the Halbach array: the magnetic field will be concentrated on the bottom of the array while canceled on the top. While this phenomenon will has been explained, it will not be investigated analytically due to simulation limitations.

The system described in this paper involves two concentric Halbach arrays with similar magnetization patterns. The first array is located on the end of the rotor, with the rotor occupying the center. The second Halbach array can be seen on the stator surrounding the rotor. It is believed that the

interaction of these two arrays will result in a horizontal restoring force as well as damping forces that should be confirmed by analysis.

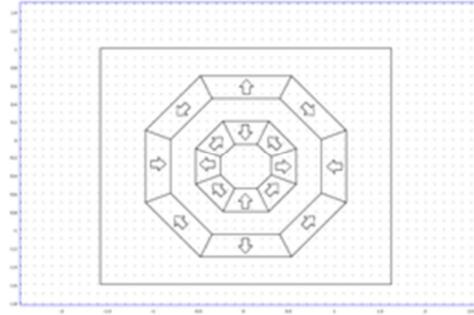


Figure 1. Double Halbach Array Configuration.

The effects of the Halbach configuration producing a restoring force in the radial direction are of the most importance. To describe the force produced by a current and magnetic field, the simple equation below is used.

$$\mathbf{F}_r = \mathbf{B}_\phi \times \mathbf{I}_z \quad (2)$$

Now an expression must be found for the magnetic field. The following equation describes the magnetic field in terms of a differential element of the field. For a circular ring magnet, the magnetic flux density  $B$  has the differential equation below, where permeability and magnetization per unit length are given by  $\mu_0$  and  $M_0$  respectively (Cheng, 1992).

$$d\mathbf{B} = \frac{\mu_0 M_0}{4\pi} \int_0^{2\pi} \frac{d\mathbf{l} \times \mathbf{R}_1}{|\mathbf{R}_1|^3} \quad (3)$$

where  $\mathbf{R}$  is a position vector describing the position of the magnet in question.  $d\mathbf{l}$  is a length vector describing a differential element of length along the magnet (Cheng, 1992).

$$d\mathbf{l} = -b \sin \phi d\phi \mathbf{i} + b \cos \phi d\phi \mathbf{j} \quad (4)$$

The drag force produced by Lorentz effects needs to be described. It can be shown that this force, which is dependent on velocity, is given by the following equation.  $B_C$  gives the field in the center of the coil wire,  $N_C$  is the number of coil turns,  $L$  is the inductance,  $R$  is the resistance, and  $w^2$  is the excitation frequency of the circuit (Han, 2000).

$$\mathbf{F}_z = \frac{v B_C^2 w^2}{2R} \frac{N_C}{\left(\frac{kvL}{R}\right)^2 + 1} \quad (5)$$

### Null-Flux Coil

The null-flux coil array uses two magnets with opposite coil windings to produce a force correcting rotor position if it is not centered between the two magnets. (Thompson M. T., 2000) A schematic of this configuration can be seen below.

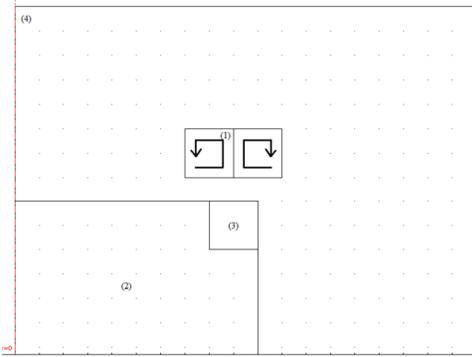


Figure 2. Null-Flux Coil Configuration.

Two forces produced by this configuration will be introduced. The first is the horizontal restoring force correcting the rotor's motion. Though background theory will not be explained here, it can be shown that this force is given by the following equation (Han, 2000).

$$\mathbf{F} = \left(\frac{1}{2}\right) \left(\frac{\alpha^2 \omega^2}{R^2 + \omega^2 L^2}\right) \mathbf{x} + \left(\frac{1}{2}\right) \left(\frac{\alpha^2 R(R^2 - \omega^2 L^2)}{(R^2 + \omega^2 L^2)^2}\right) \dot{\mathbf{x}} \quad (6)$$

In the equation above,  $\alpha$  represents the magnitude of the magnet's flux linkage with the lift magnets,  $\omega$  is the frequency at which the magnet passes the passive magnetic bearings, and  $R$  and  $L$  are the resistance and inductance of the bearings respectively. The  $x$  variable represents the position of the permanent magnet while  $\dot{x}$  can be thought of as its velocity.

A drag force is also produced by this array. Fortunately, the drag force falls at higher speeds. Though a drag force can be helpful in stabilizing the rotor under high speeds, an excessively high drag force can result in too much energy loss (Pilat). This drag force can be seen below (Han, 2000).

$$\mathbf{F} = \left(\frac{\pi}{21}\right) \left(\frac{\alpha^2 x^2 R \omega}{R^2 + \omega^2 L^2}\right) \quad (7)$$

In practice the materials chosen may deviate from those described here, but for the sake of analysis and simulation some materials will be assumed. Aluminum 2024 is assumed for the material making up the bulk of the rotor shaft. This is a common material chosen for such an application. In practice, however, a composite could be chosen to increase potential energy storage.

For the permanent magnet attached to the top of the rotor, which is used in combination with the null-flux coils to correct position, basic iron was used as the material. This basic ferromagnetic material should react appropriately under the influence of a magnetic field to correction rotor position.

The passive magnetic bearings to be utilized for the Halbach array and null-flux coils are assumed to be neodymium. Attractive features of this material

include a high resistance to being demagnetized, as well as high saturation magnetization. The particular makeup of this material, along with its properties, can be seen in the appendix. In practice, wound coils may be used to create the passive magnetic field, with options for active control later.

### 3. Analysis and Simulation

#### Halbach Array

The Halbach array was modeled with a top-down view, with the rotor protruding out of the page. To simplify analysis and to aid in possible manufacturing of this rotor system, eight Halbach magnets were modeled in a cylindrical configuration with the rotor occupying the center.

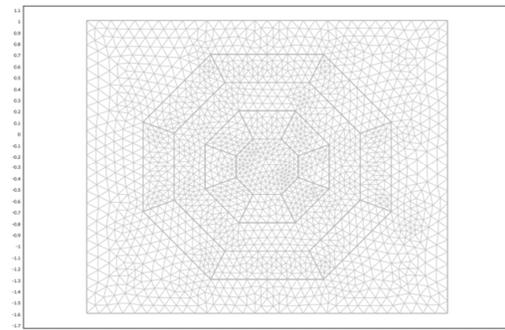


Figure 3. Halbach Array FE Model.

Another Halbach array magnetized in similar directions is located around the perimeter of the array attached to the rotor. The magnetic field produced by this configuration should result in forces correcting the position of the array attached to the rotor.

The material chosen for the magnets was neodymium. This material, which is commonly used for passive magnets, has a magnetic strength value of roughly 750,000 A/m. For the rotor core, the material was chosen to be iron. This ferromagnetic material should respond accordingly to the induced magnetic field. The finite element mesh of this simulation can be seen below.

The simulation created was a parametric model. To this end, the velocity was varied and a magnetic field solution was plotted for each. Maxwell stress tensors were also output for each of these cases. Specifically, the velocity varied from 0 m/s to 100 m/s with a 10 m/s increase for each case. The velocity in question describes the tangential angular velocity of the inner Halbach ring. The magnetic field solution for the case with a velocity of 100 m/s can be seen below.

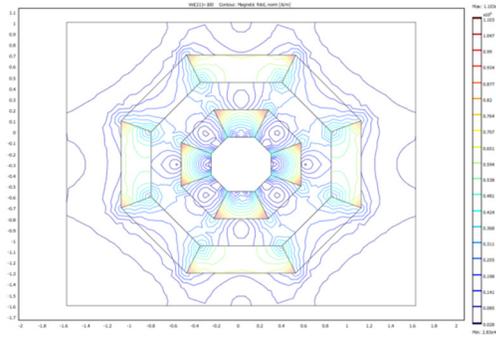


Figure 4. Magnetic Field for Halbach Array.

The magnetic flux density was also investigated. Along with the Maxwell stress tensor forces acting on the inner Halbach ring, it was plotted as a contour over the geometry. These results assumed a velocity of 100 m/s. It can be seen that the forces act to restore the rotor to its central position.

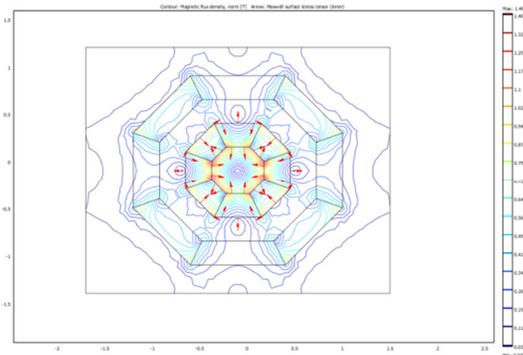


Figure 5. Magnetic Flux Density and Inner Ring Surface Stress Tensors.

The velocity was chosen to act in a clockwise direction. It stands to reason then that the damping forces, otherwise known as Lorentz forces, act in the opposite direction. While these will reduce energy from the system, this damping force is critical to stabilizing the rotor. This is especially true at high speeds. The Lorentz forces were plotted graphically for the inner Halbach ring at a velocity of 100 m/s along with the magnetic flux density.

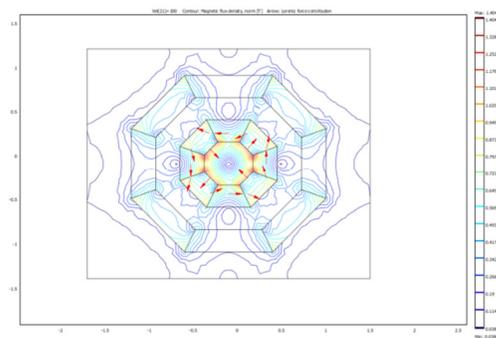


Figure 6. Magnetic Flux Density and Inner Ring Lorentz Forces.

Unlike the Maxwell stress tensor forces and the magnetic flux density, the Lorentz forces depend on the angular velocity of the Halbach ring. To see this relationship, a cross section contour plot of one of the magnets making up the inner Halbach ring was created. The Lorentz forces on this section were compared for the various velocities used. This plot can be seen below.

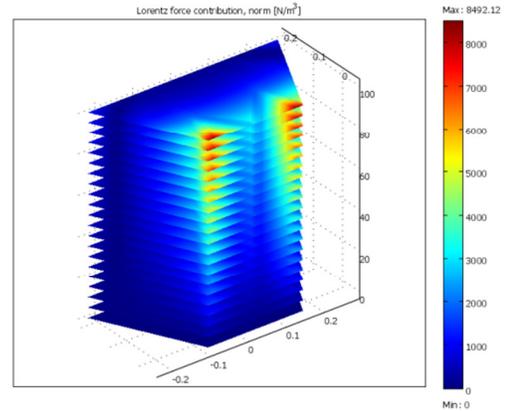


Figure 7. Lorentz Forces based on Velocity.

Further information based on this model can be seen in the appendix. While more data was collected, those presented here is considered the most pertinent to the analysis.

### Null-Flux Coil

The null-flux coil configuration was modeled using a radially symmetry, two-dimensional simulation. In the simulation considered, each magnet in this array has a square cross section measuring 0.2"x0.2". Though this is a very small magnet, they could be scaled up in practice. Choosing the material to be neodymium, a conservative magnetic field strength value of 750,000 A/m was chosen. A full table of neodymium properties can be seen in the appendix. An iron ring with dimensions of 0.1"x0.1" was placed on the upper and outer edge of the rotor. The finite element mesh of this configuration can be seen below.

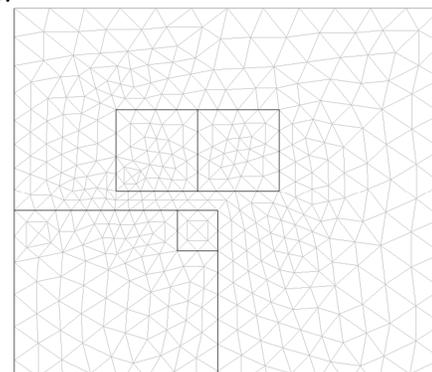


Figure 8. Null Flux Coil FE Model.

In the finite element mesh above, the dark lines represent boundary surfaces while the gray lines represent meshing elements. For this simulation, parabolic triangular elements were used.

The magnets were then magnetized as described in the theory; the rightmost magnet was magnetized with a value of 750,000 A/m into the plane and the leftmost magnet was magnetized with a value of 750,000 A/m into the plane. These magnetizations arise from the currents running through the wound coils. Maxwell surface stress tensor boundary variables were then computed. The forces on the rotor due to the magnetic field, along with a contour of the magnetic potential, can be seen below. As the model is axially symmetric about the left edge, these arrows indicate a radial force pointing inward, hence correcting rotor motion.

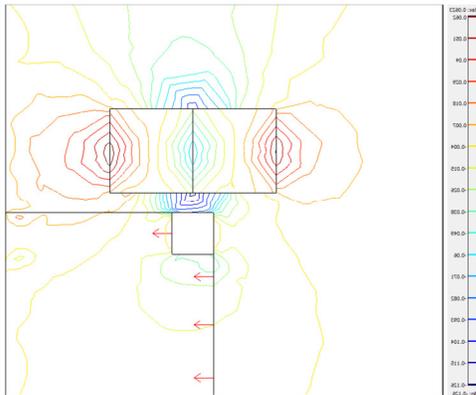


Figure 9. Magnetic Field and Forces on Rotor.

The value of the Maxwell stress tensor, which produces the force on the rotor, was also investigated along the length of the iron magnet boundary. Below can be seen a plot of how the magnitude of this radial correcting force varies over the magnet's length.

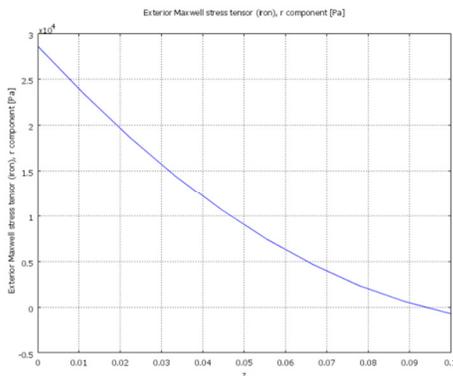


Figure 10. Maxwell Stress Tensor on Iron Magnet.

Further information based on this model can be seen in the appendix. While more data was

collected, those presented here is considered the most pertinent to the analysis.

#### 4. Conclusion

In this paper, two configurations were considered for use as magnetic bearings in a flywheel suspension system. The radial Halbach array utilizing two concentric rings was considered for levitation of the rotor as well as lateral stability. A null-flux coil configuration was also considered for the top of the structure to further correct lateral motion. The Lorentz forces arising from the angular motion of the rotor were then investigated for a number of speeds. The null-flux coil array was then simulated. Using an axially symmetric 2-D model, the magnetic field and forces correcting to the rotor motion were modeled. The analysis results show the effectiveness of the proposed magnetic bearing configuration to stabilize the lateral dynamics of a flywheel energy storage system.

#### Acknowledgement

This work was partially supported by the National Research Foundation of Korea Grant funded by Korean Government (MEST) (KRF-2009-220-D00034).

#### 5. References

1. Cheng, D. K. (1992). *Field and Wave Electromagnetics*. Reading, MA: Addison-Wesley.
2. Han, Q. (2000). *Analysis and Modeling of the EDS Maglev System Based on the Halbach Permanent Magnet Array*. Orlando, FL: University of Central Florida.
3. Juha Pyrhönen, T. J. (2009). *Design of Rotating Electrical Machines*. John Wiley and Sons.
4. Pilat, A. K. (n.d.). *Active Magnetic Suspension and Bearing*. AGH University of Science and Technology.
5. Sze Kwan Cheah, H. A. (2008). *Novel Eddy Current Damping Mechanism for Passive Magnetic Bearings*. *Journal of Vibration and Control*.
6. Thompson, M. T. (2000). *Eddy Current Magnetic Levitation*. IEEE Potentials.
7. Thompson, W. K. (2006). *Three-Dimensional Field Solutions for Multi-Pole Cylindrical Halbach Arrays in an Axial Orientation*. Cleveland, OH: Glenn Research Center.

# A Grid-Connected Desalination Plant Operation

Won Ko

Gachon Energy Research Institute, Kyungwon University  
Bokjung-Dong, Sungnam-Si, 461-701, Korea

Jong-hyun Lee

Doosan Heavy Industries & Construction  
1303-22, Seocho-Dong, Seoul, 137-920, Korea

and

Chan Ham

Dept. of Mechatronics Eng. , Southern Polytechnic State University  
1100 South Marietta Parkway, Marietta, GA 30060-2896

## ABSTRACT

In this paper, a grid-connected desalination plant operation approach is suggested. To produce potable water, large amount of energy is needed to operate pump and motor; hence most of energy is electricity. For this reason, the largest part of the operation costs is electricity charges. A suggested approach is a grid-connected desalination plant operation, which is operated based on electricity tariff rate schedule. To show the result of suggested approach, Electric tariff rate of Korea is used. The result shows that total cost reduction rate is calculated about 1.6% of annual total electric plant operation cost.

**Keywords:** Desalination, Smart Grid, Cost reduction.

## 1. INTRODUCTION

Desalination is a very attractive approach to solve water shortage matter in the world. Seawater desalination plants are rapidly expanding as natural water resources are limited and depleting [1]. Seawater desalination is highly energy intensive (consuming thermal and/or electrical energy) and, as such, energy costs are the major production costs in all large-scale seawater desalination methods including reverse osmosis (RO), multi-effect distillation (MED), and multi-stage flash (MSF) [2–4]. Traditionally, thermal processes (MED and MSF) were the most popular and economical methods for seawater desalination. However, SWRO (Sea Water Reverse Osmosis) is quickly gaining popularity as the most economical method of desalination. In SWRO, the largest portion of the operation cost is the electricity cost [3]. Currently, total SWRO Plant is about 500 million m<sup>3</sup>/day in the world. In order to reduce the energy cost, nowadays considerable attention is given to the optimization of operation and management of energy in various systems, including seawater desalination systems, e.g., [5]. A total Cost for SWRO plant operation are shown in Figure 1. From Figure 1, 23% of total cost is

energy. So various techniques are investigated regarding energy cost reduction matter. But most techniques are only considered energy efficiency problem in plant itself. In this paper, a novel concept of Grid-connected SWRO plant is explained and then suggested the approach for energy cost reduction using smart grid technology.

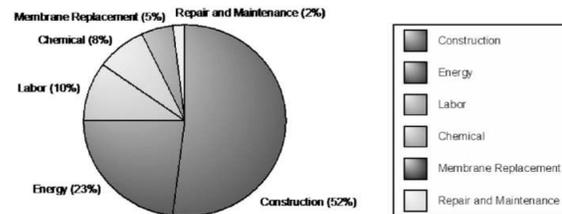


Figure 1. SWRO plant operation cost rate

## 2. GRID-CONNECTED DESALINATION PLANT

### Reverse Osmosis [6]

Reverse osmosis (RO) is a filtration method that removes many types of large molecules and ions from solutions by applying pressure to the solution when it is on one side of a selective membrane. The result is that the solute is retained on the pressurized side of the membrane and the pure solvent is allowed to pass to the other side. To be "selective," this membrane should not allow large molecules or ions through the pores (holes), but should allow smaller components of the solution (such as the solvent) to pass freely. In the normal osmosis process the solvent naturally moves from an area of low solute concentration, through a membrane, to an area of high solute concentration. The movement of a pure solvent to equalize solute concentrations on each side of a membrane generates a pressure and this is the "osmotic pressure." Applying an external pressure to reverse the natural flow of pure solvent, thus, is reverse osmosis. The process is similar to membrane filtration. However, there

are key differences between reverse osmosis and filtration. The predominant removal mechanism in membrane filtration is straining, or size exclusion, so the process can theoretically achieve perfect exclusion of particles regardless of operational parameters such as influent pressure and concentration. Reverse osmosis, however, involves a diffusive mechanism so that separation efficiency is dependent on solute concentration, pressure, and water flux rate. Reverse osmosis is most commonly known for its use in drinking water purification from seawater, removing the salt and other substances from the water molecules.

**Smart Grid [7]**

A smart grid is a form of electricity network utilizing digital technology. A smart grid delivers electricity from suppliers to consumers using two-way digital communications to control appliances at consumers' homes; this could save energy, reduce costs and increase reliability and transparency if the risks inherent in executing massive information technology projects are avoided. The "Smart Grid" is envisioned to overlay the ordinary electrical grid with an information and net metering system. Smart grids are being promoted by many governments as a way of addressing energy independence, global warming and emergency resilience issues. The idea of two way communications from suppliers to consumers to control appliances is not new, and systems have been implemented using analog technology for many years. The growth of an extensive digital communication network for the internet has made it practical to consider a more sophisticated type of smart grid. The increased data transmission capacity has made it conceptually possible to apply sensing, measurement and control devices with two-way communications to electricity production, transmission, distribution and consumption parts of the power grid at a more granular level than previously. These devices could communicate information about grid condition to system users, operators and automated devices, making it possible for the average consumer to dynamically respond to changes in grid condition, instead of only utilities and very large customers. A widely overlooked fact about the smart grid is that it is not a substitute for a real grid, but only an enhancement. The construction of a larger and better infrastructure of high-voltage transmission lines for the efficient delivery of electric power is a prerequisite to the construction of an effective smart grid.

**Grid-Connected**

A Grid-Connected desalination plant architecture and concept is shown in Figure 2. To design a Grid-connected system, a gateway portal concept is used. A Gateway portal is a smart bilateral communication and service infrastructure component between Grid and Plant. Gateway Portal is an infrastructure component to collect information and control devices in Desalination plant. Gateway Portal Concept is summarized:

- 1) The Gateway portal has two parts device portal and service portal (platform).
- 2) The device portal plays a role to collect information from devices and equipment in desalination plant.
- 3) Multiple device portals are able to communicate each other in peer-to-peer method to exchange information and service applications and also send/receive various data, meta-data, information, and control & management commands required by service portal.

The gateway portal will give the strategy to reduce electric cost through Demand Response. Peak shifting, Peak-shaving are the main technology of Demand Response.

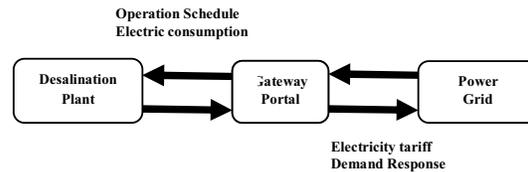


Figure 2. Grid-Connected desalination plant

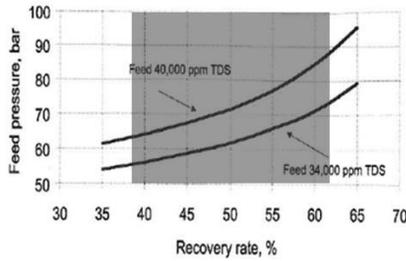
**3. ELECTRICITY PRICE REDUCTION**

In general, electricity tariff is a different as a season and usage hour. Table 1. is a electricity price from Korea Electric Power COoperation (KEPCO). Electricity prices are high during the peak electricity demand hours and low during off-peak load time period. From this electricity price table, High-Voltage B, Option III is used in order to show Grid-connected desalination plant operation cost reduction. If desalination plant operation can be scheduled using the estimation production demand and electricity price, operation cost can be reduced compare to traditional operation.

Table 1. Electricity price, Korea

Industrial Service (C)  
Contract demand of 1,000kW or more

Classification	Demand charge (won/kWh)	Time Period	Energy charge (won/kWh)		
			Summer	Spring, Fall	Winter
High-Voltage A	option I	off-peak load	46.70	45.70	50.10
		mid-load	89.30	85.30	84.00
		peak-load	182.50	89.30	116.50
	option II	off-peak load	42.00	42.00	45.40
		mid-load	84.80	81.10	75.30
		peak-load	147.80	84.50	111.80
High-Voltage B	option I	off-peak load	45.70	45.70	49.00
		mid-load	87.60	84.50	82.40
		peak-load	169.30	87.50	114.00
	option II	off-peak load	42.80	42.80	45.30
		mid-load	84.50	81.50	75.30
		peak-load	146.80	84.50	110.90
option III	off-peak load	40.30	40.30	44.20	
	mid-load	82.80	59.50	77.60	
	peak-load	145.10	82.30	109.20	
High-Voltage C	option I	off-peak load	45.10	45.10	49.30
		mid-load	87.40	84.50	81.50
		peak-load	149.30	87.40	113.70
	option II	off-peak load	41.10	41.10	44.30
		mid-load	83.40	80.30	77.50
		peak-load	145.30	83.40	103.70
option III	off-peak load	40.20	40.20	43.40	
	mid-load	82.50	53.40	77.00	
	peak-load	144.40	82.50	100.00	



TDS : Total Dissolved Solids

Figure 3. Recovery rate curve

**Recovery Rate**

The Desalination recovery rate is the product water recovery relative to the input water flow. Figure 3. shows recovery rate curve as feed pressure, the production can be controlled either change feed pressure or recovery rate during TDS is constant. Increasing pump head and flow rate, the production of desalination plant will be increased. But the membrane will be harmed by high pressure. Lowering pump head and flow rate, the production of desalination plant will be decreased. But the membrane will be maintained by low pressure. In general, recovery rate is between 48% and 50%. Therefore, if recovery rate can be controlled based on the electricity price, desalination operation cost can be saved. For the purpose of this study, electric consumption of pump in desalination plant is used. Table 2. shows the power usage in test-bed desalination plant in Korea. This test bed is composed 2MIGD (Million Imperial Gallons per Day) and 8MIGD capacity.

Table 2. Electric Consumption, Pump in plant

name	Flow	% rec	Head	density	gravity	Hy eff	Mo eff	power
Energy for 2MIGD HP Pump	834	48	386.11	1024	9.81	85	95	1112.7
Energy recovery Pelton drive	432		503.53	1024	9.81	75		455.24
Energy for 8MIGD HP Pump	1612	48	548.31	1024	9.81	85	95	3054.3
	1612	62	647.82	1024	9.81	85	95	3608.6
	1612	37	468.7	1024	9.81	85	95	2610.9
Energy recovery PX drive	1725		502.53	1030	9.81	100		2433.1
8MIGD booster Pump	1725	48	45.509	1030	9.81	80	95	289.92
	1725	62	45.509	1030	9.81	80	95	289.92
	1725	37	45.597	1028	9.81	80	95	289.92
2nd Pass RO pump	1070	91	133.49	1000	9.81	80	95	512.13
	1344	88	112.09	1000	9.81	80	95	540.16
	864	94	152.85	1000	9.81	80	95	473.51

Hy eff : Pump efficiency (%)  
 Mo eff : Motor efficiency (%)

Table 3. Electric consumption, recovery rate

Recovery Rate(%)	37	48	62
Total energy for RO (kWh)	4,969	5,551	4,487
Pump energy for RO (kWh)	2,596	2,402	2,830

Table 3. is calculated electric consumption based on recovery rate. Calculated value is applied to Table 1. Electricity price. This approach is the Grid-connected desalination operation. When electricity price is high, 37% recovery rate is used. However, electricity price low, 62% recovery rate is used. Figure 4. shows Grid-connected desalination operation based on electricity price. In Figure 4. total production is same but the production rate is different from electricity price. From table 1, peak load charge is expensive so the production is smallest than other time period. Figure 5. illustrates a comparison between traditional desalination operation and Grid-connected operation. Traditional desalination operation cost is about 3.3 billion KRW (3 million USD) and Grid-connected total saving cost is about 53,000,000 KRW/yr. This is 1.6% of total electric consumption cost.

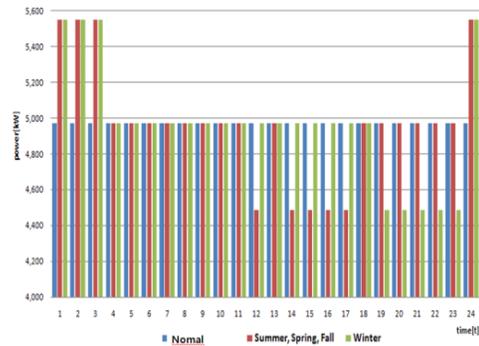


Figure 4. Operation based on electricity price

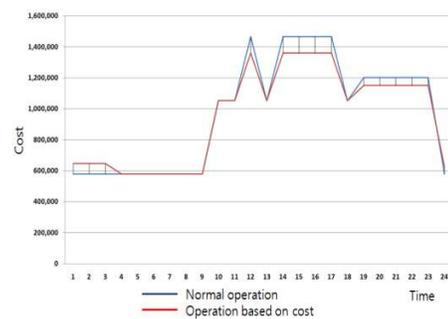


Figure 5. A comparison of normal operation and based on electricity price

#### **4. CONCLUSION**

A grid-connected desalination plant operation approach is suggested. A suggested approach is operated based on electricity price rate schedule. To show the result of suggested approach, Electric price rate of Korea is used. The result shows that total cost reduction rate is calculated about 1.6% of annual total electric plant operation cost.

#### **ACKNOWLEDGEMENT**

This work was partially supported by a grant for the Plant Technology Advancement Program funded by the Ministry of Construction & Transportation of the Korean Government (10-seaheroB04-1)

#### **REFERENCES**

- [1] P. Gleick, *The World's Water 2006–2007, the Biennial Report on Fresh water Resources*, Island Press, Chicago, 2006.
- [2] H. Ettouney, H. El-Dessouky, R. Faibish, P. Gowin, Evaluating the economics of desalination, *Chemical Engineering Progress* 98 (12) (2002) 32–39.
- [3] D. Akgul, M. Cakmakci, N. Kayaalp, I. Koyuncu, Cost analysis of seawater desalination with reverse osmosis in Turkey, *Desalination* 220 (1–3) (2008) 123–131.
- [4] A. Poullikkas, Optimization algorithm for reverse osmosis desalination economics, *Desalination* 133 (1) (2001) 75–81.
- [5] A. Scrivani, Energy management and DSM techniques for a PV-diesel powered sea water reverse osmosis desalination plant in Ginostra, Sicily, *Desalination* 183 (1–3) (2005) 63–72.
- [6] [http://en.wikipedia.org/wiki/Reverse\\_osmosis](http://en.wikipedia.org/wiki/Reverse_osmosis)
- [7] [http://en.wikipedia.org/wiki/Smart\\_grid](http://en.wikipedia.org/wiki/Smart_grid)
- [8] Amin Ghobeity, Alexander Mitsos, “Optimal time-dependent operation of seawater reverse osmosis”, *Desalination* 263(2010), pp 76-88, July 2010

# A Novel Design of Wave Energy Harvest Device with Flywheel Energy Storage System

**Kuo-Chi LIN**

Dept. of Mechanical, Materials,  
and Aerospace Engineering  
University of Central Florida,  
Orlando, FL 32816, USA

**Young Hoon JOO\***

Dept. of Control, Robot,  
and Systems Engineering  
Kunsan National University,  
Kunsan, 573-701, Korea

**Steven HELKIN**

Dept. of Mechanical, Materials,  
and Aerospace Engineering  
University of Central Florida,  
Orlando, FL 32816, USA

**Chan HAM**

Dept. of Mechatronics Engineering  
Southern Polytechnic State University  
Marietta, GA 30060, USA

## ABSTRACT

This paper describes a novel design of a wave energy harvest device that utilizes a flywheel energy storage (FES) system to yield increased power generation. The buoy design is moored to the ocean floor via a cable; as the buoy is heaved vertically by ocean waves, the cable rotates a pulley which in turn drives the rotor of an onboard generator. A ratchet within the pulley allows the rotor to only be turned in one direction. To prevent large tensions from being imposed on the cable by the back torque from the generator, a flywheel energy storage system is used. As the buoy is heaved vertically by incident waves, the electric load on the generator is removed, resulting in all of the energy extracted by the buoy to be stored in the flywheel system. Consequently the buoy is less restricted by high tensions and able to closely follow the motion of the waves, even while using a large generator with a high back torque coefficient. As the buoy moves downward after being heaved, the load is re-coupled to the generator, transferring the energy stored by the flywheel to the generator. Essentially the FES system trades power generation time for improved buoy motion. The focus of the paper is not to optimize the design of the buoy, but rather demonstrate the effectiveness of the FES system for the buoy design with arbitrary parameters. Simple simulations for a small buoy confirm the effectiveness of the proposed flywheel energy storage system—without it the wave energy harvest device produced only 90.0 watts of power, but with it the device produced 180.3 watts—an improvement of 100%. This improvement is based on a small generator with low back torque coefficient; for a large-scale design and stronger generator, the benefits are expected to be even greater.

**Keywords:** Flywheel Energy Storage, Ocean Power, Buoy, Wave Energy Harvester, Hydrodynamic Simulation

## 1. INTRODUCTION

The renewable energies, such as wind, wave, ocean current, solar energy, hydrogen generation, have received a great deal of attention in the past few years. Ocean waves represent an energy form created by wind passing over open water and gravitational pulls of the sun and moon, i.e. tides. A white paper by U. S. Department of the Interior [1] has estimated that the wave energy density levels can exceed 1,000 kW/m of wave crest length. The total annual average wave energy off the U.S. coastlines (including Alaska and Hawaii), calculated at a water depth of 60 m has been estimated at 2,100 Terawatt-hours (TWh) ( $2,100 \times 10^{12}$  Wh). [2] If even a small fraction of it can be harvested, the energy crisis in the world will be greatly alleviated.

Because renewable energy from both wind and wave sources are intermittent, the periodic energy collected should be stored for more efficient output. A flywheel stores kinetic energy by rotating a disk or rotor on its axis. The use of flywheels to provide smoother power outputs is an old technology and has been applied to steam and automobile engines for over a century. In 2006, the U. S. Department of Energy published a Federal Technology Alert [3], in which describes the new application domains of the Flywheel Energy Storage (FES) systems, in particular, the uninterrupted power system (UPS) as an effective alternative to batteries.

---

\* Professor and author of correspondence.

## 2. WAVE ENERGY HARVEST SYSTEM MODEL

Figure 1 shows the basic design of the wave energy harvest device, which converts the vertical linear motion of the wave into rotational motion. The device is mounted inside a buoy floating on the ocean surface.

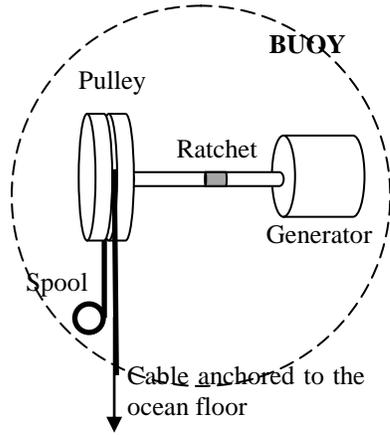


Figure 1. Wave Energy Harvest Device

An inextensible cable wraps around the pulley, with one end anchored to the ocean floor and the other end connected to a spool that keeps the cable wound up via a torsion spring. As the buoy is heaved upward by the wave, the cable comes out from the spool and rotates the pulley and the axial shaft. When the buoy moves down, the cable is retracted into the spool while the ratchet gear in the pulley prevents the shaft from rotating in the reverse direction.

To analyze the performance of the device, the equations of motion of the system are formulated based on a simple model that examines forces in the vertical direction. The motion of the buoy in the wave can be derived by

$$F_{wave} - T = ma \quad (1)$$

where  $F_{wave}$  is the sum of the forces produced by the wave,  $T$  is the tension in the cable,  $m$  is the sum of the mass of the buoy and added mass displaced by the buoy, and  $a$  is the vertical acceleration of the buoy. The tension  $T$  is positive only when the cable is driving the shaft; otherwise when the shaft is ratcheting—that is, when the driving angular velocity of the pulley is less than the actual angular velocity of the shaft—the cable is relaxed and the tension is treated as zero. The mechanical device inside the buoy can be modeled by

$$T \cdot r_{pul} - M_{gen} - M_{spo} = I\alpha \quad (2)$$

where  $r_{pul}$  is the radius of the pulley,  $M_{gen}$  is the back torque of the generator,  $M_{spo}$  is the torque produced by the spool that is constant and small,  $I$  is the total mass moment of inertia of the rotating elements, and  $\alpha$  is the angular acceleration of the axis. If there is no slippage in the pulley when the buoy moves upward, the vertical and angular accelerations can be related by

$$a = \alpha \cdot r_{pul} \quad (3)$$

It is clear that while the pulley is driving the generator rotor—i.e. while the buoy is heaved—the acceleration is limited by the generator back torque,  $M_{gen}$ . This effect of this torque can be removed completely by removing the electrical load from the generator, but this in turn causes the generator to not produce any power. However, by removing this torque while the pulley drives the system, and reapplying the load to collect energy while the pulley is ratcheting—at which time the tension in the cable is no longer affected by the back torque—the buoy system is able to achieve greater acceleration while still producing power. Thus, the use of a flywheel energy storage system to work with the wave energy harvest device is suggested.

## 3. FLYWHEEL ENERGY STORAGE SYSTEM

The flywheel energy storage system (FES) stores energy in the form of rotational kinetic energy. These storage systems lose energy from two sources: bearing friction and aerodynamic drag. However, new technologies can limit the effect of both of these factors. For example, magnetic bearings [4, 5] or even superconducting magnetic bearings [6] can levitate the shaft from physical contact with the bearing and hence reduce the friction to minimum. The aerodynamic drag can be minimized by enclosing the rotor in an air-tight vacuum chamber. [7] This type of set up is particularly suitable for marine applications, in which the system is subject to a corrosive, saltwater environment.

The purpose of this paper is not to design a new FES system; instead, with the already available technologies, the authors suggest a system that can improve the effectiveness of the wave energy harvest device detailed in the previous section. The suggested FES system uses the mechanical torque as the input to charge kinetic energy into the rotor, and draws energy from it through the generator to power the magnetic bearings. [8] The integrated harvest device-FES system works in the following way:

1. When the wave forces the buoy up, kinetic energy is charged into the FES system. Simultaneously, an onboard sensor recognizes the positive torque applied by the pulley to the rotor and causes the load on the generator to decouple via a switch, resulting in no current passing through the generator and a back torque of zero. The tension in the cable will be lower and allow the buoy to follow the motion of the wave more closely than if power is drawn by the generator during the upward motion.
2. When the pulley is ratcheting—that is, the angular velocity of the pulley is slower than that of the rotor—the load on the generator is re-coupled and causes the energy stored by the FES system to be drawn out as electrical power. As

the FES system discharges its stored energy through generator, the rotor slows down until the next wave crest approaches.

Similar FES systems have been demonstrated to be both feasible and practical. Hybrid performance vehicles, such as the Porsche 911 GT3 R Hybrid, utilize flywheel storage systems to capture kinetic energy produced by the engine, acting in place of a traditional battery. The flywheel system captures a portion of the energy dissipated during braking, and then transmits it back to the electric motor when the vehicle is accelerated. These flywheels have been shown to be effective even during bumpy road conditions; even though the conditions imposed on the buoy may be more severe, the angular velocity of the flywheel in the buoy will be only several hundred rpm, as opposed to 40,000 rpm in the hybrid vehicles [9]. Additionally, the fact that the flywheel systems generally dissipate a majority of their stored energy within about 10 seconds is not a limitation for the buoy, as the energy storage is only required for the time between incident wave crests.

#### 4. SIMULATION METHOD

A mathematical simulation was developed to model the buoy motion and angular velocity of the generator rotor of the aforementioned system. The simulation creates a second-order differential equation produced by equating the sum of the forces acting on the buoy to the product of the buoy mass and its acceleration. For simplicity, the width of the waves are considered to be substantially longer than the width of the buoy so that the problem may be treated as two-dimensional, neglecting any forces acting orthogonal to both the wave propagation and the vertical directions. Additionally, the motion of the buoy is assumed to be constrained so that motion occurs only in the vertical direction, thus requiring only one differential equation to be solved.

The purpose of the simulation is to validate the ability for the suggested FES system to improve performance of the buoy system—that is, the objective is to demonstrate an increase in power output from the system when the FES system is applied to an already-working buoy system, keeping everything else constant. As a result, several of the design parameters can be given relatively arbitrary, albeit realistic, values (the values chosen are based off of the design of a related experiment that will not be discussed here as it falls outside of the scope of this paper).

The general equation of motion for a floating body under heave contains several terms, including a hydrostatic restoring force (modeled as a spring) and the vertical heaving forces imposed by the wave. The simulation models the wave as purely sinusoidal that propagates perpendicularly to the motion of the buoy. The function

describing the wave height as a function of horizontal position  $x$  and time  $t$  is

$$y_{wave}(x, t) = a \cdot \sin\left(2\pi \cdot \left(\frac{t}{T_w} - \frac{x}{\lambda_w}\right)\right). \quad (4)$$

Here  $a$  is the wave amplitude in meters,  $T_w$  is the period of the wave in seconds, and  $\lambda_w$  is the wavelength in meters. The hydrostatic force is part of the driving motion for the wave, as given by

$$F_s = \rho \cdot g \cdot V_{submerged} - m \cdot g. \quad (5)$$

Density of the water is given by  $\rho$ , measured in  $\text{kg/m}^3$ . The first term is the buoyancy force,  $F_b(z)$ , which is a function of the buoy position with respect to the wave, as given by Archimedes' principle. The second term is the weight,  $W$ , given in Newtons.

The other portion of the driving forces for the buoy is the excitation forces caused by the interaction of the waves with the buoy. These forces include the Froude-Krylov [10], wave diffraction, and radiation forces. The latter two forces are substantially difficult to solve and may be neglected under the assumption that the wavelength of incident waves is much larger than the size and motion of the buoy. For simplification of the simulation, these terms are indeed excluded, and the wave excitation force is equal to the Froude-Krylov force, given by the surface integral

$$F_{exc} = -\rho \iint_{Sub} \frac{\partial \Phi_I}{\partial t} \hat{n} \cdot dS. \quad (6)$$

The integral is taken over the *submerged* surface of the buoy, and  $\hat{n}$  is the normal vector to the surface of the buoy. The term  $\Phi_I$  is the incident wave potential, and for deep sea conditions it is given as

$$\Phi_I = \frac{a \cdot \omega_w}{k} e^{kz} \text{Re}\{ie^{i(\omega_w t - kx)}\}. \quad (7)$$

The  $\omega_w$  term is the angular frequency of the wave, given in radians per second; the  $k$  term is the wavenumber, defined as  $2\pi$  divided by the wavelength;  $i$  is the imaginary number, equal to  $\sqrt{-1}$ ; and  $Re$  is the real operator that yields only the real portion of the complex value inside of the brackets.

Additionally, the buoy experiences a drag force as it moves through the fluid, which opposes the direction of the buoy velocity. This drag force can be calculated using the vertical velocity of the buoy,  $\dot{z}$ , and the vertical velocity of the wave,  $\dot{y}_{wave}$  found through (4), and for high Reynolds number [11], is given by

$$F_d = -\frac{1}{2} \rho \cdot A_{submerged} \cdot C_d \cdot (\dot{z} - \dot{y}_{wave}) \cdot |\dot{z} - \dot{y}_{wave}| \quad (8)$$

The submerged area,  $A_{submerged}$ , is the orthogonal projection of the cylinder (i.e. a rectangle) that is submerged. The drag coefficient,  $C_d$ , has been shown to

be approximately equal to unity for Reynolds number up to  $10^5$  [12]. A damping force caused by mechanical friction is not included as the FES system is assumed to minimize frictional losses in a fashion described in the previous section. Thus, the equation of motion takes the form

$$m \cdot \ddot{z} = F_b(z) + F_{exc}(z) - W - F_d(\dot{z}) - T. \quad (9)$$

The level of tension,  $T$ , can be calculated using (2) and (3) for when the pulley is driving the rotor, and can be taken to be a low value for when the pulley is ratcheting. The back torque imposed by the generator when an electric load of  $1 \Omega$  is applied can be found experimentally. Using a GL-PMG-500A permanent magnet alternator manufactured by Ginlong Technologies [13], the back torque and power output can be modeled to closely match experimental results using

$$M_{gen} = 1.146 \cdot \omega, \quad (10)$$

$$P_{out} = \frac{1}{4} \cdot \omega^2. \quad (11)$$

These equations are in basic SI units, with  $\omega$  being the angular velocity of the rotor, measured in radians per second.

Because the differential equation is nonlinear as a result of the drag term, the differential equation given by (9) is approximated using a fourth-order Runge-Kutta method [14]. The accuracy of the approximation can be improved by choosing a suitably small time-step for evaluation.

### 5. SIMULATION RESULTS

In order to evaluate the effectiveness of the proposed FES design, two simulations were ran—one that simply extracts power throughout the entire motion, and one that decouples the electrical load on the generator in the previously described fashion. Both simulations model the ocean waves as a sine function with 1-meter amplitude and 5-second period. The buoy used is a horizontal cylinder with a 0.3-meter radius and unit length, the mass of the system is 200 kg, the moment of inertia is  $0.25 \text{ kg}\cdot\text{m}^2$ , and the radius of the pulley is three centimeters. Again, these values are arbitrarily chosen based off of expected values for real design.

The two systems are completely identical except for the application of the load on the generator. For the system without the proposed FES concept, the resistive back torque from the generator is always active, creating additional tension in the mooring cable as the pulley drives the rotor. For the system with the FES device, the load on the generator is decoupled as the pulley drives the shaft, removing the additional tension cause by the moment described in (8). The tradeoff, however, is that during this time the generator does not draw any power. The simulation demonstrates that even though there is

less time at which power is produced, the FES system yields a greater overall average power.

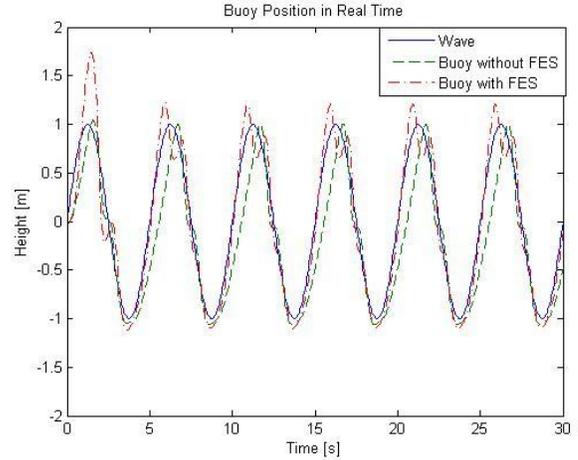


Figure 2. Simulation buoy height with respect to time for both the system without FES and the system with FES

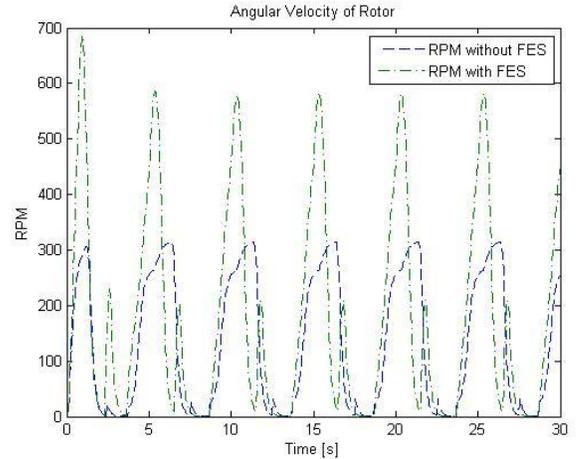


Figure 3. Simulation rotor velocity for both the system without FES and the system with FES

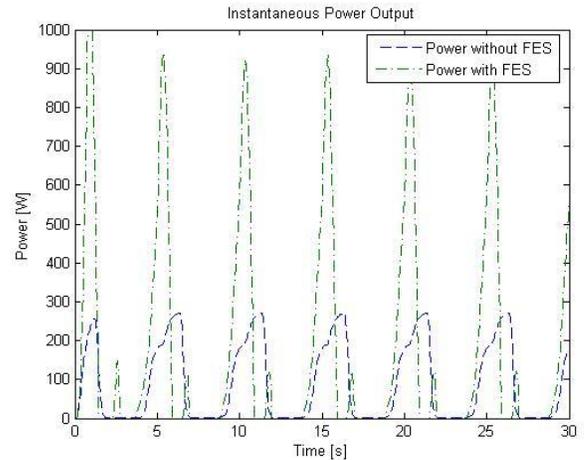


Figure 4. Simulation instantaneous power output by the generator for both the system without FES and the system with FES

The previous figures demonstrate the results of the simulation. Figure 2 presents the buoy height for both scenarios over a 30-second simulation, as well as the height of the wave for reference. Figure 3 presents the angular velocity of the generator rotor as a function of time for both systems and Figure 4 depicts the instantaneous power output by the generator for both systems.

The average power produced by the energy harvest device without the FES system was 90.0 watts, while the average power produced by the buoy with the FES was 180.3 watts. As such, the simulation demonstrates that more energy can be drawn from the system by sacrificing time during which the load is applied to the generator for higher angular velocities of the rotor. The effectiveness of the proposed flywheel energy storage system is dependent on many of the system parameters, but will generally be more valuable for larger generators that have higher startup torques back torque coefficients.

## 6. DISCUSSION AND CONCLUSION

The simulation results in the previous section demonstrate the effectiveness of utilizing a FES system to control power output in a wave energy harvest device. Figure 2 illustrates the physical effect that removing the electrical load during the heaving motion has on the movement of the buoy. The device without the FES system lags behind the wave crest position, remaining mostly submerged as it is heaved by the hydrostatic wave force. This limits the velocity that it achieves, and in turn, reduces the instantaneous power developed, as shown in Figures 3 and 4. The device with the FES system, on the other hand, follows the wave crest position much more closely. In fact, it is projected a small distance out of the water at the peak of the wave; although, realistically, the overshoot is undesirable, the optimized device will correct for this. The ability for the buoy to follow the wave closely allows for it to achieve peak rotor velocities of 600 rpm—double of that for the device without the FES system.

Figure 4 illustrates the tradeoff imposed by the suggested FES system—time for power to be generated. While the device without the system generates power while it has a positive angular velocity imposed on the motor, the device with the FES does not—instead much of the time the power output is zero. Nevertheless, the increase in buoy velocity is such that the FES system provides for triple the instantaneous power output peaks. As a result, the average power of the system is higher overall, experiencing an increase by 100% over the system without the FES system.

After comparing the power produced by the system with the total power available from a wave, it is clear that the device can benefit greatly from optimization. The total

energy density per unit horizontal area for a purely sinusoidal wave can be described by

$$E = \frac{1}{2} \cdot \rho_{water} \cdot g \cdot a^2 \quad (12)$$

where  $a$  is the wave amplitude [14]. For a deep ocean wave (depth greater than half of the wavelength), the wavelength and group wave velocity can be written as:

$$\lambda = \frac{g}{2\pi} \cdot T^2, \quad (13)$$

$$v_g = \frac{g}{4\pi} \cdot T. \quad (14)$$

For the wave modeled,  $a$  is 1 meter,  $T$  is 5 seconds, and the density of water is taken as 997 kg/m<sup>3</sup>. The power available in the wave per unit width of wave is

$$P = E \cdot v_g. \quad (15)$$

Thus, the total power available to be captured by the cylindrical buoy (width 1 meter) is 19.1 kW. The 180 watts produced by the buoy is less than 1% of the energy available.

The effectiveness of the FES system is dependent upon the amount of back torque produced by the generator. Conceptually, as the back torque imposed by the generator increases, the effectiveness of the proposed FES system will increase as well. The simulations model a small generator rated for 500 watts of power output, which has a low back torque coefficient of 1.146 kg-m<sup>2</sup>/s. However, with a larger generator the power output can increase substantially. Moreover, high gear ratios, larger-sized flywheels, larger buoys, and higher waves can all improve power output dramatically for the wave energy harvest device and integrated flywheel energy storage system.

## Acknowledgement

This work is supported by the Korea Research Foundation Grant funded by the Korean Government (MEST) (KRF-2009-220-D00034).

## 7. REFERENCES

- [1] "Wave Energy Potential on the U.S. Outer Continental Shelf", U.S. Department of the Interior Technology White Paper, May 2006.
- [2] Bedard, R., et al., 2005, "Final Summary Report, Project Definition Study, Offshore Wave Power Feasibility Demonstration Project", EPRI Global WP 009 – US Rev 1, Jan. 14, 2005.
- [3] Federal Technology Alert, "Flywheel Energy Storage", Federal Energy Management Program, U. S. Department of Energy, April 2006.
- [4] Filatov, A.V.; Maslen, E.H., "Passive magnetic bearing for flywheel energy storage systems", **IEEE Transactions on**

- Magnetics**, Volume 37, Issue 6, Nov 2001, pp. 3913 – 3924.
- [5] Zhua, K.Y., Xiaob, Y. and Rajendra, A. U., “Optimal control of the magnetic bearings for a flywheel energy storage system”, **Mechatronics**, Volume 19, Issue 8, December 2009, Pages 1221-1235
- [6] Kim, J.-S. et al, “Dynamic characteristics of a flywheel energy storage system using superconducting magnetic bearings”, **Supercond. Sci. Technol.**, vol. 16, 2003, pp. 473-478.
- [7] Zhou, L., Qi, Z., “Review of Flywheel Energy Storage System”, **Proceedings of ISES Solar World Congress 2007**, September 18-21, 2007, Beijing, China, pp. 2815-2819.
- [8] Tsao, P., Senesky, M., Sanders, S.R., "An Integrated Flywheel Energy Storage System with Homopolar Inductor Motor/Generator and High-Frequency Drive", **IEEE Transactions on Industry Applications**, vol. 39, no. 6, Nov-Dec, 2003, pp. 1710-1725.
- [9] “Porsche 911 GT3 R Hybrid using Williams Flywheel KERS”, Green Car Congress, Feb 20, 2010, <<http://www.greencarcongress.com/2010/02/gt3r-20100211.html#more>>.
- [10] Techet, A.H., “Froude Krylov Excitation Force”, Design Principles for Ocean Vehicles, March 29, 2005, <<http://web.mit.edu/13.42/www/handouts/reading-froudekrylov.pdf>>.
- [11] Hicks, D.C., Pleass, C.M., “Physical and Mathematical Modelling of a Point Absorber Wave Energy Conversion System with Nonlinear Damping”, University of Delaware, Newark, Delaware, USA, 1985.
- [12] Tritton, D.J., **Physical Fluid Dynamics**, Clarendon Press, Oxford, Great Britain, 1988.
- [13] Ginlong Technologies, Inc., 2006. <<http://www.ginlong.com/wind-turbine-pmg-pma-permanent-magnet-generator-alternator.htm>>.
- [14] Butcher, J.C., **Numerical Methods for Ordinary Differential Equations**, John Wiley & Sons, Ltd., West Sussex, England, 2003.
- [15] Pires, H.N.O., Pessanha, L.E.V., “Wave Power Climate of Portugal”, Instituto Nacional de Meteorologia e Geofísica, Portugal, 1985.

## **Rotor Dynamic Analysis and Experiment of 5kWh Class Flywheel Energy Storage System**

Cheol Hoon PARK      Sang-Kyu CHOI      Sang Yong HAM  
Nano Convergence and Manufacturing Systems Research Division,  
Korea Institute of Machinery and Materials,  
Yuseong-gu, Daejeon 305-343, Korea

### **ABSTRACT**

FESS(flywheel energy storage system) is a kind of mechanical energy battery which can collaborate with various electric energy sources such as wind power generator, regenerative brake system and so on. Generally, flywheel rotor of FESS is mounted on the magnetic bearings so as to minimize the air friction loss. The magnetic bearings should have enough stiffness and damping to suppress the imbalance response of flywheel rotor, and the rotor should be designed to have as high first bending critical speed and small unbalance response as possible to minimize the use of electric energy by magnetic bearings. The critical speed analysis of the rotor mounted on the magnetic bearings should include the magnetic bearings, its controller and FEM model of rotor. This paper presents the design procedure and rotor dynamics analysis of flywheel rotor for 5kWh class FESS mounted on the magnetic bearings. The designed flywheel rotor has succeeded to run stably up to 15,000rpm with small unbalance response. The experimental results such as open loop FRF, rigid body critical speeds are presented and compared with the analysis results as well.

Keywords: Flywheel Energy Storage System, Rotor Dynamics, Critical Speed, Magnetic Bearings and Finite Element Method.

### **1. INTRODUCTION**

FESS(Flywheel Energy Storage System) is a kind of mechanical energy storage system which can store electric energy in the form of kinetic energy and convert kinetic energy to electric energy again when necessary. Recently the application of FESS becomes various and wide in the area of, for example, uninterrupted power supplies (UPS), voltage regulator

of wind power plant, vehicles and railways and so on [1]. Especially, the exhaustion of fossil fuels is accelerating the development of FESS for the wind power plant as an alternative energy. In order to stabilize the output and maximize the operating efficiency of the wind power plant, it needs to be connected to the energy storage system [2][3]. 5kWh class FESS has been developed to meet this kind of needs. FESS is an electro-mechanical battery having a great deal of advantages of high energy density, long life and affinity for the environment compared to other forms of energy storage systems such as the chemical battery [4].

The developed FESS here has been designed to output 5kW power at 15,000rpm and the operating test has been done successfully. FESS mainly consists of non-contacting active magnetic bearings(AMB) that provide very lower frictional losses and smaller maintenance costs compared to the mechanical bearings, a composite flywheel rotor of high energy density and high mechanical strength, a motor/generator that converts mechanical energy to electric form, and vice versa, and a vacuum chamber that minimizes windage losses. Since AMB is open-loop unstable, the feedback controller is absolutely required for the stable operation. AMB controller for 5kWh FESS was developed by using MATLAB/xPC Target which is the PC-based real time control system of The MathWorks, Inc[3]. This paper presents the design concept of 5kWh class FESS and the experimental results for the operational test for the flywheel mounted on AMB up to 15,000rpm.

### **2. SYSTEM DESCRIPTION**

Figure 1 is the schematic diagram of FESS. The electric energy is converted to the kinetic energy and stored into the flywheel by the motoring operation of 20kw class motor/generator located at the lower shaft

area. If the motor/generator stops the motoring operation and changes its operating mode to the generating mode, the kinetic energy of the flywheel is converted to the electric energy. The axial and radial displacements of flywheel shaft are measured by the axial and radial gap sensors and controlled by the forces from the thrust and radial AMB actuators. The axial weight of the flywheel is mainly supported by the attractive force of the permanent magnet which is located at the upper thrust AMB actuator.

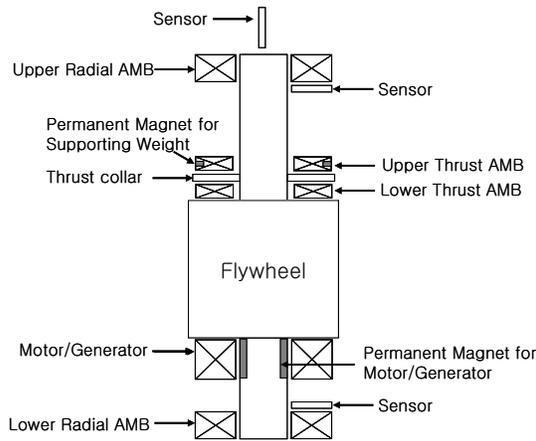


Figure 1. Schematic diagram of FESS

In order to minimize the energy losses by the control input to the AMB actuators, the control input for the unbalance response of the flywheel should be minimized. The selection of  $I_p/I_t$  is very critical to the unbalance response of the flywheel. Because the rotational kinetic energy stored to the flywheel is linearly proportional to the polar moment of inertia,  $I_p$  and the thick disk shaped flywheel is preferred to minimize the mass of the flywheel relative to the stored kinetic energy,  $I_p/I_t$  of the flywheel for FESS is generally designed to be more than 1. When  $I_p/I_t$  of the flywheel rotor here is designed to be more than 1, the conical critical speed was placed at around 19,000rpm. Consequently, the unbalance response increases continuously in proportional to the rotational speed until it reaches to the target speed 15,000rpm and the control input to the radial AMB actuator must be increased continuously and the bandwidth of the AMB controller must be over 250Hz. It's very difficult to realize the AMB actuators and controller having this kind of performance. It's not the best choice to increase the bandwidth of the AMB actuators and controller recklessly. Instead of that, if the conical critical speed can be passed through at the low rotational speed where the AMB actuators can produce

the enough forces, the rotational speed of the flywheel can be increased to the target speed keeping the displacement of the flywheel shaft due to the unbalance response minimized[6]. This is sort of easy method to realize. From this kind of reason,  $I_p/I_t$  of the flywheel here is designed to be around 0.5 so that the conical critical speed could be passed through at the rotational speed less than 3~4,000rpm.

The thrust AMB is designed to be the hybrid double acting type which has the permanent magnet inserted in the upper thrust AMB actuator and attracts the flywheel upward.

The flywheel shaft is designed to have its 1<sup>st</sup> bending mode frequency above 450 Hz so that the freedom to design the AMB controller could be ensured even for the frequency area over 250Hz corresponding to the target speed, 15,000rpm. In order to achieve this, the heavy thrust collar is located as near to the center area of the flywheel shaft as possible so that the additional mass effect could be minimized. The permanent magnet for the motor/generator is located at the lower shaft area. By this way, the burden due to the additional mass can be uniformly distributed to the upper and lower shaft. The material for the flywheel shaft is SUS304 which is relatively cheap and easy to manufacture.

The composite material is used for the flywheel rims to resist the tension due to the centrifugal force. Figure 2 shows the detailed configuration of the designed flywheel and Table 1 shows the specification of the flywheel.

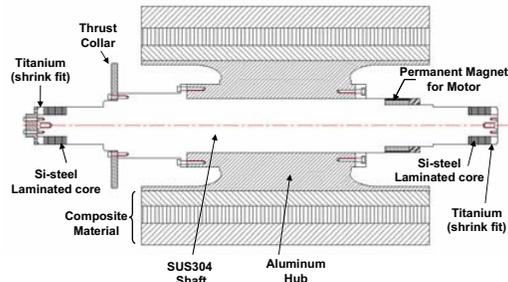


Figure 2. Detailed configuration of flywheel

Table 1 Specification of flywheel

Item	Value
Material	SUS304
Mass[kg]	407
Length[mm]	1119
Diameter[mm]	580
$I_p/I_t$	0.51

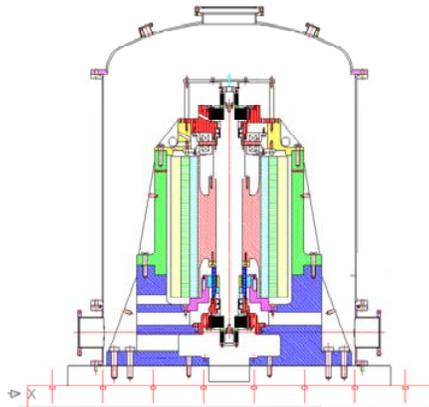


Figure 3. FESS in vacuum chamber

The casing of FESS must be designed to have the strong structure so that the radial AMB actuators, thrust AMB actuator and motor stator could be supported strongly and the flywheel rotor could rotate stably about the rotational center. Moreover, after assembling the upper and lower radial AMB actuators, the misalignment of them should be minimized. The casing of FESS must have the stiffness so that the 1<sup>st</sup> bending mode frequency of the casing could be placed above 300Hz which is above the target speed 15,000rpm, that is, 250Hz. Figure 3 shows the draft of the assembled FESS in the vacuum chamber. Figure 4-5 shows the manufactured AMB actuators, flywheel and the assembled FESS.

### 3. ROTOR DYNAMIC ANALYSIS

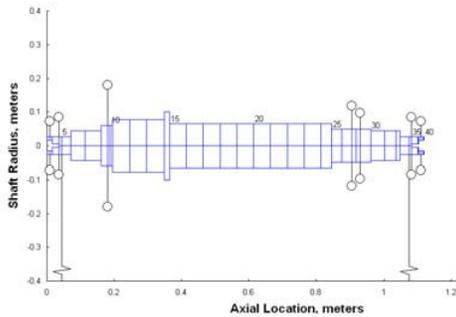


Figure 4. Rotor dynamic model for shaft only

Composite material and aluminum hub are removed from flywheel in figure 2 and flywheel shaft are modeled by using xltrotor as shown in figure 4. Titanium and Si-steel part, thrust collar and permanent magnet for motor are modeled as added masses which do not help for the stiffness of shaft. It was predicted that first and second bending mode frequencies of shaft

in free-free condition would be 518Hz and 1143Hz respectively. Figure 5 shows the rotor dynamic model for full flywheel. It was predicted that first and second bending mode frequencies of flywheel would be 612Hz and 1184Hz respectively. Aluminum hub strengthened the stiffness of shaft and increased bending frequencies. 612Hz is sufficiently high first bending frequency considering the operating speed of 300Hz. Imbalance response was predicted as figure 6. After passing through the rigid body mode with the peak response of less than 20 $\mu$ m, it would be possible to increase the rotational speed without the additional effect of imbalance.

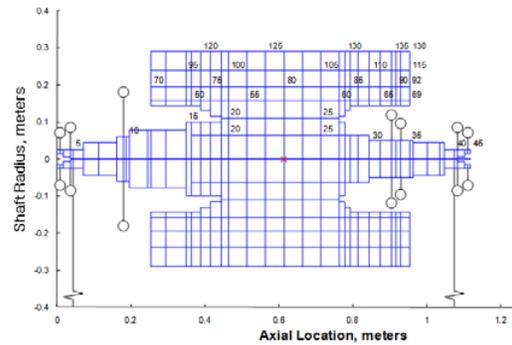


Figure 5. Rotor dynamic model for flywheel

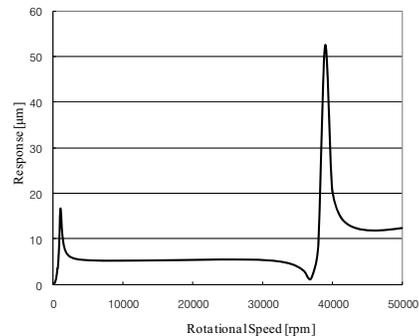


Figure 6. Prediction of imbalance response

Impact test was performed after fabricating the shaft only and full flywheel as shown in figure 7 and 8. First and second bending frequencies of shaft in figure 7 are measured to be 549Hz and 1180Hz in the upper shaft and 518Hz and 1200Hz in the lower shaft respectively. The impact test results are very similar to the analytical results, which means that the shaft was fabricated well and meet the design requirement. First bending frequencies of flywheel in figure 8 are measured to be 620Hz (figure 9) and 762Hz for the upper and lower part, respectively.



Figure 7. Impact test for shaft only



(a) Flywheel

(b) FESS

Figure 8. Fabricated flywheel and FESS

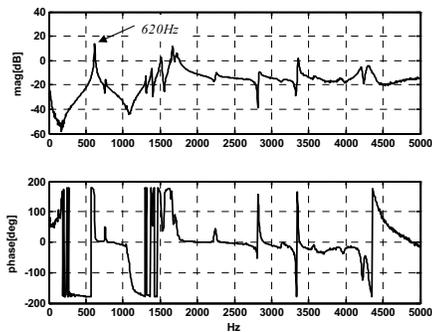


Figure 9. Impact test result of flywheel (Upper part)

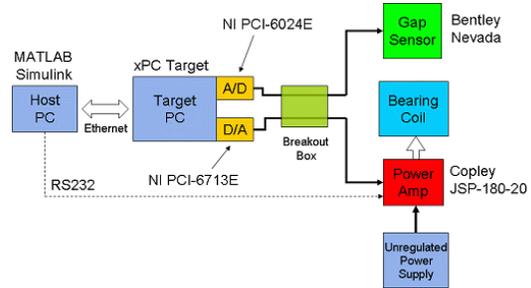


(a) Thrust Actuator

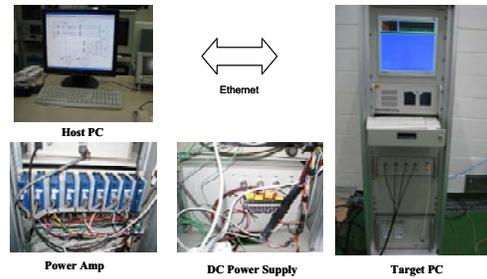
(b) Radial Actuator

Figure 10. Fabricated actuators

$$\begin{aligned} \dot{x} &= A(\Omega)x + Bu \\ y &= Cx + Du \end{aligned} \quad (1)$$



(a) Block diagram of AMB controller



(b) Each component of AMB controller

Figure 11. Hardware of AMB controller using xPC Target

#### 4. SYSTEM IDENTIFICATION

Bias linearization method is employed to get the linearized model of the radial AMB dynamics and a state-space equation is arranged as equation(1)[7].

By using the state-space model described above, the initial PD controller can be designed and FRF (Frequency response function) of AMB can be measured. The control system for AMB is implemented by using xPC Target of MathWorks, Inc which is the real time digital control system [5]. Figure 11 shows the configuration of control system using xPC Target. The control block diagram is written by using MATLAB/Simulink in Host PC and it is downloaded into Target PC. During the control code is running in Target PC, Host PC can change the control variables in Target PC and get the each signal data easily from Target PC through Ethernet. Target PC reads the value of the position sensor through A/D and calculates the control input and sends it to Power Amp through D/A. Power Amp transfers the current corresponding to the control input to the coil of AMB actuators.



confirmed that the performance of AMB is satisfactory and there is enough radial margin between the flywheel shaft and the touchdown bearings which is located from the flywheel shaft with the gap of 300  $\mu\text{m}$ .

1x runout component which is most of the overall runout is due to the imbalance response of the flywheel rotor. The magnitude of the imbalance response changes depending on the rotational speed reflecting the rotor vibration mode. Figure 16 shows the magnitude change of 1x component in the upper AMB from standstill to 15,000rpm. It can be clearly observed that the cylindrical and conical mode exist at 1,200rpm and 3,000rpm each. Zero-to-peak runout is less than 17  $\mu\text{m}$  even when the rotational speed passes through the cylindrical critical speed. It can be also seen that after the rotational speed passes through the conical critical speed, AMB runs stably with the runout less than 10  $\mu\text{m}$ .

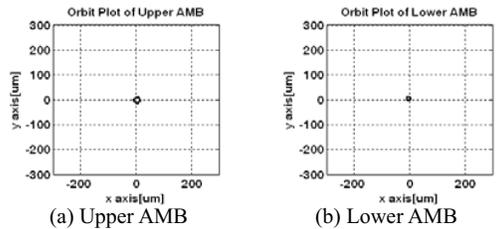


Figure 15 Orbit plot of AMB at 15,000rpm

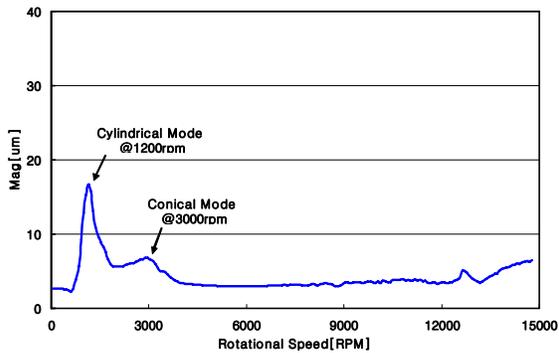


Figure 16 Magnitude of 1x component versus the rotational speed in upper AMB

## 6. CONCLUSION

In this paper, the design procedure including rotor dynamic analysis of the 5kWh class FESS mounted on AMB were introduced and it was shown that AMB have run stably up to target speed, 15,000rpm by using simple PD controller and notch filters. Rotor dynamic analysis was performed for the shaft and full flywheel

respectively and it was verified that the analytical results were sufficiently similar to the impact test results. The validity of the analytical results was confirmed by the measured FRF as well. It was also shown that the flywheel design to make its  $I_p/I_t$  0.5 was effective to minimize the unbalance response of the flywheel and reduce the control input to the AMB actuators. The controller design and implementation procedure using Matlab /Simulink and xPC Target was very simple and straightforward. In the experiment, FRFs were measured easily and the monitoring signal was gathered handily by using xPC Target, too.

## REFERENCES

- [1] R. Okow, A.B Sebitosi, A. Khan and P. Pillay, "The Potential Impact of Small-scale Flywheel Energy Storage Technology on Uganda's Energy Sector", **Journal of Energy in Southern Africa**, Vol. 20, No. 1, 2009, pp 14-19.
- [2] P. Gyuk and S. Eckroad, "EPRI-DOE Handbook Supplement of Energy Storage for Grid Connected Wind Generation Applications", EPRI, Palo Alto, CA, USA, 2004.
- [3] I. Hadjipaschalis, A. Paullikkas and V. Efthimiou, "Overview of Current and Future Energy Storage Technologies for Electric Power Applications", **Renewable and Sustainable Energy Reviews**, Vol. 13, 2009, pp. 1513-1522.
- [4] C.H. Park, S.K. Choi, J.P. Lee and Y.H. Han, "On the Dynamic Behavior of a 5kWh FESS Mounted on AMBs", **the 11<sup>th</sup> International Conference on Mechatronics Technology**, Ulsan, Korea, 2007, pp.416-420.
- [5] "xPC Target<sup>TM</sup> User's Guide", The MATLAB Works, Inc., Natick, MA, USA, 2007.
- [6] J. M. Vance, "RotorDynamics of TurboMachinery", A Wiley-Interscience Publication, USA, 1988, pp. 116-130.
- [7] S.Sivrioglu and K. Nonami, "Active permanent magnet support for a superconducting magnetic-bearing flywheel rotor", **IEEE Transactions on Applied Superconductivity**, Vol. 10, 2000, pp 1673-1677.
- [8] D. Clark, "Developing improved servos for the multiple mirror telescope", **American Control Conference, Proceedings of the 2005**, Vol, 6, 2005, pp. 3773-3775.
- [9] M. Ahrens, L. Kucera and R. Larsonneur, "Performance of a magnetically suspended flywheel energy storage device", **IEEE Transactions on control systems technology**, Vol. 4, 1996, pp. 494-502

## Controller Design and Simulation for Tracking Mount of Movable SLR, ARGO-M

Cheol Hoon PARK\* Young Su SON Sang Yong HAM Sang Yong HAM Sung Whee Lee  
Nano Convergence and Manufacturing Systems Research Division,  
Korea Institute of Machinery and Materials,  
Yuseong-gu, Daejeon, 305-343, Korea

\*Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology,  
Yuseong-gu, Daejeon, 305-701, Korea

Hyung Chul LIM  
Space Science Division, Korea Astronomy & Space Science Institute,  
Yuseong-gu, Daejeon 305-348, Korea

### ABSTRACT

Controller design procedure for prototype tracking mount of Movable SLR (Satellite Laser Ranging), ARGO-M is presented. Tracking mount of ARGO-M is altitude-azimuth type and it has two axes of elevation and azimuth to control its position. Controller consists of velocity and acceleration feed-forward controller, position controller at outer loop, velocity controller at inner loop. There are two kinds of position control modes. One is pointing mode to move from one position to the other position as fast as possible and the other one is tracking mode to follow SLR trajectory as precise as possible. Because the requirement of tracking accuracy is less than 5 arcsec and it is very tight error budget, a sophisticated controller needs to be prepared to meet the accuracy. Especially, ARGO-M is using the cross-roller bearing at each axis to increase the mechanical accuracy, which requires add-on controller DOB (Disturbance observer) to suppress friction load and low frequency disturbances. The pointing and tracking performance of the designed controller is simulated and visualized using MATLAB/Simulink & SimMechanics and the experimental results using test are presented as well.

Keywords: Satellite Laser Ranging, Tracking Mount, SimMechanics, Tracking accuracy.

### 1. INTRODUCTION

Currently, Korea Astronomy and Space Science Institute (KASI) is concentrating on a study to develop

an altitude-azimuth type movable satellite laser ranging (SLR) and its tracking mount is under development by Korea Institute of Machinery & Materials (KIMM) [1][2][3]. For the development of a robust and precise high-speed tracking mount, the core components including motors, encoders and bearings were selected, and the proto type mount was designed based on them. In this paper, we present the proto type of ARGO-M tracking mount and the procedure of controller design. The verification of the designed proto type mount and controller would be confirmed through the simulation using MATLAB/ Simulink and SimMechanics.

### 2. SYSTEM DESCRIPTION

The mounts used for astronomical telescopes are classified into altitude-azimuth types and equatorial types. ARGO-M tracking mount adopted altitude-azimuth type which is used in most of SLR systems that have been developed earlier. Figure 1 shows the preliminary design of ARGO-M tracking mount.

The main parts that consist of tracking mount are the mount base and frame, optical interface and axis-driving parts. Tracking mount supports the reception optical system of 40 mm-caliber and transmission optical system of 10 mm-caliber. In order to avoid the damage to the optical system due to the abrupt motion, driving speeds for azimuth and elevation axes are limited to 20 deg/sec and 10 deg/sec, respectively. Driving range for azimuth axis is  $\pm 350$  deg which is almost two revolutions from one side rotational limit to the other one, and driving range for elevation axis is 0–180 deg which is a half revolution from one horizontal

limit to the other one. ARGO-M will track satellites which are located in the altitude of 300-25,000 km and the pointing accuracy should be less than 5 arcsec, or 5 over 3,600 deg. Table 1 describe the specifications of ARGO-M tracking mount.

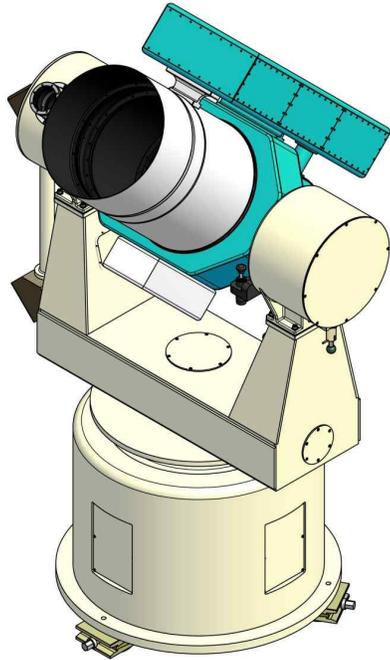


Figure 1. Proto type of ARGO-M tracking mount

Table 1 Specifications of ARGO-M tracking mount

Item	Specifications
Mount type	Alt-Azimuth (EL over AZ)
Telescope caliber	Separable transmission-reception type (10 cm, 40cm)
Mount driving speed	Azimuth: Max 20 deg/sec Elevation: Max 10 deg/sec
Mount driving acceleration	Azimuth: 5deg/sec <sup>2</sup> or higher Elevation: 2deg/sec <sup>2</sup> or higher
Mount driving range	Azimuth : ±350deg Elevation: 0-180deg
Satellite altitude range	300 – 25,000 km
Mount pointing accuracy	Less than 5 arcsec

Core components such as motors, encoders and bearing were selected to meet the operating specifications. The continuous torque of motors for AZ and EL axes are 344 N·m and 138 N·m , respectively, and the encoder resolution for AZ and EL axes are 0.020 arcsec and 0.032 arcsec, respectively. The cross roller bearing with the precision class of P2 was selected for both axis.

### 3. CONTRLLER DESIGN

Before fabricating the mount, it was necessary to predict the tracking and pointing accuracy of ARGO-M tracking mount based on the designed proto type mount. The simulation was performed by using MATLAB/Simulink shown as Figure 2. Generally, plant model for the dynamic simulation is constructed from the derivation of the equation of motion. In this study, the plant model was constructed by using Simulink/SimMechanics [4]. By using SimMechanics, three-dimensional CAD model which is constructed using software such as Solidworks can be converted to a multi-body dynamic model, and it can be included in MATLAB/Simulink model and the control simulation can be easily conducted. Figure 3 is the SimMechanics block diagram of ARGO-M tracking mount which has 2 mass bodies and 2revolute joints.

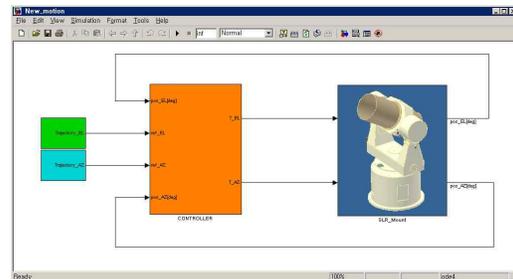


Figure 2. Simulink block diagram to simulate the control performance of ARGO-M tracking mount

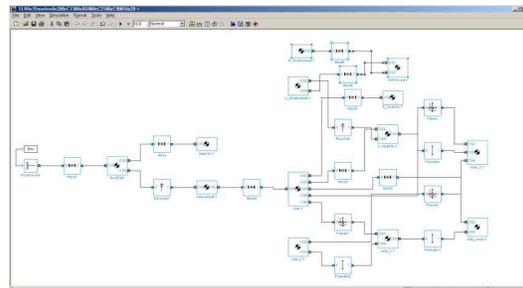


Figure 3. SimMechanics block diagram of ARGO-M tracking mount

The controller was designed for both of pointing motion and tracking motion. For the pointing motion, the controller tries to rotate the mount from one location to another as quickly as possible without residual vibration. There is a trajectory generator to produce the position, velocity and acceleration trajectory for the purpose of feed-forward. The structure of controller is a dual loop, which is same for each AZ and EL axis. The dual loop consists of an inner velocity controller and outer position controller and the control method is the proportional-integral (PI) control and disturbance observer (DOB) to suppress the friction disturbances of bearings. Figure 4 shows the controller block diagram for both of AZ and EL axis. After selecting PI gains for each velocity and position controller of each axis, they are optimized by using Simulink/Optimization toolbox.

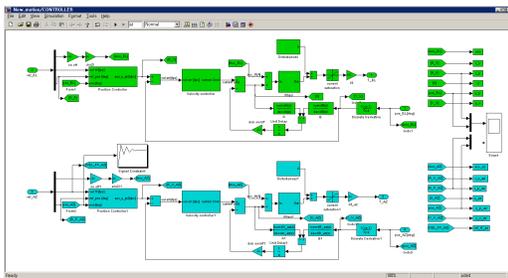


Figure 4. Simulink block diagram of controller for ARGO-M tracking mount

#### 4. SIMULATION RESULTS

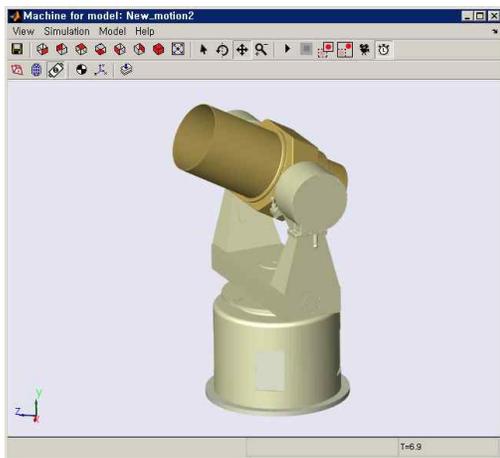


Figure 5. Animation in SimMechanics to show the motion of ARGO-M tracking mount

The one of merits which can be obtained using SimMechanics is that it is easy to visualize the motion

of tracking mount during simulation shown as Figure 5. It can help the designer improve the feeling of plant motion, select the right operating directions and make it easy to explain the simulation results to the other persons.

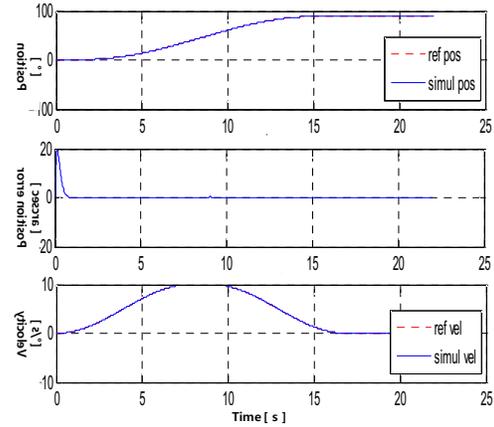


Figure 6. Simulation results of EL axis pointing motion

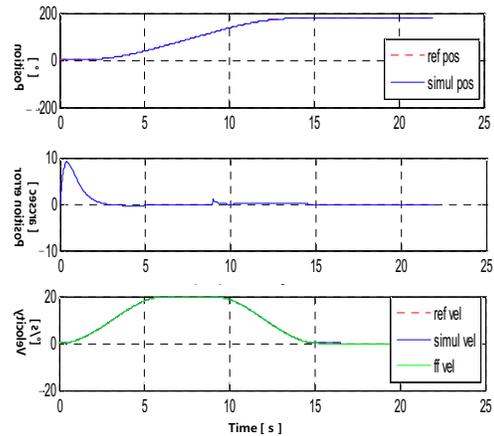


Figure 7. Simulation results of AZ axis pointing motion

Firstly, pointing motion simulation for the tracking mount was performed. Pointing reference position was given from 0 to 90 deg for elevation axis, and from 0 to 180 deg to azimuth axis. The simulation results are shown in figure 6 and 7. During pointing operation, max velocities for EL and AZ axes are limited to 10 and 20 deg/sec, respectively, as described in the specifications of Table 1. Reference position angles from starting point to end point are given with the integral of velocity trajectories. The position errors are converged less than 0.1arcsec within 3sec and the

controller keeps the small position errors under the pointing error specification of 5 arcsec with enough margins for both axes.

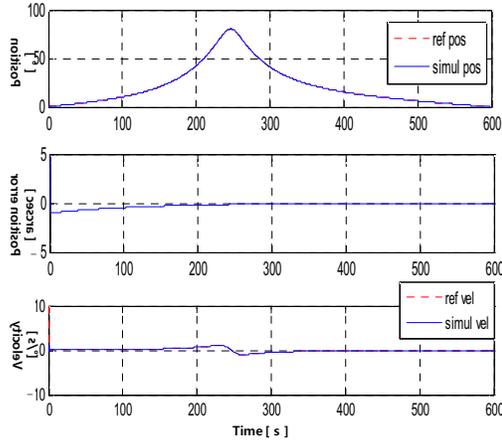


Figure 8. Simulation results of EL axis tracking motion

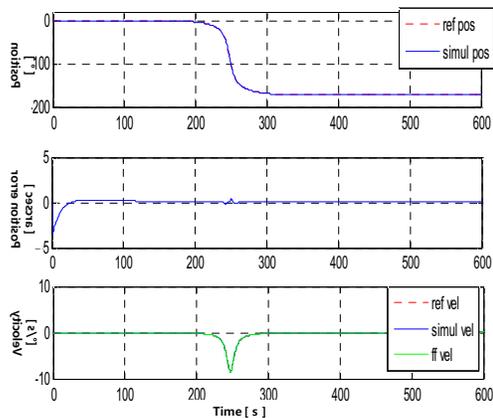


Figure 9. Simulation results of AZ axis tracking motion

Secondly, tracking motion simulation for the tracking mount was performed using the trajectories of International Space Station (ISS). The average altitude of the ISS is 373.7 km, and it is the representative of the satellites at the lowest altitude in comparison with the altitude specification of the satellites to be tracked by the ARGO-M (300 – 25,000 km). Therefore, the trajectories of this satellite can be the representative for all the required speed ranges along each axis to track satellites from the lowest speed to the highest speed. Velocity feedforwards for each axis were injected using the derivative of the position trajectories of ISS. The maximum velocities for EL and AZ axes during tracking simulation are 1 and 8.5 deg/sec, respectively.

It was predicted that the position errors are less than 1 arcsec for the whole tracking motion with sufficient margins for both axes.

### 5. CONCLUSION

In this study, we have presented the controller design procedure for the proto type design of ARGO-M tracking mount. Plant model for tracking mount is constructed using MATLAB/SimMechanics, and the motion simulation is performed using MATLAB/Simulink. From the simulation results, it was predicted that both of pointing and tracking motion could meet the accuracy specification with sufficient margin. However, this simulation has a limitation that it can reflect the rigid body motion only although there is a couple of flexible modes in the real tracking mount. Currently, the proto type mount is under construction and we will complement the controller with a couple of filters to suppress the flexible mode vibration in the stage of real controller implementation.

### REFERENCES

- [1] J.H. Jo, et al. "The design concept of the first mobile satellite laser ranging system (ARGO-M) in korea", *Journal of Astronomy and Space Science*, Vol 28, No 1, 2011, pp 93-112.
- [2] H.C. Lim, et al. "Tracking capability analysis of ARGO-M satellite laser ranging system for STSAT-2 and KOMPsat-5", *Journal of Astronomy and Space Science*, Vol 27, No 3, 2010, pp 245-252.
- [3] C.H. Park, Y.S. Son, and B.I. Kim "Preliminary design of tracking mount for movable SLR", *Journal of Astronomy and Space Science*, Vol 27, No 2, 2010, pp 135-144.
- [4] MathWorks, "SimMechanics User's Guide", the MathWorks., Natick, MA, USA

# The Development of a Propellantless Space Debris Mitigation Drag Sail for LEO Satellites

Michael Pfisterer, Kevin Schillo, Christopher Valle, and Kuo-Chi Lin  
Department of Mechanical, Materials, and Aerospace Engineering, University of Central Florida  
Orlando, FL 32816, United States

and

Chan Ham  
Department of Mechatronics Engineering, Southern Polytechnic State University  
Marietta, GA 30060, United States

## ABSTRACT

The KnightSat II is the University of Central Florida's entry in the sixth iteration of the University Nanosatellite Program. The principle objective of the satellite is to deploy a gossamer sail, which will dramatically increase the aerodynamic drag acting on the satellite and reduce the time it will take to deorbit. A prototype sail and deployment system has been successfully built and tested. The other satellite subsystems have undergone extensive design and testing, and have been integrated into a working satellite prototype.

**Keywords:** Nanosatellite, Space Debris, Drag Sail, Mitigation, Propellantless Deorbit

## 1. INTRODUCTION

The issue of space debris is a growing concern. The U.S. Space Surveillance Network tracks over 19,000 objects greater than ten centimeters in diameter, and hundreds of thousands of smaller objects. All of these objects pose a serious threat to both commercial and scientific spacecraft [1].

A recent event that highlights this danger was the destruction of the Iridium 33 communications satellite when it collided with the defunct Russian Kosmos-2251. This collision generated over two thousand pieces of new debris [2]. The International Space Station has been forced to execute maneuvers to avoid collisions with debris, which further highlights the danger presented by these objects [1].

These concerns have led to the more strict enforcement of the NASA Procedural Requirements for Limiting Orbital Debris, which requires LEO spacecraft to limit their orbit lifetimes to twenty-five years [3]. NASA Technical Standard 8719.14 states that objects in LEO should be removed or placed in a storage orbit within a twenty-five year period. Objects orbiting at altitudes of less than six hundred kilometers are of little concern, as atmospheric drag typically limits their orbital lifetimes to no greater than twenty-five years [3]. Objects in altitude greater than seven hundred kilometers are of the greatest concern, as they may remain in orbit for centuries.

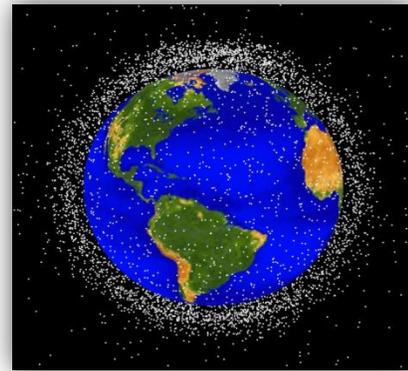
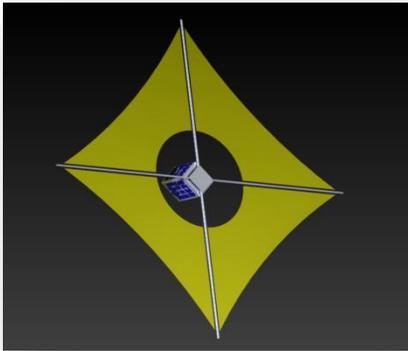


Figure 1 : A Simulation showing the larger debris in LEO.  
(NASA Orbital Debris Office)

Traditionally, large spacecraft have used onboard propulsion systems to maneuver in orbit and position themselves for removal. Smaller spacecraft generally lack these propulsion systems, and are restricted to an orbit of six hundred kilometers or lower to ensure that they will deorbit within the twenty-five year limit [4]. This has led to the development of propellantless orbital maneuvering systems suitable for smaller spacecraft. One of the simplest of these systems works by increasing the cross-sectional area of the spacecraft. This leads to a dramatic increase in the aerodynamic drag acting on the satellite, effectively increasing the rate of orbital decay. NASA's recent Nanosail D deployed a small experimental solar sail, which has potential to be developed into a breaking system. The future CNES mission MICROSCOPE will utilize a gossamer wing structure in an attempt to reduce its expected orbit time from sixty-seven years to twenty-five years [3]. While these systems work in their specific applications, they either cannot be scaled up to larger spacecraft with their current design, or they do not permit stable flight.

The proposed Attitude Control and Aerodynamic Drag Sail (ACADS) system will allow for a drastic increase in cross-sectional area from a compact, lightweight storage system utilizing the unique folding and deployment of a gossamer sail. This system has been tested using sails as small as 2x2 meters and as large as 20x20 meters, allowing it to be used on spacecraft of various sizes. The uniform deployment of ACADS, along with its perimeter magnetic torque coil will allow for stable flight and provide the potential for future development of propellant-less multi-orbit missions. Simulation

results show that the ACADS system to be very effective in reducing the deorbit time of the satellite, which will be described in detail in a later section of this paper.



**Figure 2: Deployed ACADS attached to KnightSat II**

The ACADS sail system was developed as the experimental payload of the KnightSat II project as part of the Air Force Research Lab University Nanosatellite Program (UNP). The UNP provides the opportunity for university student to work with professors and industry professionals on an amateur spacecraft while competing for a launch opportunity. Students were required to design all the spacecraft subsystems from determining system requirement to delivering functioning prototypes.

This paper will detail the design development of the Knightsat II spacecraft, with the primary focus being the ACADS payloads, covering the design requirements and overviews of the support subsystems and the spacecraft structure. This will be followed by a discussion of the simulations used to determine the effectiveness of the sail, as well as the ability of the spacecraft structure to meet UNP requirements.

## 2. DEVELOPMENT OF ACADS AND KNIGHTSAT II

### Missions and Requirements

The missions and primary requirements of the KnightSat II spacecraft were based on the payload, the ACADS sail system. The mission of the KnightSat II team is follows:

“The KnightSat II team seeks to design and build a satellite for research and demonstration of a novel propellantless propulsion and attitude control system utilizing a deployable gossamer sail.”

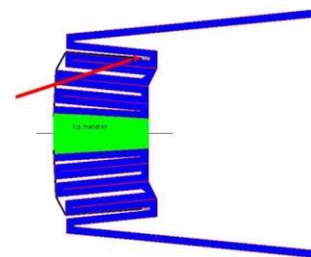
The specific demonstration for this system would be to facilitate rapid deorbit of the spacecraft with the ACADS sail. The ACADS system achieves this effect by way of aerodynamic drag. This is accomplished by drastically increasing the cross-sectional area of the spacecraft, which will lead to a massive increase in aerodynamic drag. Research was conducted to determine the requirements for drag enhancement devices on LEO spacecraft. Changing the cross sectional area of the spacecraft so drastically, in this case over 30 times, also

increases the likelihood of collision. NSS-1740-14 addresses this issue and states the area time product of the system must be reduced [5]. With this information, the requirement for acceptable time for de-orbit can be determined. In the case of KnightSat II, the orbit time must be reduced by greater than 30 times. Simulations had to be conducted to ensure the design met these requirements.

NASA Technical Standard 8719.14 also mentions the use of drag enhancement devices such as ACADS must significantly reduce the collision risk of the spacecraft and generating large debris [3]. With ACADS’ increase in cross-sectional area, there is an increased likelihood of impact with orbital debris. With this in mind, it must be taken into consideration that the vast majority of that area consists of extremely thin Kapton, and that any collision with it will likely only create a hole in the sail, and not further debris. Initial findings and interpretations lead the KnightSat team to believe that the increased area due to the sail can be disregarded, and only the boom lengths considered. If the sail component can be disregarded, the area-time product and collision probability would be affected drastically. This requirement is still being researched, and once further explored with input from technical experts from the NASA Orbital Debris Program Office its impact on the final design and deployment of ACADS can be determined.

### ACADS Design

The sail system consists of a square stowage box with 4 doors that open outward and permit deployment of the sail. Within the box is an electronics package, pressure vessel, sail, and booms. The electronics control the sequence of deployment, and ensure that the booms deploy in a uniform manner. The pressure vessel provides nitrogen gas with which to inflate the booms. The sail is stored in four segments and each segment is stored on one side of the box. The sail segments are attached to four booms that are conically folded within themselves at each corner of the box. At the perimeter of the sail are ten turns of 30 gauge wire. The wire leads will be laid down the inside of the booms and run through the stowage box and interface with the satellite’s C&DH and power systems. These perimeter coils will be used for attitude control experiments, utilizing the large surface area provided by the sail. The sail itself is made of 3µm Kapton, which is 70% emissive and will allow for light to reach the solar panels and minimize the effect of solar pressure on the satellite. The booms are made of Sub-Tg resin impregnated Vectran fabric, which is pliable when heated and becomes rigid once cooled. Heaters will be used to prepare the boom segments for deployment and disabled when deployment is complete. The booms will be painted in reflective coatings to prevent loss of rigidity once deployed. This material has successfully been utilized in spaceflight.



**Figure 3: Conical folding of boom segments**

The stowed dimensions of the box are 35x35 centimeters. The height is 5cm. The deployed dimensions of the stowage box are 45x45 centimeters and will fit flush with the top surface of the spacecraft. Deployed, the sail will have an area of 10 square meters. The projected mass of the entire sail system is expected to be less than 5 kilograms. The dimensions of the sail are represented below:

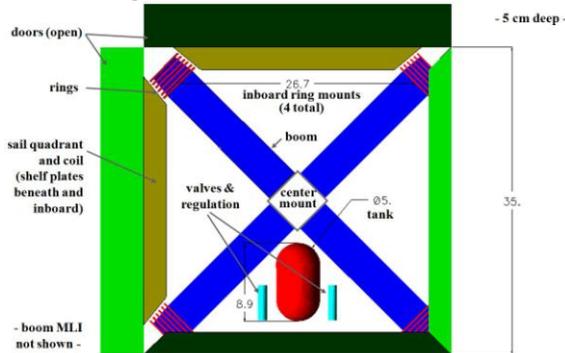


Figure 4a: Stored ACADS sail

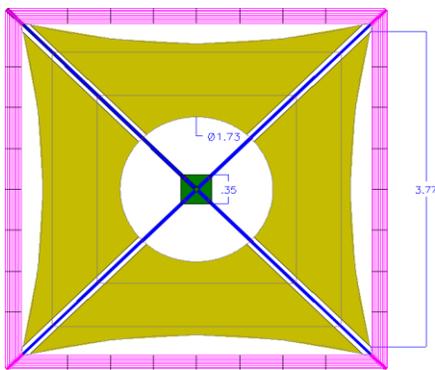


Figure 4b: Deployed ACADS sail

The construction of the sail is being contracted to L'garde Inc., who have fabricated and tested larger 400 square meter sail prototypes that deployed successfully in vacuum chambers and function in the same fashion as KnightSat II's payload. They have also successfully deployed inflatables in space, such as the Spartan-207 Inflatable Antenna Experiment deployed from STS-77 [6].

### Safety

Safety is a major concern with ACADS because of the use of pressurized gas in its deployment. The major concern with ACADS is premature deployment, either on the ground or during launch. To address this, the booms and pressure vessel will be vented to the atmosphere/space while on the ground and during the ascent of the launch vehicles. Not until KnightSat II has been released from the spacecraft, charged, and capable of sending power to ACADS would the ascent valves be closed and deployment possible. Electrical inhibits and redundant closed valves between the pressure vessel and the regulator supplying pressure to the booms provide additional mechanisms preventing unwanted deployment.

It has been recommended by NASA personnel that ACADS have at least a two fault tolerant system. Currently research is being conducted as to exactly which components, such as valves and regulators, would be suitable for use in the

system. The current concept for the fault tolerance system is presented in the diagram below:

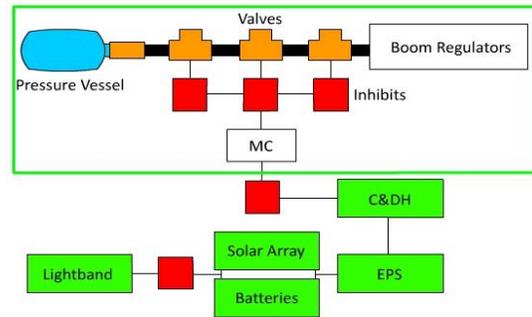


Figure 5. Sail fault tolerance system.

## 3. SPACECRAFT SUBSYSTEMS

### Command and Data Handling

The spacecraft utilizes an Atmel AVR-32 bit MCU on a custom designed board for control of the spacecraft and health monitoring. The Knightsat II software solution will consist of a main loop that prioritizes and calls other programs or functions. The power, communications, and thermal programs have highest priority. In addition to basic spacecraft control, the flight computer will also be responsible for the deployment of the ACADS system.

### Attitude Determination and Control System

Initial stabilization of the spacecraft prior to the deployment of ACADS will be achieved by using three magnetic torque rods. Each of the rods is composed of EFi alloy 50, and wrapped with more than two thousand turns of 28 AWG wire. An RCM5700 processor is used to operate the ADCS. The CPU is equipped with three angular rate gyros, three orthogonal DC accelerometers, three orthogonal magnetometers, a multiplexer, a 16-bit analog/digital converter, an embedded microcontroller, and six pulse width modulators. The pulse width modulators are connected to the torque rods, and can be used to manipulate the magnitude of the current that is supplied to the rods.

An H-bridge circuit was constructed using twelve 2n2222A transistors. The H-bridge is able to control the direction of the current to be controlled, enabling complete manipulation of the magnetic torque generated by the rods. A 3DM-GX1 Inertial Motion Unit and a sun sensor provide the necessary data for attitude determination, such as Euler angles, angular velocities, and magnetic field vectors. Data from the IMU and sun sensor will be relayed to the primary flight computer for compression and transmission to the ground.

### Communications

The communications system will utilize two CalAmp JSLM2 transceivers connected to a Kantronics 9612+ modem, and will be responsible for sending data to and receiving commands from the ground station.

### Electrical Power System

The EPS will perform three primary functions to support the objective of KnightSatII: generate, store, and distribute all electrical power to be used by the spacecraft. Power will be

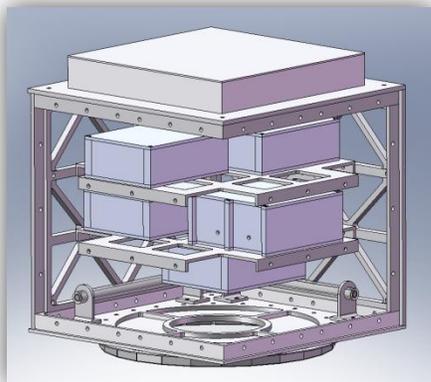
generated through photovoltaics mounted on nonstructural plates on four sides of the spacecraft. Rechargeable batteries will be used for power storage to be utilized during the eclipse portion of the orbit, supplement solar arrays, and maintaining the 9.6V bus for the spacecraft. The primary task for the distribution system will be converting the bus voltage to an appropriate operating voltage, either increase or decrease, for specific components on the spacecraft. The EPS will not be controlling power directly; this role has been passed on to the C&DH subsystem.

**Thermal Control**

KnightSat II will utilize a mostly passive thermal control scheme. Electric resistance heaters will be used to keep the batteries at a suitable temperature, and preliminary analysis shows the rest of the spacecraft will be at sufficient temperature.

**Structure**

The structure of KnightSat II was designed to be modular so it would be capable of accepting changes in the subsystems and be easy to work with during assembly. This was accomplished by using an isogrid pattern. The large pattern allowed ease of access to the internals and the ability for the structure to support itself with multiple walls removed during integration. The solar panels are designed to be mounted with standoffs so that the panel itself could be added or the wall and panel can be its own subassembly. The structure was composed of 6061 T6 Aluminum to save cost and chemical film coated to meet standards. Since no launch vehicle has been selected for this mission, specific UNP6 requirements called for the structure to have a fundamental frequency greater than 100 Hertz, and be able to withstand a load of twenty g's in each axis so that as a secondary payload it may be mounted in any way and survive the stresses and vibrations of launch.



**Figure 5: KightSatII structure cutaway**

**4. ANALYSIS AND SIMULATION**

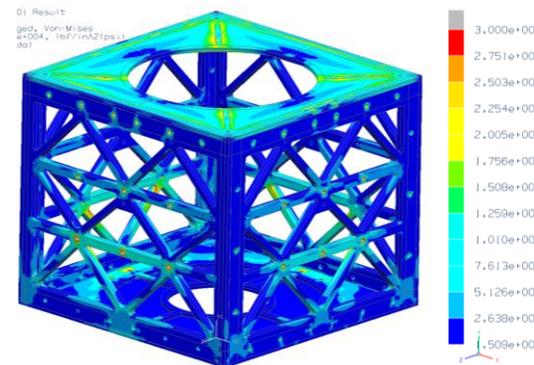
**Structure**

Analysis was conducted on the KnightSat II structure to ensure that it met UNP requirements in both static loading and fundamental frequency. Masses simulating spacecraft components and ACADS were applied prior to simulation. The

static load analysis showed the structure was capable of withstanding the required acceleration of twenty g's in any direction.

**Table 1: Static load simulation results**

Load Case #	Load	Maximum Vonmises (ksi)	Minimum Margin of Safety, Ultimate
1	20Gs (+x)	3.2	4.0
2	20Gs (-x)	3.2	4.0
3	20Gs (+y)	7.3	1.2
4	20Gs (-y)	7.3	1.2
5	20Gs (+z)	3.1	4.2
6	20Gs (-z)	3.1	4.2

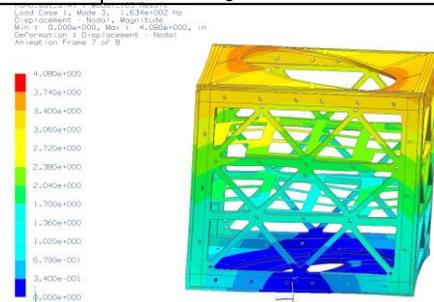


**Figure 6: Static load simulation results**

The modal analysis conducted showed that the structure and all of its elements met the fundamental frequency requirements set by the UNP.

**Table 2: Modal analysis result**

Mode #	Frequency (Hz)	Description
1	107	First bending of lower mounting plate.
2	145	First bending of top plate.
3	163	General twisting motion about attachment location.



**Figure 7: Modal analysis results**

**ACADS Capability: Drag**

A number of simulations have been conducted by students at the University of Central Florida to predict the performance of the spacecraft in orbit with the deployment of the sail. These simulations showed promising results, in that the

sail produced a significant enough drag force to reduce the orbit time by a large margin. The effects are significant enough to justify the large increase of cross-sectional area when the sail is deployed.



Figure 8: Simulation of KnightSat II without ACADS

The same simulation was conducted with the ACADS system deployed, which resulted in a significant increase in the rate of orbital decay.

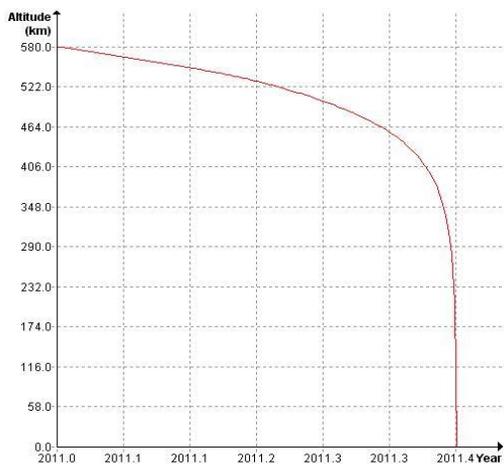


Figure 9: Simulation of KnightSat II without ACADS Deployed

The simulation shows the spacecraft falling from orbit in under six months, and an increase in the possible operating ceiling from the previous six hundred kilometer limit to nine hundred kilometers. This shows that the deployment of the sail brings the spacecraft out of orbit fifty times faster, while creating a surface area thirty-three times greater than the satellite without the sail. This means that the sail successfully decreased the area-time product of the spacecraft as required by NSS-1740-14.

### ACADS Capability: Stabilization

In addition to functioning as an effective deorbiting mechanism, the sail is also envisioned as providing an integral aspect of the attitude control system. The perimeter of the sail will be laced with ten turns of magnetic coils, which, combined with two magnetic torque rods located within the satellite, will provide magnetic attitude control.

A simulation of this was created in Simulink that implemented six-degrees-of-freedom equations of motion with respect to body axis. The results of this simulation show that it would take about two hours for the satellite to stabilize. Clearly this is not a system for a satellite mission in which rapid stabilization is required, but for the mission objectives of the KnightSat II, it is sufficiently effective.

## 5. CONCLUSION

The KnightSat team has completed the majority of the top level research and design concept work for the sail. Additionally, the analysis conducted by the team shows that the system has the capability of meeting its objectives and providing an effective means of altering the orbit of KnightSat II and eventually removing it from LEO completely at an accelerated rate.

A future goal of the project is to use the ACADS system in conjunction with an electrodynamic tether as a propellantless orbital maneuvering system for LEO spacecraft. This can enable long term, inexpensive multi-orbit missions using the sail to lower the orbit, and current running through the tether to increase it.

## ACKNOWLEDGEMENT

This research is sponsored by Air Force Office of Scientific Research (AFOSR) under the 6th Air Force University Nanosatellite Program (UNP-6), Contract Number: FA95500910122

## REFERENCES

- [1] Madler, Ronald A. "Orbital Debris—A Space Hazard," in *Space Mission Analysis and Design*, Third Edition.
- [2] Liou, J.-C. and Shoots, D. "Satellite Collision Leaves Significant Debris Clouds." *Orbital Debris Quarterly News*. Volume 13, Issue 2. April 2009.
- [3] "NASA Technical Standard 8719.14, Process for Limiting Orbital Debris." National Aeronautics and Space Administration, Washington, DC.
- [4] Wertz, James R. "Earth Coverage," in *Space Mission Analysis and Design*, Third Edition.
- [5] NASA Safety Standard: Guidelines and Assessment Procedures for Limiting Orbital Debris. (NSS 1740.14, August 95).
- [6] "Spartan 207/Inflatable Antenna Experiment Flown on STS-77." Preliminary Mission Report." NASA Goddard Space Flight Center.

# Qualification of Electrical Ground Support Equipment for New Space Programs

Felix A. SOTO TORO, Ph.D.  
 Electrical Engineering Division, NASA  
 Kennedy Space Center, FL 32899, USA

and

Bruce T. VU, Ph.D.  
 Mechanical Engineering Division, NASA  
 Kennedy Space Center, FL 32899, USA

and

Mark S. HAMILTON  
 Materials Science Division, NASA  
 Kennedy Space Center, FL 32899, USA

## ABSTRACT

With the Space Shuttle program coming to an end, the National Aeronautics and Space Administration (NASA) is moving to a new space flight program that will allow expeditions beyond low earth orbit. The space vehicles required to comply with these missions will be carrying heavy payloads. This implies that the Earth departure stage capabilities must be of higher magnitudes, given the current propulsion technology. The engineering design of the new flight hardware comes with some structural, thermal, propulsion and other subsystems' challenges. Meanwhile, the necessary ground support equipment (GSE) used to test, validate, verify and process the flight hardware must withstand the new program specifications.

This paper intends to provide the qualification considerations during implementation of new electrical GSE for space programs. A team of engineers was formed to embark on this task, and facilitate the logistics process and ensure that the electrical, mechanical and fluids subsystems conduct the proper level of testing. Ultimately, each subsystem must certify that each piece of ground support equipment used in the field is capable of withstanding the strenuous vibration, acoustics, environmental, thermal and Electromagnetic Interference (EMI) levels experienced during pre-launch, launch and post-launch activities. The benefits of capturing and sharing these findings will provide technical, cost savings and schedule impacts information to both the technical and management community.

**Keywords:** Qualification; Testing; Ground Support Equipment; Electromagnetic Interference Testing; Vibration Testing; Acoustic Testing; Power Spectral Density.

## 1. INTRODUCTION

There are many risks associated with Space Flights design and development. Many of these risks are known, but some of them are unknown. New space programs have the benefit of learning from previous programs, but there are disadvantages. The new program requirements bring questions and problems that must be addressed using the best judgment possible. Building and performing multiple tests have proven to mitigate many unknowns. Test facilities play a vital role in the determination of limits to which both ground and flight hardware are tested. However, budget limitations may preclude from performing test on every single piece of hardware and software. Thus, engineering analysis and judgment are the next best level of conformance.

## 2. CLASSIFICATION OF HAZARD LOCATIONS

In accordance with the Occupational Safety and Health Administration (OSHA), the National Electrical Code (NEC) defines hazardous locations as those areas "where fire or explosion hazards may exist due to flammable gases or vapors, flammable liquids, combustible dust, or ignitable fibers or flyings." Table 2.1 classifies each of those categories<sup>1</sup>.

CLASSES	GROUPS	DIVISIONS	
		1	2
I Gases, vapors, and liquids (Art. 501)	A: Acetylene B: Hydrogen, etc. C: Ether, etc. D: Hydrocarbons, fuels, solvents, etc.	Normally explosive and hazardous	Not normally present in an explosive concentration (but may accidentally exist)
II Dusts (Art. 502)	E: Metal dusts (conductive, and explosive) F: Carbon dusts (some are conductive, and all are explosive) G: Flour, starch, grain, combustible plastic or chemical dust (explosive)	Ignitable quantities of dust normally are or may be in suspension, or conductive dust may be present	Dust not normally suspended in an ignitable concentration (but may accidentally exist). Dust layers are present.
III Fibers and flyings (Art. 503)	Textiles, wood-working, etc. (easily ignitable, but not likely to be explosive)	Handled or used in manufacturing	Stored or handled in storage (exclusive of manufacturing)

The NASA's Kennedy Space Center (KSC) facilities and operational areas comply with these standards and classify its activities in accordance with the following categories:

- 1) Hazardous Location
  - A - Class 1 Div 1
  - B - Class 1 Div 2
  - C - Non-hazardous

Thus, when designing avionics, electrical, fluids, mechanical and pneumatic systems, personnel need to consider the worse environment to which their systems can be exposed. Hence, systems engineers must test and qualify their hardware to minimize the risk of failure.

## 3. QUALIFICATION METHODS

The NASA location at the Kennedy Space Center (KSC) in Florida, the agency designs, develops, integrates, tests, processes and operates the Ground Support Equipment (GSE) necessary to achieve safe Space Flight Readiness. Despite being aware of the numerous "unknowns" in the process's Life Cycle, it is understood that one way to mitigate and/or reduce the probability of failure of its GSE, is to perform multiple tests to the multiple systems. The location of use of each subsystem's assembly and components is predetermined based on the architectural design. Depending on the location, the environmental conditions and intensity experienced by similar components may be different. Therefore, in order to conduct sound engineering assessment, each component is ground tested to the specifications of its worse load environment. For example, similar components are used in two completely different environments, then such GSE component is qualified by testing to the more severe environmental load levels.

There are budgetary constraints when complying with this testing methodology. Component qualification testing can become very expensive very quickly. To obtain the best test results, each test must be conducted independently from other tests. Each test may require different personnel skill set, different test equipment and different facilities. A secondary concern is experienced when the system's design personnel is supporting the Qualification Testing phase, instead of supporting another project's phase such as design, development, procurement, assembly, analysis, functionality testing, integration, verification and validation. Additionally, there is a possibility that the GSE under testing may not be approved for use in the field. The GSE units that underwent qualification tests can only be used in the field if a panel of experts determines that there is no risk associated with using a unit that underwent thru a Qualification Test.

There are no set sequences for testing GSE, only suggested sequences<sup>11</sup>. It may be a matter of resources availability or the determination to conduct test to destructive limits. Ideally, all tests are performed using the same set of GSE components and all components survive all tests. The flowchart depicted in Figure 3.1 shows a typical sequence of events that should be considered during subsystems' component qualification of GSE for new Space Programs.<sup>2</sup>

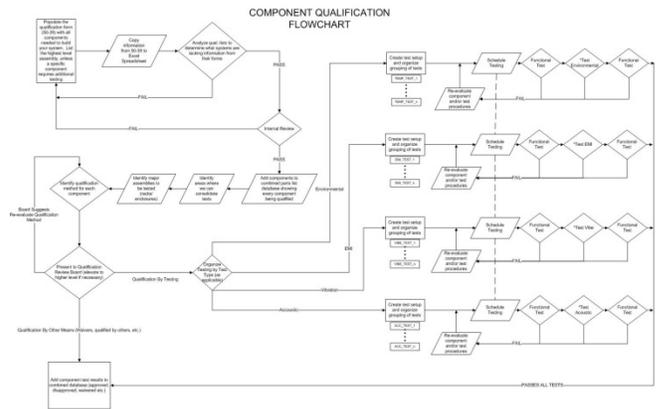


Figure 3.1: Component Qualification Flowchart of Electrical Ground Support Equipment.

Typical methods of qualification<sup>3</sup> defined in KSC-STD-G-0003 include, but are not limited to:

- a) Qualification by Testing
- b) Qualification by Similarity
- c) Qualification by Legacy (Prior Qualification)
- d) Qualification by Usage and Analysis
- e) Qualification by Higher Level Assembly Testing

For the purposes of this paper, the main tests considered are: Environmental, Electromagnetic Interference (EMI), Vibration and Acoustics testing. Each test, along with some general test requirements, is described in the following sections.

### 3.1 Vibration / Acoustic Testing

Although the Vibration and Acoustic tests should be performed independently, for the purposes of this paper it was decided to categorize them together. The four (4) levels below categorize the location where the hardware is installed. Here are the levels:

#### Vibration/Acoustic Level

- A - Mobile Launcher, around exhaust well, no shock mount
- B - Mobile Launcher, Launch Umbilical Tower
- C - Mobile Launcher, Shock Mounted or minimal grms
- D - Launch Pad

Thus, hardware is ground tested to the highest “worst” levels exposed according to its installed location. Another test design consideration relates to the operational scenarios of each system. Some systems such as Communication and Tracking require being operational before, during and after a space vehicle launch. However, some systems (such as Sound Suppression Water) are not required to be operational after the space vehicle has cleared the launch tower.

The KSC-NE-8764 Volume I – entitled “Crew Launch Vehicle (CLV) Mobile launcher solid rocket motor exhaust plume induced environment Acoustic and Vibration” document depicts the Vibration and Acoustic environment in the near and far field during lift-off<sup>4</sup>. The closer to the exhaust well, the higher the Grms levels experienced by the hardware. The metric of “grms” is typically used to specify and compare the energy in repetitive show vibration systems<sup>5</sup>.

When evaluating Overall Sound Pressure Levels (OASPL) both the Mean (50% Confidence Level) and the Specification (97.7% Confidence Level) values must be considered. The Power Spectral Density (PSD) analysis of most legacy space vehicles are publicly available. Therefore, for ground acoustic testing of new space vehicles, it is acceptable to scale down (or up) the Octave Band Sound Pressure Levels (OBSPL) levels of previous space vehicles to determine the expected acoustic levels of the new space vehicle. Predicting these values is critical to ground testing requirements verification and validation. Additionally, the search for test facilities may be limited if the required test levels are abnormally high.

To determine the Sound Pressure Level (SPL) Spectrum we use a logarithm representation of the given frequency:

$$\text{Thus, } \text{SPL}(f) = 10 \text{ Log}_{10} S(f) \quad [\text{units: dB, re: } 2 \times 10^{-5} \text{ Pascals (Pa)}] \quad (1)$$

where  $S(f)$  is the square of the pressure ratio  $\{ p_{rms}^2 / p_{ref}^2 \}$

Then the Overall Sound Pressure Level (OASPL) states that the total energy contained in the spectrum<sup>6</sup> is given by:

$$E = \int S(f) df \quad (2)$$

which is then integrated over all resolved frequencies<sup>7</sup>. Then the OASPL becomes:

$$\text{OASPL} = 10 \text{ Log}_{10} (E) \quad [\text{units: dB}] \quad (3)$$

If we use as reference 0 [dB] or  $2 \times 10^{-5}$  Pa, then the OASPL can be calculated follows<sup>6</sup>:

$$\text{OASPL} = 10 \text{ Log} \{ p_{rms}^2 / p_{ref}^2 \} \quad (4)$$

$$\text{Equation (4) can be re-written as } \text{OASPL} = 20 \text{ Log} \{ p_{rms} / p_{ref} \} \quad (5)$$

where  $p_{rms}^2 = \sum_k p_k^2$  and  $p_{ref}$  is ostensibly the audible limit of the human ear

A typical calculation at different Octave bands given its Sound Pressure level is shown in Table 3.1 below<sup>7</sup>.

Table 3.1 Sound Pressure levels at various Octave bands

Octave Band	Sound Pressure Level (SPL) (dB) Ref: 0 [dB] = $2 \times 10^{-5}$ (Pa)	Sound Pressure $p_k^2$ ( $\text{Pa}^2$ )
31.5	124	$1.005 \times 10^3$
63	130	$4.000 \times 10^3$
125	135	$1.265 \times 10^4$
250	139	$3.177 \times 10^4$
500	134	$1.005 \times 10^4$
1000	128	$2.524 \times 10^3$
2000	124	$1.005 \times 10^3$
4000	120	$4.000 \times 10^2$
8000	116	$1.592 \times 10^2$
	$p_{rms}^2 = \sum_k p_k^2$	$6.356 \times 10^4$
Overall Sound Pressure Level (OASPL)	$\text{OASPL} = 10 \text{ Log} \{ p_{rms}^2 / p_{ref}^2 \}$	142 [dB]
Overall Sound Pressure Level (OASPL)	142	

The Voyager I spacecraft mission was launched from Cape Canaveral Air Force Station in Florida on a Titan IIIE / Centaur vehicle<sup>8</sup> in September 5, 1977. Figures 3.1(a) and Figure 3.1(b) show the configuration of each spacecraft.

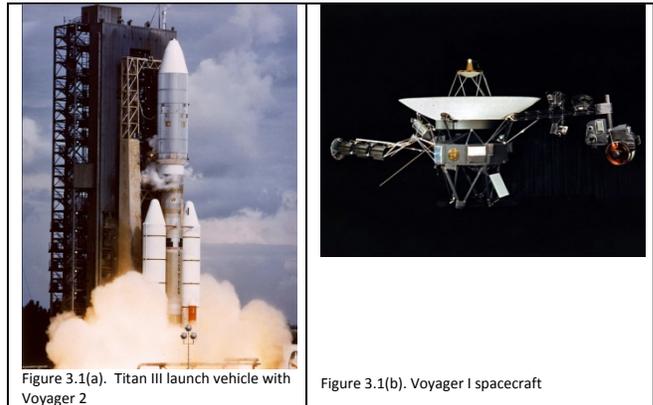
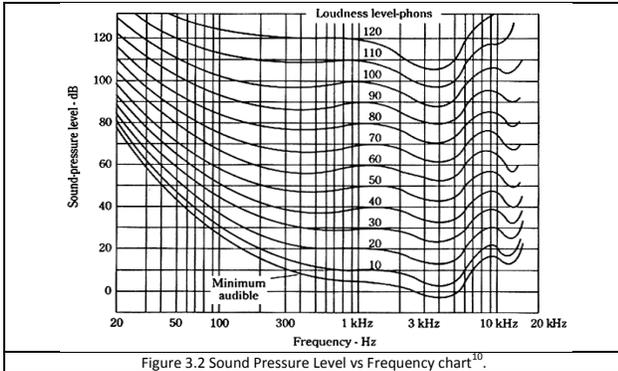


Figure 3.1(a). Titan III launch vehicle with Voyager 2

Figure 3.1(b). Voyager I spacecraft

It is very important to observe that the values of Table 3.1 can be used as a guide to scale (up or down) acoustic levels of new space launch vehicles in the perimeter of its launch platform. It should be noted that the value of OASPL is greater than any individual sound pressure level in the specification, because it represents an intensity of the spectrum as a whole.<sup>9</sup> Therefore, when designing ground acoustic and vibration tests, the test team must:

- a) Estimate the Average Sound Pressure Levels (SPL) for each location. Typical test locations are as close as the Mobile Launch Tower (MLT) and as far as 15,000 ft (Far Field) from the MLT.
- b) Consider the various frequencies at each of those locations. Typical Octave Band Center Frequencies range from 2 Hz – 8,000 Hz. See Figure 3.2 for a generic SPL vs Hz chart.



### 3.1.1 Vibro-Acoustic Test Level Requirements

Sound Pressure Levels of Heavy lift vehicles can surpass higher than normal limits. For these extreme cases, testing are conducted at facilities that can reach levels over 130 dBs and that can fit full racks, enclosures and the entire spacecraft if possible. The OASPLs values below<sup>4</sup> were taken from the Saturn V and the Space Shuttle program:

- 1) (Ref. KSC-NE-8764, Appendix A-3)
  - a) ML Base “except vicinity” (away) of Exhaust Hole
    - i. Mean Overall Acoustic Sound Pressure Level (OASPL) @ 50% Confidence Level = 141.7 dB.
    - ii. Specification (OASPL) @ 97.7% Confidence Level = 147.2 dB.
- 2) (Ref. KSC-NE-8764, Appendix A-5)
  - a) ML Base, Compartments around (internal/near) exhaust well
    - i. Mean Overall Acoustic Sound Pressure Level (OASPL) @ 50% Confidence Level = 150.4 dB.
    - ii. Specification (OASPL) @ 97.7% Confidence Level = 161.9 dB.
    - iii. [\[Ref. KSC-NE-8764 Appendix A-7\] “Induced Environment Acoustic and Vibration”](#)
  - b) ML Tower – All Levels
    - i. Mean Overall Acoustic Sound Pressure Level (OASPL) @ 50% Confidence Level = 165.6 dB.
    - ii. Specification (OASPL) @ 97.7% Confidence Level = 170.5 dB.

### 3.1.2 Test duration:

- 1) ([Ref. KSC-STD-164B, p.22](#)) “Environmental Test Methods for GSE Standard...”
  - a) Time: minimum test is 10 seconds.
  - b) For each additional launch, add 5 seconds.
  - c) Do not exceed 3 minutes.
  - d) 3 minutes = 100 launches.
  - e) Assumptions: => 10s + 5s\*(5 launches – 1st test launch) = 30seconds.  
Therefore, for 36 launches, the maximum length of a test at any particular frequency is calculated as follows:  
10s + 5s\*(35 launches – 1<sup>st</sup> test launch) = 180seconds = 3 minutes (MAX)

### 3.1.3 Vibration Test Setup

Accelerometers and other sensors are placed in strategic locations throughout the hardware considered for testing. Ideally, racks, enclosures and panels are tested at least twice; once with no contents (empty configuration) and then loaded. Additionally, sensors are moved from test to test, to ensure the critical areas are considered as hardware experiences different effect at different locations under different loads. Figures 3.1.3(a) and 3.1.3(b) depict a typical test configuration. The hardware is to be tested in the expected operational configuration. Therefore, if dampers are included in the installation design, then it must be tested as such.



Figure 3.1.3(a): Placement of accelerometers in an empty enclosure.

Figure 3.1.3 (b): Instrumented enclosure ready for vibration testing.

### 3.1.4 Acoustic Test Setup

Microphones and other sensors are placed in strategic locations throughout the hardware considered for testing. Ideally, racks, enclosures and panels are tested at least twice; once with no contents (empty configuration) and then loaded. Additionally, sensors are moved from test to test, to ensure the critical areas are considered as hardware experiences different effect at different locations under different loads. Figures 3.1.4(a) and 3.1.4(b) depict a typical test configuration. The hardware is to be tested in the expected operational configuration. Therefore, if sound isolators are included in the installation design, then it must be tested as such.



Figure 3.1.4 (a): Camera and cables feed from Test facility to Control room at the Johnson Space Center Acoustic Test Facility.

Figure 3.1.4(b): Audio sensors mounted over a wall panel prior to testing.

### 3.2 Environmental Testing

Qualification testing of both ground support equipment and flight hardware is expected to be performed in accordance to the exposed/induced environment where it will be used. There are many variables to each test condition, but for the purposes of this paper, the three environmental scenarios are considered.

The following Environmental Control Levels are defined in KSC-STD-164 (except EMI, acoustic, vibration, explosion and lift-off blast)<sup>11</sup>:

- A - Outdoor, extended temperature (-25C to +85C, Humidity, Rain, Icing, Fungus, Salt fog, Sand and Dust)
- B - Outdoor (0 to +70 C, Humidity, Rain, Icing, Fungus, Salt fog, Sand and Dust)
- C - Indoor (0 to +60 C, Humidity)
- D - Climate Controlled

### 3.3 Electromagnetic Interference Testing

Whether or not the GSE will be used or installed near the Mobile Launch Tower, Electromagnetic Interference (EMI) Tests are performed during equipment qualification testing. EMI, also called radio frequency interference or RFI is a disturbance that affects an electrical circuit due to either electromagnetic induction or electromagnetic radiation from an external source<sup>12</sup>. Signals can be affected by EMI leaks. Proper cable connections play a huge role in minimizing signal disturbance. Although, there is not a set way on the order of qualification testing, it is advisable to conduct EMI qualification test after a vibration test has passed inspection. Cables connections may come loose after a strenuous motion test, thus an EMI leak might not be detected if the EMI qualification test was performed prior to the Vibration and/or Acoustic test.

## 4. REPORTS

It is recommended to capture all pertaining component qualification information in a database<sup>13</sup>. There may be multiple stakeholders that benefit from efficient reporting tools. For example, subsystems engineers, test design engineers, test conductors, budget analysts, logistics control, configuration management, travel coordinators and even managers benefit from the collected data.

Figure 4.1 below shows a one page report that includes assembly information for components used and tested by various subsystems and their expected location of use. The expected test levels correspond to the Hazard Location, Vibro-Acoustic location and its Environmental control levels.

Figure 4.1: One page report of component qualification test information.

### 5. CURRENT AND FUTURE DEVELOPMENTS

The next step in the component qualification testing of the new space program is to be able to conduct functional tests using the equipment as configured in the design documents. The component qualification test board will play a critical role as they will decide which methods are acceptable to qualify the new costly electrical ground support equipment. Concurrently, Mechanical, Fluids systems are having similar decisions about their Ground Support Equipment.

### 6. ACKNOWLEDGEMENTS

The author of this paper was able to provide the above information due to the involvement with the Kennedy Space Center’s GSE component qualification team.

- Member from NASA include:  
 Chris Brown  
 Steve Bryan  
 Zachary K. Cline  
 Erik C. Denson  
 Reza M. Fotros  
 Mohamed Kaouk  
 Paul J. Mackey  
 David M. Stark

- Members from the NASA contractors include:  
 Okon Anwana  
 Thomas Downs  
 Wayne Devoe  
 Joshua Duncan  
 Ronald J. Sup  
 Rory Thomas

The contributions by each team member proved to enhance the component qualification test process of previous and future space programs.

### 7. REFERENCES

[1] Classification of Hazard Locations OSHA  
 United States Department of Labor: Occupational Safety & Health Administration  
<http://www.osha.gov/doc/outreachtraining/htmlfiles/hazloc.html>

[2] Component Qualification Flowchart of Electrical Ground Support Equipment.  
 Kennedy Space Center, FL Component Qualification Test Form 50-39 – Instructions, 2010.

[3] KSC-STD-G-0003 “Qualification of Launch Support and Facility Components, Standard For Engineering Development Directorate”. January 2003.

[4] KSC-NE-8764 Volume I “Crew Launch Vehicle (CLV) Mobile launcher solid rocket motor exhaust plume induced environment Acoustic and Vibration”. February 2008.

[5] Doertenbach, Neill: “The Calculation of Grms” QualMark Corp.  
[http://www.google.com/url?sa=t&source=web&cd=1&ved=0CByQfIAA&url=http%3A%2F%2Fwww.dfrsolutions.com%2Fuploads%2Fservices%2FHALT\\_grms\\_calculation\\_ndoertenbach.pdf&rct=i&q=grms%20level&ei=FvD0TdS0Lce2taf2\\_7WXBw&usq=AFQjCNHTmwb9CqbW4oU4C1J7g7PufmLxw](http://www.google.com/url?sa=t&source=web&cd=1&ved=0CByQfIAA&url=http%3A%2F%2Fwww.dfrsolutions.com%2Fuploads%2Fservices%2FHALT_grms_calculation_ndoertenbach.pdf&rct=i&q=grms%20level&ei=FvD0TdS0Lce2taf2_7WXBw&usq=AFQjCNHTmwb9CqbW4oU4C1J7g7PufmLxw)

[6] Acoustic Data Definitions: “Mach Wave Elimination”  
[http://supersonic.eng.uci.edu/mwe/mwe\\_noise\\_data.htm](http://supersonic.eng.uci.edu/mwe/mwe_noise_data.htm)

[7] Calculation of Overall Sound Pressure Level (OASPL), p. 63  
[http://books.google.com/books?id=6cyTiKviOUY&pg=PA53&pg=PA53&dq=OASPL+calculation&source=bl&ots=JFQat5xft5&sig=8exDSz1ge2gOo3c71e2luc8415Q&hl=en&ei=sID5TZ\\_HKerq0gH04ryWAw&sa=X&oi=book\\_result&ct=result&resnum=3&ved=0CCQ6AEwAg#v=onepage&q=OASPL%20calculation&f=false](http://books.google.com/books?id=6cyTiKviOUY&pg=PA53&pg=PA53&dq=OASPL+calculation&source=bl&ots=JFQat5xft5&sig=8exDSz1ge2gOo3c71e2luc8415Q&hl=en&ei=sID5TZ_HKerq0gH04ryWAw&sa=X&oi=book_result&ct=result&resnum=3&ved=0CCQ6AEwAg#v=onepage&q=OASPL%20calculation&f=false)

[8] Voyager 1 mission facts. [http://en.wikipedia.org/wiki/Voyager\\_1](http://en.wikipedia.org/wiki/Voyager_1)

[9] Acoustic Noise Requirement – May 1996, pp. 2-3.  
[http://www.google.com/url?sa=t&source=web&cd=1&ved=0CByQfIAA&url=http%3A%2F%2Fklabs.org%2FDEI%2FReferences%2Fdesign\\_guidelines%2Fdesign\\_series%2F1259.pdf&rct=i&q=OASPL%20calculation&ei=BdH6TzSvFMHm0QHk9WzAw&usq=AFQjCNFa4EkdmlcztgvlX2pN8f7SE5VzVw](http://www.google.com/url?sa=t&source=web&cd=1&ved=0CByQfIAA&url=http%3A%2F%2Fklabs.org%2FDEI%2FReferences%2Fdesign_guidelines%2Fdesign_series%2F1259.pdf&rct=i&q=OASPL%20calculation&ei=BdH6TzSvFMHm0QHk9WzAw&usq=AFQjCNFa4EkdmlcztgvlX2pN8f7SE5VzVw)

[10] Typical Sound Pressure Level vs Frequency Chart.  
<http://www.educylopedia.be/electronics/Images/Fletcher-Munson.gif>

[11] KSC-STD-164B “Environmental Test Methods for Ground Support Equipment, Standards for John F. Kennedy Space Center, NASA. May 16, 1989. pp. 5-6, 38-39.

[12] Electromagnetic Interference (EMI) definition.  
[http://en.wikipedia.org/wiki/Electromagnetic\\_interference](http://en.wikipedia.org/wiki/Electromagnetic_interference)

[13] Report of Component Qualification Test. Kennedy Space Center, Florida. March 2011.  
[http://en.wikipedia.org/wiki/Electromagnetic\\_interference](http://en.wikipedia.org/wiki/Electromagnetic_interference)

# Development of a Novel Wireless Electric Power Transfer System for Space Applications

Gabriel VAZQUEZ RAMOS  
 National Aeronautics and Space Administration  
 Launch Services Program (Mail Code: VA-H3)  
 John F. Kennedy Space Center, FL 32899  
[Gabriel.VazquezRamos-1@nasa.gov](mailto:Gabriel.VazquezRamos-1@nasa.gov)

and

Dr. Jiann-Shiun YUAN  
 University of Central Florida  
 Department of Electrical Engineering and Computer Science  
 Orlando, FL 32816-2362  
[yuanj@mail.ucf.edu](mailto:yuanj@mail.ucf.edu)

## ABSTRACT

This paper will introduce a new implementation for wireless electric power transfer systems: space applications. Due to the risks that constitute the use of electrical connector for some space missions/applications, a simple wireless power system design approach will be evaluated as an alternative for the use of electrical connectors. This approach takes into consideration the overall system performance by designing the magnetic resonance elements and by verifying the overall system electrical behavior. System characterization is accomplished by executing circuit and analytical simulations using Matlab® and LTSpiceIV® software packages. The design methodology was validated by two different experiments: frequency consideration (design of three magnetic elements) and a small scale proof-of-concept prototype. Experiment results shows successful wireless power transfer for all the cases studied. The proof-of-concept prototype provided ~4 W of wireless power to the load (light bulb) at a separation of 3 cm from the source. In addition, a resonant circuit was designed and installed to the battery terminals of a handheld radio without batteries, making it turn on at a separation of ~5 cm or less from the source. It was also demonstrated by prototype experimentation that multiple loads can be powered wirelessly at the same time with a single electric power source.

**Keywords:** Magnetic Resonance, Wireless Power Transfer, Magnetic Coupling, Space Systems, Power Amplifier.

## 1. INTRODUCTION

For Space Systems, power connectors constitute a vital and complicated component required for the success of every space mission. Some of these missions require crucial connector mate and de-mate operations in environments full of contaminants and/or performed by automatic (unmanned) systems. These operations can constitute a risk to the mission due to connector deterioration after been exposed to environmental contaminates (i.e. lunar regolith, also call moon dust) or by bended pins due to misaligning during a connector mate operation [1]. To mitigate these concerns, this paper is proposing the use of a wireless power transfer (WPT) system for space applications. This will be accomplished by the evaluation and

implementation of the magnetic resonance concept for space systems.

In classical physics, mechanical resonance has been widely demonstrated in various applications and experiments (i.e. identical tuning forks resonating by only impacting one). In 2007, a research team from the Massachusetts Institute of Technology demonstrated a similar principle on electric circuits called magnetic resonance (also known as magnetic coupling) [2].

The WPT system proposed in this paper will transfer electric power without the need of having physical electrical connection. The intent is to have a modular design that can be used as a “wireless power connector” in a wide range of space exploration applications (i.e. automatic docking systems for space systems rendezvous, wireless sensors for launch vehicles, wireless battery charging systems, robotic mission charging stations on other planets, moons and asteroids, etc.). The main scope of this paper is to provide a simple design approach for the implementation of a WPT and its feasibility for space applications.

## 2. ANALYSIS

Based on technical literature review, peer reviewed papers, and component level testing, a simple design approach will be delineated for the analysis and fabrication of WPT devices. A block diagram is provided in Fig. 1, showing the main elements required for the WPT system implementation.

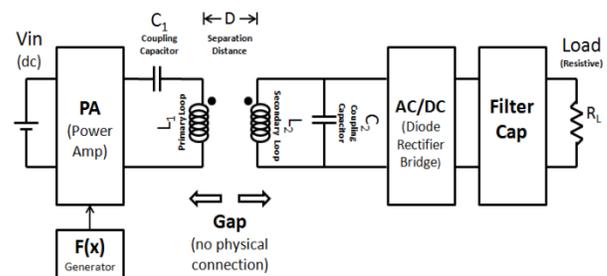


Fig. 1. WPT System Block Diagram

### Overall System Operation

An electric source circuit generates a sinusoidal signal (represented by a power amplifier, PA, in Fig. 1) inducing magnetic pulsation signals at the primary loop ( $L_1$ ) of the source resonating circuit ( $C_1$  and  $L_1$ ). The secondary loop ( $L_2$ ) will receive the magnetic pulses due to the fact that is part of the load resonating circuit ( $C_2$  and  $L_2$ ). The load resonating circuit is tuned to the same frequency as the primary circuit. This magnetic energy induces a sinusoidal electric signal in the secondary [2]. The alternate current (AC) signal is then rectified by a diode H-bridge and a capacitor to provide a direct current (DC) signal to the load. This process will transfer energy wirelessly from the primary circuit ( $V_{in}$ , source) to the secondary circuit (Load) [2].

The overall system analysis and design was performed by a systematic series of simulations using the combination of Matlab® and LTSpiceIV® software packages. This integrated system circuit simulation combines different technical topics evaluated for the design. These topics were divided into two main design simulations: magnetic elements simulation and system circuit simulation. Each of these simulations provides an important contribution for the overall understanding of the system's electrical behavior.

### Magnetic Elements Simulation

The magnetic elements (Fig. 2) will be defined as the primary and secondary loops with their respective coupling capacitor for resonant circuit. To determine the magnetic element parameters, a Matlab® based code was generated using simple electromagnetic and circuit equations. This tool helped us to understand how to design the inductor loops and their corresponding coupling capacitor for the desired operating frequency.

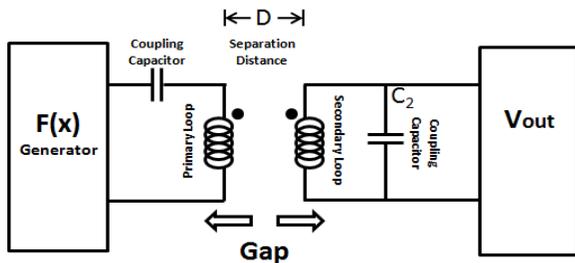


Fig. 2. WPT System Magnetic Element Definition

The Eq. (1) defines the self inductance ( $L$ ) of an inductor loop with no magnetic core [3].

$$L = N^2 * R * \mu_0 * [\ln\left(\frac{8*R}{r}\right) - 1.75] \quad (1)$$

Where:

- L = inductance of the loop
- N = number of turns in the loop
- R = radius of the loop
- r = loop conductor radius
- $\mu_0$  = permeability of vacuum

In order to adequately transfer power wirelessly, the primary and the secondary magnetic loops need to be tuned to the same resonant frequency ( $\omega_0, f_0$ ). This tuning is accomplished by connecting a capacitor in series to the primary circuit and in parallel to the secondary circuit (also known as series-parallel coupling) [4]. The Eq. (2) is utilized to determine the capacitor value required to generate the resonant circuit with the inductor [5]:

$$C = \frac{1}{\omega_0^2 * L} \quad (2)$$

Where:

- C = coupling capacitance required
- $\omega_0$  = frequency of oscillation [rad/sec]
- L = inductance of the coil

An additional important calculation to be considered is the mutual inductance between the two loops. This value will be calculated for the required separation distance of the vacuum gap and then simulated in LTSpiceIV® to determine the overall circuit response. Before calculating the mutual inductance, it is required to determine the magnetic coupling coefficient using the Eq. (3) [6]. By using the magnetic coupling coefficient, a characterization of the mutual inductance can be obtained with Eq. (4) [7].

$$k = \frac{1}{\left[1 + 2^{2/3} \left(\frac{D}{\sqrt{R_1 * R_2}}\right)^2\right]^{3/2}} \quad (3)$$

$$L_M = k * \sqrt{L_1 * L_2} \quad (4)$$

Where:

- k = mutual coupling coefficient
- D = physical distance between  $L_1$  and  $L_2$
- $R_1$  = radius of the loop 1
- $R_2$  = radius of the loop 1
- $L_M$  = mutual inductance
- $L_1$  = inductance of the loop 1
- $L_2$  = inductance of the loop 2

A simple proof-of-concept prototype was assembled to demonstrate the proposed design methodology. The parameters used in the proof-of-concept prototype are listed in Table 1. These parameters will be later used to calculate the coupling capacitors, self inductance of the loops, and mutual inductance parameters. The calculated parameters using the developed Matlab® code are shown in Table 2.

Table 1: Proof-of-Concept Prototype Characteristics

Parameter	Symbol	Value	Units
Number of turns	$N_1$ & $N_2$	8	turns
Radius of the Loops	$R_1$ & $R_2$	0.24	m
Loop conductor radius	$r_1$ & $r_2$	0.001	m
Loop conductor resistance (in DC)	a	0.064*	Ohm
Frequency of operation	$f_0$	15	kHz
	$\omega_0$	94.25	k Rad/s
Vacuum Permeability	$\mu_0$	$4 * \pi * 10^{-7}$	H/m
Separation distance	D	0.03	m
Load (light bulb)	$R_0$	4**	Ohms

\* The loop conductor resistance (litz wire cable) was calculated using reference [8]  
 \*\* Measured resistance for the filament. However, it was noted after system simulation and prototype experimentation that the light bulb filament was behaving as a 40 Ohms resistor.

Table 2: Calculated Prototype Circuit Parameters

Parameter	Symbol	Value	Units
Self Inductance	$L_1$ & $L_2$	49.38	$\mu H$
Mutual Inductance	$L_M$	42.85	$\mu H$
Coupling Capacitor	$C_1$ & $C_2$	2.14*	$\mu F$
Leakage Inductances	$L_{lk1}$ & $L_{lk2}$	6.533	$\mu H$

\* Mylar capacitors of 2.2uF (5% accuracy) were used. After circuit testing, the coupling frequency was found to be 15.5 kHz.

### System Circuit Simulation

To characterize the overall circuit performance, it was required to find a circuit model that can help predict the performance of the overall system. The model used to emulate the behavior of the two hollow inductor loops (under magnetic resonance) is the non-ideal transformer equivalent circuit [9]. Fig. 3 illustrates the equivalent circuit of the two loops in magnetic resonance configuration.

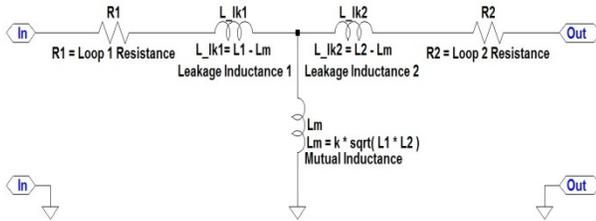


Fig. 3. Equivalent Circuit for Two Hollow Inductors (N1 = N2)

By implementing the non-ideal transformer equivalent circuit, the overall circuit was simulated in LTSpiceIV® using the parameters previously calculated (Table 2). Fig. 4 illustrates the coupling configuration that was simulated as a “series-shunt” coupling [5]. A generic sinusoidal voltage generator was used to simulate the power provided by the PA.

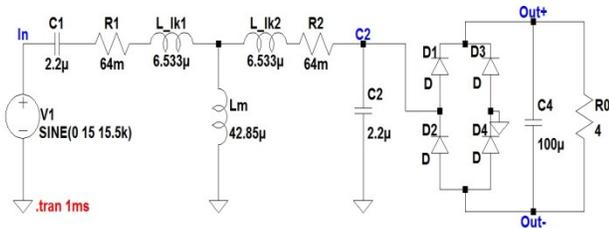


Fig. 4. Wireless Power Transfer Model with Calculated Loop Inductances with a Separation Distance of D = 3 cm

The simulation outputs of the circuits shown in Fig. 4 are illustrated in Fig. 5. The green trace is the function generator/power amplifier voltage source (VIN), the blue trace is the secondary voltage before rectification (VC2) and the red trace is the output voltage (VOUT+, OUT-) of the circuit.

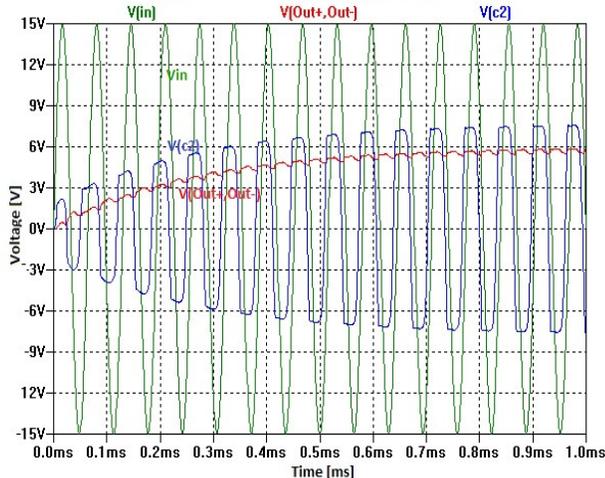


Fig. 5. Simulation Output of the WPT Model with Calculated Loop Inductances from Fig. 4

### 3. RESULTS

The analysis previously discussed was validated using two different experiments to better characterize the magnetic resonance concept: frequency considerations and a small scale proof-of-concept prototype.

#### Experiment #1: Frequency Considerations

To analyze the frequency of operation effects, three circuit configurations were designed with the same inductor loop characteristics (Table 3). The three frequencies of operation studied were: 84 kHz, 839 kHz and 1.757 MHz. The self inductances of the magnetic loops were determined with the Eq. (1) and coupling capacitors were selected to match the frequency of operation according to the Eq. (2). The testing configuration is illustrated in Fig. 2. The main objective of this test was to preliminary study the magnetic element behavior at different frequencies and validate that magnetic resonance can be achieved with the design approach described in the previous section (Analysis).

Table 3: Magnetic Loop Design Characteristics

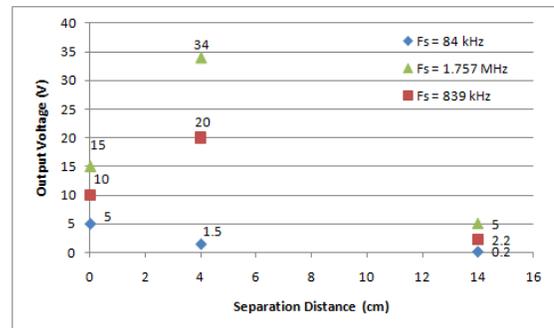
Characteristic	Loop 1	Loop 2
Radius [cm]	7.5	7.5
Number of Turns	7	7

The frequency consideration test results are listed in Table 4 and illustrated in Fig. 6. The testing circuit (Fig. 2) was powered by a function generator. The settings of the function generator remained unchanged throughout these tests for all three cases. The main objective of this experiment was met (to validate the magnetic coupling design approach described in the Analysis section). In addition, it was also noticed for the frequencies studied that magnetic coupling is greater at higher frequencies of operations. Further analysis/testing will be required to determine which frequency is the optimal.

Table 4: Frequency Consideration Test Results

Frequency of Operation	Separation Distance (cm)	Vout (Vpeak)
f <sub>s</sub> = 84 kHz	D <sub>1</sub> = 0	5 V
	D <sub>2</sub> = 4	1.5 V
	D <sub>3</sub> = 14	200mV
f <sub>s</sub> = 839 kHz	D <sub>1</sub> = 0	10 V
	D <sub>2</sub> = 4	20 V
	D <sub>3</sub> = 14	2.2 V
f <sub>s</sub> = 1.757 MHz	D <sub>1</sub> = 0	15 V
	D <sub>2</sub> = 4	34 V
	D <sub>3</sub> = 14	5 V

Note: The settings of the function generator (V<sub>in</sub>) remained unchanged throughout these tests (all three cases).



Note: Only three separation distances were measured per frequency.

Fig. 6. Frequency Consideration Test Results

### Experiment #2: Small Scale Proof-of-Concept Prototype

A small scale proof-of-concept prototype was assembled with the values provided on Table 1 and Table 2 to validate the design approach discussed in the *Analysis* section. Using the block diagram in Fig. 1 as the design guideline, a series of commercial parts were evaluated, individually tested and selected to complete the prototype design (Fig. 7).

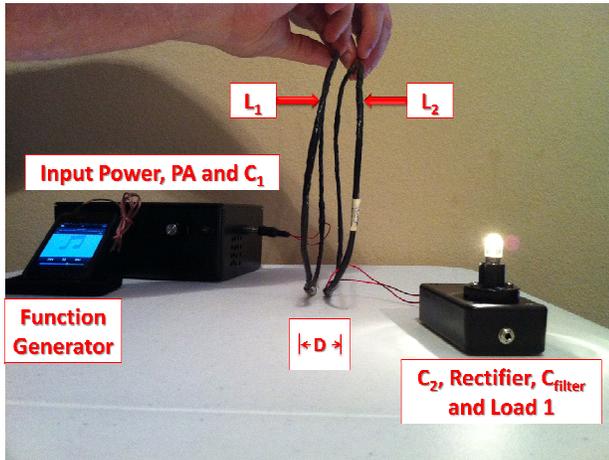


Fig. 7. Small Scale Proof-of-Concept Prototype

Due to the fact that the loops inductors were manually built, variably on the inductance value is expected. The Agilent LCR Meter 4980A was used to measure the inductances of the loops at the prototype operating frequency (15.5 kHz). Table 5 lists the measured loop inductances and the calculated mutual inductance for a separation distance of 3cm. Fig. 8 illustrates the inductors built for the small scale prototype.

Table 5: Measured Loop Inductances and Mutual Inductance

Parameter	Symbol	Value	Units
Self Inductance #1	$L_1$	44.1	$\mu H$
Self Inductance #2	$L_2$	54.3	$\mu H$
Mutual Inductance*	$L_M$	42.46	$\mu H$
Leakage Inductance 1*	$L_{lk1}$	1.638	$\mu H$
Leakage Inductance 2*	$L_{lk2}$	11.84	$\mu H$

\*Value calculated based on the measured self inductances ( $L_1$  and  $L_2$ ) and using a separation distance of 3 cm.

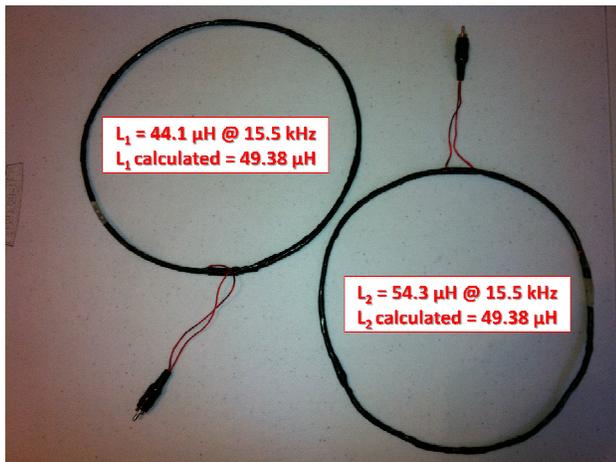


Fig. 8. Loop Inductors Built for the Small Scale Prototype

The WPT LTSpiceIV® model was updated with the loop measurements acquired with the LCR meter (Fig. 9) and the simulation output is illustrated in Fig. 10. The green trace is the function generator/power amplifier voltage source ( $V_{IN}$ ), the

blue trace is the secondary voltage before rectification ( $V_{C2}$ ) and the red trace is the circuit output voltage ( $V_{OUT+, OUT-}$ ).

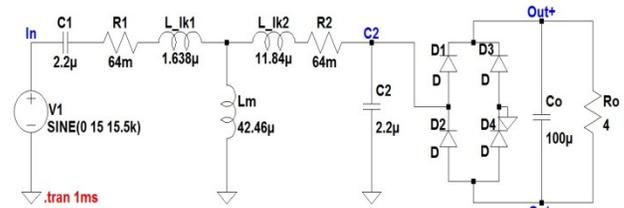


Fig. 9. Wireless Power Transfer Model with Measured Loop Inductances with a Separation Distance of  $D = 3$ cm

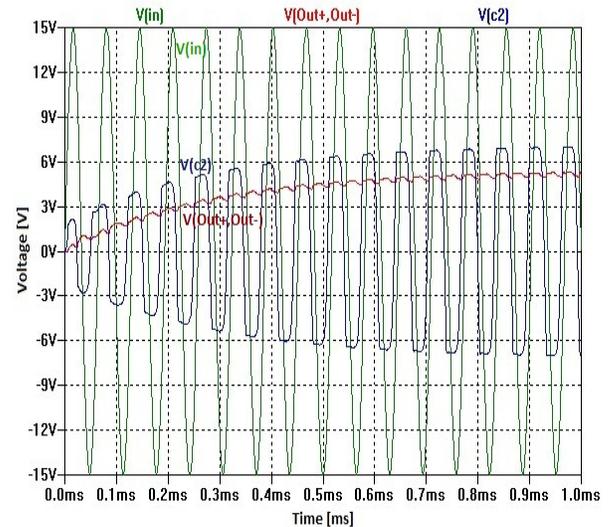


Fig. 10. Simulation Output of the WPT Model with Measured Loop Inductances (Fig. 9)

The system operating frequency ( $f_0$ ) was targeted to be 15 kHz; however, after further calibration and tuning, the system demonstrated to be magnetically coupled at ~15.5 kHz. This discrepancy is attributed to the coupling capacitor tolerance of 5%, the loop conductor radius measurement and the variability of the manually built inductor loops. An accurate representation of the loop conductor radius and coupling capacitor values are required for an accurate system coupling according to Eq. (1), Eq. (2) and Eq. (3).

Two different alternatives were implemented/designed to provide the frequency of operation (sinusoidal at 15.5 kHz) required by the PA: XR2206 function generator integrated circuit [10] and by programming an audio signal. The audio signal was programmed in Matlab® as a sine wave with the required frequency, an audio file format (Windows Media Audio or WMA) was generated using Matlab®'s "wavwrite()" command and then the audio file was downloaded and played using the iPhone music player. The iPhone's output is connected to the PA as the function generator (Fig. 11). The amplitude of this signal was controlled by the volume of the iPhone. Fig. 11 illustrates the output voltage supplied by the iPhone using the audio signal programmed in Matlab®.

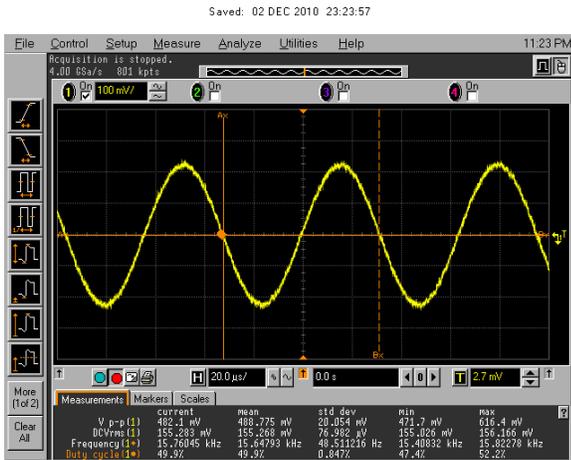


Fig. 11. iPhone Output Used as a Function Generator (PA Input Signal)

The  $V_{DC}$  power supply was selected to be the Lexmark z33 printer power adapter (30 Vdc and up to 0.5 Amp). By the use of voltage regulators (LM7809, LM7812 and LM7815), the voltage was stepped down to supply the required voltage to all the electronics subsystems and components used on the prototype. For a simple PA implementation, the frequency of operation was selected to be within audible range (20 Hz- 20 kHz). This enables us to use the “off-the-shelf” commercial PA: Velleman-Kit K4001 [11].

As the WPT system block diagram illustrate in Fig. 1, the PA will supply the required power of the source resonant circuit ( $C_1$  and  $L_1$ ). The source loop ( $L_1$ ) voltage will induce the magnetic resonance effect into the load loop ( $L_2$ ) through the load resonant circuit ( $C_2$  and  $L_2$ ) [2]. The source and load loops voltages (at  $D \approx 0$  cm) are illustrated in Fig. 12 and Fig. 13 respectively.

An AC-to-DC converter (rectifier) was design to transform the received AC signal (at  $L_2$ ) to the DC signal required by the load. The rectifier was design using 1n5817 Schottky barrier diodes connected on an H-bridge configuration. Schottky barrier diodes where selected due to their fast recovery time. Based on simulation trials, the filter capacitor was selected to be  $100\mu F$ .

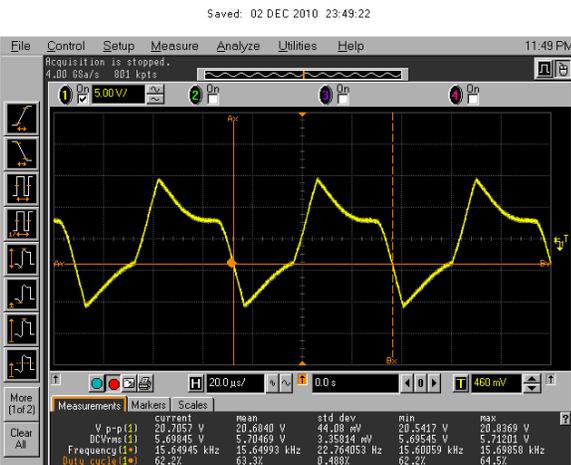


Fig. 12. Source Loop ( $L_1$ ) Voltage

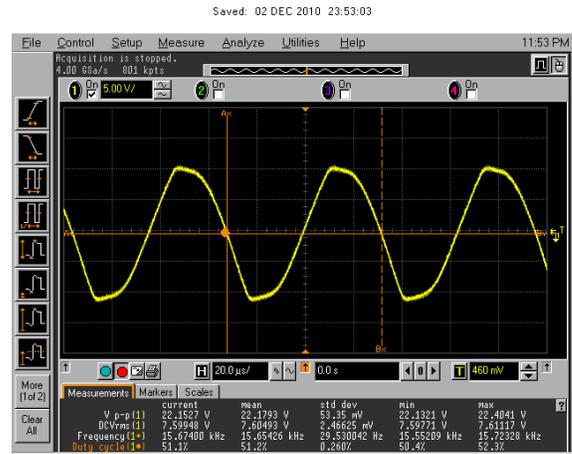


Fig. 13. Voltage Induced at the Load Loop ( $L_2$ )

The prototype performance was carefully characterized by measuring and plotting the output root mean square (RMS) voltage, current and power for different source/load separation distances (Fig. 14, Fig. 15 and Fig. 16). It is noticed that the optimum power transfer separation is about 3 cm (~4W). The peak power transfer distance result concurred with a similar study performed on magnetic resonance [12].

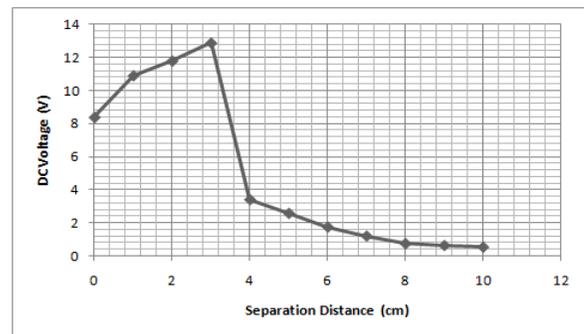


Fig. 14. Prototype Output Voltage vs. Separation Distance

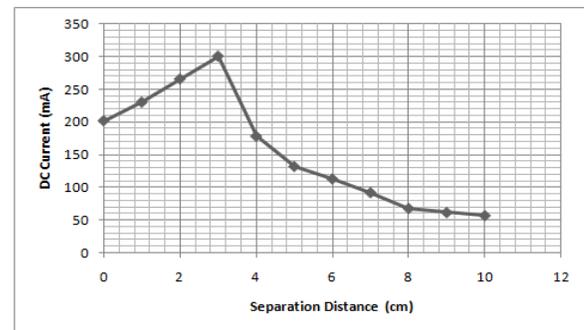


Fig. 15. Prototype Output Current vs. Separation Distance

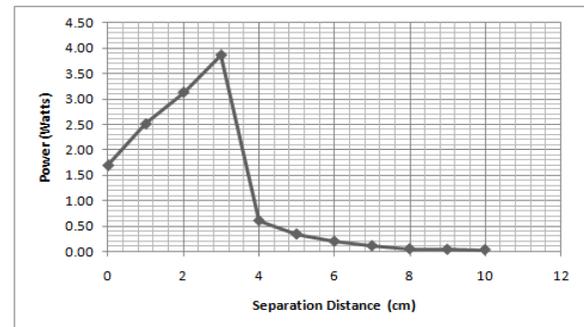


Fig. 16. Prototype Output Power vs. Separation Distance

It is noticed that the theory presented in the *Analysis* section provides similar results than the small scale proof-of-concept prototype experimental results. However, it is also noticed that due to parasitic effects and unwanted capacitive coupling, internal to the inductor windings, provides an unpredicted response in the test data. Further analysis and experimentation will be required for an overall more accurate electrical characterization of the system.

An additional test performed to the prototype was to add a non-conductive material between the source element and the load element (i.e. wood, plastic, etc.). As expected, magnetic coupling was achieved even through a non-conductive material.

A second load was designed to operate using the prototype source (Fig. 17). A receiving magnetic element was designed and connected to the battery terminals of a handheld radio (Durabrand PR-355 AM/FM Sports Radio) to resonate at the same frequency as the wireless power source previously described ( $f_0 = 15.5\text{kHz}$ ). The WPT receiver device designed to the handheld radio replaced the use of 3 AA batteries. The radio turns on within a proximity of  $\sim 5$  cm from the source. It was also noticed that the radio and the light bulb can receive power wirelessly at the same time using the same power source.

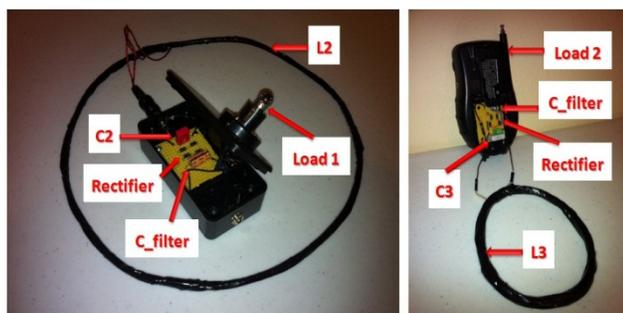


Fig. 17. Loads: Light Bulb (left) and Handheld Radio (right)

#### 4. CONCLUSION

Theory analysis and experimental results suggest that magnetic resonance is a feasible and reliable technology that represents an alternative to conventional electric connectors. Preliminary testing also showed that magnetic resonance can be achieved through non-conductive materials (i.e. wood, plastic, etc.). Multiple mobile devices can be powered/recharged at the same time using a single source. This was accomplished by tuning the source and all the loads to the same operating frequency. In addition, it was demonstrated by prototype experimentation that magnetic resonance can be achieved within audible frequencies ( $20\text{Hz} \leq f_0 \leq 20 \text{kHz}$ ). This can represent a major industrial advantage due to the fact that there are currently a wide range of options available for audio power amplifiers.

More testing and experimental data will be required for the implementation of a reliable and efficient wireless power system for space applications. In theory, the magnetic resonance principle is believed to operate in the space environment. However, using the principles discussed in this paper and additional analysis, a more advanced design will be evaluated for compliance to NASA and other military systems standards [13][14][15][16]. Design compliance to these standards will be highly dependent on the system application (i.e. launch vehicles, spacecraft, rover, wireless battery charger, docking systems, etc.)

#### 5. BIBLIOGRAPHY

- [1] National Aeronautics and Space Administration *Workmanship Problems Pictorial Reference/Wiring and Cabling NASA Electric Parts and Packaging Program* (Online) URL: <http://nepp.nasa.gov/index.cfm/14257>
- [2] Aristeiidis Karalis, J.D. Joannopoulos, Marin Soljacic. "Efficient Wireless Non-Radiative Mid-Range Energy Transfer", *Annals of Physics*, Vol. 323, 2008, pp. 34-48.
- [3] Chunbo Zhu; Kai Liu; Chunlai Yu; Rui Ma; Hexiao Cheng; , "Simulation and experimental analysis on wireless energy transfer based on magnetic resonances," **Vehicle Power and Propulsion Conference, 2008. VPPC '08. IEEE**, 3-5 Sept. 2008, pp.1-4
- [4] Stielau, O.H.; Covic, G.A.; "Design of loosely coupled inductive power transfer systems," **Power System Technology, 2000. Proceedings. PowerCon 2000. International Conference on**, vol.1, 2000 pp.85-90
- [5] Chunbo Zhu; Chunlai Yu; Kai Liu; Rui Ma; "Research on the topology of wireless energy transfer device," **Vehicle Power and Propulsion Conference, 2008. VPPC '08. IEEE**, 3-5 Sept. 2008, pp.1-5
- [6] Mur-Miranda, J.O.; Fanti, G.; Yifei Feng; Omanakuttan, K.; Ongie, R.; Setjoadi, A.; Sharpe, N.; , "Wireless power transfer using weakly coupled magnetostatic resonators," **Energy Conversion Congress and Exposition (ECCE), 2010 IEEE**, 12-16 Sept. 2010, pp.4179-4186,
- [7] Hee-Seung Kim; Do-Hyun Won; Byung-Jun Jang; , "Simple design method of wireless power transfer system using 13.56MHz loop antennas," **Industrial Electronics (ISIE), 2010 IEEE International Symposium on**, 4-7 July 2010., pp.1058-1063,
- [8] New England Wire Technologies **Litz Wire Technical Information 2005** New England Wire Technologies, NH
- [9] Judek, S.; Karwowski, K.; , "Supply of electric vehicles via magnetically coupled air coils," **Power Electronics and Motion Control Conference, 2008. EPE-PEMC 2008. 13th**, 1-3 Sept. 2008pp.1497-1504
- [10] EXAR Corporation. **XR 2206 Function Generator IC**, EXAR Corporation, 2007.
- [11] Velleman Components NV. **7W Mono Amplifier, Kit**. Velleman Components NV. K4001.
- [12] Shen, F. Z.; Cui, W. Z.; Ma, W.; Huangfu, J. T.; Ran, L. X.; , "Circuit analysis of wireless power transfer by "coupled magnetic resonance"," **Wireless Mobile and Computing (CCWMC 2009), IET International Communication Conference on**, 7-9 Dec. 2009, pp.602-605
- [13] National Aeronautics and Space Administration. **Electrical Bonding for NASA Launch Vehicles, Spacecraft, Payloads, and Flight Equipment**, Washington, D.C., NASA, NASA-Std-4003.
- [14] United States Department of Defense. **Requirements of the Control of Electromagnetic Interference Characteristics of Subsystems and Equipment**. Wright-Patterson Air Force Base, OH, United States Department of Defense, Mil-Std-461.
- [15] United States Air Force, **Electromagnetic Compatibility Requirements for Space Systems**, Los Angeles, CA, USAF Space Division, Mil-Std-1541.
- [16] United States Department of Defense. **Product Verification Requirements for Launch, Upper Stage and Space Vehicles**, El Segundo, CA, Space and Missile Systems Center, Mil-Std-1540.

# Evaluation of Corporate Governance Influence on Stock Performance of CEE Companies

**Julia BISTROVA, Natalja LACE**  
**Engineering Economics and Management Faculty,**  
**Riga Technical University,**  
**Meza 1/7, Riga, LV-1007, Latvia**

## ABSTRACT

Corporate governance (CG) becomes a very essential factor to consider prior to investing in the company. A number of studies proved its importance on the developed equity markets. However, intuitively corporate governance should gain more importance due to high degree of uncertainty because of the unstable environment. In order to assess the influence of corporate governance quality on Central and Eastern European companies' stock performance, the CG assessment model, which includes 21 evaluation criteria, was developed. Based on the model rating, the companies with the highest CG quality (top 25%) outperformed companies with the worst CG quality (bottom 25%) by 0.98% on a monthly basis during the period of 2008 - 2010. Study demonstrate that companies with good CG quality are able to offer lower risk.

**Keywords:** corporate governance, information disclosure, board of directors, stock returns, management team

## 1. INTRODUCTION

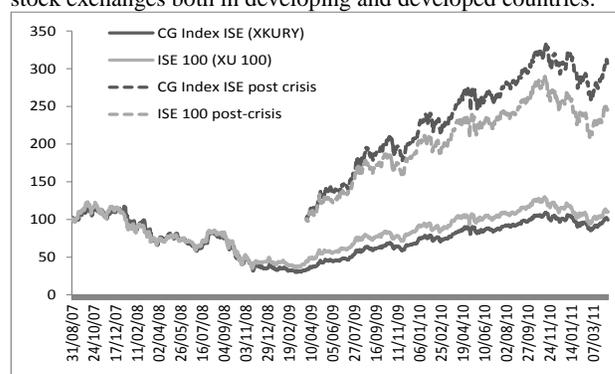
Out of the world's largest 100 economic entities 44 are corporations and 56 are states. The proportion becomes more prominent if to consider 150 largest entities, where the share of corporations increases to 59%, according to Keys and Malnight [13]. The company is like a state with its own regulations, its supervisory and executive bodies, which obviously need to run the company according to the certain rules and procedures that ensure value-based management [7]. Establishment of the quality corporate governance ensures significant limitation of the agency problem and is intended to maximize shareholders' as well as other interested parties' wealth. High quality of corporate governance (CG) is a guarantee of the long-term trust between shareholders and the management of the company.

Corporate Governance and ethical behavior problem and has escalated after the corporate scandals of *Enron*, *Parmalat*, *Worldcom* and was vastly discussed again during the global liquidity crunch with regards to financial entities.

The discussion on Corporate Governance covers mainly the trade-off between the benefits provided by the best corporate governance practice and the tangible and intangible costs for disclosure of information and corporate governance system establishment. As various empirical and theoretical researches suggest in the long-term companies definitely benefit when establishing good corporate governance practice. The tangible positive results of good governance are evident in the booming economy and markets but they appear to be clearer during the economic downturn. Having CG, the company is

able to soften the sharp decline of share price as it was seen during the last global financial crisis.

The evidence of corporate governance positive influence on company's value and stock return has been proved by the various researchers [10]. The topic becoming more popular, there have been created a number of stock indices on various stock exchanges both in developing and developed countries.



**Figure 1.** Performance of Turkish Corporate Governance Index vs. general market (ISE 100)

For example, Istanbul Stock Exchange has created CG index composed of companies, which comply with ISE CG recommendations. Figure 1 shows the comparison of Turkish CG index and Turkish 100 largest companies. Since the period, when CG index was launched, it has underperformed 100 Turkish companies. But if to consider post-crisis period, companies complying with CG recommendations demonstrate better than the market performance.

Though the stock exchanges in Central and Eastern European countries have not created index of this kind yet, the majority of them have published the codes for best practice for voluntary disclosure and implementation for the public companies. Significant number of the companies included in the "blue-chip" indices in this region manage to follow the best practice, but the level of corporate governance is very different from country to country, which is analyzed later in this paper. As several corporate governance experts and market participants state, the quality of corporate governance has massively improved in the past decade.

The question remains still whether good corporate governance by increasing shareholders' trust exerts positive influence on stock return. Therefore, the study's principal hypothesis is that high quality of corporate governance is praised by the superior stock returns. Besides, there were checked minor hypotheses, which were mainly focusing on the separate factors of the corporate governance such as board

independence, information disclosure, frequency of meetings, board and management team turnover etc. Therefore, the aim of the study is to provide an overview of corporate governance level in CEE listed companies and to figure out the relationship between stock performance and the level of corporate governance.

The methods used in the research involve model work-out based on the previous research and exchange recommendation as well as judgment on the quality of CG of CEE companies. The authors have used also quantitative methods (correlation, regression, etc.) to find out the relevance of CG influence on stock prices.

## **2. IMPORTANCE OF CORPORATE GOVERNANCE FACTOR**

A number of studies conducted on the developed markets state that corporate governance has strong influence on the stock market returns. Gompers, Ishii and Metrick constructed "Governance Index" which covered the assessment of shareholders' rights at 1500 companies in 90-ties [10]. Based on the index they have modeled the portfolio strategy that would long companies with strongest rights (lowest decile) and short companies with weakest rights (highest decile). As a result, the investor could earn 8.5% outperformance. The similar study was done by Drobetz in Germany showing the monthly difference in performance of well and poorly governed firms of 1.73% [9].

The significant correlation of such factors as CG index, CEO-Chairman separation and independence of board members with stock performance was found by Bhagat and Bolton [3]. But they did not find any evidence of quality of CG being a proxy for future stock performance. The findings of their study show also that given low quality of corporate governance of a certain entity and given its poor performance there is a high probability of management turnover.

Positive correlation between firm value and quality of corporate governance in case of largest 300 European companies (FTSE Eurotop 300) has been indicated by Dutch scientists Bauer, Guenster, Otten [2]. But when adjusting for country difference the relationship is weakening.

The contrary situation was discovered in Japan by Aman and Nguyen, who discovered that poorly governed firms outperform well-governed firms [1]. However, the results were statistically insignificant, but the study has clearly showed that significantly higher risk is attributed to the poorly governed firms.

Some researches have been made by considering separate factors which determine the quality of corporate governance. The significant relationship was indicated between equity performance and board independence (Hermalin and Weisbach [11, 12], Bhagat and Black [3]), stock ownership of board members (Bhagat, Carey, and Elson [5]), CEO and Chairman separation (Brickley, Coles, and Jarrell [8]).

The story in emerging markets is a bit different: due to often concentrated ownership structure companies have rather low motivation to disclose the information to outsiders. The need in minority shareholders obviously is less compared to the situation in developed markets. The regulations regarding corporate governance are less strict than they are in developed stock markets. In most of the cases these are just the recommendations imposed by the local stock exchanges. Anyway, the question of corporate governance influence becomes more topical. In case of favourable outcome (positive correlation with stock returns), the obtained results proved by the empirical research can be used to persuade the companies to stick to the best practice.

The available related literature provides the evidence of well-governed companies' outperformance also in the emerging markets. Roy Kouwenberg says the corporate governance matters with regards to Thai public companies: stock return of the best 20% companies according to the CG score in the period 2003-2005 was 19% p.a. year better than the stock return of the weakest 20% companies [14].

Indian market represented as NIFTY 50 has been studied by Samontaray, who found significant relationship between share price and in such independent variable as EPS, sales, net fixed assets and corporate governance factors [17].

There have been several studies on CEE stock markets, but the studies were done rather on macro level or considered separate factor which determine CG quality. Research made on 151 CEE companies by Mueller and Peev indicates that the firms' which are controlled mainly by foreign shareholders are overdoing their counterparts with mainly locals represented in the ownership structure [15]. Another study on ownership influence on CEE companies' performance considered mainly the type of ownership structure: strategic, state, financial, founder/family [6]. The results of the study indicated that the best-performing companies have state representation in their ownership, which were followed by the family/founder controlling.

Pajuste has been also researching ownership and shareholders' rights in CEE stock markets for the period of 1994-2001 [17]. Her findings provide the evidence of significant controlling shareholder influence on the performance of the company and that minority shareholders' rights are often abused making the market absolutely inefficient and risks are not justified by the returns, which are lower than average.

## **3. CORPORATE GOVERNANCE ASSESSMENT MODEL**

Corporate Governance evaluation is often done in the developed markets, where several rating agencies provide their scores for the public company's governance. In emerging market and specifically in CEE countries the term of corporate governance has appeared relatively recently so no centralized CG assessment for CEE companies was available at the time the study was made.

Therefore, in order to evaluate the quality of corporate governance of CEE companies, corporate governance assessment model was created (1.Table). The framework for the model was developed according to the CG recommendations to the listed companies provided by the local stock exchanges (Nasdaq OMX, Warsaw, SET etc.). Besides, the list was expanded by adding the most important factors defining CG quality, which are widely recognized and adopted in the mature financial markets.

The model consists of four pillars, where each is dedicated to a certain set of factors (supervisory board, management team, investor relations, information disclosure) defining the quality of corporate governance. The maximum score a company can get is 21, which is obtained by summing up all points in each segment. If the information regarding management team or board of directors was not available, then the neutral rating was assigned to the certain criteria when evaluating CG of a particular company.

### **Supervisory Board/BoD**

Major attention in the model is paid exactly to the board of directors, which is usually representing shareholders' interests and has mitigating role between shareholders and management.

Table 1

**Corporate Governance Assessment Model**

	Worst		Neutral		Best	
<b>Supervisory Board/BoD</b>						
Independence of directors	0-50%	0	50-75%	0.5	75-100%	1
Diversified skills of directors	concentrated	0	average	0.5	diversified	1
Frequency of meetings	<4 and >10 meetings	0	4-5 and 9-10 meetings	0.5	6-8 meetings	1
Performance-based compensation	no	0	partial	0.5	for all members	1
Frequency of Elections	once in more than 3 years	0	once in 3 years	0.5	once in 1 year	1
Stability of BoD	30% and more change	0	10-30% member change	0.5	0-10% member change	1
Committee structure	No committees	0	over 3 comm.	0.5	3 committees	1
CEO and Chairman positions are separated	no	0			yes	1
<b>Management Team</b>						
Logical and clear organization	no	0	not very plausible	0.5	yes	1
Organization supports reporting structure	no	0	partly supports	0.5	yes	1
Stability	30% and more	0	10-30% member change	0.5	0-10% member change education and experience relevant	1
CEO Background	irrelevant	0	notvery relevant	0.5		1
<b>Investor Relations/AGM</b>						
Conference calls, webcasts, presentations	no	0	yes, but not perfect	0.5	yes	1
Dividend policy	no	0	not 100% clear	0.5	yes	1
Information regarding AGM	no	0	not explicit	0.5	yes	1
<b>Disclosure of Information</b>						
Annual Reports	no	0	limited for past years	0.5	yes	1
Quarterly reports	no	0	half-year	0.5	yes	1
Info on mgmt	no	0	not explicit	0.5	yes	1
Info on BoD	no	0	not explicit	0.5	yes	1
SRI report/part of AR	no	0	not explicit	0.5	yes	1
Ownership Structure	no	0			yes	1
<b>Total</b>		<b>0</b>		<b>9.5</b>		<b>21</b>

CG assessment model includes such crucial evaluation factors as board independence, directors' skill diversification rate, CEO and Chairman roles separation, directors performance-based remuneration. Empirical research has proved that these factors influence shareholder's return. Besides, the authors added the model with meeting frequency assessment, election frequency, board stability and committee structure, which intuitively can impact entity's governance and performance.

**Management**

Basically all sources providing CG evaluation models pay much less attention (if at all) to the management team organization and so does the model. There are just four points included to make a judgment on the executive team as a part of entity's CG. Organization of the management team is the key in the assessment of the quality of management team. Clear responsibility assignment ensures good governance and process management. Relevance of CEO background becomes crucial in the entrepreneurial environment, which is excessively present in the developing CEE countries.

**Investor Relations**

The quality of investor relations gains importance when the company seeks to attract investors as well as support existing ones with up to date information. Shareholders of the particular company are usually curious to understand the

dividend payout policy, be acknowledged with the latest financial results as well as informed about the pending annual general meeting (AGM) and its agenda.

**Information Disclosure**

The degree of information disclosure is of utmost importance in order eliminate the principal-agent problem. Nowadays for the majority of investors only source of information is basically the company's website, where financial reports for various periods are stored. Besides, the authors included assessment of availability of information on management team and supervisory board as well as information on ownership structure and corporate social responsibility. The latter becomes very topical, which is proved by creation of multiple SRI indices (FTSE4Good, Dow Jones Sustainability Index).

**Research Methodology**

Relevance of relationship between share price and quality of corporate governance was tested by using the above explained assessment model. The higher the rating the better is the governance of the company, which was put in relation to share price with the help of linear regression. The authors have also tested whether the market recognizes better governed companies, which was tested by looking at company's valuation.

Overall the authors have obtained the data for 116 companies quoted on stock exchanges in 10 CEE countries: Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Romania, Poland, Slovakia, Slovenia. The corporate governance assessment was based mainly on the latest available annual report (for FY 2009) and the information published on corporate homepages. The share prices for the last 3 years were extracted from the data provided by the CEE stock exchanges, and two periods were considered in the study: 1 year (2010) and 3 years (2008-2010). Turnover of the executive team and board of directors was not considered for the longer (3Y) assessment period due to data irrelevance. However, assessment of other factors was applied retrospectively.

#### 4. RESEARCH RESULTS

##### General Overview of CEE Corporate Governance

The financial markets in CEE are yet in the development phase and so is the attitude towards corporate governance and best practice implementation. The level of corporate governance is very different from country to country (figure 2). Highest overall score was received by Estonian, Lithuanian and Slovenian companies, which have very good information disclosure and excellent investor relations. The lowest score was obtained by the Romanian companies, which are very weak in providing the information, thus making it almost impossible to consider the company as an investment target for a foreign investor.

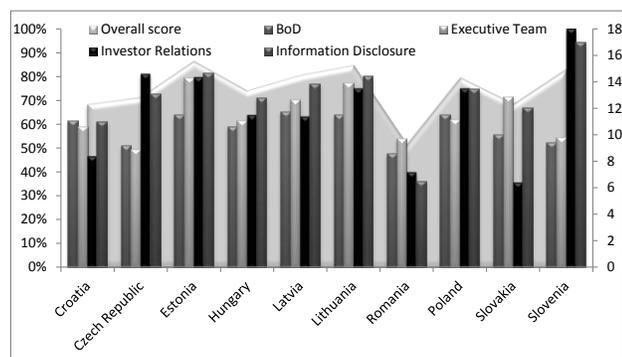


Figure 2. Average CG assessment of CEE companies by countries

It is interesting that in majority of cases Board of Directors (BoD) scores were the lowest compared to other categories. Partial explanation is found in the frequency of elections, where almost all companies were penalized due to not having annual elections (as considered in best practice standards). BoD is being elected once in 3-4-5 years, which makes the assessment of each BoD member's activity and contribution inefficient. Besides, the companies in CEE have established fixed remuneration system for BoD members, while the performance-based remuneration has been proved to be value-creating strategy. Highest scores in executive team evaluation were obtained by the companies from Estonia, Latvia, Lithuania and Slovakia mainly thanks to the high stability of the management team and to its logical organization structure. Surprisingly, Czech and Slovenian companies, though having high overall scores, have rather weak ratings of the executive teams.

The reason for that is unclear executive structure, which often does not correspond with the reporting structure (e.g. regional management organization, while reporting is by

divisions). Moreover, CEO education and experience often is not relevant to the business essence of the company.

Besides, the quality of corporate governance to a great extent depends on shareholding structure. If the company has strategic shareholding of Western European origin (e.g. 51% of *Magyar Telekom* held by *Deutsche Telekom*, 62% of *TEO LT* held by Swedish *Teliasonera*), then the company is significantly influenced by its shareholders and is forced to implement also Western European CG standards. The companies, which have as controlling shareholder local individuals, usually do not bother about complying with recommendations of the local stock exchanges.

Figure 3 chart shows the assigned ratings across the categories. Almost 90% of the companies are disclosing shareholder's structure and have separated roles of CEO and Chairman. The companies are very active in publishing the minimum set of documents (annual and quarterly reports) required by investors, but are not very willing to make additional reporting: presentations, webcasts, CSR reports.

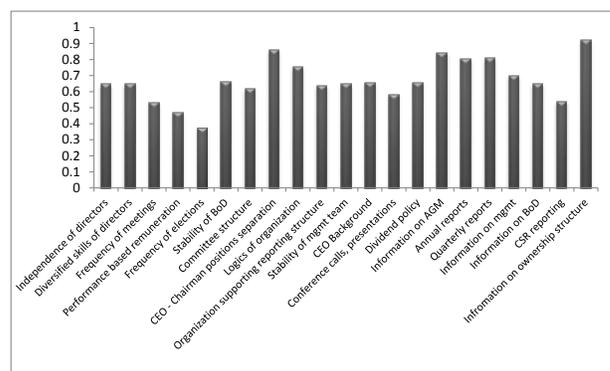


Figure 3. Average CG assessment of CEE companies by criteria

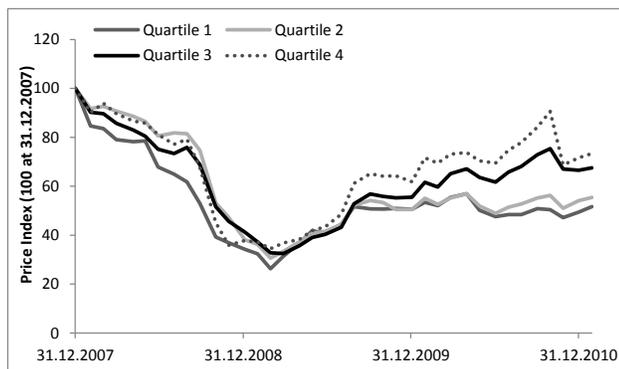
The encouraging trait of CEE companies is BoD independence, its stability and directors skill diversification. Regarding BoD independence it needs to be mentioned, that this factor greatly depends on the shareholding structure: in case of controlling and significant influence of one shareholder, the BoD in most cases represents the interests of this shareholder.

Supervisory board meeting frequency was assessed setting as a best case scenario best practice meeting frequency of 6-8 times a year: not be deeply involved in the business and still manage to understand the company's development and consult executive team. CEE companies tend to have more often BoD meeting, which are usually exceeding 10 times a year with only some rare instances of 3-4 meetings a year. Perhaps, for a dynamic and often uncertain environment which is overwhelmingly felt in CEE countries, often meetings are necessary to be able to timely make decisions.

Management scores for CEE companies are above average in all four categories. One can distinguish also logical and clear organization of the management team, which surely is an advantage of CEE companies.

##### CG Influence on Stock Performance

The companies analyzed have been divided into quartiles according to the Corporate Governance rating. The price index was calculated for each quartile as seen on the chart (figure 4).



**Figure 4.** CEE companies performance according to CG quartiles (1 quartile – worst quality of CG)

The results show that CEE companies with above average quality of corporate governance outperform their peers with weaker CG ratings. As it is seen on the chart, the outperformance starts to be obvious in the recovery phase after the global liquidity crunch. So, most probably during the steep decline on stock markets and high degree of uncertainty CEE region investors recognized the value added of better governed companies and risk associated with poor information disclosure. The same research approach applied to one year period shows that during the last year investor largely praised better governed companies, as top 25% of companies managed to outperform the worst 25% of companies by 0.99% each month. It is also worth noticing that there is a large gap of 1 and 3 year period between stock returns of the best two CG quartiles and worst two CG quartiles. The statistical significance tests show that the relationship between corporate governance quality and stock returns is significant but only for three year time period. Multiple R (correlation) for one year period is 13%, while for three years is 22%, both of which are on a rather level.

Table 2

Regression Parameters	Regression Statistical Significance	
	Time Period	
	1 Year	3 Years
Multiple R	0.13	0.22
R Square	0.02	0.05
Standard Error	0.80	0.43
t-stat	1.16	1.98
F test	1.34	3.94

The hypothesis regarding riskiness of investments could be that investing in companies with high CG scores one is exposed to lower risk. The table 3 provides risk measures by beta and volatility for two time periods.

Table 3

Nr. of Quartile	Risk Figures of CG Quartiles					
	Price Index		Volatility		Beta	
	1Y	3Y	1Y	3Y	1Y	3Y
1	105.2	51.7	9.7%	15.8%	0.91	0.82
2	109.8	55.5	12.9%	15.6%	0.96	0.93
3	121.7	67.6	11.7%	14.8%	0.92	0.97
4	118.6	73.4	14.7%	16.2%	0.59	0.75

Volatility figures are relatively similar for all quartiles especially if we consider longer period. Beta results, basically,

support the hypothesis about riskiness of investments. Best 25% companies, according to CG rating, have lower beta than the companies with weaker score on CG.

Therefore, both during longer and shorter term well-governed companies offer relatively low risk in relation to the market.

## 5. CONCLUSIONS

Corporate Governance term is relatively new for the companies operating in Central and Eastern Europe but in a quite short time period lots of companies were able to offer investors explicit information on their governance system. In many cases the process of establishing CG system was strongly influenced by the controlling shareholder either positively or negatively. Highest CG ratings have been obtained by Baltic as well as Slovenian companies, while Romanian companies are distinguished by very weak corporate information disclosure. The majority of companies analyzed provide extensive disclosure but, it seems, that the institute of BoD is not well understood yet: staggered board elections, high degree of BoD involvement in routine business management, compensation is not linked to performance.

Though the culture of equity investing is still in its development phase in Central and Eastern European countries, the value of good corporate governance is being recognized by the investor. The findings of the present study prove the hypothesis that there is a significant influence of corporate governance quality and stock returns. The results are statistically significant during the longer term period of 3 years. It is also worth noticing that better managed companies are able to offer low risk as measured by stock price beta.

The authors of the study have also checked almost all separate criteria in the model developed (e.g. board independence, meeting frequency). One of the most influential factors on the stock performance is information disclosure quality. Besides, skill versatility of members of BoD had also significant positive influence on the share performance of the analyzed CEE companies.

Research findings clearly show that corporate governance in CEE countries cannot be neglected, so the primary recommendation, based on the study, to CEE investors would be consider the quality of corporate governance, transparency and information reliability. CEE listed companies are also advised to pay decent attention to CG, information disclosure as well as continue developing investor relations at high pace as this can improve corporate stock performance.

## 6. REFERENCES

- [1] H. Aman, P. Nguyen, "Do Stock Prices Reflect The Corporate Governance Quality Of Japanese Firms?", **Journal of The Japanese and International Economies**, Vol. 22, No.4, 2007, pp. 647-662.
- [2] R. Bauer, N. Gunster and R.Otten, "'Empirical Evidence on Corporate Governance in Europe", **Journal of Asset Management**, Vol 5, No. 2, 2004, pp. 91-104.
- [3] S. Bhagat, B. J. Bolton, "Corporate Governance and Firm Performance", **Journal of Corporate Finance**, Vol. 14 (2008, Special Edition on Corporate Governance), pp. 257-273.
- [4] S. Bhagat, B. Black, "The non-correlation between board independence and long term firm performance", **Journal of Corporation Law**, Vol. 27, 2002, pp. 231-274.

- [5] S. Bhagat, D. Carey, Ch. Elson, “Director ownership, corporate performance, and management turnover”, **The Business Lawyer**, Vol. 54, 1999.
- [6] J. Bistrova, N. Lace, “Ownership Structure in CEE Companies and its Influence on Stock Performance”, **Economics and Management**, No.15, 2010, pp. 880-886.
- [7] E.F. Brigham, M.C. Erhard, **Financial Management; Theory and Practice**, South-Western College Pub, 2004.
- [8] J. A. Brickley, J. L. Coles, G. Jarrell, “Leadership structure: Separating the CEO and chairman of the board”, **Journal of Corporate Finance**, Vol. 3, 1997, pp. 189-220.
- [9] W. Drobetz, A. Shillhofer, H. Zimmermann, “Corporate governance and expected stock returns: Evidence from Germany”, **European Financial Management**, Vol. 10, 2003, pp. 267–293.
- [10] P. Gompers, J. Ishii, A. Metrick, “Corporate governance and equity prices”, **The Quarterly Journal of Economics**, Vol. 118, No.1, 2003, pp. 107-155.
- [11] B. Hermalin, M. Weisbach, “Boards of directors as an endogenously determined institution: A survey of the economic literature”, **Economic Policy Review**, 2003, pp. 7-26.
- [12] B. Hermalin, M. Weisbach, “Endogenously Chosen Boards of Directors and Their Monitoring of the CEO”, **American Economic Review**, Vol. 88, No. 1, 1998, pp. 96-118.
- [13] T. Keys, T. Malnight, **Corporate Clout: The Influence of the World’s Largest 100 Economic Entities**, Strategy Dynamics Global Limited ([www.globaltrends.com](http://www.globaltrends.com)), August 2010.
- [14] R. Kouwenberg, “Does Voluntary Corporate Governance Code Adoption Increase Firm Value in Emerging Markets? Evidence from Thailand”, **Research paper**, Mahidol University, College of Management, 2006.
- [15] D. Mueller, E. Peev “Ownership Structures and Investment in Transition Economies”, **Financial Development, Integration and Stability**, pp. 322-336.
- [16] A. Pajuste, “Corporate Governance and Stock Market Performance in Central and Eastern Europe: A Study of Nine Countries, 1994-2001”, **Working Paper**, 2002.
- [17] D. P. Samontaray, “Impact of Corporate Governance on the Stock Prices of the Nifty 50 Broad Index Listed Companies”, **International Research Journal of Finance and Economics**, No. 41, 2010, pp. 7-18.

# **The Model of Sustainable Performing of SMEs in Context of Company's Life Cycle**

**Guna CIEMLEJA, Natalja LACE**

**Faculty of Engineering Economics and Management, Riga Technical University  
1/7 Meza Str., Riga, LV-1007, Latvia**

## **ABSTRACT**

The goal of the research is, studying the performance of SMEs and the factors affecting performance achieving, to propose performance measurement concept encouraging the sustainable development of SMEs and to suggest performance evaluation approach according to company's life cycle. The object of the study is SMEs from the sub-sector of manufacturing industry in Latvia. The authors create a model of sustainable performing of SMEs on the base of the results obtained from qualitative and quantitative research. External and internal business environment factors influencing effective performance of the enterprise and performance indicators that are to be supervised principally, according to the enterprise life cycle phases are included in the model. The developed model of SME sustainable performing has been tested in the sector of printing in manufacturing industry. Calculations and data processing were carried out using Microsoft Excel and EVIEWS software.

**Keywords:** SME, sustainable performance, life cycle, model.

## **1. INTRODUCTION**

An enterprise as a core of any economic system has a great impact on sustainable development of a state or region. Small and medium-sized enterprises (SMEs) are socially and economically important for their national economies, since they represent about 99 per cent of all active enterprises in the European Union. The importance of SMEs to the EU economy indicates a need to assess their performance in order to find appropriate performance measurement and management tools. These issues become very topical during the period of global economic recession.

The most popular form of business in Latvia is the limited liability company, which according to LURSOFT data, in the period from 1991 until 2010, accounts for approximately 60% of the total number of registered companies. Statistical data on the registration and liquidation of a specific form of commercial activity for the last 18 years reveal the issue of viability of the enterprises. For December 2009 25% of the enterprises registered in 2002 as limited liability companies did not survive 7 years. More than 70% of the companies registered earlier (in 1992 and 1993) were liquidated [10].

The problems within an enterprise arise from improper activities, incompetence or even negligence. Performance measurement system can significantly influence and support SMEs' organizational development. Effectiveness and efficiency shall be manifested in all business processes of an enterprise.

Previously conducted research works in the field of performance evaluation of SMEs have not solved all the arising problems due to the specific industry issues. Besides, several disputable questions exist, for instance, whether large companies' performance measurement models can be applied for the needs of small and medium-sized enterprises. It indicates

the necessity for continuation of studying the above-mentioned issues in order to find practical solutions.

## **2. DEVELOPING THE SYSTEM OF AN ENTERPRISE SUSTAINABLE DEVELOPMENT**

Basing on various scholars' researches on issues of sustainable development [2, 3, 8, 11, 13, 17] the authors conclude, that the main point of a concept of «sustainable development» is - coordinated and systemic advancement of economic subject towards the aim takes place only when all three dimensions of sustainability: social, economic and ecologic - are incorporated simultaneously into subject's activities. In this respect the following factors of sustainable development, that are common to all enterprises, can be mentioned: 1) income, which is formed by consumers (clients) utilizing products and services, produced by the enterprise; 2) financial stability and positive dynamics of profitability; 3) competences and skills of the personnel; 4) inclusion of ecological issues in the enterprise's management process; 5) positive attitude of the society towards the performance of the enterprise.

The authors conclude that the sustainability of the enterprise depends on the management system of the enterprise, which provides effectiveness and efficiency of sub-systems, taking into consideration deviations from the state of equilibrium. It demands concretization of the possibilities of practical application of sustainable development concept in the enterprise, taking into consideration that all the processes supporting sustainability of the enterprise are mutually connected, interact, and functional process of each management level is being implemented through dimensions of sustainability. Thus, the quality of the enterprise management influences total result, taking into consideration innovative potential of the enterprise, which includes: management systems, finances, employees, technologies and production.

On his turn, a human being as a special element and factor of the enterprise system complicates functionality of the system with his social expressions, because only a human being can create an idea in this system and implement it. The authors conclude that viability of the enterprise in a long-term period depends on the innovative potential, which is based upon a creative approach that is being implemented by all the stakeholders of the enterprise – not only employees, but also shareholders and customers [9]. It is justified by the results of the implemented activities – discussions in the enterprises.

Important factors for the sustainability of the enterprise are being formed in the social environment. Social capital [7, 19, 21, 25, 26, 27] can be considered as one of the potentials of the enterprise development, which increases return from the use of other capitals. Therefore, to provide a possibility to acknowledge the linkage of the social capital with the enterprise performance in the context of sustainable development, management of social and customers' capital has become of vital importance. It influences productivity, competitiveness and sustainable development of the enterprise

(minimizes operative expenditures for obtaining information, accelerates circulation of information, lessens asymmetry of information and enhances development of new knowledge).

In order to improve efficiency of the system, which results from both enterprise's management and government efficiency, also small enterprises shall seriously turn to evaluation of its performance [4, 5, 15, 18, 22, 23]. Performance measures characterize the fulfillment of goals, but they can be used also as a strategic tool of the enterprise management.

Completing Stafford Beer's [6] idea about the significance of enterprise indicators in providing sustainability of the system, the authors consider that three levels of performance are being formed in the enterprise: *actual*, *target (planned)* and *standard*.

*Actual performance* of the enterprise is being formed in the current time as an actual return from the utilization of existing resources, taking into consideration existing restrictions. *Target (planned) performance* corresponds to the return that the enterprise plans to obtain from the utilizing of existing resources, taking into consideration existing restrictions. On its turn, *standard performance* of the enterprise is an eventual return that can be obtained by the enterprise if it develops existing resources, takes off the restrictions and uses the opportunities, which can be achieved taking into consideration influence of factors maintaining sustainable development of the enterprise. Potential performances of the enterprise include unused opportunities, which are the subject of possible development through using innovations and competent enterprise management.

The authors consider that it is possible to improve efficiency by utilizing the enterprise's performance measurement system, which includes dimensions of sustainable development in combination with the elements – processes supporting sustainable development of the enterprise: *production process, sales process, personnel management process, financial management process, accounting process*.

As a result of interaction between management levels and functional fields and taking into consideration management level, the following issues and characteristics incorporated in performance measures are being formed: 1) strategic level – strategic layout of production machines, choice of placement, development of new products, planning of labour force long-term development, providing profit, selection of accounting technical solutions; 2) administrative level – management of production flows and schedules, development of production price policy and sales promotion campaigns, providing wages, social benefits and acquisitions, analysis, budget planning, supervision of expenses and income, supervision of prime costs; 3) knowledge level – development and designing of new products, analysis of the market situation, research, identification of clients, forecasting employees development and careers, analysis of customers cash flow and survey of possibilities for decreasing risks, forming investment portfolios; elaboration of accounting methodology in the enterprise; 4) operational level – performance of production machines, load control, quality and material consumption analysis; resources and time invested in customers' service; personnel training and environment maintenance expenses, supervision of customers' cash flow and accounting operations.

The authors demonstrate their approach to enterprise performance measurement in the context of sustainable development, which is incorporated in the levels of enterprise management and functional areas, in a pyramid shape (Figure 1).

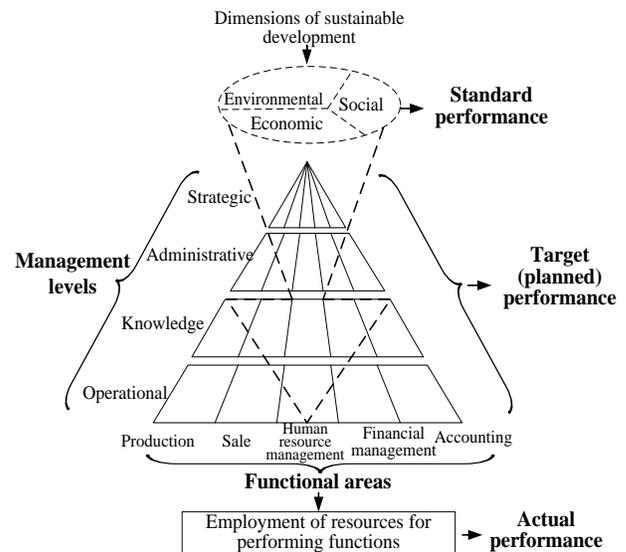


Figure 1 Application of an enterprise sustainable performance measurement system

During the period of its existence, the enterprise as a social-economic system passes through different functional stages, which in literature are defined as enterprise life cycle. Concept of the enterprise life cycle (ELC) [1, 8, 12, 14, 16, 20 24] was created within the enterprise management theory with an aim to explain changes in the enterprise in the context of time.

In literature there is no uniform division of ELC in stages, as well there is no uniform approach regarding number of stages and methodological grounds for defining the stages. A common feature of all the models of the ELC is that cyclic development of the enterprise can be foreseen, and the enterprise shall be able to function under the existing conditions with a future vision, where development possesses consequent and qualitative changes.

Therefore, enterprise management, which is oriented towards solving problems, which are characteristic to the particular stage (phase) of ELC, shall be considered as a condition enabling the enterprise to maintain sustainable performance.

### 3. EMPIRICAL RESEARCH

The authors carried out empirical research in 2010 with an aim to find out opinion of respondents (experts) about external and internal factors influencing the enterprise's performances, significance of performance measures in different stages of the ELC, and how the process of achieving goals is influenced by cooperation with business partners.

Survey questions were divided into 6 groups: the 1<sup>st</sup> group characterized critical influence of macroeconomic external environment and demand factors on the enterprise performance in different phases of ELC (questions 1–6); the 2<sup>nd</sup> group comprised questions about level of skills and abilities to be identified in the internal environment of the enterprise, which define forming of the enterprise offer and influence also performance (questions 7–21); questions of the 3<sup>rd</sup> group are about performance features, which are the basis for indicators describing performance (question 22–28); the 4<sup>th</sup> group represents questions connected with the social dimension of the

enterprise sustainable development, which is characterized by social relations between employees and customers and the influence of these relations on achieving goals of the enterprise and its performance (questions 29–50); questions of the 5<sup>th</sup> group are about acknowledging significance of various resources according to investment of these resources into the final product (questions 51–57); the 6<sup>th</sup> group contains questions, which are about identifying unfavorable factors that are being formed in both external and internal environment of the enterprise and on the level of social relations (questions 58–63).

Data obtained from the questionnaires were generalized, grouped and ranged by using *MS Excel PivotTable* tools. In order to obtain ordinal scale (rank) measurement, a bigger or smaller rank has been assigned to each factor on the ordinal scale. According to answers given by respondents, evaluation is made according to a 5-point system according to 5 criteria: *does not influence at all* – 1 point; *does not influence significantly* – 2 points; *partly does not influence* – 3 points; *influences* – 4 points, *influences a lot* – 5 points. The author accepts that separate phases of the cycle form the total life cycle of the enterprise, and obtained evaluations are gathered to assess influence of each element over the whole life cycle of the enterprise.

As in analysis process of separate factors, conditional evaluation, which is based on determining ranks, was used, for

defining interaction (linkage closeness) between separate factors and features, the authors carried out a correlation test. Coherence of rank features are defined using Spearman's ( $r_s$ ) and Kendall's ( $r_k$ ) rank correlations coefficients (using EViews 6.0 software). In order to obtain statistically valid determination of interaction between separate external environment factors and other manifestations influencing effectiveness of enterprise performance over different phases of ELC, the authors select those pairs of factors, which are characterized by Spearman rank correlation coefficient  $r_s$  at the  $n$  number of observations ( $n=23$ ), with the degree of freedom  $\nu = n-1$ , if the following conditions is fulfilled: 1) coherence is statistically significant at the two-sided significance level with validity level ( $\alpha = 0.05$ ), if  $r_{s\text{ computed}} \geq r_{s\text{ critical}}$ , where  $r_{s\text{ critical}} = 0.428$ ; 2) coherence is statistically significant at the one-sided significance level with validity level  $i$  ( $\alpha = 0.05$ ), if  $r_{s\text{ computed}} \geq r_{s\text{ critical}}$ , where  $r_{s\text{ critical}} = 0.368$ .

Results of the empiric study confirmed the results of the theoretic research results and the approach for assessing effectiveness of SME performance according the ELC phases. External and internal factors influencing performances of the enterprise and performance indicators to be supervised principally according to the phases of the ELC in correspondence with their significance were justified (see Table 1).

Table 1 Factors influencing the enterprise performance and performance indicators to be supervised principally corresponding to the phases of ELC according to their significance

ELC Growth phase	ELC Maturity phase	ELC Decline phase
<i>Factors of external environment influencing performances according to their significance:</i>		
<i>External macro-environment</i>		
Tax laws	Tax laws	Tax laws
<i>External micro- environment</i>		
Consumer purchasing power Qualified labour force Contacts with business partners in external environment Resource access	Consumer purchasing power Qualified labour force Obtaining new information Equal partners	Consumer purchasing power Qualified labour force Availability of external financial resources Relations with clients
<i>Factors of internal environment influencing performances according to their significance:</i>		
<i>Social</i>		
Quality level of clients' servicing Secure and stable relations with clients Intercommunication among employees	Quality level of clients' servicing Secure and stable relations with clients Intercommunication among employees	Secure and stable relations with clients
<i>Environmental</i>		
Ability to improve products	Ability to improve products Wide assortment of goods and services Ability to react to changes in market Ability to introduce innovations	Ability to improve products Wide assortment of goods and services Ability to react to changes in market Possibilities to improve manufacturing processes
<i>Goal achievement:</i> Value system of the enterprise, which is suitable for business partners; enterprise-wide uniform value system; mutual trust of employees		
<i>Performance features according to their significance</i>		
Rate of asset turnover Ability to provide revenues Cost structure	Rate of asset turnover Ability to achieve goal (productivity) Ability to provide revenues	Ability to achieve goal (profitability) Rate of asset turnover Cost structure
<i>Performance indicators to be supervised principally according to their significance</i>		
Liquidity Marginal revenues	Liquidity Productivity	Profitability Liquidity
Indicators of social and environmental factors according to the specifics of the sector		

Taking into consideration significance of the performance features in each phase of ELC, a totality of performance indicators that are to be supervised principally is being formed, as well as model of sustainable performing of SMEs, taking into consideration phases of the enterprise life cycle, was elaborated.

### 3. THE MODEL OF SMEs SUSTAINABLE PERFORMING

The model of SME sustainable performing comprises two stages: 1) Determining of phases of the enterprise life cycle and 2) Continuous improvement process, which is directed towards sustainable development.

To determine the phase of the enterprise life cycle identification parameters have to be applied. Basing on the theoretical study of ELC, the authors suggest using the following parameters for identifying the phases: 1) turnover dynamics, which is connected with forming of customers' basis as the source of enterprise revenues; 2) profit dynamics 3) balance of cash flow; and 4) level of processes' formalization (Table 2).

Table 2 Identification parameters in the context of enterprise life cycle

ELC Growth phase	ELC Maturity phase	ELC Decline phase
Sector growth rates (Ts) compared with the enterprise commercial turnover growth rate (Te)		
Te>Ts	Te ≈ Ts	Te<Ts
Dynamics of enterprise profit (Pe)		
Te>Pe	Pe>Te	Pe dynamics is negative
Balance of cash flow (CF) (operative - OCF, financial - FCF, investments - ICF)		
OCF+ FCF+ ICF-	OCF+ FCF- ICF-	OCF+/- FCF+/- ICF+/-
Level of formalization		
Low (till 20% of total processes)	High (40-80% of total processes)	Very high (above 80% of total processes)

The second phase of the model "Continuous improvement process" consists of four sub-phases: 1) the assessment of actual, standard and planned indicators which should be supervised principally, 2) determining the unused efficiency, 3) elaborating of action mechanism to improve enterprise unused efficiency and effectiveness and 4) correction of target values of performance measures.

### 4. TESTING OF THE MODEL ON AN ENTERPRISE OF PRINTING INDUSTRY

To test the model of the sustainable performing of SMEs the authors applied it in the medium-sized enterprise from the printing sector of the manufacturing industry.

In order to define enterprise compliance to a certain ELC phase, the authors took into consideration regularities of enterprise and sector development. For featuring dynamics of performances of the printing sector of manufacturing industry, the indicator of chain changes is being used. The authors consider that performance of the enterprises of printing sector can be characterized using data gathered by the Association of

Latvian Printing Companies (ALPC) about turnover dynamics of the members of the Association. In order to enable an enterprise to form the budget for the next periods, taking into consideration the tendencies of the sector, the ALPC forecasts for the near future (1 to 3 years) can be used. Forecasts are being obtained based upon the evaluation of the sector experts and taking into consideration processes and tendencies in the European and global printing production market. Using data of the particular printing enterprise, the authors determine the enterprise compliance to a certain phase of ELC according to the model of SMEs sustainable performing and identification parameters in the context of the enterprise life cycle (Table 2).

According to the carried out analysis the particular enterprise is in its decline phase and according to the performance features and measures (Table 1) of the decline phase of ELC actual level of the indicators to be supervised principally (profitability and liquidity) shall be determined, as well as social and environment influence indicators shall be determined according to the specifics of the printing sector.

In order to determine standard value of performance indicators to be supervised principally, it is necessary to use characteristics of the sector, which are connected to the turnover per employee or productivity.

To observe ecological indicators in the printing industry, the indicator of effective use of paper, which characterizes level of technology, was chosen.

In the group of social impact indicators, the authors choose dynamics of changes in number of sick-leave days, which according to employee's length of service and long-term sick-leave can be related to occupational sickness, as well as dynamics of expenditures allocated to compulsory health checks. As an important indicator, the authors mention expenditures allocated to improvement of professional skills of employees of the enterprise, per employee per year, because it indicates a possibility of enterprise sustainable development using innovations and knowledge as an element of support mechanism.

Basing on data provided by the ALPC and the printing enterprise, standard, target (planned) and actual performance measures of the enterprise according to the decline phase of ELC are summarized in the Table 3.

Table 3 Indicators for the printing enterprise in 2009

Name of the indicator	Value of indicator		
	Standard	Planned	Actual
Turnover, LVL	2 502 900	10% decrease of enterprise turnover in 2008 i.e. 1 100 000	838 928
Profitability, %	3.85	1.00	-0.29 (= 0)
Liquidity, coefficient	Sufficient liquidity =2.01	Sufficient liquidity =1.76	Actual liquidity = 1.41
Coefficient of effective use of paper	0.85	0.79	0.82
Expenditures to improvement of professional skills per employee, per year, LVL	Aspiration to max. rational possible level	25.00	12.89

Determining the unused efficiency is necessary to identify reserves that exist in the enterprise in the context of sustainable development. The relevant levels of the enterprise indicators to be supervised principally for 2009 are presented in Figure 2; 3 and 4.

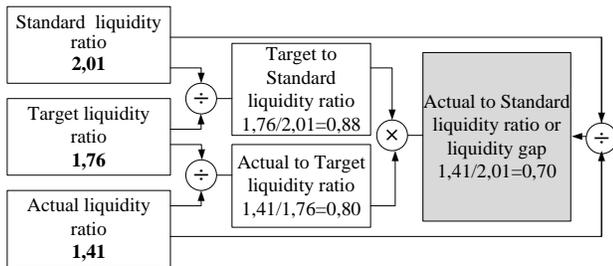


Figure 2 Assessment of the enterprise liquidity indicator

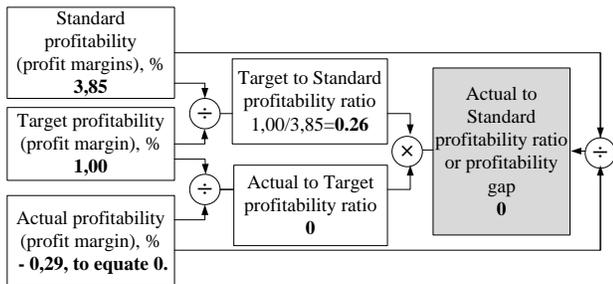


Figure 3 Assessment of the enterprise profitability indicator

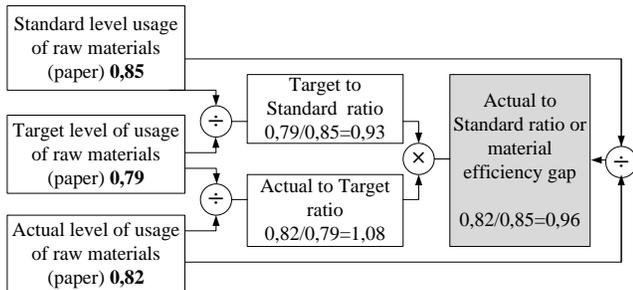


Figure 4 Assessment of the indicator of useful consumption of the paper by the enterprise

According to data presented in the Table 1, before the aging phase enterprise gets through the maturity phase with its characteristic features and indicators to be supervised principally – liquidity and productivity. Monitoring of these indicators, and especially that of productivity according to standard, planned and actual levels, helps the enterprise in maintaining effective management. The goal of the above mentioned might be prolonging maturity phase as much as possible in order to transfer into the new quality – in the growth phase again, and not to transfer to the decline phase. Problems of the decline phase, which occur in connection with low levels of profitability and liquidity, limit abilities of the enterprise to receive external financing support (bank credit, leasing) which is necessary in this phase to renew technologies and re-structure the enterprise. The author considers that enterprise position in the decline phase indicates the competence level of the enterprise management.

Activities to be implemented to improve enterprise unused efficiency and effectiveness, according to the author shall be connected with: 1) improvement of the value created by existing and new clients, and revealed by the commercial profitability; 2) improvements, connected with rising productivity, in use of labour force as the basic element of social capital; 3) optimization of cost structure with an aim to reach the lowest possible level of marginal revenues; 4) increasing efficiency of assets utilization. Commercial profitability of the enterprise may change if terms of enterprise income tax (tax rate; algorithm, which is used in tax calculations; cases when enterprise has obtained durable items – equipment and utilities) change. However, this indicator reflects how profit is being formed during manufacturing and sales of products and services. Knowledge and skills are of crucial importance in promoting innovations. Therefore, innovations, which currently cannot be envisaged without a social cooperation among enterprises and other market players, have to be considered as an element of the possible action mechanism, which provides an opportunity to improve the current level of effectiveness.

The printing enterprise in the decline phase can improve its performances by investments in innovations and technologies, taking into consideration that also the structure of production shall be changed. It can also be done by increasing customer equity, which is connected with revision of relations with existing clients and development according to the chosen strategy.

## 5. CONCLUSIONS

During developing the model of sustainable performing of small and medium enterprises in the context of ELC, the following conclusions have been made:

1. Ability of the enterprise to function in a coordinated and systemic manner, without losing capacity of performance in indefinite future, shall be connected with sustainable development, which incorporates all three dimensions: social, economic and ecologic. The basis for implementing sustainability is formed by the enterprise management systems, which provide functional efficiency and effectiveness of sub-systems, taking into consideration principles of sustainable performance.
2. Significant factors providing sustainability of the enterprise are formed in the social environment, because, in the context of an enterprise performance, management of knowledge, human and social capital is the «process of value creation», which shall be maintained taking into consideration the peculiarities of human resources management.
3. Development of the enterprise is cyclic. It is being created as a totality of stages, which forms uniform phases characterized by specific goals and tasks. Fulfillment of which fully drives the enterprise towards sustainable development.
4. An enterprise sustainable performance is connected with certain parameters, which change along the transfer from one stage of the life cycle to another. These parameters change, because goals, strategy, organizational structure, processes, technology and culture change. Thus the enterprise management, which is directed towards solution of the problems that are characteristic to the respective phase of the enterprise life cycle, is to be considered as a pre-condition enabling enterprise sustainable performance.

Several proposals how to increase efficiency and effectiveness of small and medium-sized enterprises were made:

1. In order to improve results achieved by small and medium-sized enterprises and maintain sustainable development, the authors recommend complementing indicators of the enterprise management system with economic, environment and social indicators, which correspond to the specifics of the sector of the particular enterprise. It shall be done to define standard and planned indicators in each dimension of sustainability.
2. Indicators, which are to be supervised principally and correspond to the phase of enterprise life cycle, shall be included in the competence of the enterprise financial and management accounting along with other indicators selected by the enterprise and characterizing enterprise performances. It shall be done to provide sustainable development and management decision-making in due time to reach this goal.
3. Taking into consideration phases of enterprise life cycle and applying the developed model of sustainable performing for small and medium-sized enterprises that can be used for performance control and management in the context of sustainable development, SMEs can improve their action mechanism according to the actual needs of the enterprise.

## 6. REFERENCES

- [1] I. Adizes, *Managing Corporate Lifecycles*, Paramus, NJ: Prentice Hall Press, 1999.
- [2] S. Banerjee, "Who Sustains Whose Development? Sustainable Development and the Reinvention of Nature", *Organization Studies*, Vol. 24, No. 1, 2003, pp. 143–180.
- [3] P. Bansal, "Evolving sustainably: a longitudinal study of corporate sustainable development", *Strategic Management Journal*, Vol. 26, 2005, pp. 97-218.
- [4] J.B. Barney, "Firm resources and sustained competitive advantage", *Journal of Management*, Vol. 17, 1991, pp. 99–120.
- [5] R. Basu, "New criteria of performance management: a transition from enterprise to collaborative supply chain", *Measuring Business Excellence*, Vol. 5, No. 4, 2001, pp. 7-12.
- [6] S. Beer, "Fanfare for Effective Freedom", 1973, Retrieved from <http://www.williambowles.info/sa/FanfareforEffectiveFreedom.pdf>.
- [7] N. Bosma, R. Thurik, G. de Wit, "The Value of Human and Social Capital Investments for the Business Performance of Startups", *Small Business Economics*, Vol. 23, No. 3, 2004, pp. 227–236.
- [8] K.S. Cameron, M.U. Kim, D.A. Whetten, "Organizational dysfunctions of decline", *Academy of Management Journal*, Vol. 30, No. 1, 1987, pp. 126-138.
- [9] G. Ciemleja, N. Lace, "The factors determining innovation-based attitude of Latvian SMEs towards sustainability", *5th International Scientific Conference "Business and Management –2008"*, May 16-17, 2010, Vilnius, Lithuania, pp. 28 – 36.
- [10] G. Ciemleja, N. Lace, N. Koleda, "Quantitative harmonious model of sustainability factors: measuring contribution of financial viability", *The 6th International Scientific Conference "Business and Management 2010"*, May 13-14, 2010, Vilnius Lithuania, pp. 104-112.
- [11] F. Dixon, "Total Corporate Responsibility: Achieving Sustainability and Real Prosperity", *Ethical Corporation Magazine*, 2003 [Electronic resource]. – Read 30 May 2009.-[www.ethicalcorp.com](http://www.ethicalcorp.com)
- [12] M. Epstein, M. Roy, "Sustainability in Action: Identifying and Measuring the Key Performance Drivers", *Long Range Planning*, Vol. 34, 2001, pp. 585–604.
- [13] J. Fiksel, "Emergence of a Sustainable Business Community", *Pure and Applied Chemistry*, Vol. 73, No. 8, 2001, pp. 1265-1268.
- [14] L.E. Greiner, "Evolution and revolution as organizations grow", *Harvard Business Review*, Vol. 50, No. 4, 1972, pp. 37–46.
- [15] R.D. Gulbro, L. Shonesy, P. Dreyfus, "Are small manufacturers failing the quality test?" *Industrial Management & Data Systems*, Vol. 100, 2000, pp. 76–80.
- [16] S.H. Hanks, C.J. Watson, E. Jansen, G.N. Chandler, "Tightening the life cycle construct: A taxonomic study of growth stage configurations in high technology organizations", *Entrepreneurship: Theory & Practice*, Vol. 1-8, No. 2, 1993, pp.5–30.
- [17] K. Hockerts, "The sustainability radar – a tool for the innovation of sustainable products and services", *Greener Management International*, Vol. 25, 1999, pp. 29-49.
- [18] E.K. Laitinen, G. Chong, "How do Small Companies Measure Their Performance?", *Problems and Perspectives in Management*, Vol. 4, No. 3, 2006, pp. 49-68.
- [19] M.L. Lengnick-Hall, C.A. Lengnick-Hall, "HR's role in building relationship networks", *Academy of Management Executive*, Vol. 17, No. 4, 2003, pp. 53-63.
- [20] E. Masurel, K. van Montfort, "Life cycle characteristics of small professional service firms", *Journal of Small Business Management*, Vol. 44 No. 3, 2006, pp. 461–473.
- [21] M.W. McElroy, "Social innovation capital", *Journal of Intellectual Capital*, Vol. 3, No. 1, 2001, pp. 30–39.
- [22] A. Neely, "The performance measurement revolution: why now and what next?", *International Journal of Operations & Production Management*, Vol. 19, 1999, pp. 205 – 228.
- [23] Neely, A. (2002). *Business Performance Measurement: Theory and Practice*, Cambridge University Press.
- [24] K.C. Robinson, P.P. McDougall, "Entry Barriers and New Venture Performance: A Comparison of Universal and Contingency Approaches", *Strategic Management Journal*, Vol.22, No. 6/7, 2001, pp. 659-685.
- [25] R.T. Rust, K.N. Lemon, D. Narayandas, *Customer Equity Management*, Prentice Hall/Pearson Education, Inc. Pub., 2005.
- [26] C. Trigilia, "Social Capital and Local Development", *European Journal of Social Theory*, Vol. 4, No. 4, 2001, pp. 427-442.
- [27] H. Westlund, *Social Capital in the Knowledge Economy. Theory and Empirics*, Springer-Verlag Berlin Heidelberg, 2006.

# **Dynamic Financing Models of Tertiary Education According to Market Needs**

**Ingars ERINŠ, Ronalds TARAŠKEVIČS**

**Faculty of Engineering Economics and Management, Riga Technical University  
1/7 Meza Str., Riga, LV-1007, Latvia**

## **ABSTRACT**

Economic methods and decentralisation are the most effective means in managing any process. Any funds of tertiary educational institutions, both earned by the institution and allocated as state budget subsidies, should be handed over directly without intermediaries to the parties who performed the work. Funds should be allocated according to the unified methodology, and the workload to be accomplished in a definite period of time is the main criterion. Workload to be accomplished is measured according to strictly defined and measurable parameters.

In actual practice, there is a tendency to add work quality as an additional criterion. It is problematic to apply the quality criterion because, firstly, the quality of any work performed by professionals cannot be below certain standards; secondly, there are no measurable, definite parameters for quality assessment if work outcome is a service. Education is a commodity, specific commodity actualised as a service.

The author offers the algorithm of calculation methodology of fund allocation that is described considering the experience of Riga Technical University, where this methodology has been implemented.

**Keywords:** higher education, state budget, tuition fee.

## **1. INTRODUCTION**

The provision of financing for higher education has always been an actual topic, for it to be appropriate. One of the most relevant questions in the world is financing for education – of which already are many articles and publications based on researches. The country – the government – always is trying to finance higher education, but it will always be insufficient and will not provide for all the functions

to work correctly [1]. A topicality has become the rational and effective use of budget resources for the provision of education process, as well as obtaining the assessment of the results.

In the overall financial management methodology the subsidies of government budget and the financing of universities themselves – it further distribution and use development issues in higher education institutions are very important, because on it depends the rational and effective use of funds, as well as from government, as from private sources [3].

As one of the most important principles is the determination of the accordingly base unit of distribution. To determine and choose in advance of which ensure the study process and realize the other activities, should be distributed financial means, therefore, who are the main beneficiaries of these funds and users of higher education institutions [2]. By means of a common methodology the funds are distributed and it results in total structural unit revenue, of which is settled the public utility payments and covered other expenses. In the end result there is created balance sheets of separate departments and university revenue – expense balance sheet.

Methodology reveals how out of the state budget distributed grants – no budget revenue – could be divided and directed to study process realizing, considering certain conditions and principles for these funds rational and more efficient use – it is based on its lifelong approbation experience in RTU.

## **2. LITERATURE REVIEW**

It is necessary to perform relevant groundwork for the establishment and implementation of decentralised automatic management system at the tertiary educational establishment:

1) To select a cornerstone for the system architecture: either an individual or a structural unit.

A structural unit was chosen as a less dynamic but stable element of the system, which can serve as a basis for a flexible management system. Further it is necessary to establish a legal basis for the structural units – the structural units are established and liquidated by decision of the Senate of the tertiary educational establishment.

2) When an array of structural units of the tertiary educational establishment is formed, it should be connected to the array of employees, thus developing an automatic management system “Personnel”. It is necessary to allocate employees such as teaching and maintenance staff to the corresponding structural units; to elaborate staff recruitment and dismissal policies, to determine financial responsibility and any additional duties in accordance with labour laws and regulations. Any necessary limitations should be incorporated into a system.

3) The register of the study courses taught at the tertiary educational establishment and used in study programme design was developed at the next stage. The programmes in their turn are offered in the Latvian and foreign tertiary education market. It was specified that study courses can be registered using a special algorithm only by professors, associate professors, assistant professors and lecturers. The registered study courses were evaluated by the Senate Commission of the tertiary educational establishment and were accepted or declined by the Department of Academic Affairs. Accepted study courses were included in the register. All study courses were divided into four categories: compulsory study courses characteristic of tertiary education, the range of which determines whether a tertiary educational establishment majors in exact sciences or humanities, compulsory study courses aiming at training skills and competences necessary for a particular profession, optional study courses, which provide opportunity to gain additional skills in the chosen professional field, elective study courses which students can choose to enrich their competence in particular fields whether because they lack certain knowledge or they want to acquire additional knowledge, skills and competences in the fields unrelated to mastering a particular profession [7]. After that it is necessary to determine the procedure of amending the study courses, and who is eligible to do that. As study courses are registered by the teaching staff of the tertiary educational establishment, and the academic staff in their turn is allocated to the structural units, the study courses are automatically allocated to the structural units. The structural units are legally

responsible for the quality of the study courses and the study process as well as for providing educational and methodological literature, facilities and equipment. Then it is necessary to solve a problem concerning a study course that arises when teaching staff switch structural units or quit their job; it should be determined how the person who registered a study course can authorize other members of the teaching staff to teach the study course in question, for example, at affiliates.

4) At the next phase the Study Programme Register of the tertiary educational establishment should be developed. After that it is necessary to determine the procedures, financing and legal responsibility for design, registration, evaluation, endorsement, licensing and accreditation of the study programmes in compliance with laws and regulations of the Republic of Latvia. Study programmes can be designed only of the study courses that are endorsed and included in the Study Course Register of the tertiary educational establishment. If there is a need for additional study courses, they have to be registered, endorsed and included in the Study Course Register, and only after that it is possible to proceed with the design of the study programme.

5) In order to fairly, independent on the subjectivism of the human factor, allocate state budget funds and student tuition fees to each structural unit using a unified algorithm, it is necessary to develop a principal calculation document [5]. Salaries and other fund limits for an academic year or semester assigned to each structural unit are calculated on the basis of this document. The head of the study programme or an authorized person develops an electronic version of the principal calculation document. The head of the study programme holds full responsibility for the validity of this document.

6) Completed and developed principal calculation document is electronically sent to the responsible financial manager of the tertiary educational establishment, who verifies and endorses it.

7) Prior to allocation of budget funds of the tertiary educational establishment, it is necessary to form 5 types of arrays of the endorsed principal calculation documents: allocation of state subsidies (state order) for an academic year (12 months), allocation of tuition fees paid by full-time students for a semester (6 months), allocation of tuition fees paid by part-time students for a semester (6 months), allocation of tuition fees paid by extra-mural students for a semester (6 months), and

allocation of tuition fees paid by full-time, part-time, extra-mural students and those enrolled at individual study programmes for a semester (6 months). State subsidies are allocated to structural units once per academic year (in October), but tuition fees – twice per academic year (in November and April). The fiscal year of the tertiary educational establishment should start on 1<sup>st</sup> October and finish on 30<sup>th</sup> September. Setting such a time period allows reducing university bankruptcy risks significantly, as state subsidies are transferred to the tertiary educational establishment accounts starting with 1<sup>st</sup> January, but tuition fees – during August and September for the first semester, and during February and March for the second semester.

8) Having formed the above-mentioned arrays of principle calculation documents, the Financier (Chancellor), in cooperation with the Head of the Planning and Finance Department, develops an electronic version of regulations on calculation of funds for each array of principal calculation documents for a particular academic year or semester. If studies are offered on a self-funded basis and are organized individually, i.e. only to the students enrolled at the study programme, separate regulations on calculation of funds for the principal calculation document of each study programme are developed.

9) When the arrays of principal calculation documents and the regulations on calculation of funds are developed for an academic year or a semester, funds can be allocated. The above-mentioned arrays and regulations can be developed both simultaneously and successively, but calculations are made only successively by fund type.

The methodology can be used only for allocation of state budget subsidies and tuition fees; other income is allocated to structural units according to other methodologies or procedures confirmed by the Senate of the tertiary educational establishment.

All funds that are transferred to the account of a tertiary educational establishment are allocated to all structural units of the establishment according to the unified methodology or procedure. The units are the actual performers of the given work. The allocation of any funds to structural units is based on the principle that funds, which are assigned for a special work to be accomplished, should be handed over directly without intermediaries to the structural units, which specialise in the implementation of the work, and are responsible for it [4]. At the time of

calculation the tertiary educational establishment has „m” number of structural units (this number is dynamic), which can be divided into two categories: the structural units which receive state subsidies, and the structural units that should earn funds themselves, because they instruct specialists at the programmes, which are not commissioned by the state. All funds, which the structural units manage to raise themselves following the code of ethics of the tertiary educational establishment, remain at the disposal of the structural units [6]. Tuition fee is set for each study programme separately following the current actual costs and the tertiary education market situation in the given year. The tuition fee is confirmed by the Senate of the tertiary educational establishment annually, and it remains fixed during the entire study period. The methodology of allocation of budget subsidies and tuition fees is confirmed by the Senate of the tertiary educational establishment; this issue is annually reviewed by the Senate, and adjusted if necessary.

The amount of budget subsidies allocated to the tertiary educational establishment by the Ministry of Education and Science or tuition fees paid by the students, which is specified in the regulations on the calculation of funds, is initially divided into two parts: salaries and other funds according to the ratio specified in the regulations. Calculation methodology is described considering the experience of Riga Technical University (RTU).

$$N_{d/a} = 0.60 N_{RTU} \text{ and } N_{p/l} = 0.40 N_{RTU} \quad (1)$$

where

$N_{RTU}$  - budget subsidies allocated to RTU by the Ministry of Education and Science or tuition fees, LVL;

$N_{d/a}$  - salary fund, LVL;

$N_{p/l}$  - other funds, LVL.

According to the decision of the Senate of the tertiary educational establishment, 1% of the above-mentioned funds is dedicated to internal (RTU) scientific projects and is allocated on a competitive basis by the Commission appointed by the Vice-Rector for Research.

$$N^z_{d/a} = 0.01 N_{d/a} \text{ and } N^z_{p/l} = 0.01 N_{p/l} \quad (2)$$

where

$N^z_{d/a}$  - salary fund for internal (RTU) scientific projects, LVL;

$N'_{p/l}$  - other funds for internal (RTU) scientific projects, LVL.

The remaining sum is calculated for further allocation:

$$N'_{d/a} = N_{d/a} - N^z_{d/a} \text{ and } N'_{p/l} = N_{p/l} - N^z_{p/l} \quad (3)$$

where

$N'_{d/a}$  - the remaining part of salary fund for further allocation, LVL;

$N'_{p/l}$  - the remaining part of other funds, LVL, which should be respectively reduced for further allocation:

$$N''_{p/l} = N'_{p/l} - 0.2409 N_{d/a} - S^t - 8,0 S_v \quad (4)$$

where

$N''_{p/l}$  - the remaining part of other fund for further allocation, LVL;

$0.2409 N_{d/a}$  - compulsory social security contribution, LVL;

$S^t$  - scholarships and transport service compensation, LVL;

$8.00 S_v$  - library funding (LVL  $8.00 * S_v$  - the number of RTU students studying on the state budget confirmed by the Ministry of Education and Science or the number of students, who pay tuition fee, LVL).

The next step is to deduct funding from  $N'_{d/a}$  and  $N''_{p/l}$  for centralized services in compliance with ratios specified, which are calculated for a year and confirmed by the Senate. Thus, the salary fund volume allocated to the structural units (departments and professor groups)  $N'_{f d/a}$  and the volume of other funds  $N''_{f p/l}$  are acquired.

In order to sustain the most qualified teaching staff, professors, and to facilitate ongoing training of teaching staff in the structural units, the value denoting a fixed portion of salary -  $N'_p = \text{LVL } 200 * P^s_{v.sk.}$  and a fixed portion of other funds -  $N''_p = \text{LVL } 6.46 * 50 m^2 * P^s_{v.sk.}$  is introduced in the model, where  $P^s_{v.sk.}$  is a number of professor positions on the permanent staff at the university at the moment of calculation. The volume of the fixed portion for each structural unit depends on the number of professors on the permanent staff. These values are calculated for a month and refer to allocation of state budget funds only; in order to calculate the total for a year, the value should be multiplied by 12.

Concurrently, the work load for an academic year or semester for each study course, study programme, faculty and university in general is calculated in the array of principal calculation document [8]:

$$D^i_{MP} = S_v^i * K P^i * b^i * a^i \quad (5)$$

where

$D^i_{MP}$  - work load at the  $i^{th}$  study course;

$S_v^i$  -  $i^{th}$  number of budget study places or students at the study course;

$K P^i$  -  $i^{th}$  number of credit points at the study course;

$b^i$  -  $i^{th}$  study course domain ratio;

$a^i$  -  $i^{th}$  study course study level ratio.

Having summarised the work load per study programme for all study courses, the  $i^{th}$  study programme work load is arrived at -  $D^i_{s/p}$ .

Having summarised the work load per all study programmes implemented by a faculty, the  $i^{th}$  faculty work load is arrived at -  $D^i_f$ .

Having summarised the work load per faculties, institutes and affiliates of the university, the work load of the university is arrived at -  $D_{RTU}$ .

Finally, salary and other expenses attributable to one study place or per credit point are calculated:

$$n'_{d/a} = \frac{N'_{f-d/a} - N'_p}{D_{RTU}} \quad (6)$$

$$n''_{p/l} = \frac{N''_{f-p/l} - N''_p}{D_{RTU}} \quad (7)$$

When the value of one study place or student credit point for a study year or a semester has been calculated according to the types of funds to be allocated, salary fund and other fund limits are calculated.

1) Resources for a separate study course are calculated within the framework of each study programme, defining the amount of salaries and other funds (expenses):

$$N^i_{MP} = D^i_{MP} * n'_{d/a} \text{ (LVL)} \quad (8)$$

$$N''^i_{MP} = D^i_{MP} * n''_{p/l} \text{ (LVL)} \quad (9)$$

where: the amount of salary and other funds attributable to  $i^{th}$ -study programme is a sum of the

values of the study courses, which the study programme is designed of.

2) The fund limits attributable to departments and professor groups, salaries and other funds are determined as a sum of calculated value of salaries and other funds attributable to the corresponding structural unit, fixed portion added (only for subsidies). Subsidies are calculated for a period of a year, tuition fees – for a semester.

Calculations are built into software and processing takes up about 2-3 hours. Eventually, three types of documents are drawn up: study programme work load, LVL; structural unit work load, LVL, as well as study course work load within a structural unit, LVL. These documents electronically as well as in the printed form are sent by the responsible financial manager or Chancellor service to the Planning and Finance Department and to structural units. Particular applications are used to input the data into accounting software. The heads of the structural units adjust these limits within a period of two weeks, reallocating resources to or receiving from other structural units according to calculated fund limits. Figures for adjusted fund limits should be submitted to the Planning and Finance Department. The number of employees and their salaries should be adjusted in accordance with new fund limits. The access to this information is differentiated: the heads of departments have access only to the data concerning funds and expenditure attributable to their department, the heads of institutes have access to the data on funds attributable to structural units within the institute, the deans have access to the data on funds attributable to structural units within the faculty; the head of the Planning and Finance Department, Rector, Vice-Rectors and Chancellor have access to the data attributable to all structural units.

## CONCLUSIONS

Education is a commodity, specific commodity actualised as a service. All funds that are transferred to the account of a higher educational establishment are allocated to all structural units of the establishment according to the unified methodology or procedure. The allocation of any funds to structural units is based on the principle that funds, which are assigned for a special work to be accomplished, should be at the disposal of the structural units, which specialise in the

implementation of the work, and are responsible for it.

It is necessary to perform relevant groundwork for the implementation of decentralised automatic management system. This methodology can be used only in allocation of state budget subsidies and tuition fees.

## REFERENCES

- [1] Bruce Johnstone D. (2006). Financing Higher Education: Global perspectives on higher education. AW Rotterdam: Sense Publishers, Vol.3, pp.292.
- [2] Bruce Johnstone D., Marcucci N. P. (2010). Financing Higher Education Worldwide. Maryland: The Johns Hopkins University Press, pp.322.
- [3] Herbst M. (2007). Financing Public Universities: Higher Education Dynamics. The Netherlands: Springer. Vol.18, pp.235.
- [4] Higher Education in the World 2006: The Financing of Universities. Guni series on the social commitment of universities. New York: Palgrave Macmillan, pp.330.
- [5] Knight J. (2009). Financing Access and Equity in Higher Education: Global perspectives on higher education. AW Rotterdam: Sense Publishers, Vol.17, pp.235.
- [6] Odebero S. (2010). Innovation in higher education financing in developing countries: The potential of helb loans in equalizing higher education opportunities empirical evidence from Kenya. Leipzig: LAP LAMBERT Academic Publishing GmbH & Co. Kg, pp.173.
- [7] Strickfaden H. M. (2009). The German Market of Private Higher Education Financing: Are the Existing Product Suitable to Satisfy Students` Needs?, pp.138.
- [8] Malceva G.I., Penchuk E.S., Fadeykina N.V. (2002). Organizaciono-ekonomicheskij mehanizm i finansovoje regulirovanije visshogo profesionalnogo obrazovaniya. Novosibirsk: Sibirskij institut finansov i bankovskogo dela (SIFBD), pp.160 (in Russian).

# **Information Technologies for Sustainable Business Development: Case of Latvia**

**Elīna GAILE-SARKANE**

**and**

**Deniss ŠČEULOVS**

**Faculty of Engineering Economics and Management, Riga Technical University, 1/7 Meza Street,  
Riga, LV-1048, Latvia**

## **ABSTRACT**

The aim of the proposed scientific research is to determine information technology's and information's utilization possibilities in company long-term development. Authors analyze and show significant importance of the information technology and information resource as its main part for the business and role production development and market positioning. Authors offer their view to the information resource utilization problems and propose several new definitions: information resource, information power (outcome etc.), information resource capacity, and revise ways how to estimate information value (importance, relevance) as mean and source for the business using methods of competition and information units. Conducted research allows evaluate importance of the information resource for the business and shows its overall relevance to the national economy.

**Keywords:** information technologies, resource, source, mean, competitiveness, sustainability, evaluation.

## **1. INTRODUCTION**

Sustainability will become increasingly important to business and management over time, and the risks of failing to act decisively are growing.

Since 1987 when the concept of "sustainable development" was publicly introduced by the Brundtland report of World Commission on Environment and Development, this notion has gained immense popularity among scientists, governments, and business people. A few years ago in a speech at an event organized by business action for sustainable development Kofi Annan, former UN Secretary-General, remarked: "And more and more we are realizing that it is only by mobilizing the corporate sector that we can make significant progress. The corporate sector has the finances, the technology and the management to make this happen" [21]. The recognition of importance of business involvement with

sustainable development has aroused a variety of business partnerships [22].

Sustainable development has broad appeal and little specificity, but some combination of development and environment as well as equity is found in many attempts to describe it [11].

Sustainable development is commonly perceived as fighting poverty while protecting the environment on a macro-level [2].

U.S. National Academy of Sciences review of the sustainability literature identified three major values to be sustained: nature, life support systems and community [3]. Sustainable development when incorporated by the organization is called corporate sustainability and it contains, like sustainable development, all three pillars: economic, ecological and social. These three dimensions interact [5]. In other words, the condition toward the direction of sustainable development can be presented as simultaneous improvement in company performance in all three dimensions, economic, environmental and social [6].

What are some things that, a business cannot do without? The most obvious thing is money. With no money, a company cannot survive for long. Demand for the company's product is also necessary. Modern organizations are under increasing pressure from stakeholders to find new ways to compete effectively in dynamic markets and changing customer preferences [12]. The marketplace has never been as dynamic and muddled as we enter the twenty-first century. This dynamism is a result of great changes in the following:

- The structure of markets.
- The mobility of individuals and the great increase in global travel.
- The growth of information technology and its impact on marketing.
- The nature of marketing segments.
- Strategic alliances and networks between organizations [14].

China International Electronic Commerce Center Director Liu Junsheng said, "The financial crisis, has posed a challenge to e-commerce. At the same time, it also brings

new opportunities” [13]. With the rapid growth of e-commerce, many transactional activities are now being conducted through the internet. It has changed the traditional marketplace into one in which the business model is highly complex and requires changes to business strategies, products and services, marketing methods, etc. [8]. Organization for Economic Co-operation and Development (OECD) officially acknowledges electronic commerce as a new way of conducting business [18]. Despite the economic crisis the world is facing today, e-commerce is proving to be a great opportunity for branding and having online presence, especially for companies that had never before considered the internet as an alternative marketing strategy.

In the new context, two major factors determine the future survival or success of organizations: electronic commerce [7], and the knowledge from customers [20], encouraging the adoption of e-commerce and the use of the internet as a platform to access and collect important knowledge from customers.

In nowadays for e-commerce success is extremely important resource is information and informational technologies where information plays the main role. The list of “resources” that a business cannot survive without goes on and on. In order for a business to continue in the long term, all of these factors must remain. For this reason, it is important for a business to be sustainable not just environmentally, but in all areas.

Therefore, the authors recommend to analyze how development and usage (utilization) of information technologies affect sustainability of a company development.

## 2. SPECIFICATIONS INFORMATION TECHNOLOGIES AND ITS MAIN ELEMENT – THE INFORMATION

The Information is one of the main resources, which together with other material resources – energy, financial and natural is included in every product or service and defines product’s competitiveness and business results.

Modern business sustainable development also depends on how effective business is using information, therefore we believe that the information could be taken as a source and mean of the modern business sustainable development.

Information resources form one of the most important markets in Latvia – information market. In terms of development information market is ahead of national economy growth in Latvia.

According to authors’ researches to characterize information resource the following definition might be used: *Information resource in business is a means or possibility of an information character, used to create specific products, services or processes* [10].

In accordance with authors’ researches information resource is:

- Transformations object: resource → result (see fig.1);

- Initial or intermediate means, which could be used to create new information resource.

Information resources affect usage of the other resources (see fig.1) and normally lead to the savings (economy, retrenchments). After utilization of information resource there are two possible outcomes: real product, service or process; and feedback in the form of information, which has significant value for the further information usage and gaining of the economic effect.

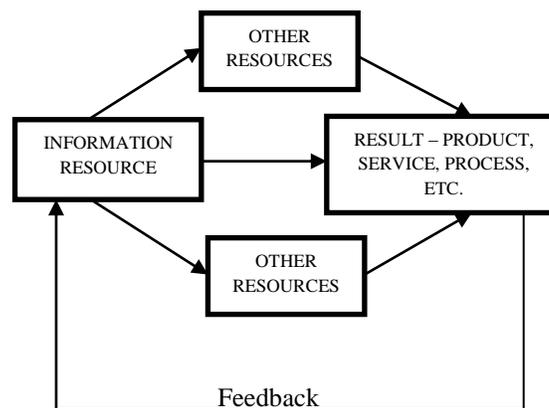


Fig.1 Information resource as transformation object in business

Let us point, that information resources are characterized by specific properties and features:

- Information resource has some level of entropy;
- It is non-material resource;
- It might be saved on various medias;
- It has fixed value (price);
- It might be transformed into managerial goals and tasks;
- It might have various indicators and degrees (ranks) of importance (technical, juridical, social, etc.);
- Information resource has the same properties as have marketing elements (marketing mix, segmenting, etc.);
- Importance degree of the information resource might be fixed or variable in various systems;
- Information resource provides stimulating effect on new idea generation;
- Utilization of information resource or its elements contributes to reduction of the uncertainty level during products manufacturing, distribution, usage, etc.;
- Information resource is not possible without information technology and its development.

Since information resource is one of the most significant resources for ensuring business’ long-term development, it is to examine economic features of the information resource. Power (outcome) of the information resource is very important characteristics, depending on economic objects (entities) informatization in the country or, in other words, how wide and effective information technology is used in companies.

Computerization level determines speed of information processing, storing and downloading, and also affects economically justified unbiased (impartial, objective) decision-making. In turn, the use of computer technology is one of the most important components to ensure information resource effective utilization.

So far the level of public computerization is comparatively good and, according to Central Statistical Bureau data, reached 62.8% for households and 95.2% for businesses in 2010 [4]. It is clear, that in near future even higher level of computerization might be expected, contributing the use of information resources.

Researching information resources and information technologies from the business point of view, it is to say that information technologies not only ensure selling and buying of the information, but also exercise access to the information, active usage and exchange by the public community, contributing its versatile development. Therefore authors conclude, that information technology effect to the business development and market economy in general is of very high importance and its proper and timely application stimulates sustainable business development.

Assessing relevance and role of the information resource for competitiveness ensuring, let us point that qualitative information is a means for a competitiveness improvement required for economic and market analysis. To assess the impact of the information to the product's competitiveness, it is advisable to adopt (use) M.Porter's five forces competitiveness model [15], what can be done on the basis of information resources system, which includes five types of resources: marketplace; consumers; competitors and their products; quality; prices - sales and consumer prices [1].

### 3. INFORMATION RESOURCE CAPACITY

According to the authors, there is a usefulness of information, i.e. that part of an information resource, which provides competitive advantages in reality. Information resource capacity is part of the information resource, which is real of service. It is necessary, mutually agreed part of the information resource, which form product according to required parameters, excluding the information gap. Information resource capacity is described by following relation:

$$I_r = \frac{I_n}{I}, (1)$$

where:  $I_r$  – information resource capacity;

$I_n$  – necessary useful part of the resource, which describes advantages of the given product;

$I$  – total information resource, which is consumed to develop and consume given product.

Thus, information resource capacity is a tool to produce competitiveness, although it, as a source of information participates in product's design. Really, information resources may be different and their quality affects both

the product's quality and total costs, connected with a product (sales and operating or consumer prices).

It is known, that if sales costs relative to consumer prices is within 0.05 to 0.2 range [17], then pricing (product's sales costs) has little influence on product competitiveness.

$$\frac{C_{sc}}{C_{cp}} = 0.05 \div 0.2, (2)$$

where:  $C_{cs}$  – sales costs;

$C_{cp}$  – consumer price.

This means that in this situation product's sales cost becomes a little sensitive competitiveness element.

From here follows the problem – make total expenses of a way, to reduce its consuming component in favor of high competitiveness advantages (qualitative characteristics). At the same time lessen expenses' sales cost component, which directly influences product's competitiveness.

As a result product's competitiveness at the market is determined by two most important indicators – price (includes sales cost and consumer price) and quality. Hence, competitiveness model might be pictured as follows:

$$\frac{Q/Q_{et}}{EXP/EXP_{et}} = CS, (3)$$

where:  $CS$  – competitiveness;

$Q$  and  $EXP$  are respectively quality of given and etalon product;

$Q$  and  $C$  min desired quality level and lower price limit.

Indicators  $Q$  and  $C$  must be modeled according to product's design stages taking into consideration their mutual interaction to ensure desired quality.

Accordingly, given the need to minimize costs expenses, increased attention during product's quality modeling process must be paid to expenses reduction during all the stages of the quality loop, ensuring that: must be paid to expenses (costs) reduction during all the stages of the quality loop, ensuring that:

$$\Delta EXP > 1, (4)$$

Parallel to quality model, pricing model is creating, accounting market, production, direct and other factors. Quality and pricing models form competitiveness multi-factor model<sup>12</sup>:

$$C = f(F_1, F_2, F_3 \dots F_n) \rightarrow \max, (5)$$

where:  $F_1, F_2, F_3, \dots F_n$  – competitiveness factors respecting their importance.

Sum of the weighted factors produces quantitative value of the competitiveness. Let us adopt qualitative competitiveness value for the base product  $CS_b$  for a unit of competitiveness and name it as competitive unit. Thus,

$$V_{cs} = \frac{CS}{CSb}, (6)$$

where: CS and CSb are designed and base product's competitiveness accordingly.

Knowing company's EXP CSb and reaching  $V_{cs} = 1$ , it is possible to plan C costs for designed product.

$$EXP_{cs} = V_{cs} \times EXP_{csb}, (7)$$

where: EXP<sub>cs</sub> and EXP<sub>csb</sub> – costs to ensure competitiveness of designed and based product respectively.

If volume of the investments for the new product implementation is known, then it is possible to calculate CS in competitive units according to factors, not taking S into consideration.

$$V_{cs} = \frac{EXP_{ks}}{EXP_{ksb}}, (8)$$

Comparing VCs with required value it is possible to conclude on pricing of the product for market entry. We believe, appropriate, based on information and products competitiveness units, costs modeling, considering influence of internal and external factors contributes to the efficient management of competitive products.

In accordance with the theory of establishing competitiveness advantages, authors suggest information resources evaluation criteria. Information resources are evaluated by comparing with each other (by mutual comparing). Information resources used to improve competitiveness will always be compared with required information (see table 1).

Table No1 **Information evaluation**

Information evaluation criteria	Remarks
$I_m \geq I_n$	$I_m$ – market information, $I_n$ – required information
$I_p \geq I_n$	$I_c$ – information on price, $I_{vaj}$ – required information
$I_{inov} \geq I_n$	$I_{inov}$ – information on innovations, $I_n$ – required information
$I_{pr} \geq I_n$	$I_{pr}$ – production (manufacturing) information $I_n$ – required information
$I_{fact} \geq I_n$	$I_{faktiskā}$ – production information, $I_n$ – required information

According to the information from Table 1, it is always necessary to compare actual or in company available information with required information.

$$\sum I_i \rightarrow CS_{max}, (9)$$

where:  $I_i$  – information types used in information resource;

CS – company's competitiveness.

Aforementioned researches confirm, that information

resource serves as a one of the most relevant resources for business sustainable development, and thus must be rational used. Summarizing it is to conclude, that an information as a resource and its qualitative utilization within a company ensure increase of company's and its products' competitiveness.

#### 4. INFORMATION TECHNOLOGY IN BUSINESS: CASE OF LATVIA

Economic activities connected with information technologies, products manufacturing and services were at level of 3.6% from GDP in 2008 and at level of 3.3% from GDP in 2009 [4]. According to Central Statistical Bureau of Republic of Latvia (CSB) data at the beginning of 2010 95.2% of all companies were using computers and 90.7% of all companies were connected to the internet [4]. At the beginning of 2010 22.5% of all companies were using wireless internet access, 26.8% ISDN lines, 35.5% - xDSL and 46.7% other broadband internet connections [4]. It is surprising, that in 2010 only 48.6% of all companies have had a home page [4]. In 2009 only 6.9% were selling [4] and only 17.1% were buying [4] goods thru the internet.

In 2010 only 55% of all companies have used automatic data exchange services. At the same year 30.9% of all companies' employees have regularly used computer at the offices and 27.9% employees have regularly used computers with internet connection [4]. All these data prove, that e-commerce in Latvia is at a very initial point.

From December 2010 till January, 2011 authors conducted a survey to ascertain the intensity of the use of information technologies by small and medium-sized enterprises and technology impact on companies' development. 319 companies' executives were questioned during this survey. Survey consisted of 23 questions of a discrete type and 3 questions of a public type.

Survey results match with a data from public statistical sources concerning computers number growth in companies starting from 2006. Study shows a tendency of substantial computers quantity growth (about 25%) exactly in sales and services sector. At the same time technical, manufacturing, printing etc. enterprises show no such a trend. Executives have had a positive assessment of the information technology (IT) impact on information flow. 68% agreed, that information flow grew along with development of IT; 18% stressed, that the use of IT haven't produce any impact on the company information flow; 14% stated, that haven't thought about it. At the same time all survey participants confirmed relevance of the internet for company development and stated that social networking will increase impact to the business results.

As a result of study authors also summarize negative factors, appearing with increase of information technology utilization by business.

Most often mentioned are as follows:

- Increase of occupational diseases;
- Growth of number of spine disease;
- Uncontrolled personal use of electronic environment;
- Duties and tasks take longer time to complete;
- Etc.

There are 1 503 400 (June, 2010) active internet users in Latvia, i.e. 67.8% of total population [9]. More than 85% of the world's online population has used the internet to make a purchase – increasing the market for online shopping by 40% in the past two years – according to the Nielsen Global Online Survey on internet shopping habits. Among internet users, the highest percentage shopping online is found in South Korea, where 99% of those with internet access have used it to shop, followed by the UK (97%), Germany (97%), Japan (97%) with the U.S. eighth, at 94%. According to Latvian CSB only 8.5% of Latvian internet users have had an experience of purchasing in the internet [19]. Based on CSB data 86.8% of all enterprises were using internet in 2009, at the beginning of the 2010 total quantity of internet users increased by almost 4% and reached 90.7%. It is interesting, that the use of internet in regions is just slightly lower than in Riga and Riga metropolitan area (by about 5%). Authors believe that this uniform indicator of internet utilization in whole territory of Latvia has a positive impact on the use of IT in business and contributes to sustainable business development.

Hosting provider GRAM Technologies in cooperation with Latvian Internet association has conducted regular quarterly study “Latvian business and internet: Internet is oxygen for about half of the businesses” 193 companies have participated in this survey from which 52% are 10 or more years on the market; 50% of respondents are owners of the business; 12% of respondents - executives; 6% - marketing directors. Results of the survey are as follows: in the event of internet loss, it will very negatively affect 37% of companies and 4% will abort their operations.

43% of respondents state, that in case of internet loss their business will face minor complications but overall this event won't produce any substantial issues. Only 7% stated that this event won't impact their business at all. So, for more than 50% of businesses participated in a study internet isn't just a sales improvement means, but “an oxygen” they can't live without.

One of the most popular sales improvement channels for businesses is Google. This statement agreed 45% of respondents, followed by e-mail marketing (38%) and commercials at advertisement portals (35%). Attention to the products promotion thru social networks pays 25% of the respondents. And only 19% uses advertisement banners [16].

Despite the recent economic crisis, businesses pointed to the positive turnover dynamics, increase of sales volumes and overall growth of competitiveness. Thus, authors conclude that the use of information technology in business ensure business development.

## 5. CONCLUSIONS

Summarizing results of the study authors came to a conclusion that information technologies positively impact more business segments and contribute to sustainable business development.

Research results clearly show that information technology mainly affects company's information flow. Considering that information resource has today become one of the prevail resources for sustainable development, it must be outlined that this resource economic evaluation becomes of the high importance.

Utilization of the information resource ensures general business competitiveness growth.

Study conducted by the authors on information technologies utilization for business development shows growth of the IT utilization volumes, especially in business segments not affected by wide use of IT before. Businessmen agreed that IT has a positive impact to the development of business results, although pointed to some shortcomings.

## 6. REFERENCES

- [1] L. Bañasová, D. Cagaňová, M. Čambál, “The Identification of Key Managerial Competencies as a Tool for Increasing Business Competitiveness”, **Proceedings of the European Conference on Knowledge Management**, 2010, pp. 46-53.
- [2] R.J. Baumgartner, D. Ebner, “**Corporate Sustainability Strategies: Sustainability Profiles and Maturity Levels of Sustainable Development**”, 2010 [Electronic resource], viewed on March 12, 2011, <http://www.interscience.wiley.com> DOI: 10.1002/sd.447.
- [3] Board on Sustainable Development. **Our Common Journey: A Transition Toward Sustainability**. Washington, DC: National Academy, 1999.
- [4] Central Statistical Bureau of Latvia, [Electronic resource], viewed on March 23, 2011, <http://data.csb.gov.lv/Dialog/Saveshow.asp>
- [5] D. Ebner, R.J. Baumgartner, “**The Relationship Between Sustainable Development and Corporate Social Responsibility**”, 2006 [Electronic resource], viewed on March 12, 2011, <http://www.crrconference.org>.
- [6] F. Figge, T. Hahn, “Sustainable value added – measuring corporate contributions to sustainability beyond eco-efficiency”. **Ecological Economics** 48 (2), 2004, pp. 173-187. [Electronic resource], viewed on

- February 20, 2011, [http://www.csiproggress2007.org/index.php?option=com\\_frontpage&Itemid=1](http://www.csiproggress2007.org/index.php?option=com_frontpage&Itemid=1)
- [7] A. Gupta, et. al., "An empirical study of consumer switching from traditional to electronic channels: A purchase-decision process perspective". **International Journal of Electronic Commerce**, vol. 8, issue 3, 2004, pp 131-161.
- [8] T.L. Hsia, J.H. Wu, E.Y. Li, "The e-commerce value matrix and use case model: A goal-driven methodology for eliciting B2C application requirements." **Information & Management**, vol. 45, issue 5, 2008, pp. 321-330.
- [9] Internet World Stats, [Electronic resource], viewed on April 23, 2011, <http://www.internetworldstats.com/stats4.htm>.
- [10] J. Karimi, T.M. Somers, A. Bhattacharjee, "The Role of Information Systems Resources in ERP Capability Building and Business Process Outcomes". **Journal of Management Information Systems**, Vol. 24 Issue 2, 2007, p221-260, [Electronic resource], viewed on February 12, 2011, <http://web.ebscohost.com>.
- [11] R.W. Kates, T.M. Parris, "Long-Term Trends and a Sustainability Transition". **Proceedings of the National Academy of Sciences**, Vol. 100, No.14, 2003, pp. 8062-8067.
- [12] P.A. Phillips, C. Wright, "E-Business's impact on organizational flexibility." **Journal of Business Research**, 2008, Article in press.
- [13] Pressmeldungen.at. "**Financial Crisis: A great Opportunity for E-commerce**", [Electronic resource], viewed on April 18, 2011, <http://www.pressemeldungen.at/69035/financial-crisis-a-great-opportunity-for-e-commerce/>.
- [14] A. Ranchhod, "**Marketing Strategies: A Twenty-first Century Approach**". Essex: Pearson Education Limited, 2004, 239 p.
- [15] M.E. Porter, "**The Competitive Advantage: Creating and Sustaining Superior Performance**". N.Y.: Free Press, 1985, 260 p.
- [16] Research: "**Internet – the „oxigen” for approximately of half Latvian companys**", [Electronic resource], viewed on April 23, 2011, [http://www.lia.lv/aktualitates/28/petiju\\_ms\\_internets\\_\\_skabeklis\\_aptuveni\\_pusei\\_kompaniju/](http://www.lia.lv/aktualitates/28/petiju_ms_internets__skabeklis_aptuveni_pusei_kompaniju/).
- [17] S. Sarkisjan, "**The theory of prediction and decision making**", Moscow: Highschool, 1977.
- [18] T. Sung, "E-commerce critical success factors: East vs. West". **Technological Forecasting and Social Change**, vol. 73, issue 9, 2006, pp.1161-1177.
- [19] The Nielsen Company, "**Over 875 Million Consumers have Shopped Online – the Number of Internet Shoppers un 40% in Two Years.**" [Electronic resource], viewed on April 23, 2011, <http://nz.nielsen.com/news/OnlineShopping.shtml>.
- [20] M. Tsai, C. Shin, "The impact of marketing knowledge among managers on marketing capabilities and business performance." **International Journal of Management**, vol. 21, issue 4, 2004, pp. 524-530.
- [21] M. Wade, "**Good company citizenship. In Governance and Sustainability: New Challenges for States, Companies and Societies**", Greenleaf: Sheffield, 2005, pp. 186–199.
- [22] World Business Council for Sustainable Development (WBCSD) 2007. "**Doing Business with the World – The World Comm. Environ. Dev. 1987. Our Common Future**". New York: Oxford University Press.

# Modeling of Consumer Behavior for Business Sophistication

Elina GAILE-SARKANE

and

Ieva ANDERSONE

Faculty of Engineering Economics and Management, Riga Technical University, 1/7 Meza Street,  
Riga, LV-1007, Latvia

## ABSTRACT

In today's competitive business environment many different indicators influence business performance. In this research the authors analyze one of most important company stakeholders – their customers. The research is based on evaluation and analysis of existing consumer behavior models. On the basis of this research the authors have developed the mathematical model for modeling consumer behavior and investments in communication with customers. Model is based on assumption about the existence of two groups – people and goods. Each group is characterized by a series of parameters. Developed model contains only 3 parameters of each variable (a consumer and a product) and, in order to perform an accurate calculation, it is also necessary to determine the coefficients of mutual impact of each parameter, what should be indicated by experts. The model elaborated by the authors, would be suitable for practical application within business in order to analyze the return of the company investments, identify perspective and less perspective areas of investment towards business sophistication.

**Keywords:** Consumer behavior, modeling of consumer behavior, mathematical approach, business

## 1. INTRODUCTION

Today's competitive business environment requires more and more attention to planning, forecasting and analyzing business, especially for small and medium companies. Many authors of scientific publications had been devoted to topics of sustainability and business sophistication. The understanding of business sophistication has been analyzed since 1990, but still this is actual topic for business researchers.

In 1998 Rue and Ibrahim has wrote: "Numerous products in academic publications as well as practitioner-oriented journals have recognized the importance of planning for small businesses. Yet there is surprisingly little empirical work that has examined the relationship between planning sophistication and organizational performance among small businesses" [15]. And the same we can say today – there are small number of researches done on planning sophistication and organizational performance.

One of today's goals for every company is to perform all activities towards business sophistication. According to the authors' point of view business sophistication is more related to the process than to the result. Sophistication is not completed process, the same as quality, management, etc. For this reason it is important to understand from which components the business sophistication will be combined of in every industry and in every company. For each company these components might be different. At the same time there are processes and parameters what are similar for all industries, like market analysis, planning, evaluation of customers, forecasting, controlling etc. This assumption can be affirmed by analysis of existing publications about business sophistication.

For example, Stupak has analyzed marketing strategies and communication for business sophistication [18], Garry and Harwood have tackled the topic of client sophistication and they have founded out that the level of client sophistication has a moderating influence in a number of key areas. These are identified as: service evaluation criteria; interdependency and power; the atmosphere in which solicitor-client interactions take place and relationships develop; the nature of joint personal relational goals and the role of trust and extent of commitment. [6] Lace and Ciemleja have analyzed different business indicators and business performance within context of business sustainability what drives us to sophistication of the business. They have pointed out the

essence of the category «sustainable development» and characterized factors impacting on company's sustainable development, applying the methodology of «Golden section» [12]. Accordingly we may conclude that business sophistication can be provided in different ways and with different approaches.

One of key areas for the business is their customers. This area can be analyzed for many different reasons but always with one aim – to find out consumer behavior under different circumstances. The success of a company depends on understanding of their customers as well as on possibility to communicate with them. Accordingly, the authors of this research presume that modeling of consumer behavior may provide company with needed information for business sophistication.

## 2. MODELS OF CONSUMER BEHAVIOR

One can come across the concept of *consumer behavior* in several realms of science, however systems of perception differ. Psychology, for instance, is searching for explanation within biological, evolutionary and cognitive processes. Sociology examines the impact of social groups and society, while social psychology explores the role of an individual within a social group.

The beginnings of consumer behavior modeling are dating back to the 60-ies of the 20<sup>th</sup> century, when it started developing alongside the theory of marketing. Tracing back for the products of various types of consumer behavior *modeling boom* in the late 60ies of the 20<sup>th</sup> century, a British market research company developed a *modeling research* group to explore consumer behavior. The group elaborated and developed several classifications for various types of models.

The group also drew distinctions between the different ways in which the models might be displayed [14]:

- Verbally – most of us will put our assumptions about consumer behavior into words in order to explain them both to ourselves and other people.
- Algebraically – some of our ideas are best transmitted via an algebraic equation. Weber's Law or the Fishbein model of attitudes would fit into this category.
- Pictorially – almost every diagram can be viewed as model - a diagram illustrating some point about the topic under discussion.

In addition, they identified categories such as descriptive, diagnostic and predictive models and a rather

more objective pair which were called successful and unsuccessful.

Perhaps the most used set of categories is that of low, medium or high-level models. In this case the level refers to the level of complexity – so a low-level model would be a relatively simple representation of the phenomenon, while a high-level model of the same event would be much more complex and detailed and include many more variables.

There are several theories explaining consumer behavior – diverse in their form, range, terms of a situation and application, as well as their level of complexity. These theories exploit all the social sciences – economy, psychology, sociology and anthropology.

The model of psychoanalysis was elaborated by Sigmund Freud. He discovered new dimensions within the consumer behavior by introducing psychological elements into the consumer's decision-making. Freud (1923) believed that a human psyche consists of three levels: ID (the deepest layer of subconscious propensities functioning in accordance with the principle of enjoyment and demanding immediate satisfaction), Ego (the aware subject functioning as a mediator between the ID and reality), and, finally Super Ego (the inner conscience of a personality representing viewpoints by adapting to social pressure – developing as a summary of ID desires rejected by the reality). [17]

In accordance with the theory of Freud, consumer behavior is the function of these three elements (1).

$$CB = f(ID, E, SE), \quad (1)$$

Where: CB – Consumer behavior,

ID – The deepest layer of subconscious propensities,

SE - The inner conscience of a personality.

The theory of Gestalt is based on the individual's perception of products and ideas. The core idea of Gestalt therapy is that a person's mental life does not depend on isolated feelings, which are mechanically connected by the principle of associations, but it rather develops as a retained entirety or an image. [5]

The theory of cognitive dissonance (1957) was developed by Leon Festinger who reckoned that a human being has a tendency to seek for coherence between his/her attitude, conviction and action. Cognitive dissonance is a sense of discomfort caused by two conflicting ideas emerging at the same time. Individuals in such situations develop motivation to reduce the dissonance by changing their attitude, conviction and behavior in order to justify and rationalize it. Dissonance

may manifest as a remorse, anger, disappointment or confusion [9, 10]. A salesperson, in compliance with this theory, should try to reduce the post-purchase dissonance of consumers by reassuring them about the correctness of their choice and by encouraging their trust into products. This can be achieved by means of advertisements, post-purchase service, product manuals and other marketing attributes [1].

Behavioristic model is based only on observations and measurable variables. Cognitive models (purchase related to cognition) of consumer behavior are related to information processing. Indicators such as activation, level of participation, emotions, motives and attitude are predominant. Information processing of long-term and short-term memory and variables – learning, reasoning and knowing – is related to decision making process – i.e. – information seeking, processing and selection [3]. All cognitive models, in their term, are part of decision making process reflected in successive stages and having various factors of influence (Blackwell/Miniard/Engel model (2001), Howard/Sheth model (1969), etc.). [2, 7]

In compliance to the theory of behaviorism, one or more hypothetical coherences between variables can be modeled by means of the so called theoretical models. According to behaviorists, all functions of an individual can be described by means of a scheme: stimulus-reaction (hereinafter – SR). It means that any type of human reaction can be achieved provided that the „right“ stimulus is applied. Even though the original opinions of the school of behaviorism have undergone a considerable development, they still exist as a version of theoretical model [16].

Section models (decision making models) are one of the commonly applied methods of consumer behavior research. The models are based on the following concept: by identifying the basis of purchase stages (preconditions), they can be arranged in a sequence, which means that the purchase decision making process can be divided into separate stages. Division into stages depends on the classic or the extensive decision making process and these processes can be modeled as follows [4]:

- the initial stage of the process,
- the stage of search and pre-selection,
- the stage of evaluation and selection,
- implementation (or realization) stage,
- post-purchase stage.

One classic example of a personal variable model is Fishbein model [11]. The model – it is summarized by the formula (2):

$$A_0 = \sum_i^n B_i a_i, \quad (2)$$

Where:  $A_0$  = the attitude towards object o;

$B_i$  = the strength of belief i about o;

$a_i$  = the evaluation aspects of B;

n = the number of beliefs.

Very popular is the Rice Perceived Value / Perceived Probability of Satisfaction model. This model can be applied in practice by combining it with the theory of value (defining various dimensions of value), which is extensively described in the literature of economics, and analyzing the mutual interaction between the consumer values and the consumption volume.

Most famous of the consumer behavior models are comprehensive (large) models. These large models are the more complex formulations of consumer behavior, and they commonly attempt to encompass all the factors and elements which the authors feel to be relevant to the behavior.

Among most popular models we can name: classic Engel, Kollat and Blackwell model, the extensive Howard/Sheth model, The Nikosia model (1996) etc.

Bettman and Jones have analyzed these models and offered the idea that all formal consumer behavior models can be classified in four groups [8]:

- Stochastic models – consisting of two important components – „laws“ intended for individual behavior and individual models;
- Information processing models – are based on an assumption that an individual constantly receives information from the environment and this information is an integral part in the process of choice;
- Experimental and linear models – have general mathematical structure (for instance, Fishbein model);
- Multisystem consumer choice models – represent a broad structure of relationships, usually verbal and contain various simplified formal models.

After examining the most crucial studies and models on consumer behavior elaborated before the beginning of the 21st century, the authors concluded that these studies mainly reflect the consumers' process of decision making, as well as view all the stages of decision making in a more or less detailed way. In the mean time, according to the authors' opinion, the factors influencing the consumer behavior and the intensity of their impact have not been analyzed sufficiently. Most of the above mentioned models characterize the purchase decision making process illustratively and, therefore, the authors

emphasize the importance of a complex approach towards assessment of consumer behavior- a combination of expert methods, marketing methods and methods of economically mathematical modeling.

Titko J. and Lace N. in their research has pointed out that the shareholders' value creation has become the primary goal of any company [20], therefore the authors of this research believe that modeling of consumer behavior and communication intensity would drive a company towards fulfillment of business performance.

### 3. MATHEMATICAL MODEL OF CONSUMER BEHAVIOR

Based on the previous studies, the authors suggest elaborate a mathematical model in order to forecast the consumption of goods and services having the impact of the external direct and indirect factors. By applying of this model the in praxis companies might reduce their expanses on communication with potential and existing customers as well as become more efficient in the market. The model is based on assumption about the existence of two groups – people and goods. Each group is characterized by a series of parameters. Besides, these parameters are temporarily fixed. It means that they can be measured. For instance, by investing goods into advertising (hereinafter within the model - parameter  $b_1$ ), it is possible to determine the impact on consumption of certain goods or groups of goods.

The essence of the model is directed towards fulfillment of the needs of a consumer while purchasing a product  $\alpha_x$ , with the characteristic parameters (all or part of the parameters). In other words, a purchase process can be described as a function of parameters characterizing an individual and a product – i.e. -  $f(\alpha; \beta)$ .

In its turn, it means that any product can obtain an indicator demonstrating its value or the price the consumer is willing to pay or pays in order to acquire the product (3.)

$$P = \sum f(\alpha, \beta), \tag{3}$$

Where: P – Price of the product;  
 $\alpha$  - Consumers and their characteristic parameters;  
 $\beta$  - Goods and their characteristic parameters.

By knowing these indicators, it is possible to model the consumption volume, considering the impact of various variable factors, and to take a decision as to

whether, for instance, advertising costs need to be increased or not. This type of modeling allows the managers to take grounded and thought-out decisions.

In order to be able to convey a variable, it is necessary to carry out derivation of a parameter (see formula 4). After attaining a positive result, it is possible to continue investing, while a negative result give a signal that a decision needs to be taken on reduction of investment within the particular parameter.

$$\frac{\partial}{\partial b} = \left( \sum_n f(\alpha, \beta) \right) \Rightarrow \max, \tag{4}$$

During the performance of modeling, in order to relieve the modeling process, the function mentioned above can be viewed as linear. Thus a matrix of parameters characterizing a product or a product can be composed and referred to as C (see formula 5).

$$\alpha^+ C \beta = \sum_{i \neq 0}^n \sum_{j=1}^m a_i C_{ij} b_j \tag{5}$$

Where: C – The matrix characterizing the parameters of a product;  
 $\alpha$  - Consumers and their characteristic parameters;  
 $\beta$  - Goods and their characteristic parameters.

In order to calculate the turnover or the profitability a company may obtain in the result of investing into a particular parameter (for instance, advertising investments), it is necessary to derive from this parameter (see formulae 6 and 7.)

$$I = \frac{\partial}{\partial b_n} = \left( \sum_{\alpha} \left( \sum_{i=1}^n \sum_{j=1}^m a_i C_{ij} b_j \right) \right), \tag{6}$$

or:

$$I = \frac{\partial}{\partial b_n} \left( \sum_{\alpha} \alpha^T C \beta \right) \tag{7}$$

Where: I – profitability of the parameter for the company

The composed model can be applied in order to carry out more profound studies of consumer behavior or purchase habits. For instance, a company may collect information on the consumers' income, their cultural level, etc. Therefore, a matrix of mutual interaction of

parameters can be composed i.e. to examine the impact of the advertisement or brand of a certain product on people with low income. Consequently, by means of derivations of certain parameters, a conclusion can be drawn as to profitability provided by advertising investments, improvement of salespersons' work, and choice of location of the company and other parameters for the company.

The above model can be examined by means of a specific product, for example, a car. Therefore, matrixes describe expenses of a person  $\alpha$  when purchasing a product  $\beta$ , i.e. a car. In order to compose a matrix, it is necessary to determine the parameters and their impact on the decision making by means of an expert method (see formula 8).

$$\alpha \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \text{ and } \beta \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \quad (8)$$

Where:  $\alpha$  - an individual  $x$ ;

$a_1$  - parameter characterizing the individual - income;

$a_2$  - parameter characterizing an individual - age;

$a_3$  - parameter characterizing an individual - gender;

$\beta$  - a product (car);

$b_1$  - parameter characterizing a car - advertising investment;

$b_2$  - parameter characterizing a car - brand popularity;

$b_3$  - parameter characterizing a car (the brand).

Further, it is necessary to determine the interrelation of these parameters (see formulae 9 and 10).

$$P_{\alpha\beta} = (a_1, a_2, a_3) \begin{pmatrix} c_{11}, c_{12}, c_{13} \\ c_{21}, c_{22}, c_{23} \\ c_{31}, c_{32}, c_{33} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \quad (9)$$

Where: P - Expenses of an individual  $\alpha$  for the product  $\beta$ ;

$a_1$ - $a_3$  - Parameters characterizing an individual;

$b_1$ - $b_3$  - A parameter characterizing the car;

$c_{11}$  -  $c_{33}$  - Mutual impact of the characteristic parameters.

By means of shortening formula 9, mutual interaction can be described as follows:

$$P_{\alpha\beta} = \sum_{i=1}^3 \sum_{j=1}^3 a_i C_{ij} b_j \quad (10)$$

In the process of research, the mutual impact of the parameters should be defined. In order to perform an accurate calculation, it is also necessary to determine the coefficients of mutual impact.

Thus, by means of derivation regularity is obtained, demonstrating the volume of revenue gain for the company, provided that the expenses will increase by 1 monetary unit. It provides company to stay more competitive among rivals within the industry as well as to guide a company towards business sophistication.

This model can also be used in situations when precise statistical data on the volume of consumption for a certain group of goods and indicators characterizing a consumer are known.

Despite the fact that a model contains only 3 parameters of each variable (a consumer and a product), the model describes the main idea of modeling and demonstrates the approach elaborated by the author as regards modeling of the consumer behavior. Since modeling requires a large amount of output data, each company planning to carry out modeling, has to be capable of providing the necessary information about their clients and their consumption habits. It, thus, means that a company has to carry out regular market and consumer research.

#### 4. CONCLUSIONS

Modeling of consumer behavior is just one step towards business sophistication but it may provide a company with possibility to become more efficient in communication with their customers.

Also the model elaborated by the authors, would be suitable for practical application within business in order to analyze the return of the company investments, identify perspective and less perspective areas of investment. This information also provides a company with information needed for further development, efficient and effective performance in the market.

There are various approaches as regards modeling of consumer behavior. Among them there are many theoretical models that cannot be put into practice.

## 5. REFERENCES

- [1] A. Bawa, P. Kansal (2008) Cognitive dissonance and the marketing of services: some issues/**Journal of Services Research**, Volume 8, Number2, October 2008-March2009, 2008 by Institute for International Management and Technology, pp. 31-51.
- [2] R.D. Blackwell, P.N. Miniard, J.F. Engel (2001) **Consumer Behavior**. 9.Auflage, FortWorth S.83
- [3] S. K. Dion (2009) Unsettling Questions: Cognitive Dissonance in Self-Deception/ **Social Theory and Practice**, Vol.35, No1, January 2009, pp. 73-106.
- [4] T. Foscht, B. Swoboda (2007) **Käuferverhalten. Grundlagen – Perspektiven – Anwendungen**. 3. aktualisierte Auflage, Wiesbaden: Gabler 2007 – S. 140
- [5] Z. Freids (2007) **Ievadlekcijas psihoanalīzē**. – Rīga: Zvaigzne ABC, 2007
- [6] T. Garry, T. Harwood (2009) The moderating influence of client sophistication on relationships within business-to-business credence service markets, **Journal of Business and Industrial Marketing** Volume 24, Issue 5, 12 June 2009, pp. 380-388.
- [7] J.A. Howard, J.N. Sheth (1969) **The Theory of Buyer Behavior** – New York 30 p.
- [8] J.M. Jones, J. R. Bettman ( 1986) Formal Models of Consumer behavior: A Conceptual Overview, **The Journal of Business** (pre-1986); Oct 1972; 45,4; ABI/INFORM Global p. 544
- [9] W. Kroeber-Riel, P. Weinberg (2003) **Konsumentenverhalten**, 8.Auflage, München
- [10] W. Kroeber-Riel (1974) Absatztheorie, verhaltensorientierte, in: Tietz, B.(hrsg.): **Handwörterbuch der Absatzwirtschaft**, Stuttgart 1974, S.159-167.
- [11] A. Kuß, T. Toczak (2004) **Käuferverhalten**, 3.Auflage, Stuttgart S.57
- [12] N. Lace, N. Koleda, G. Ciemleja (2010) Quantitative harmonious model of sustainability factors: measuring contribution of financial viability // **The 6th International Scientific Conference: «Business and Management 2010», 13 - 14 May 2010, Vilnius, Lithuania: selected papers / Vilnius Gediminas Technical University**. – Vilnius: Technika, 2010, pp.104 - 112. – ISBN 978-9955-28-311-9.
- [13] J. Mazanec (1978) **Strukturmodelle des Konsumverhaltens**, Wien, 1978
- [14] Ch. Rice (1997) **Understanding customers** – Second edition. – Butterworth Heinemann, 1997., 312 p.
- [15] L.W. Rue, N.A. Ibrahim (1998) The relationship between planning sophistication and performance in small businesses, **Journal of Small Business Management**, Volume 36, Issue 4, October 1998, pp. 24-32., SCOPUS
- [16] M.R. Solomon (2006) **Consumer behavior: buying, selling, an being** – 7th ed., 2006, 654p.
- [17] Dž. Stengrūms (2008) **Filosofija – Zvaigzne ABC**, 2008
- [18] R.J. Stupak (2000) Perception management: An active strategy for marketing and delivering academic excellence, business sophistication, and communication successes, **Public Administration and Management**, Volume 5, Issue 4, 2000, pp. 250-260., SCOPUS
- [19] **The Hutchinson Dictionary of Ideas** (1994) – Helicon Publishing, 1994, 659p.
- [20] J. Titko, N. Lace (2010) Performance Measures for a Business Unit in Latvian Retail Banking // **The 6th International Scientific Conference: «Business and Management 2010», 13 - 14 May 2010, Vilnius, Lithuania: selected / Vilnius Gediminas Technical University**. – Vilnius: Technika, 2010, pp. 255 -262. – ISSN 2029-4441

# Determining the Significance of the Criteria Describing Enterprise Marketing

**Romualdas GINEVIČIUS**  
**Vilnius Gediminas Technical University**  
**LT-10223 Vilnius, Saulėtekio al. 11, Lithuania**

and

**Valentinas PODVEZKO**  
**Vilnius Gediminas Technical University**  
**LT-10223 Vilnius, Saulėtekio al. 11, Lithuania**

and

**Adomas GINEVIČIUS**  
**Vilnius Gediminas Technical University**  
**LT-10223 Vilnius, Saulėtekio al. 11, Lithuania**

## ABSTRACT

Enterprise performance largely depends on the effective enterprise marketing, which may be achieved only if the effective marketing management is used. The latter involves a possibility of quantitative evaluation of the state of an enterprise at any moment of time. The number of the criteria describing the state of an enterprise is determined by the marketing mix, widely used in developing the marketing strategy, which includes such components as product, price, promotion and place. These criteria are in turn subdivided into subcriteria describing marketing activities.

Enterprise marketing is characterized by a great number of various criteria, therefore, multicriteria evaluation methods are well suited for its quantitative evaluation. To use these methods, the significance and weights of the criteria should be known. Experts determine the criterion weight in two steps. First, the estimates elicited from experts are ranked and their consistency is checked, then, the weights of the criteria are determined. These methods are illustrated with case studies involving real calculations.

**Keywords:** Enterprise marketing activities, Determination of the criterion weights, MCDM.

## 1. INTRODUCTION

Enterprise performance largely depends on the effective enterprise marketing, which may be achieved only by the effective marketing management. The latter requires a possibility of quantitative evaluation of the state of an enterprise at any moment of time. There are several reasons for this. First, enterprises usually invest much money in marketing, but often do not know what methods may be used for determining its effectiveness. Second, quantitative evaluation of marketing strategies provides a possibility for their improvement. Third, the use of methods of quantitative evaluation of marketing allows managers to analyse its effect on enterprise performance [1].

Enterprise marketing may be referred to complex phenomena. These phenomena differ from other issues because we can hardly find a single criterion (or value) that could describe all their aspects, which can be observed in reality [1, 2].

The analysis of the literature on the problem shows that marketing may be described in various terms.

The marketing mix is probably the most famous marketing term. Its elements are the basic, tactical components of a marketing plan. Also known as the Four Ps (4Ps), the marketing mix elements are described as follows:

- 1) Product is an article or service that you are selling;
- 2) Price shows how much you are charging your product/service;
- 3) Promotion means how and what you tell people about your offer, i.e. your product/service and price;
- 4) Place denotes how and where people can buy your product.

Marketing decisions generally fall into the above four controllable categories. These four Ps are the parameters that the marketing manager can control, subject to the internal and external constraints of the marketing environment. The goal is to make decisions that centre the four Ps on the customers in the target market and is similar to the idea of mixing a cake. A baker will alter the proportions of the ingredients in a cake depending on the type of cake he wishes to bake.

Möller [3] notes that the wide appreciation of the marketing mix among field marketers is the result of their profound exposure to this concept during the years of studies, since most introductory marketing manuals define it as “the heart of their structure” [4] and identify the 4Ps as the controllable parameters likely to influence the consumer’s buying process and decisions [5, 6]. He also points out that another significant asset of the mix is the fact that it is a concept easy to memorise and apply. David Jobber [7] noted: “The strength of the 4Ps approach is that it represents a memorable and practical framework for marketing decision-making and has proved useful for case study analysis in business schools for many years”. Enjoying

large-scale appreciation, it is not surprising that the 4Ps became even synonymous to the very term marketing, as this was formulated by the American Marketing Association [8].

The development of the marketing mix has received considerable academic and industry attention. Numerous modifications to the 4Ps framework have been proposed and the most concerted criticism has come from the services marketing area. A study by Rafiq and Ahmed [9] suggested that there is a high degree of dissatisfaction with the 4Ps framework.

People: all people involved with consumption of a service are important. For example, workers, management, consumers, etc.

Process: a procedure, mechanism and flow of activities by which services are used.

Physical evidence: the environment in which the service or product is delivered. The one which helps to communicate is tangible and the knowledge of the people around us is intangible.

The concept of 4Ps has been criticized as being a production-oriented, rather than customer-oriented definition of marketing [10]. It is referred to as a marketing management perspective. Lauterborn [11] claims that each of these variables should also be seen from a consumer's perspective. This transformation is accomplished by converting product into the customer's solution, price into cost to the customer, place into convenience, and promotion into communication, or the 4Cs [12].

The author of this paper performed an analysis and discovered that most of Eastern and Central European companies (even the most successful ones) are still using the traditional marketing mix of 4Ps. Therefore, the marketing mix consisting of product, price, promotion and place will be analysed further in the paper.

The analysis of the experience of the Eastern and Central European countries in the considered area shows that the considered 4Ps model is still widely used for describing marketing there. In general, the criteria describing the particular aspects of marketing may have various dimensions and be oppositely directed [13]. It means that for one criterion the increase of a particular value may show a better situation, while for another criterion the increase of this value may denote that the situation is worsening. To perform quantitative evaluation of marketing, all the criteria expressed in various dimensions as well as the oppositely directed criteria should be integrated into a single generalizing criterion. To solve this problem, multicriteria methods commonly used for quantitative evaluation of socio-economic systems should be applied [14-16].

The application of these methods requires that a set of criteria describing enterprise marketing as well as their values and weights should be defined [17].

A set of criteria describing enterprise marketing may be determined by using the 4Ps model. The analysis of the data on enterprise performance shows that the information about the values of the considered criteria is missing. Therefore, expert evaluation is required to determine these values. Another important problem is associated with the determination of the criteria weights, which completely relies on expert evaluation [18].

Usually, several experts perform the evaluation, therefore, the determination of the criteria weights used in marketing analysis is made in two stages. At the first stage, the estimates elicited from experts are ranked and checked

for consistency, while at the second stage the criteria weights are determined.

## 2. RANKING OF THE SIGNIFICANCES (WEIGHTS) OF THE CRITERIA DESCRIBING ENTERPRISE MARKETING AND DETERMINING THE CONSISTENCY OF ESTIMATES ELICITED FROM EXPERTS

The sets and subsets of criteria describing various aspects of enterprise marketing activities form an hierarchical structure (Fig 1). The formation of this structure makes sense, when the number of evaluation criteria is so large that it exceeds the admissible limits, which does not allow an expert to determine the criterion weight accurately [19].

In our case, there are 22 evaluation criteria. This is more than the maximum number of criteria suitable for expert evaluation, which makes 10-12 criteria [20]. By arranging them into an hierarchical structure, we get that the largest number of criteria describing the particular groups of marketing activities makes nine. Therefore, expert evaluation may be performed rather accurately.

The results of expert evaluation is the matrix  $E = \|e_{ik}\|$  ( $i = 1, \dots, m; k = 1, \dots, r$ ), where  $m$  is the number of the criteria compared and  $r$  is the number of experts. At the first stage, the experts rank the criteria according to their significance.

Usually, the estimates of various experts differ, therefore, the data obtained in the evaluation can be used in the research if the estimates elicited from all experts are consistent. To determine the consistency of experts' estimates, the so-called concordance coefficient  $W$  is used. The coefficient is calculated based on the data presented in the matrix of the criterion ranking  $E$ .

The concordance coefficient was defined by M. Kendall [21]. Taking into account the judgements of all experts, we calculate the sum  $e_i$  of the  $i$ -th criterion ranks, the mean value  $\bar{e}$

$$\bar{e} = \frac{\sum_{i=1}^m e_i}{m} = \frac{\sum_{i=1}^m \sum_{k=1}^r e_{ik}}{m}, \quad (1)$$

and the square sum  $S$  of the deviation of the values  $e_i$  from the mean value  $\bar{e}$ :

$$S = \sum_{i=1}^m (e_i - \bar{e})^2, \quad (2)$$

The concordance coefficient  $W$  is calculated by the formula:

$$W = \frac{12S}{r^2 m(m^2 - 1)}. \quad (3)$$

The concordance coefficient can be practically used if its limiting value, indicating that the estimates provided by the experts may be still considered consistent, is determined. M.Kendall proved [21] that when the number of the objects is  $m > 7$ , the significance of the concordance coefficient could be determined, using  $\chi^2$  criteria. The random value

$$\chi^2 = Wr(m-1) = \frac{12S}{rm(m+1)}, \quad (4)$$

is distributed according to the distribution  $\chi^2$  with the degree of freedom  $\nu = m - 1$ . According to the chosen significance level  $\alpha$  (which is usually equal to 0.05 or 0.01) we can find the critical value  $\chi_{kr}^2$  from the table of  $\chi^2$  distribution with the degree of freedom  $\nu = m - 1$ . If the value of  $\chi^2$  obtained by formula (4) is larger than  $\chi_{kr}^2$ , the estimates of the experts are considered to be consistent.

Eleven highly qualified experts evaluated the significance of subcriteria describing the activities associated

with each component of the enterprise marketing system (i.e. *product, price, promotion, place*). Each criterion was assigned the rank which could range from one (if the criterion was important) to  $m$ , if the considered criterion was least important (with  $m$  denoting the number of the criteria considered). The results of the evaluation of subcriteria describing *product* are given in Table 1.

Table 1. The results obtained in ranking subcriteria describing *product*

Criterion	Expert											Sum of ranks	Place assigned
	1	2	3	4	5	6	7	8	9	10	11		
1 Range of goods (products)	8	1	1	1	5	1	8	5	5	8	4	47	4
2 Product design	3	7	5	7	6	4	4	4	3	3	6	52	5
3 Innovations	5	4	2	3	3	3	3	2	1	2	1	29	2
4 Quality	1	2	3	2	2	2	2	1	2	1	3	21	1
5 Brand / trademark	2	3	4	4	1	5	1	3	4	4	2	33	3
6 Packing (form, size, etc.)	4	8	6	8	8	7	7	7	7	5	7	74	7
7 Extra services	6	5	7	6	7	6	5	6	6	6	5	65	6
8 Warranties	7	6	8	5	4	8	6	8	8	7	8	75	8

Based on the data presented in Table 1, the concordance coefficient  $W = 0.592$ , and chi square value  $\chi^2 = 45.58$  were calculated by the formulas (1-4), while the critical value  $\chi_{kr}^2$ , taken from the distribution table with the degree of freedom  $\nu = 8 - 1 = 7$  and the significance level  $\alpha = 0.05$ , was equal to 14.07. The obtained  $\chi^2$  value

is considerably larger than the critical value, therefore, the estimates of experts are considered to be consistent.

The subcriteria describing the second main component of the marketing model *price* were ranked in a similar way, and the consistency of experts' estimates was determined. The results obtained in the evaluation of subcriteria describing *price* are presented in Table 2.

Table 2. The results obtained in ranking subcriteria describing *price*

Criterion	Expert											Sum of ranks	Place assigned
	1	2	3	4	5	6	7	8	9	10	11		
1 Initial price	1	2	1	1	3	3	1	2	1	1	2	18	1
2 Special offers / discounts	6	1	2	2	1	4	2	1	3	2	1	24	2
3 Terms of payment	4	6	6	6	7	7	5	6	4	6	5	62	7
4 Responsibilities	7	4	5	3	2	6	3	5	2	3	4	44	3
5 Price differentiation	3	5	4	4	5	1	6	4	7	4	6	49	5
6 Pricing strategies	6	3	3	5	6	2	4	3	6	5	3	46	4
7 Crediting, payment conditions	2	6	7	7	4	5	7	7	5	7	7	64	6

Based on the data presented in Table 2, chi square value  $\chi^2 = 35.63$  were calculated by the formulas (1-4). The obtained  $\chi^2$  value is considerably larger than the

critical value ( $\chi_{kr}^2 = 12.59$ ), therefore, the estimates of experts are considered to be consistent.

The estimates of the experts referring to subcriteria describing the third main component of the marketing system, *promotion*, are given in Table 3.

Table 3. The results obtained in ranking subcriteria describing *promotion*

Criterion	Expert											Sum of ranks	Place assigned
	1	2	3	4	5	6	7	8	9	10	11		
1 Advertising	2	2	3	1	1	2	2	2	1	1	2	19	1
2 Increase of sales, promotion	1	3	2	2	4	3	1	1	1	2	1	21	2
3 Planning and organisation of business communication	3	1	1	4	6	1	3	3	4	5	3	34	3
4 Personal communication (relationships)	6	4	5	3	2	5	5	4	6	4	4	48	4-5
5 Brand (trademark) management	5	5	4	6	3	4	4	5	3	3	6	48	4-5
6 Corporate identity	4	7	6	7	5	7	7	6	7	6	7	69	7
7 Information and communication with the public	7	6	7	5	7	6	6	7	5	7	5	68	6

Based on the data presented in Table 3, the square sum  $S = 2486.9$  of the deviations from the mean value, as well as the concordance coefficient  $W = 0.734$ , and chi square value  $\chi^2 = 48.45$  were calculated by the formulas (1-4), while the critical value  $\chi_{kr}^2$ , as mentioned above, was

equal to 12.59. The obtained  $\chi^2$  value is considerably larger than the critical value, therefore, the estimates of experts are considered to be consistent.

The results of the expert evaluation of subcriteria describing the fourth main component of the marketing system, *place*, are given in Table 4.

Table 4. The results obtained in ranking subcriteria describing *place*

Criterion	Expert											Sum of ranks	Place assigned
	1	2	3	4	5	6	7	8	9	10	11		
1 Place of sales	2	3	1	3	2	1	1	2	1	2	1	19	1
2 Direct sales	1	2	4	1	1	2	2	1	3	1	2	20	2
3 Indirect sales	4	1	3	2	4	6	4	3	4	3	5	39	4
4 Sales online	5	6	5	4	5	3	5	5	5	5	4	52	5
5 Channels of sales / distribution, mediators	3	5	2	5	3	4	3	4	2	4	3	38	3
6 Storing of products and equipment	6	4	6	6	6	5	6	6	6	6	6	63	6

Based on the data presented in Table 4, the concordance coefficient  $W = 0.711$ , and chi square value  $\chi^2 = 39.10$  were calculated by the formulas (1-4). The obtained  $\chi^2$  value is considerably larger than the critical value ( $\chi_{kr}^2 = 11.07$ ), therefore, the estimates of experts are considered to be consistent.

predicts the least possible  $i$ -th criterion value  $e_{ik(\min)}$  and its largest value  $e_{ik(\max)}$ . The mean values  $\bar{e}_{ik} = (e_{ik(\min)} + e_{ik(\max)}) / 2$  of evaluation intervals  $(e_{ik(\min)}, e_{ik(\max)})$  correspond to expert evaluation in points. In this case, quantitative evaluation (weight) of  $i$ -th criterion significance  $\omega_i$  is obtained by the formula  $(\sum_{i=1}^m \omega_i = 1)$ :

### 3. DETERMINING THE SIGNIFICANCE (WEIGHTS) OF THE CRITERIA DESCRIBING ENTERPRISE MARKETING

$$\omega_i = \frac{\sum_{k=1}^r \bar{e}_{ik}}{\sum_{i=1}^m \sum_{k=1}^r \bar{e}_{ik}}, \quad (5)$$

When the consistency of expert estimates is proved, it is possible to undertake the task of the second stage of evaluation, i.e. to determine the weights (significance) of the criteria. Experts evaluated the significances (weights) of the four main components (criteria) of the marketing system [22] including the *product*, *price*, *promotion* and *place*, as well their subcriteria. The indeterminate character of the evaluation values is clearly demonstrated when the range of the assigned values is determined, i.e. the  $k$ -th expert

The significance of the four main components (criteria) of the marketing system expressed in per cent, as well as the limits of variation of the estimates provided by eight experts, their mean values and weights calculated by formula (5) are presented in Table 5.

Table 5. The estimates of the significance of the main marketing criteria

Criterion	Expert	1	2	3	4	5	6	7	8	Mean values of evaluation intervals	The sum $\sum_{k=1}^r \bar{e}_{ik}$ and weights $\omega_i$
1 Product (P <sub>1</sub> )		40;60	50;70	100;100	20;80	50;70	40;100	50;100	50;80	50.0;82.5 66.3	530 0.282
	Mean value $\bar{e}_{1k}$	50	60	100	50	60	70	75	65		
2 Price (P <sub>2</sub> )		20;30	50;70	100;100	50;50	25;40	30;90	20;100	50;80	43.1;70.0 56.6	452.5 0.240
	Mean value $\bar{e}_{2k}$	25	60	100	50	32.5	60	60	65		
3 Promotion (P <sub>3</sub> )		10;20	40;50	100;100	50;50	30;50	30;80	30;100	40;60	41.3;63.8 52.5	420 0.223
	Mean value $\bar{e}_{3k}$	15	45	100	50	40	55	65	50		
4 Place (P <sub>4</sub> )		40;60	40;50	100;100	50;50	30;50	30;100	40;100	40;80	46.3;73.8 60	480 0.255
	Mean value $\bar{e}_{4k}$	50	45	100	50	40	65	70	60		

The calculations show that the experts estimate the first criterion (*product*) as most significant, while the significances (weights) of the three other criteria (*price*, *promotion*, *place*) do not differ considerably.

In a similar way, the values of the estimates' variation, their limits and weights were calculated by

formula (5) for all subcriteria describing the marketing system (Fig.1).

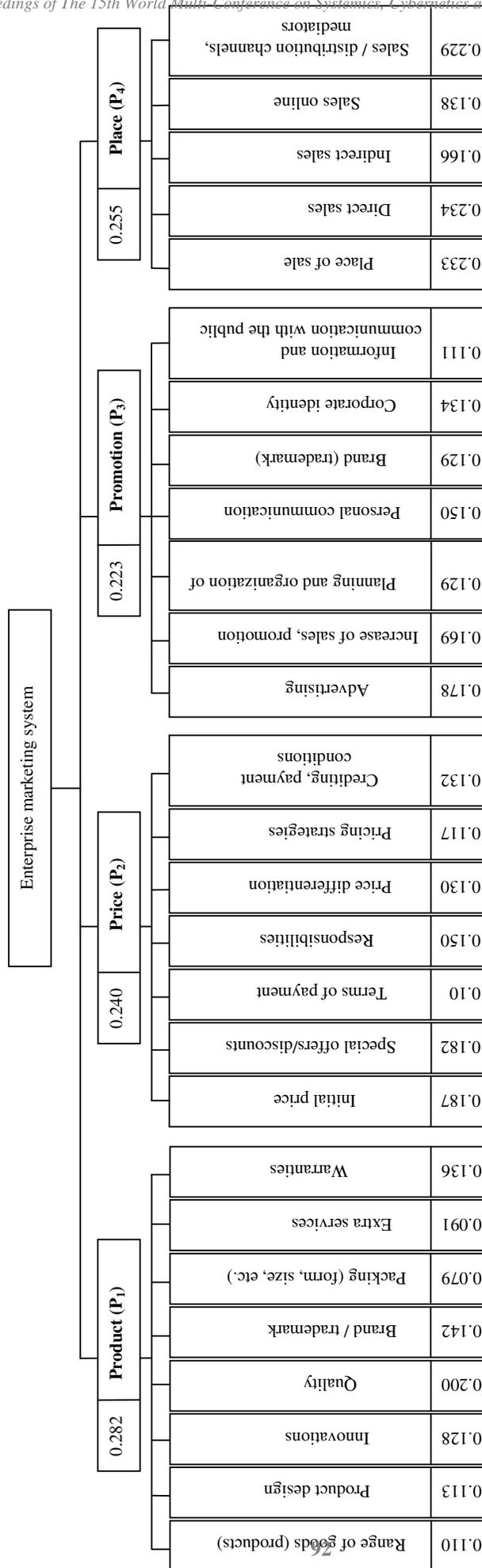


Fig 1. Weights of the components of enterprise marketing system described by the hierarchical set of criteria

Given the weights of the criteria, describing the enterprise marketing system, the quantitative evaluation of the state of this system may be continued at some other levels, including the determination of the criteria weights. Then, the appropriate multicriteria evaluation method may be chosen for the final evaluation of the system.

#### 4. CONCLUSIONS

The effectiveness of enterprise performance largely depends on its marketing strategies. These marketing activities require proper management, which may be achieved only if their effectiveness can be quantitatively evaluated at any moment of time.

Enterprise marketing in the countries of Eastern and Central Europe is usually described by the 4Ps model, including such elements (or evaluation criteria) as product, price, promotion and place. In its turn, each of these components may be subdivided into subcriteria, describing them.

Enterprise marketing may be referred to complex phenomena whose multiple aspects can be observed in reality. To perform quantitative evaluation of marketing activities, all the criteria describing them should be integrated into a single generalizing criterion. To solve these problems, methods of multicriteria evaluation can be used, but for this purpose the values and weights of the criteria should be known.

The weights of the criteria are determined by experts, therefore, this process consists of two stages. First, experts' estimates are ranked and their consistency is checked. Then, the weights of the criteria are determined. In multicriteria evaluation, the weights of the criteria are determined for each set of criteria at any level of their hierarchical system.

#### 5. REFERENCES

- [1]. A. Ginevičius, "Quantitative evaluation of enterprise marketing effectiveness", **Technological and Economic Development of Economy**, Vol. 13, No.1, 2007, pp. 19-23.
- [2]. R. Ginevicius and V. Podvezko, "Housing in the context of economic and social development of Lithuanian regions", **Int. J. Environment and Pollution**, Vol. 35, No. 2/3/4, 2008, pp. 309-330.
- [3]. K. Möller, "The Marketing Mix Revisited: Towards the 21st Century Marketing by E. Constantinides", **Journal of Marketing Management**, Vol. 22. No. 3, 2006, pp. 439-450.
- [4]. D.W. Cowell, **The Marketing of Services**, Institute of Marketing and the CAM Foundation, Heinemann Professional Publishing, 1984.
- [5]. P. Kotler, **Marketing Management**, 11th Edition, Prentice Hall International Editions, 2003.
- [6]. F. Brassington and S. Pettitt, **Principles of Marketing**, Third Edition, Prentice Hall / Financial Times, 2003.
- [7]. D. Jobber, **Principles and Practice of Marketing**, Third edition, McGraw Hill, 2001.
- [8]. P.D. Bennet, **Dictionary of Marketing Terms**, Chicago: American Marketing Association, 1995.
- [9]. M. Rafiq & P.M. Ahmed, "Using the 7Ps as A Generic Marketing Mix: An Exploratory Survey of UK and European Marketing Academics". **Marketing Intelligence & Planning**, Vol. 13, No. 9, 1995, pp.4-15.
- [10]. D. Popovic, "Modelling the Marketing of High-Tech Start-Ups", **Journal of Targeting, Measurement and Analysis for Marketing**, Vol. 14, No. 3, 2006, pp. 260-276.
- [11]. B. Lauterborn, "New Marketing Litany: Four Ps Passé: C-Words Take Over", **Advertising Age**, Vol. 61, No. 41, 1990, p.26.
- [12]. C.L. Goi, "A Review of Marketing Mix: 4Ps or More?" **International Journal of Marketing Studies**, Vol. 1, No. 1, 2009.
- [13]. R. Ginevičius, V. Podvezko and S. Bruzė, "Evaluating the Effect of State Aid to Business by Multicriteria Methods", **Journal of Business Economics and Management** Vol. 9, No. 3, 2008, pp.167-180.
- [14]. J. Figueira, S. Greco and M. Ehrgott, Eds. **Multiple Criteria Decision Analysis: State of the Art Survey**. Springer, 2005.
- [15]. W.K.M. Brauers, R. Ginevičius and V. Podvezko, "Regional development in Lithuania considering Multi-Objective by the Moora Method", **Technological and Economic Development of Economy**, Vol. 16, No. 4, 2010, pp. 613-640.
- [16]. R. Ginevicius, R. Martinkute. and V. Podvezko, "Multicriteria Decisions in Capital Markets", **The 10th World Multi-Conference on Systemics, Cybernetics and Informatics**, 2006, Orlando, Florida, USA, Vol. 1, 2006, pp. 36-39.
- [17]. A. Ginevicius, V. Gineviciene and V. Podvezko, "The Effectiveness of Enterprise Marketing System", **International Conference in Management and Marketing Sciences (ICMMS 2008)**, Book of Abstracts, 23-25 May 2008, Athens, Greece, pp. 31-32.
- [18]. V. Podvezko, "Determining the level of agreement of expert estimates", **International Journal of Management and Decision Making**, Vol. 8, No.5/6, 2007, pp. 586-600.
- [19]. R. Ginevicius and V. Podvezko, "Complex assessment of sustainable development of state regions with emphasis on ecological and dwelling conditions", **Ekologija**, Vol.53, Supplement, 2007, pp. 41-48.
- [20]. R. Ginevicius and V. Podvezko, "Assessing the Accuracy of Expert Methods", **Engineering Economics**, Vol. 5 (40), 2004, pp. 7-12.
- [21]. M. Kendall, M., **Rank correlation methods**, Fourth edition, London, Griffin, 1970.
- [22]. R. Ginevičius and A. Ginevičius, "Sustainability Decisions in Marketing Complex Cost Optimization Process", **12th World Multi-Conference on Systemics, Cybernetics and Informatics/ 14th International Conference on Information Systems Analysis and Synthesis**, JUN 29-JUL 02, 2008, Orlando, Florida, USA, Vol. V1, Proceedings, pp. 35-40.

# Sustainability: Environmental, Social and Corporate Governance Performance in Czech SMEs

**Alena KOCMANOVA, Marie DOCEKALOVA**  
**Brno University of Technology, Faculty of Business and Management,**  
**Department of Economics, Brno, Czech Republic**

and

**Petr NEMECEK**  
**Tomas Bata University in Zlín, Faculty of Management and Economics,**  
**Department of Management, Zlín, Czech Republic**

and

**Iveta SIMBEROVA**  
**Brno University of Technology, Faculty of Business and Management,**  
**Department of Management, Brno, Czech Republic**

## ABSTRACT

The paper focuses on environmental, social and corporate governance performance in relation to business sustainability in small and medium enterprises in the Czech Republic. Corporate Governance is understood as a key element when achieving economic performance as well as growth enhancing the investors' trust. It also makes it possible to create structures supporting the determination, checking and achievement of the objectives of the enterprise. It creates suitable initiatives for the members of the administrative bodies and the management. In the integration with sustainability, featured as the business strategy, the long-term corporate objectives, efficiency, performance, competitiveness are pursued through the inclusion of economic, environmental and social aspects to the corporate management. The sustainability of small and medium enterprises in the Czech Republic may bring about its effects in practice through the analyses of environmental, social and corporate governance performance, increased interests of uninvolved groups in the issues, competitiveness, increased market share etc. The inclusion of sustainability in the corporate management of an enterprise calls for a change in business. The target of the paper is the empirical analysis of the environmental, social and corporate governance performance in small and medium enterprises in the Czech Republic.

**Keywords:** sustainable development, sustainability, environmental performance, social performance, corporate governance

## 1. INTRODUCTION

In the Czech Republic, the sustainable development was defined in 1992 by Act No. 17/1992 Sb., on the Environment. From the original definition accept as a key element the theory of sustainable development meaning that the present generation should live in a way that does not restrict the life of the generations to come. These restrictions stem mainly from the

depletion of restricted resources. These may be economic, environmental and also social resources. Currently, the sustainable development is understood in a comprehensive way and „three pillars“ are considered: **society, economy and the environment**. Regardless the context, the crucial idea remains the same - people, environment and economic systems are interlinked [5].

## 2. SUSTAINABLE DEVELOPMENT AND BUSINESS ENVIRONMENT IN THE CZECH REPUBLIC

It can currently be observed that there is a growing importance of the support of sustainable development on the part of enterprises in the Czech Republic. The business sector, particularly small and medium enterprises, plays an irreplaceable role in terms of employment.

The importance of the business sector in market economy is quite obvious. However, an important role is also played by the business support. Research and empirical data demonstrate that the small and medium enterprises sector has a major share in the employment, in general. At the same time, small and medium enterprises are highly sensitive to the quality of the business environment. Small companies are mainly characterized by a weaker financial background. The basic task of such an enterprise is to keep or increase its own market value and to generate (reasonable) profit [15].

What is therefore of importance for the enterprise is that the market system should provide enough incentives and possibilities to make it behave in accordance with sustainable development. In order to make an enterprise competitive in the domestic and foreign markets, it should develop the principles of sustainable development. The enterprises are required, similarly to all other social stakeholders, to strongly boost the productivity of resources, while creating at least the same values, to reduce the consumption of materials, water, energy, to optimize services and products required by the customer.

An enterprises setting out sustainable development as

a strategic objective will sooner or later encounter a question of how to *measure sustainability of the enterprise, how to set the objectives* and what measures and procedure should be applied to meet the objectives. Therefore, it is necessary to collect, record, analyze and hand over information on the economic consequences of environmental and social activities. This information need must be satisfied and new tools must be developed to support the decision-making processes. Indicators, management analysis tools, communication and information system are employed by the enterprises to determine the sustainable development objectives.

What is emphasized is the use of existing tools in the enterprise, support of strengths and reduction of weaknesses [4].

Indicators used to measure sustainable development in enterprises are continuously developed by various international organizations in order to achieve an internationally recognized standard intended to ensure mutual comparability of national economies, individual fields of industry and individual enterprises

The most important international activity is Global Reporting Initiative-GRI focusing on standardizing the sustainable development report (Sustainability Report).

The objective of these voluntary activities is to win or strengthen the trust of a business entity with investors and other stakeholders.

The enterprise is under the pressure of *economic, environmental* and currently also *social* legislation, ethical codes and conditions of consumers, suppliers and therefore profit maximization ceases to be the only criterion. The decision-making must also include voluntary approaches related to sustainable development. The objective is to achieve a balanced relationship between the environmental, economic and social pillars - to ensure sustainable management at the level of the enterprise. It is also necessary to define the sustainable corporate governance, i.e. such management that results in the sustainable development of the enterprises as well as the whole society.

Enterprises that wish to succeed in the advanced international as well as domestic markets declare the quality of their business management by introducing relevant environmental management systems (ISO 9000, ISO 14 000, EMAS, Cleaner production, LCA, BOZP, SCR, etc.) and the ownership of applicable certificates. This guarantees the business partners, as well as other stakeholders (public administration, the general public) a certain standard of

corporate culture and the level of development of the relevant enterprise.

To measure the sustainable development at the corporate level abroad, a number of economic, environmental and social indicators are available, documenting the development of changes in the care of the enterprise in the individual environmental domains over the defined period of time. The indicators may be absolute and relative, financial parameters may be used for the indication.

The evaluation of sustainable development keeps being developed at the international level by various international organizations in order to achieve an internationally recognized standard aimed at a mutual comparison of national economies, the individual branches of industries and individual enterprises.

Micro enterprises, small and medium enterprises form an important part of every economy, as they account for a majority in the structure of companies. Over the past two decades, their number has been on the rise and their influence in the business environment, as well as social environment where they act as an element strengthening the social stability has been increasing. In relation to the environmental aspects the research indicates that the environmental management systems can also be introduced in small and medium enterprises. The environmental management system is a highly motivating voluntary tool for small and medium enterprises. A great incentive for the enterprises is the obtaining of a certain certificate in order to make their efforts at improving the management of the enterprise and its stance to the environment more visible.

The small and medium enterprise sector is a driver of the business domain, growth, innovation as well as competitiveness. It is also an important tool when creating job opportunities. As of 31<sup>st</sup> December 2010, there was a total of 1 029 871 small and medium enterprises in the Czech Republic, representing 99.83 % of all economically active entities. They had a 62.33 % share in the total employment rate and a 36.22 % share in the GDP.

The SMEs also have a major share in the total export and import where both these values reach a minimum of 50 %. A detailed view of the small and medium enterprise development in the Czech Republic in 2000 - 2010 is provided in Figure 1. It also provides a division into natural persons and legal entities.

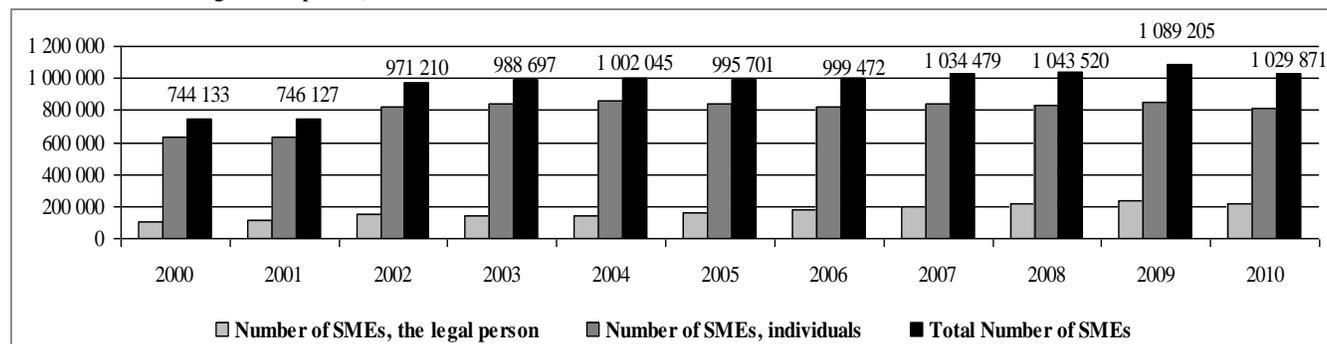


Figure 1 Development of the number of active SME in the CR  
(Source: our own processing of data at the MPO [16])

The importance of small and medium enterprises in the economy is not challenged by anyone. Their ability to react fast to changed conditions or ability to absorb free labour is irreplaceable. As regards their functioning (their number, share in the employment rate and added value), the Czech economy is

standard.

The corporate governance of small and medium enterprises needs a procedure to measure, control and improve the performance of the enterprise. Such a procedure for effective and efficient management should be simple and

efficient and it should integrate these various viewpoints of the performance in environmental, economic and social terms [3].

### 3. EMPIRICAL ANALYSIS OF ENVIRONMENTAL, SOCIAL AND ECONOMIC PERFORMANCE IN SMALL AND MEDIUM ENTERPRISES IN THE CZECH REPUBLIC

Sustainability of small and medium-sized businesses in the Czech Republic may bring effects in the capturing the practice in the form of analyses of the environmental, social and economic performance, increasing the interest of stakeholders in the issues, etc. Surveys in small and medium-sized enterprises focused on the mapping of sustainable development situation in 2010. A total of 280 companies from the processing industry, construction, trade and services selected from a corporation database were contacted. Of the total, 27.4 % were companies with under 250 employees, 28.2 % with under 50 employees and 17.7 % companies with fewer than 10 employees. In the sample of companies analyzed, the most frequently represented industries were the processing industry, trade, construction industry and services.

Among small and medium-sized enterprises, differences in the interest in, and intensity of response to, the issue of sustainable development were expected. It follows from the survey that 57% of companies include a reference to sustainable development, and include it into their strategic goals, and 43% of companies do not include a reference to sustainable development and it is not a part of their strategic goals, either.

In the survey, respondents were offered a choice of four statements describing sustainable development and were asked to choose the one that best described their experience and practice. There may be a number of explanations why the content of the concept is not well-understood in small and medium-sized enterprises, e.g., employees are inadequately informed about sustainable development, failure to include strategic and social goals, etc. A small percentage of respondents from small and medium sized enterprises agreed on the basically textbook definition of sustainable development, the best-know definition is that sustainable development will provide for a "balance between three pillars" (8.6 %), and the "future generations" definition was second (18.5 %) best characterizes

Table 1 Link to different areas of sustainable development

	Definitely Not	Rather Not	Rather Yes	Definitely Yes
in environmental field	5.6%	6.4%	16.9%	13.7%
in economic field	3.2%	6.5%	15.3%	21.0%
in social field	4.8%	7.3%	16.9%	12.9%

(Source: own processing of empirical research)

In their opinion, the small and medium enterprises contribute most to sustainability in the economic field. On the other hand, they view their contribution as smallest in the social field.

Implementation of sustainable development in small and medium-sized enterprises may be advantageous for several reasons, most importantly because it may enhance eco-efficiency of technological processes, mitigate negative impacts on health and the environment, lead to an economic success of the company, facilitate negotiations with state authorities, local

governments and other stakeholders, enhance positive attitudes of employees towards their company and thus achieve significant economic effects, increase attractiveness for would-be investors and for international cooperation, etc.

The Czech Republic is quite successful in joining in the worldwide efforts at sustainable development of human society. As regards small and medium enterprises in the Czech Republic, there is an effort to mitigate the negative impacts on the environment by introducing more environment-friendly technologies aimed at reducing the volumes of waste to the minimum. There is a very strong involvement of voluntary initiatives and activities focusing on environmental protection in these companies by introducing voluntary tools in the environmental and social fields.

The effort of the enterprises is to focus on sustainability. Thus, sustainability is a strategy of the sustainable development process. Sustainability combined with the business environment has become known due to the environmental approaches in the enterprises. Sustainability has three important dimensions for the enterprises: economic growth, social responsibility, and responsibility for the environment. The sustainability of the enterprise is a strategic approach concentrating on the efficiency, efficacy and performance of the enterprise as well as on the creation of values for the owners (competitiveness) resulting from the environmental, economic and social dimensions. The definition of sustainability is related to the conception of strategy marked as sustainable development strategy in relation to the enterprise [11], [12].

The sustainability strategy of an enterprise currently encompasses a broad approach related to the integration of environmental, social and corporate governance, understood as key elements to achieve economic performance and to enhance investors' trust.

### 4. ENVIRONMENTAL PERFORMANCE

The vision of the sustainable development of an enterprise is currently introduced as a broad approach concerning the integration of economic, environmental and social characteristics.

In the Czech Republic, the environment-oriented management system is based on the ČSN EN ISO 14 000 standard represented mainly by the ČSN EN ISO 14001 standard. The ČSN EN ISO 14 000 standard series govern the implementation of the environmental management systems and their audits. The basic idea behind these international standard is to encourage the enterprises in all aspects towards active and independent dealing in environmental matters. The ČSN EN ISO 14 001 standard may be applied in enterprises of all types and sizes, while respecting internal and external conditions.

To evaluate the environmental performance, information about the environmental profile of the enterprise is necessary. The environmental profile evaluation (EPE) according to ČSN EN ISO 14 031 is yet another important internal tool providing continuously reliable and verifiable information to the corporate governance, which makes it possible to determine whether the environmental profile of the enterprise meets the criteria set by the corporate governance.

Attention should mainly be paid to environmental operations that contribute to the steady growth of values for the owners; therefore, it is necessary to concentrate on the future and long-term effects supporting sustainable development of the enterprise.

The empirical research indicates that the priority aspect for small and medium enterprises in environmental area is the consumption of energy and heat (41.5%) followed by waste (34.7%), see Figure 2.

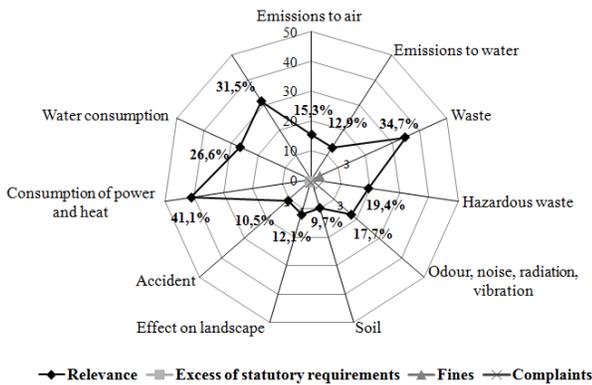


Figure 2 Significant environmental aspects (Source: own processing of empirical research)

The SMEs most often improve their environmental profile by minimizing waste (35.5%), improving the environment (31.4%). On the contrary, they pay least attention to sustainable traffic, see Figure 3.

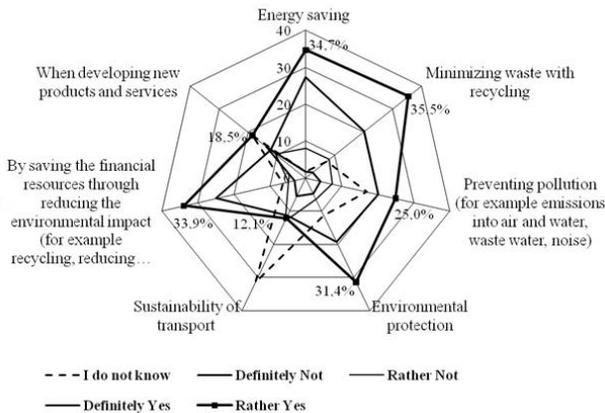


Figure 3 Ways of reducing the environmental impact (Source: own processing of empirical research)

Czech SMEs have a number of voluntary environmental tools. A total of 37.9% of the contacted companies uses the Quality management system according to ISO 9000, 26.6% use the environmental management system; acc. to ISO 14000, 17.7% of companies have introduced cleaner production. The least applied is the Environmental management accounting and Environmental management systems based on EMAS.

### 5. SOCIAL PERFORMANCE

Social performance of the enterprise is another important component of the economic and environmental performance of the enterprise. Conversely, it could also be said that good economic performance forms solid foundations for ecological and social aspects. Therefore, it is generally good for an enterprise to process and adopt principles in the field of health

and safety of the employees, human resources and ecology. Corporate social performance could be considered as employee care.

An important element of the social pillar of sustainable development is occupational health and safety (OHS). The right to the protection of health is one of the fundamental human rights guaranteed in the developed countries by the constitution and the charter of fundamental human rights and freedoms. To this end, the EU has adopted a number of directives. For the corporate governance of each enterprise this indicates the need of permanent creation of such conditions at work that guarantee a high degree of safety for the employees of the enterprise as well as its customers and surroundings. The trend underscoring the social aspects of sustainable development is the concept of corporate social responsibility (CSR).

Other key issues related to the corporate social responsibility are: human rights, employees' rights, involvement of municipalities and relationships with suppliers, information policy including issues such as releasing information, transparency, educating the consumers and anti-corruption measures. The importance of the corporate social aspects is mainly linked to the human resources. From the economic point of view it is possible to evaluate the performance of a person by e.g. the volume of work made over a certain time, the work performance may be measured by productivity of labour. From the social point of view, the person is willing and able to achieve an optimal performance (i.e. achieve optima work performance), if s/he uses his/her full potential, feels self-fulfillment and satisfaction and develops his/her personality.

Based on the conducted research, 48.4% of all contacted companies take care about the feedback with their customers, suppliers and other partners. It comes as a surprise that the SMEs do not support the local labour markets and do not provide work for people from the community. The employment in this way is offered by as little as 4% of the companies. 29% of companies provide funds to support public activities and projects, particularly for charity purposes. On the other hand, as little as 5.6% of companies, encourage their employees to participate in public activities and projects. Only 12.9% of companies are able to employ the ecological parameters of their products and services to gain competitive advantages although 26.6% of companies consider the environmental impacts of their products in the course of their development, see Figure 4.

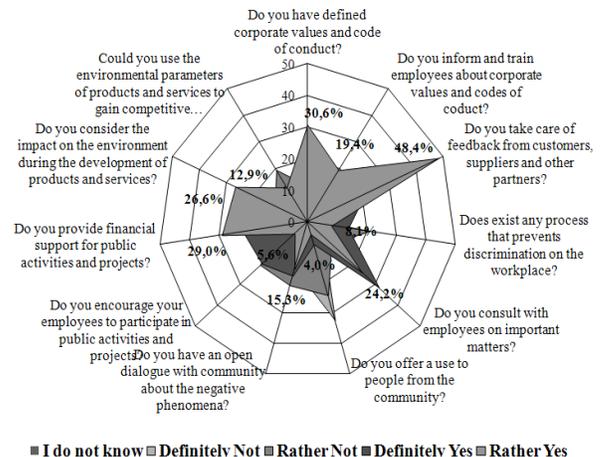


Figure 4 Responsible corporate behavior in the social field (Source: own processing of empirical research)

The most frequently used tool is occupational health and safety, this tool is implemented in 78.2% of cases. This is followed by safe company (16.9%), checklist for the safety of suppliers (14.5%) and Corporate Social Responsibility (13.7%), while the CSR will become a priority tool in the future, see Table 2

The previous text clearly indicates the effect of the corporate social performance on the economic performance. Despite that, there is still a lack of theoretical and empirical knowledge dealing with the relations between the social and

economic performance.

The common rule applicable to most enterprises is that improvement in the field of social performance is only acceptable if it pays off. To pay adequate attention to social performance and its management, it is first necessary to identify the linkage between the economic and social performance. If it turns out that the linkage is strong enough, the social performance will start being perceived as a major competitiveness factor.

Table 2 Tools used in the social field

	Used	Will be implemented	Implementation in the future	Now important	Barriers to implementation
Social Responsibility of Companies – CSR	13.7%	8.8%	10.5%	10.4%	6.5%
Occupational Health and Safety Management System (HSMS)	78.2%	4.0%	2.4%	4.8%	4.0%
Check-list for security of contractors (SCC)	14.5%	8.8%	7.3%	31.5%	8.1%
Safe company	16.9%	14.6%	3.2%	33.0%	5.6%

(Source: own processing of empirical research)

The remaining key question is whether social performance may mean a competitive advantage. The answer depends on the current and future competitive standing of the enterprise.

## 6. CORPORATE GOVERNANCE

The importance of Corporate Governance consists in its contributing to not only corporate prosperity, but also to responsibility. Along with the development of global markets investors' activity increases, with them demanding higher standards of responsibility, conduct and performance. Investors tend to seek opportunities outside their domestic markets ever more often. The companies trying to gain resources on the international capital markets, however, often find that capital is only available for those who conform to the internationally accepted standards of Corporate Governance and publishing of information. These are only some of the reasons leading to the worldwide improvement of the Corporate Governance standard and, in some degree, to its convergence. The defining of Corporate Governance is not a matter of unified terminology. In the evaluation of the corporate governance is described with the following quotes: a system through which companies are managed and controlled [10].

The statutory bodies are responsible for corporate management. The responsibility of a body covers the setting of a company's strategic goals, the management keeping check on realization of the goals, supervision of the management and informing shareholders about the performance of duties of a steward [1]. According to another description of Corporate Governance it is a process through which companies respond to the rights and requests of stakeholders [2].

Corporate Governance is understood as the key element in the effort to reach economic efficiency and a growth

justifying increase in the investor trust. It encompasses a broad range of problems arising from the relationships between the corporate management, the administrative authorities, shareholders and the other stakeholders [6].

In the Czech Republic, on the other hand, the dualistic model is used; it consists of the board of directors as the executive body and the supervisory board as the body of supervision. Based on the KPMG research a board of directors in the Czech Republic typically comprises three members, followed by a five-member setup. A typical supervisory board comprises three or six members. An average number of members of the administrative bodies (9.6) does not reach the numbers common in the EU with the average equalling (12.5) members [9].

In the Czech Republic a Code of supervision and management of companies based on the OECD principles was introduced as early as 2001 and was last updated in 2004.

The KPMG survey addressed 500 largest Czech companies. The objective of the survey was to understand some procedures of corporate governance and working methods used by the Boards of Directors and Supervisory Boards of companies in the Czech Republic. A special emphasis was placed on the ability of the members of the corporate bodies to gain information and work with it efficiently. The results of the survey indicate that most interviewed companies have only set the corporate governance rules to an extent meeting the basic statutory. This minimalist approach means that some procedures and practices are a far cry from the recognized best practice [8].

According to the conducted SMEs research, the most frequently published information is environmental and social information, not in the form of separate reports but as a part of another summary report, such as, for example, a chapter in the annual report (16.9%). The least frequent form is a sustainable development report, see Table 3.

Table 3 Forms of voluntary reports published

	I do not know	Definitely Not	Rather Not	Rather Yes	Definitely Yes
Environmental Report	9.6%	34.7%	14.5%	6.5%	8.1%
Health, Safety and Environmental Report	14.6%	36.3%	12.1%	2.4%	8.1%
Corporate Social Responsibility Report	18.6%	33.9%	14.5%	1.6%	4.8%
Sustainable Development Report	15.4%	37.9%	12.1%	5.6%	2.4%
Accrued report	12.1%	27.4%	8.9%	8.1%	16.9%

(Source: own processing of empirical research)

Corporate Governance is a system through which business companies are managed and controlled. The structure of corporate governance defines the division of rights and duties between the individual stakeholders in a company and lays down detailed rules and procedures for the decision-making on business matters of a company. On this basis a structure is created that establishes the company goals and the means of reaching the goals and monitoring performance“.

## 7. CONCLUSIONS

The enterprise is thus exposed to the pressures of economic, environmental and currently also social legislation, ethical codes and conditions of the consumers, suppliers and the profit maximization therefore stops being the only criterion. The decision-making must also include voluntary approaches related to sustainable development. The target is to achieve a balanced relationship between the environmental, economic and social pillars- to secure sustainable management at the corporate level. As it has been stressed several times in the article, sustainability has three important dimensions for an enterprise and although there may be compromises between these dimensions , social responsibility (towards the employees and other stakeholders) and environment-friendly behavior lead generally towards the strengthening of trustworthiness, and therefore this approach also makes sense from the business point of view.

The crucial element in achieving the targets in all the aforesaid areas is the Corporate Governance. Corporate Governance should consider the interests of the stakeholders and affect the development of the business strategy through its activities and its implementation in order satisfy the justified interests of these parties. Simply speaking, Corporate Governance should integrate the efforts of all three areas – environmental, social and economic and bring them to life through a suitable strategy and means.

### Acknowledgements:

This paper is supported by The Czech Science Foundation. Name of the Project: Construction of Methods for Multifactor Assessment of Company complex Performance in Selected Sectors. Reg. Nr. P403/11/2085

## 8. REFERENCES

- [1] A. Cadbury, **The Committee on the Financial Aspects of Corporate Governance and Gee and Co. Ltd.** Report. 1992, p.61.
- [2] A. Demb, F.F. Neubauer, **The Corporate Board-Confronting The Paradoxes.** New York, Oxford University Press 1992.
- [3] A. Kocmanová, P. Němeček, **Economic, environmental and social Issues and Corporate Governance in Relation to Measurement of Company Performance.** In Liberec

- Economic Forum 2009. Liberec: Technická univerzita v Liberci, Fakulta ekonomiky. 2009, pp. 177-187.
- [4] A. Fedorová, A. Kocmanová, J. Hyršlová, M. Hájek, **Environmentální management podniku: Environmentální účetnictví a rozhodovací úlohy.** Brno: CERM,s.r.o., 2004.
  - [5] B. Moldan, T. Hák, H. Kolářová, **K udržitelnému rozvoji České republiky: vytváření podmínek. Svazek 1, Zdroje a prostředí.** Praha: Univerzita Karlova v Praze, Centrum pro otázku životního prostředí. 2002, pp.12.
  - [6] J. KLÍROVÁ, **Corporate Governance-správa a řízení obchodních společností.** Praha, ManagementPress, 2001.
  - [7] J. Klírová, P. Kavalíř, **Demand for Education in Corporate Governance on Increase.** Acta Oeconomica Pragensia, No. 5, 2004, pp.13-21.
  - [8] KPMG ČR. **Corporate Governance v České republice.** Praha, 2005.
  - [9] P. Haspeslagh, R. Slagmulder, M. Bloemhof, **Strategy and performance management at DSM.** INSEAD Case 01/2004 - 5165. This case is available at [www.ecch.com](http://www.ecch.com) and also published in *Strategy: Analysis and Practice*, McGee J. et al., McGraw-Hill, 1st ed., 2005, pp. C437-457.
  - [10] P. Kavalíř, **Hodnocení corporate governance v České republice.** Centrum ekonomických studií VŠEM, Working Paper CES VŠEM No 14/2005.
  - [11] S. Harts, **A natural resource-based view of the firm.** Academy of Management Review, 20(4). 2005, .966-1014.
  - [12] W.E. Stead, J.G. Stead, **An empirical investigation of sustainability strategy implementation in industrial organizations.** In D Collins&M.Starik (Eds.) *Research in corporate social performance and policy.* Supplement 1 pp.43-66.
  - [13] OECD, **Principles of Corporate Governance.** Available at: <http://www.oecd.org/dataoecd/32/18/31557724.pdf>
  - [14] Global Reporting Initiative. Available at: <http://www.globalreporting.org>.
  - [15] F. JEŽEK, **Meaning entrepreneurial environment and his brief characteristic in conditions CR 2008.** Available at: <http://mpa.ub.unimuenchen.de/12467/1/PodnikProstrediR EPEC.pdf>.
  - [16] Report on development Czech SMEs and his support in the year 2010. Available at: <http://mpo.cz>.

# **Portfolio Management Modified by GARCH-Type Models and Moving Average Correlation in the Global Currency Market**

**Konstantins KOZLOVSKIS, Jurijs LACIS**

**Faculty of Engineering Economics and Management, Riga Technical University  
1/7 Meza Str., Riga, LV-1007, Latvia**

## **ABSTRACT**

The development of information and financial technologies offers a wide spectrum of opportunities in Internet trading and its automation in the global financial markets. An investor can use different approaches in his trading. Nowadays, there are lots of information sources about how to trade. Besides, available modern software can help in mathematical model building including econometric models providing wider their practical application. In this paper some GARCH-type models are used as a modification in risk and return evaluation in Markowitz portfolio building approach. Consequently a bibliography of research papers on GARCH-type models and other econometric ones during 30 years since they were introduced by Engle in 1982 would run to hundreds of pages, and it is hard to find the work that is most relevant to practitioners and especially in the field of speculative trading.

In this paper the authors use widely known Markowitz portfolio and modify it by GARCH-type models for operating expected risks and returns in the portfolio traded in the global currency market for speculative purposes. Also moving average correlation is used in the portfolio to predict interdependences between two currency pairs. Such modifications are compared with naive portfolio diversification.

**Keywords:** global currency market, portfolio management, GARCH-type model, n-period moving correlation

## **1. INTRODUCTION**

Trading opportunities in financial markets are constantly broadened due to development of different factors. Extension of financial services opens new potentialities of Internet trading merging several financial markets under one trading platform. Financial innovations offer new and more interesting financial instruments and products, which allow investors to get round difficulties related to some limitations in classical financial markets, for instance, trading short positions in stock markets (solved by CFD contracts) or new gold trading accounts in which all trades are recalculated in terms of gold (it helps to protect trading accounts against inflation and

unpredictable negative fluctuations). Modern information technologies and solutions in the field of software-to-software interface are capable to integrate different information systems with trading platform providing more qualitative analysis of financial data by other software.

Modern trading platforms and achievements in information technologies give unique possibility to automate trading financial instruments. Free competition between brokerage firms leads to increased leverage and splitting trading lots. If a classical trading lot is equal to 1.0, today it is possible to trade smaller lots of 0.1 and even 0.01, for instance, in the global currency market. Of course, this splitting decreases potential profit and risks, but on the other hand, it allows managing complex portfolios with lesser capital.

The Markowitz portfolio is a portfolio, which takes into account the expected return and risk of assets as well as interdependencies between chosen assets through covariation decomposed in correlation and standard deviations. The presence of such variables gives a good opportunity to apply different mathematical approaches to modify a portfolio built by Markowitz approach. In fact, not all mathematical modifications can be easy applied in trading. For example, modifications with copula [11] are effective too, but its implementation is too labor-intensive [10].

In this paper the authors used some econometric models like AR (autoregression) models, GARCH-type models, n-period moving average correlation in portfolio trading in the global currency market implementing automated trading for speculative purposes.

## **2. METODOLOGY**

The main problem is automated trading in the financial markets is programming opportunities and trading strategy complexity. If a trading strategy is too complex, it cannot be programmed or programming can be too complicated to be realized. On the other hand, there should be a possibility to integrate programmed module in trading software. For example, trading software called Metatrader has such possibility. Analyzing opportunities of portfolio trading in the global currency market, the

authors apply relatively simple econometric models to diversify risks of currency pairs using portfolio management. The authors use the 3 types of currency portfolios mentioned below.

**2.1. Simple portfolio with equal weights**

Despite the sophisticated theoretical models developed in the last 50 years and the advances in methods for estimating the parameters of these models, investors continue to use such simple allocation rules for allocating their wealth across assets (for example, [4]). In this case of currency portfolio capital is distributed between financial instruments in equal proportions calculated as 1/N, where N is number of financial instruments. This portfolio strategy is able to show the forecasting power of used models applied for forecasting expected return without taking into account any dependences between currency pairs. Autoregressive models of time series are representations of a time series by functions of its own lags. The authors use the following autoregression model to predict future returns:

$$p_t = \sum_j a_j p_{t-j} \tag{1}$$

The authors chose the autoregressive process because it is often used to model and predict various types of natural phenomena. Its application is widely distributed in financial data analysis and it is easy to implement in programming software.

**2.2. The Markowitz portfolio modified by GARCH-type models**

If a traditional widely known Markowitz portfolio can be expressed as (1),

$$\left\{ \begin{array}{l} \mu_p = \sum_j x_j \mu_j \\ \sigma_p^2 = \sum_j \sum_k x_j x_k \rho_{jk} \sigma_j \sigma_k \\ \sum_j x_j = 1 \end{array} \right. \tag{2}$$

the modified by GARCH-type model in terms of this work can be written as (3).

$$\left\{ \begin{array}{l} \mu_p = \sum_j x_j \mu_j^{AR(n)} = \hat{\mu}_p \\ \sigma_p^2 = \sum_j \sum_k x_j x_k \rho_{jk} \sigma_j^{GARCH} \sigma_k^{GARCH} \rightarrow \min \\ \sum_j x_j = 1 \end{array} \right. \tag{3}$$

where

- $\mu_p$  – the expected portfolio return;
- $\hat{\mu}_p$  – the expected investor’s return;
- $\mu_j^{AR(n)}$  – the expected j-th currency pair return estimated by autoregression model AR(n);
- $\sigma_p^2$  – the expected portfolio risk;
- $\rho_{jk}$  – the correlation between j-th and k-th currency pairs;
- $\sigma_j^{GARCH}$  – the expected currency pair risk estimated by GARCH-type model;
- $x$  – the weights.

Markowitz postulates that an investor should maximize expected portfolio return ( $\mu_p$ ) while minimizing portfolio variance of return ( $\sigma_p^2$ ) [14]. In the original Markowitz’s work the expected return of each asset is a constant for chosen period and based on the previous return data. The same is with the expected risk of an asset. Typical behavior of any financial instruments is shown in the Figure 1. Actually, such dynamics can be described quite well with ARCH/GARCH-type models.

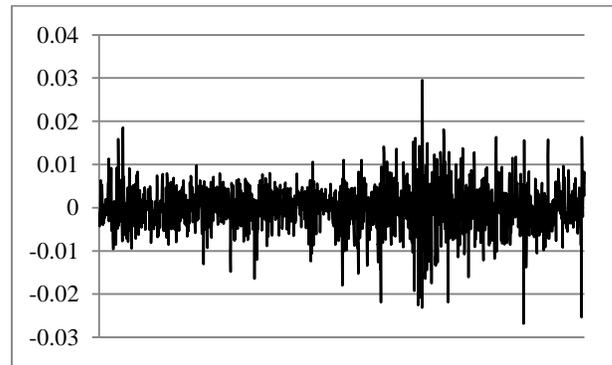


Figure 1 Typical example of security’s return (USDJPY)

In 1982 Engle introduced a new class of models called ARCH [6]. An important property of ARCH models is their ability to capture volatility clustering in financial data, i.e. the tendency for large and small swings in returns to be followed by larger and smaller swings of random direction. Generalized ARCH(p,q) model was first suggested by Bollerslev [5]. ARCH/GARCH effects is widely documented in the financial literature, for instance: for index returns [1], for futures markets [16], for individual stock returns [8], for bond yields and returns of US Treasury Bills [17], etc. In nowadays there are lots of widely used ARCH-type models, the most known of which are EGARCH (introduced by Nelson) [15], ARCH-M (introduced by Engle, Lilien and Robins) [9], TGARCH (introduced by Zakoian) [18], etc. The authors of this paper used the simplest models from the estimated ones.

In the Markowitz portfolio modified by GARCH-type model the expected return of a currency pair can be calculated by conditional mean expressed by any

model(in our case by AR(n)) and the expected portfolio risk can be evaluated by GARCH-type models. Thus, the currency pair's risk will be dependent on news about volatility from the previous period, measured as the lag of the squared residual from the mean equation (the ARCH term) and last period's forecast variance (the GARCH term). Such modification could evaluate expected risks more careful and build more suitable portfolio structure.

**2.3. Markowitz's portfolio modified by GARCH-type models and moving average correlation**

In fact, interconnections between financial instruments are not fixed practically in any financial market. The same situation is observed in the global currency market. Too many factors have impact on currency pairs' movements. In long-term perspective the correlation coefficient between two currency pairs is more stable than in short-term perspective [11]. The typical situation in changes in correlation between two currency pairs is shown in the Figure 2 and Figure 3. As it shown the correlation coefficient is widely distributed within interval [-1;1]. The main idea is to use this dynamics of correlation forcing the Markowitz's portfolio.

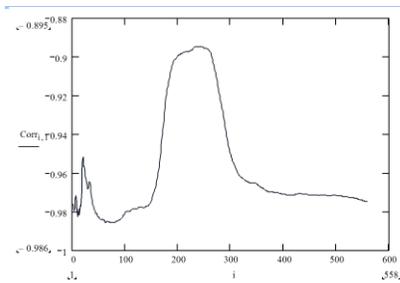


Figure 2 Correlation coefficient in long-term perspective

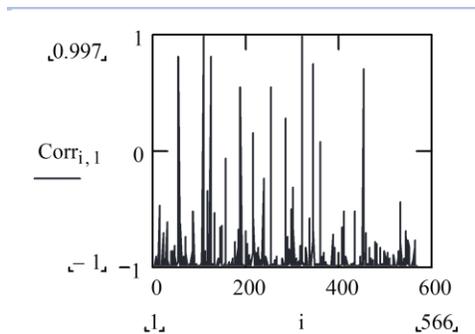


Figure 3 Correlation coefficient in short-term perspective

In 2000 Engle introduced dynamic conditional correlation as a simple class of multivariate GARCH models [7]. To simplify the approach the authors used a simple moving average correlation coefficient which can be calculated as

$$\rho_{ij,t}^{MA} = \frac{\sum_{k=1}^n \rho_{ij,t-k}}{n}, \tag{4}$$

wheren is a period of moving average.

Thus, the most complicated portfolio can be written as

$$\left\{ \begin{aligned} \mu_p &= \sum_j^N x_j \mu_j^{AR(n)} = \hat{\mu}_p \\ \sigma_p^2 &= \sum_j^N \sum_k^N x_j x_k \rho_{jk}^{MA} \sigma_j^{GARCH} \sigma_k^{GARCH} \rightarrow \min \\ \sum_j^N x_j &= 1 \end{aligned} \right. \tag{5}$$

Assuming that in the next period the correlation between two currency pairs is equal to average value based on previous periods, such portfolio structure could be able to take into account both volatile variance (GARCH-type component in the portfolio) of currency pairs and their constantly changing interdependence (moving average correlation component in the portfolio).

**3. EMPIRICAL RESULTS**

The authors analyzed the most liquid currency pairs in the global currency market. According to the statistical information from the Bank for International Settlements, the most tradable currency pairs are USDJPY, USDCHF, USDCAD, GBPUSD, EURUSD, EURJPY, EURGBP, EURCHF, AUDUSD. The average prices were used for each instrument represented by daily data. The time period was chosen from January 1, 2000 to April 1, 2011. Any models used in the paper were estimated on the data of 2004 (total 263 observations) and implemented on the data from 2005 (total 1621 observations).

**3.1. Portfolio management with naive diversification**

Analyzing the average daily data of 9 currency pairs, the authors found that forecasted average prices can be very simply described by autoregressive process. The Table 1 represents evaluated autoregressions for analyzed currency pairs.

Table 1 Statistics of AR(n) models

No	Currency pair	Model	R-squared	DW stat
1	USDJPY	AR(3)	0,9759	1,9203
2	USDCHF	AR(3)	0,9835	1,9393
3	USDCAD	AR(4)	0,9940	1,9804
4	GBPUSD	AR(4)	0,9705	1,9722
5	EURUSD	AR(3)	0,9854	1,9391
6	EURJPY	AR(3)	0,9615	1,9458
7	EURGBP	AR(4)	0,9841	2,0104
8	EURCHF	AR(3)	0,9861	1,9464
9	AUDUSD	AR(4)	0,9850	1,9625

In the Table 1 it is shown that autoregression models describe initial data of 2004 very well. Using these

models for forecasting average prices and realizing the corresponding strategy (if forecasted average price in next period is greater than in the previous then long position is open, otherwise, short), return on each currency pair without their weights in portfolio and with them were calculated (see Table 2).

Table 2 Profits using naive portfolio diversification

No	Currency	Profit, pips	
		w/o weights	with weights 0.11
1	USDJPY	-1522	-168
2	USDCHF	1830	201
3	USDCAD	-286	-31
4	GBPUSD	373	41
5	EURUSD	4588	505
6	EURJPY	-830	-91
7	EURGBP	2020	222
8	EURCHF	1511	166
9	AUDUSD	-3490	-384
10	Portfolio	4192	461

Despite of good forecasting power total result is not so good. There are both results: negative and positive. Total result is positive but it cannot be considered as a predicted results based on econometric models. The Figure 4 shows, that the dynamics of cumulative portfolio return resembles behavior of any financial asset, that is total result is dependent on present situation in market and trends. On the hand, the graph eliminates lots of figures of technical analysis which can be interpreted very simply. Thus, cumulative portfolio return can be used as a benchmark for 9 currency pairs. Trading in the trends and using favorable fluctuations in such benchmark, it could be possible to make profit.

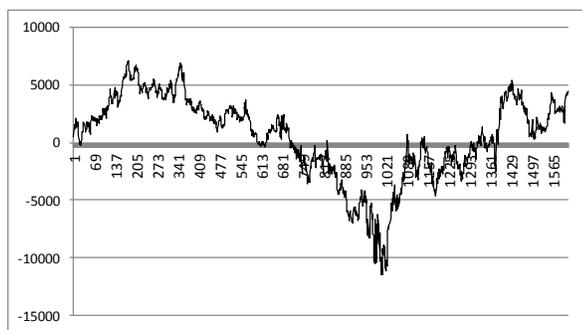


Figure 4 Cumulative portfolio return, pips

Bad implementation of autoregression models can be explained by forecasting average prices. In previous period an investor cannot open position by average price because it is not known yet while trading session is open. Probably, it is worth to forecast average prices based on the lagged values from the past, for instance, by such autoregression model:

$$p_t = \sum_j a_j p_{t-2} \tag{6}$$

Forecasting average prices for periods t+1 and t+2 using one autoregression model or combined two models, investor could open short or long position in next period by forecasted average price taking into account where the next average price in time period t+2 is forecasted to be.

### 3.2. Portfolio management forced by GARCH-type models

Implementation of evaluated GARCH-type models in forecasting the variance of currency pair in Markowitz’s portfolio showed impossibility to diversify portfolio as it was planned. Using 9 currency pairs with similar expected returns and risks leads to equally weighted portfolio structure. It means that the weights for each currency pair are equal to 0.11 (1/9). To avoid this effect the authors grouped currency pairs in 3 portfolios with 2 currency pairs in each.

Estimated GARCH-type models and profits using the strategy of the second portfolio are shown in the Table 3. For this portfolio the expected investor’s return was calculated as 0.1% (average daily return on analyzed currency pairs in 2004).

Table 3 The comparison of portfolio modified by GARCH

No	Currency	GARCH-type model	Profit, pips	
			w/o weights	with weights
1	USDJPY	AR(3)-EGARCH(1,1)	-1115	-109
	USDCHF	AR(3)-IGARCH(1,1)	2097	1695
	USDCAD	AR(4)-IGARCH(1,1)	-422	-716
2	GBPUSD	AR(4)-IGARCH(1,1)	1197	-1817
	EURUSD	AR(3)-IGARCH(1,1)	3467	3660
	EURJPY	AR(3)-EGARCH(0,1)	-333	1590
3	EURGBP	AR(4)-EGARCH(1,1)	2085	165
	EURCHF	AR(3)-GARCH(1,1)	1556	2587
	AUDUSD	AR(4)-IGARCH(1,1)	-3577	-1414
Total			4956	5312

Analyzing the dynamics of cumulative portfolio return (see Figure 5) that total behavior is improved. There are no large drawdowns as in the first portfolio.

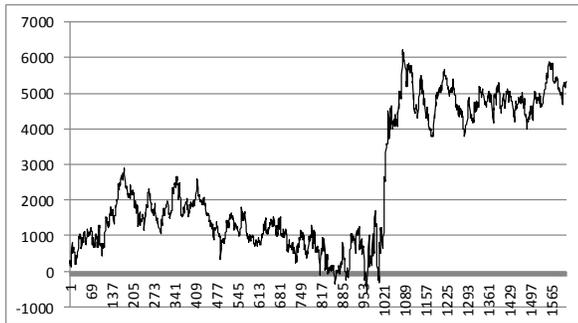


Figure 5 Cumulative portfolio return applying GARCH-type models

Implementation of GARCH-type models in the Markowitz portfolio improved risk management significantly and profitability non-significantly. As it is shown, despite of improvements in the Markowitz portfolio, there is still no possibility for speculative return.

### 3.3. Portfolio management forced by GARCH-type models and moving correlation

For this portfolio the authors also used the expected investor’s return equal to 0.1% and the same GARCH-type models as in the previous portfolio. 3-, 5- and 10-period moving average correlations were calculated between each currency pairs. The total result of applying GARCH-type models and changing-over-time correlation is shown in the Table 4.

Table 4 The impact of moving average correlation on total results

No	Currency	Profit, pips (with weights and using n-period moving average correlation)		
		3-period	5-period	10-period
1	USDJPY	-191	-357	-1115
	USDCHF	1253	1202	2097
	USDCAD	-603	-997	-422
2	GBPUSD	1198	1198	1198
	EURUSD	3468	3468	3468
	EURJPY	-333	-333	-333
3	EURGBP	2085	2085	2085
	EURCHF	1556	1556	1556
	AUDUSD	-3577	-3577	-3577
	Total	4856	4245	4957

Applying GARCH-type models and 3-, 5- and 10-period moving average correlation in the Markowitz portfolio, total result is similar to the previous results. Common dynamics of cumulative portfolio return (see Figure 6) actually repeats the dynamics of the previous portfolio modified only by GARCH-type models.

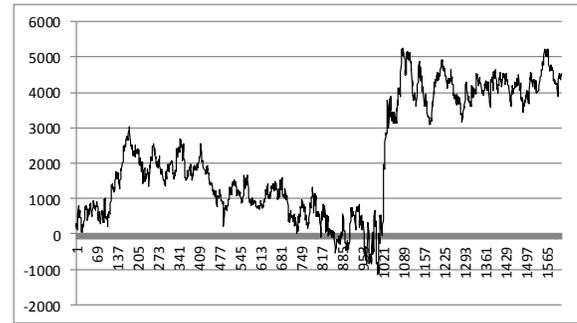


Figure 6 Dynamics of cumulative portfolio return applying GARCH-type models and 10-period moving average correlation

Comparing drawdowns of two portfolios, it seems that the best result belongs to the second portfolio. Using moving correlation in calculations it was noted that values of this correlation is more volatile in USD pairs. In other currency pairs correlation coefficient is more stable or does not have significant impacts on portfolio weights. In any case, in USD pairs the impact of moving average correlation was significant. Probably, this effect can be used in some more specific strategies.

## 4. CONCLUSIONS

1. Portfolio management methods based on interdependences between currency pairs are more effective than so called naive diversification.
2. Implementation of GARCH-type models in Markowitz’s portfolio improves its quality. GARCH-type models can capture clusters in currency pairs’ returns and impact on portfolio structure trying to regulate risks (conditional variance) on impact from previous periods. Such modification can decrease total drawdown in cumulative portfolio return.
3. Moving average correlation shows some interesting facts on interdependences of currency pairs. Significant impact is not discovered, but this effect, possibly, could be used in high frequency data, for instance, in 4-hour, 1-hour or even 15-minutes data.
4. Comparing all results, the authors can conclude that despite on positive impact of implemented modifications in the Markowitz portfolio there are no reasons to use this methodology in automated Internet trading for speculative purposes based on average daily data. Probably, some corrections in limiting losses should be made.

## 5. REFERENCES

[1] V. Akgiray, (1989) Conditional heteroscedasticity in time series of stock returns: evidence and forecasts, **Journal of Business**, 62, 55-80.

- [2] C. Alexander, **Market Models**, John Wiley & Sons, 2003.
- [3] Bank for International Settlements, December, 2010, Triennial Central Bank Survey/ Report on global Foreign exchange market activity in 2010 <http://www.bis.org/publ/rpfx10t.pdf>.
- [4] S. Benartzi, R. Thaler. (2001). Naïve Diversification Strategies in Defined Contribution Saving Plans. **American Economic Review** 91:79–98.
- [5] T. Bollerslev, (1986) Generalised autoregressive conditional heteroskedasticity. **Journal of Econometrics**, 31, 307-327.
- [6] R.F. Engle, (1982) Autoregressive conditional heteroscedasticity with estimates of variance of UK inflation, **Econometrica**, 50, 987-1008.
- [7] R.F. Engle, (2000) Dynamic conditional correlation – a simple class of multivariate GARCH models. PDF version retrieved from <http://weber.ucsd.edu/~mbacci/engle/391.pdf>
- [8] R.F. Engle, C. Mustafa, (1992) Implied ARCH models from option prices, **Journal of Econometrics**, 52, 289-311.
- [9] R.F. Engle, D. Lilien, R. Robins, (1987) Estimation of time varying risk premiums in the term structure, **Econometrica**, 55, 391-408.
- [10] V. Jansons, K. Kozlovskis, N. Lace, (2006). Portfolio Modeling Using the Theory of Copula in Latvian and American Equity Market. **Simulation In Wider Europe**, 19th European Conference on Modelling and Simulation ECMS 2005
- [11] K. Kozlovskis, N. Lace, (2006) Challenges of Decision Making in the Latvian Equity Market // The 10th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2006), 41-44.
- [12] K. Kozlovskis, (2007) Moving Slope as a Long-Term Indicator in Foreign Exchange / K. Kozlovskis // International Scientific Dual-Conference Towards Knowledge Based Economy & Enterprise Management: Diagnostics, Strategy, Effectiveness". 12-13 April, Riga, Latvia. Conference Proceedings. – CD. – 6 pages.
- [13] H.M. Markowitz, 1952. **Portfolio Selection**. *Journal of Finance* 7:77–91.
- [14] H.M. Markowitz, (1991). “Foundations of Portfolio Theory”, **The Journal of Finance**, Volume 46, Issue 2, January, 469 – 477.
- [15] D.B. Nelson, (1991) Conditional heteroskedasticity in asset returns: A new approach. **Econometrica**, 59, 347-370.
- [16] G.W. Schwert, P.J. Seguin, (1990) Heteroscedasticity in stock returns, **Journal of Finance**, 45, 1129-1155.
- [17] A. Weiss, (1984) ARMA models with ARCH errors, **Journal of Time Series Analysis**, 5, 129-143.
- [18] J.M. Zakoian, (1994) Threshold heteroscedastic models, **Journal of Economic Dynamics and Control**, 18(5), 931-955.

# Changing of Topicality of Human Competencies within Companies' Life Cycle

Inga LAPINA and Daira ARAMINA

Riga Technical University, Faculty of Engineering Economics and Management  
1/7 Meza Street, Riga, LV-1007, Latvia

## ABSTRACT

The aim of the paper is to describe the theoretical aspects of company life cycle and model of competencies and present the analysis of the role of human competencies in the context of sustainable development of the company. Research methods such as analysis and synthesis of the relevant scientific literature and logical and comparative analysis are applied as well as the focus group method is used to assess the findings. Based on both the theoretical research and empirical analysis the authors propose their characteristics of the basic groups of competency model, show the level of importance of these groups in different stages of company development and draw conclusions regarding the interconnectedness of organizational culture and human competencies.

**Keywords:** Company Life Cycle, Human Competencies, Company Development, Human Capital, Organizational Culture.

## 1. INTRODUCTION

Human competencies and human capital development issues should be evaluated in the global and local context. In order to make human capital competitive, it is of vital importance to be aware of and be able to overcome the obstacles and seek new possibilities for developing and realizing of human competencies. To this end analysis of issues of developing and realizing human competencies that directly influence labour productivity is required. Hence, company growth issues related to the impact of human competency are studied in two aspects: in terms of socio-economic progress and in terms of factors and circumstances which are objectively necessary for the development and growth of society.

Human capital is comprised of a wide and complex set of competencies of a human being as an element of the society. The set of these qualities includes various intrinsic parameters of an individual. For better analysis and assessment of these parameters it would be useful to start with recognizing their components being classified as:

- those referring to a human being only as a physical individual;
- those referring to a human being as an intellectual individual.

The latter ones may be subdivided into qualities and competencies referring to knowledge, thinking and ability to act in a certain way and a particular area.

Alongside with globalization and new competency-based society developments today, new issues related to sustainable development of a company also arise, such as:

- movement to a new dimension of public life that envisages the human being in the centre of socio-economic mechanism (as a producer and consumer) has created the need to develop new approaches to economic and management theories that are closely related to labour and development of human competencies;

- attention is focused on two socio-economic areas – intellectual sphere and consumption, resulting in the need to assess or reassess the importance of human competency for the economic development and progress of society, because it directly affects the formation of qualified workforce and sustainability of every individual company .

In the context of company development compliance with labour market requirements is also to be taken into consideration. It consists of two components:

- set of competencies needed currently for retaining a post;
- competencies ensuring the necessary qualification changes.

The authors link the aspects mentioned to effective and high quality personnel management at different stages of company life cycle and identification of development trends of personnel competencies in the circumstances where vitally important technological, economic and value changes take place both inside the company and in the external environment.

## 2. MODEL OF COMPETENCIES

Some researchers consider D. McClelland to be the founding father of competency approach. His article in the journal "American Psychologist" in 1973 initiated the development of competency model in USA. D. McClelland believed that traditional examinations and test results cannot predict well enough whether a human is able to perform his job well and that there exist some other ways of how to define competencies ensuring job achievements. He used the concept „competency” to replace a narrower term „skill”, including in it the behavioural aspects and technical abilities. The methodology of behavioural event interviewing and criterion-referenced assessment is also D. McClelland's contribution to the development of competency approach. [13]

One of the most widely used definitions of the concept of competence in the scientific literature published in Latvia states: „competence is a set of knowledge, skills and attitudes that qualifies for completion of a task of a particular kind or level [19]”.

There is no unequivocal agreement about what a competency is or is not. In 1980 D. Clamp defined job competency as „qualities attributed to an individual that manifest themselves as effective or excellent job performance”. He also pointed out that „competencies are indications which are linked by causative reasoning with effective or outstanding job performance.” [8] Extending this definition L.M. Spencer and S.M. Spencer in 1993 described competency as „traits inherent to an individual that are linked by causative reasoning with effective or outstanding job performance or other situations and may be characterized by certain criteria.” [21] They explained that competencies include motives, character traits, self-image, knowledge and skills.

The competency model is a set of responsibilities and tasks that is a part of work process including standards required for this process. In different sources of literature competencies are grouped and structured, creating different competency models

appropriate for each situation. Each competency model is made of individual competence groups, which are defined and described in a greater detail; moreover, each competence may be diagnosed by specifically designed assessment methods. Since each group of competencies covers a wide range of individual skills and abilities, it is possible to assess each professionally active person and determine a person's strengths and progress, stating which component is more advanced, as well as which competencies are dominant to achieve appropriate results in his/her work.[4] Thus, it is possible to determine personality type, stating whether a person is focused on professionalism, knowledge, interpersonal skills, activities, innovation and leadership, or on personality traits. The notion of competence is rather new in Latvia; thus it is still rather difficult to define its structure and practical significance. In fact, human abilities, knowledge and skills are all called competencies. Authors have divided them into basic groups consisting of qualities defined in greater detail (see Tab.1).

Table 1. Characteristics of competency model groups

<b>Professional competencies</b>
Self-organization of a person for intellectual and physical activities to address business and other issues. Presentation, project and process management, as well as quality management skills. Orientation towards goals and objectives. Analytical skills. Skills to operate devices and equipment of varying complexity. Problem solving skills, competence judgements. Knowledge management, awareness of results. Ability to synthesize new and strategic ideas and strategic decision-making skills.
<b>Social and communicative competencies</b>
Cooperating with others, developing new plans, setting tasks and targets. Ability to express ideas in a well-structured and logical manner. Ability to influence and take responsibility, to motivate, delegate to others and maintain emotional stability in difficult situations. Ability and readiness to act fairly. Integrity, ethical behaviour, objectivity, empathy, autonomy. Ability to instantly integrate various items and solve problems. Identify and tackle conflict situations.
<b>Personal and responsibility competencies</b>
One's ability to self-assess. Willingness to develop oneself as a personality, develop one's talent. Take responsibility for the quality of the work done. Being able to positively affect the team. Ability to resolve conflicts. Self-control and openness to change. Productive work, adequate behaviour, motivation, general intelligence, operational intelligence, mental flexibility, complex thinking skills, concentration, mindfulness and alertness, good memory.
<b>Innovative, learning and leadership competencies</b>
Individual's self-organization for active communication with others. Ability to combine emotions, motivation, abilities and experience. An individual seeking and accepting opportunities to learn and seek advice. Takes responsibility for his/her own training or training of subordinates, as well as understands the need for continuing education and provides such opportunities to others. Actively participates in mentoring and training new employees. Able to make decisions, willing to support innovation. Consistency, persistence and purposeful activity. Planning and management skills, creativity.

Lately in Latvia the concept „competency” has been more widely used in some other aspects as well. For example, „competency” concept is included in ISO standards. The standard states „Performers influencing product quality should

be competent on the basis of adequate education, training, skills and experience”. [18]

Competency model in a company may be used to analyze every employee's individual performance including it into certain competency and development models, thus becoming aware of strengths and weaknesses of the company that would, in turn, help to effectively position and employ the existing human resources.

It may be concluded that not only different definitions of the term „competency” but also different groupings of competencies have been used in literature and in previously done researches. For example, M.A. Armstrong defines competencies taking into consideration two aspects: the action competencies (soft skills) that are characterized as an opinion on how to act to perform one's job well and the technical and functional competencies (hard skills) that individuals have to know and exercise to perform their jobs in a proper way. [3] In its essence a competency model is a list of the competencies required for a particular position. Working with the competency models provides the possibility to think about the criteria for how to identify the competencies and at what level of a particular position the relevant competency should represent itself.

When exploring the quality of education, competencies are grouped taking into consideration three criteria:

- to know and understand (theoretical knowledge, ability to understand);
- to know how to act (practical skills and action skills and their implementation in practice);
- to know how to live (values, attitudes, responsibility).

Recently research papers have highlighted general skills and employment core competencies. [22]

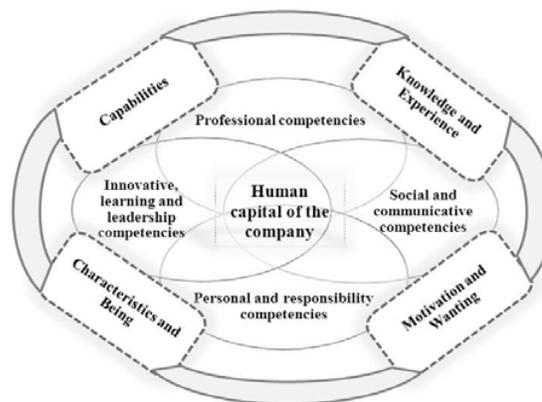


Figure 1. Basic groups of competency model and the related components [created by authors]

Summarizing the findings of scientific literature review the authors classify the competencies in the context of human capital and company development into four groups (see Fig.1.). Wanting, Being, Capable to, and Knowledge are all important for competence/competent behaviour [7]. Figure 1 also shows the components that together form an individual's competency. Analyzing a number of scientific literature sources, it may be concluded that the concept of competence has undergone definite development stages. Fast changes in public life and education make us look at the concept of „competence” from a different point of view, extending its contents and adapting the nature of competence to contemporary requirements.

If before, there existed a rather narrow opinion that the notion of competence implies a set of responsibilities and rights of an official stipulated by law and regulations in a specified field, lately the understanding of this notion has become much broader. For example, considering the skills for a specific real-life standard situation to be insufficient, entrepreneurs ask for competency where not only individual skills of every person acquired through training are included, but also social behaviour, team work ability, initiative and readiness to take risks, as well as combination of knowledge and subjective qualities, which managing directors of the companies often call „life skills”.

If an individual performing a job has the education, knowledge and relevant skills required, he/she may be considered a qualified person. If an individual demonstrates an ability to achieve the desired results he/she may be considered to be **competent**. Thus, qualified and competent are not synonyms. Qualified personnel sometimes may be unable to deliver the results required for a job as he/she may have knowledge and skills, but not apply them properly. Based on this, competency should be understood as a proven ability to apply knowledge and skills in practice.

Summarizing all the above the authors come to conclusion that the principles for assessing competencies and employee “professionalism” or “qualification” have considerably changed. If before it was assessed on the basis of the specific education and qualification level that encompassed mostly technological knowledge and skills related to a particular speciality, then now competency is treated as an ability and skills to solve all issues arising in production or commercial activities, not dealing only with technological issues.

Thus, the authors conclude that competence is not only the knowledge learned, acquired and accumulated, skills and attitudes, but also the ability to apply and develop those further. Competence is linked to individual’s abilities to learn, communicate and cooperate in a changing environment. Based on this approach and considering the rapidly changing environment it can be inferred that competence has in fact become one of the key human characteristic features and is closely related to company’s life cycle. Therefore, the authors deem it necessary to analyze changes of topicality of human competencies in the process of formation and development of a company.

### 3. COMPANY LIFE CYCLE

In economics and management science the „life cycle approach” is widely used for description and analysis of process development. The concept life cycle would ordinarily include the description of a process, product, system or another existing structure in a dynamic development.

Most widely known is the product life cycle approach by Ph. Kotler describing the changes in the volume of products sold. [14] I. Mootee, the business strategist and innovation specialist describes the life cycle of a brand by analyzing “the brand power” changes in time. [17] J. Gido and J. Clements analyze the project lifecycle – changes of project efforts over a time span. [11] G. Moore describes the technology adoption life cycle according to types of personalities. [16]

In his publications A.M. Ahmed mentions the business life stage model analyzing company success (its changes and complexity) versus time and efforts. [2] L. Greiner divided enterprise growth into five phases of starting-up, surviving, growing by authorization, taking-off and full-grown, forming the five-phase model. [12]

E. Gaile-Sarkane analyzes particularities of different factors and their role in consumer behaviour [9], that forms and influences product consumption and company life cycle features in different countries.

According to J. Bratton, the evaluation of human resources is related to the procedures and processes which are used in order to evaluate the results of human resource management practices in the enterprise. [5]

In fact, we may say that, just like a human being, any company has its own life, life span and life cycle with different stages, starting with foundation and ending with closing the company. Between these two extremes many different phases of development still exist – childhood, growing up, zenith, stability and aging. Conducting company performance analysis it is found that the average life cycle of a company is 20 years. Companies develop, sometimes are sold, change ownership, merge, cooperate, integrate, restructure and often enough liquidate. Business is a process of perpetual changes.

Human competency is a part of human capital (see Fig.2). The total potential of the company consists of Human capital, Social capital, Innovation capital, Spiritual capital and Financial capital [15], where the significance of every type of capital varies in different stages of the company life cycle.

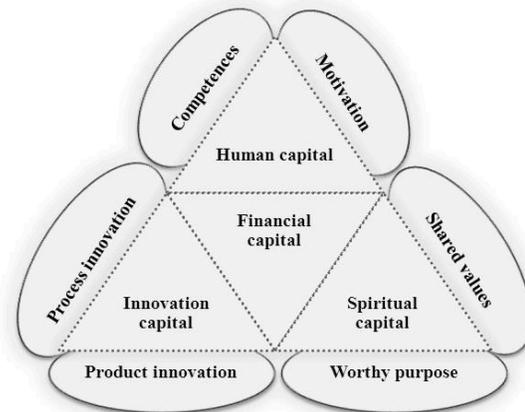


Figure 2. Structure of company capital [adapted from 15]

Having analyzed the different approaches to the life cycle theory it is unequivocal that the life cycle theory may be applied to describe different processes and systems defining precisely the variable parameters in time. It should be noted that life cycle is traditionally depicted by a bell shape curve. However, there are some observations testifying that the character of the curve may vary and correspond to several functions. It may be said that the life cycle is changes of one or several parameters over time. Therefore, the authors present a general and simplified life cycle formula, see Eq. (1):

$$LC = fQ_{1-n}(t), \quad (1)$$

where:

LC – life cycle

$Q_{1-n} \dots (t)$  – changes of the scope of a certain parameter over time.

In this case the parameters could be, for example, macroenvironment, industry, resources, products, etc. Assuming that the scope of performance is directly dependant on organization’s effectiveness that is in turn defined by rational use of resources (or efficiency), and knowing that one of the

most significant resources in a company is employees, the formula mentioned above can be changed into, see Eq. (2):

$$CLC = fHC(t) \rightarrow \max, \quad (2)$$

where:

CLC – company life cycle

HC(t) – changes of human competencies over time.

Different favourable and unfavourable company performance processes may develop during the company life cycle. The length of this cycle is influenced by how well the founders and managers of a company are able to analyze and forecast the processes taking place.

E.H. Schein analyzes the development of the organizational culture and points out three development stages: founding and initial development, average age, maturity and decline. [20] The criteria for this classification are manager's and subordinates' roles in the process of value formation. During the first stage it is mostly the management that influences the organizational culture of the company, while in the second stage, which is characterized by strong organizational culture and a new manager who is not linked to the previous stage of the development process, mutual impact of both the manager and subordinates on the consolidation process of basic values becomes pronounced.

One of the specific features of Latvian business activities, when compared to Western European countries, is that during the first years of independence very many people who were involved in establishing and managing companies without having acquired professional education in management relied on their own intuition and everyday experience. Part of them were forced to leave business due to mistakes made, but some companies reached great results and have become leaders in their industry. Nowadays due to competition the managers of the successful companies more than ever before need to understand the role and principles of their organization. This understanding may be reached by analyzing the organizational culture.

Scholars of organizational culture are of the opinion that basic values are formed, internalized and consolidated during the company development process or in any of the stages of its life cycle. K.S. Cameron and R.E Quinn tackle the changes of the organizational culture in close connection with its growth. [6]

Knowing that changes of human competencies are associated with the changes of company culture the authors of this research offer representation of organisational changes in a company (see Fig.3) directly reflecting changes of human resources and culture at different stages of company life cycle.

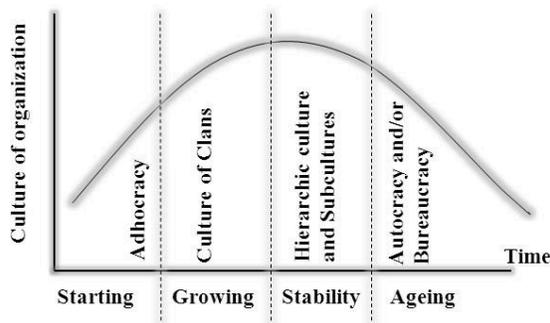


Figure 3. Organizational Culture in different Companies' Life Cycle stages [created by authors]

In the earliest stages of development, organizations tend to be dominated by “adhocratic” culture, while during the

development stage it is supplemented by clan culture. Increasing the number of workforce and extending the scope of activities creates the need to form substructures and increase formal control to create unified performance standards, thus intensifying the significance of the hierarchic culture. Later the culture is gradually supplemented by elements of market culture, enabling the company to increase its competitiveness. The framework of the organizational culture diagnostics introduced by the above mentioned scholars „photographs the situation”, but does not show how the model identified has been formed and what role has management played in the formation of the organizational culture model.

During the stability and ageing stages the organizational structure frequently becomes dysfunctional. To avoid this, managers should manage organizational culture changes paying particular attention to the development of the most important employee competencies for a given stage of company life cycle. When characterizing manager performance efficiency in different company life cycle stages, the American scholar I.Adidzes presents the concept of ten-staged organizational life cycle which, from the authors' point view, allows for a more complete understanding of the development process of organizational culture and its management possibilities in Latvia's context than does the approach of scholars mentioned earlier. I. Adidzes divided the enterprise lifecycle into gestating, nestling, step-learning, adolescence, adult, stabilizing, peering, early bureaucracy, bureaucracy and degenerating, forming the ten-phase model. [1]

The authors are of opinion that I. Adidzes approach makes it possible to analyze the human resource competency changes corresponding to the different company life cycle stages. Thus, for example, when the approach to company management changes (referring to flexibility), the human competencies will change as well.

#### 4. CHANGES IN TOPICALITY OF COMPETENCIES

On the one hand, employee competency is a prerequisite for a company's development – it determines the company's ability to be innovative and consequently, the development and future prospects of the company. On the other hand, competency is the result of innovation because employee qualification is directly affected by the changes due to technological, organizational and social innovations.

If human resources are not adequately trained, lack of competencies and qualification may become a bottleneck for the company when introducing the innovation processes, mainly in terms of technological aspects.

Far-sighted and systematic development of knowledge-based society and human competency changes requires active adjustments in the company with the aim to ensure the competitive edge of the company. Human resources become the decisive factor for the company development and innovation process. [10]

Which factors determine employee competence requirements in different stages of company development? These are changing market conditions, such as:

- increasing market globalization, high competition;
- short product life cycle, increasing rate of innovation;
- increasing market segmentation and consumer orientation;
- increasing quality requirements;
- increasing complexity, dynamism and reliance on technologies;
- ageing of knowledge and necessity to acquire new knowledge;

- high requirements on specific issues and employee's methodological and social knowledge;
- language skills;
- higher degree of responsibility, self-control, etc.

All these factors are a great challenge for innovation strategies in companies that, in turn, determine the requirements for employee competencies.

Human activities take various forms depending on internal and external conditions affecting the company. Moreover, organization may find itself in different stages of development, and so the competence of people in the company should be consistent with the life stage of the given organization.

Focus group method was used to validate results of the authors' research on human competences and company life cycle. A discussion of this focus group, that involved a number of entrepreneurs, revealed that most of them cannot precisely define topicality of specific competencies in a dynamically changing environment where new challenges and problems to be solved appear constantly. They mostly make routine decisions and pay insufficient attention to changes in human competencies. According to the authors, the reason for the situation is a competence gap which can be defined as the difference between individual (managerial) and institutional (organizational) interests. When both of those types of interest are not coherent, a conflict arises among individuals and the company. Managers and staff often do not possess the knowledge required by the company because management do not systematically perform the diagnosis of their company and, as a result, do not know which stage of life cycle the company has entered. This lack of knowledge often leads to company failure.

Investigating this problem with focus group method the authors have concluded that professional competencies and personal and responsibility competencies are crucial at the beginning stage of company life cycle (see Fig.4), when companies are open to changes and are flexible and able to manoeuvre, but at the same time are comparatively more difficult to control.

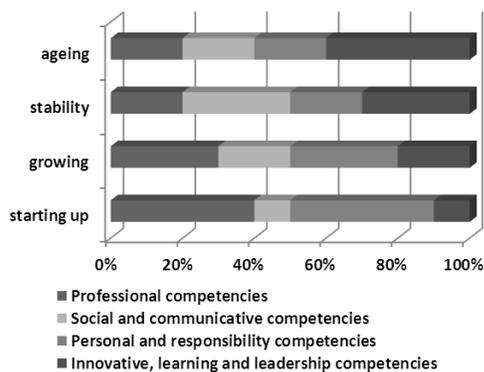


Figure 4. The level of importance of competencies in different stages of company life cycle [created by authors]

A new organization is able to change in a relatively simple way (similar to the desire to develop oneself as a personality, to perfect one's talent and to take responsibility for the quality of performance), however, its actions may be unexpected due to lack of control.

During the company development stage professional competencies and personal and responsibility competencies continue to retain their leadership role. At this moment the company enjoys the advantages of the starting up and growing stages which means elasticity and controllability. That is,

organization finds itself in a superior or highest development stage when innovative, learning and leadership competencies and social and communicative competencies, are not yet relevant and necessary. The main characteristic features of this stage are as follows:

- functionally well-established system and organizational structure;
- result-oriented performance, focussing on customer needs;
- company results are considered good.

With the next life cycle stages being stability and ageing, it is important for every company to reach this stage and be able to stay here as long as possible.

During the stability stage social and communicative as well as innovative, learning and leadership competencies occupy the leading role leaving professional, personal and responsibility competencies behind. Company performance is controllable, but it lacks flexibility and is not so apt to changes and adaptations. This stage may be described as constant, permanent, established and stable process that can not cause great deviations in company's performance.

The characteristic features of the company ageing stage may be self-satisfaction, lack of interest and low job motivation. In this phase company may be described as having great money resources, but lack of good proposals how to best use them. It is the beginning stage for bureaucracy development and sharp decline, so innovative, training and leadership competencies become most important. It is the time when an individual's self-organization for active performance together with others impact future viability of the company. An individual (company) seeks for and accepts further training opportunities, consulting possibilities and chances for further development, understands the need for and ensures the future possibilities of continuing education, as well as takes active part in preparing and training new employees (the company seeks innovative solutions). Consistency, perseverance and purposefulness may bring the company back to life.

The small and medium-sized companies of Latvia should not yet be afraid of ageing problems because, notwithstanding the hard work, they are still in the growing stage. A long period of time will pass while Latvia as a state and its companies would reach their youth and zenith. The key prerequisite is that company founders and managers constantly follow which stage of development their company has reached and take the right decisions at the right moment.

## 5. CONCLUSIONS

In contemporary circumstances it is most important that all employee competency elements would not exist separated from each other, but would supplement each other and inheritance of knowledge and experience would take place so that an employee, finding himself in a new stage of company life cycle, would be able to feel his/her real professional possibilities and qualification level and make competent decisions for the benefit of company development.

There are many objective factors that affect human development, formation of a free intellectual personality and high quality workforce. The authors conclude that human capital formation in a company depends to a great extent on the stage of company life cycle and increase in efficiency of its systems and processes. In the context of competency-based, sustainable company development, the competencies, according to the grouping developed by the authors, can be seen in the following way:

- as the key workforce formation factor;

- as accumulated intellectual potential generating favourable conditions for the scientific and technological progress, increase of labour productivity and transformation of innovations for production development;
- as a factor for improving organizational culture that contributes to improved production relations through coordination of inquiries and actions and thus both enhances the production process and ensures scientific organization and management of company performance;
- as a factor grounded in intentional, creative and free from suspicions environment that creates conditions for interest harmonization in the company which manifests as a necessity for everyone to be aware of his/her place in the company;
- as characteristics of workforce quality – everything that makes an individual a personality and is affected by education as a specific work process.

The role of human competencies in the company development may be assessed by other criteria as well, which the authors propose to group the following way:

- cognitive (ensuring knowledge and competencies);
- methodological (skills to apply knowledge and competencies into practice);
- practical (acquisition of skills and competencies to practically organize and manage processes, i.e., professional competency);
- ideological (identifying the awareness level of human behaviour which directly links education with community interests and aims);
- social (formation of personality, its mental and moral development, personality socialization);
- economic (formation of qualified and competent workforce).

Since any process analysis takes into consideration different aspects and uses various scientific methods, the tendency of integration and synthesis of basic and additional knowledge has become the dominant feature in human capital and human competency development in the contemporary world. Many-sided vision of the process and its examination in conjunction with surrounding „natural” and „artificial” environment is of great significance. If previously branches of science were formed based on object features – i.e., when new areas and activity types arise in the cognitive process, the contemporary science is shifting from an object to problem orientation where new knowledge is created through raising specific significant theoretical or practical multidisciplinary problem. It means that by accelerating the science differentiation the scientific and technological process has become an integrating factor in the field of fundamental and applied sciences.

In conclusion the authors agree with the world experience saying that science cannot contribute to production without education and workforce competencies. In all industrialized countries educating of human resources with the aim to enhance human competency is not only the means to reach other aims but an independent aim in itself. Enhancement of education and human competencies reaching an adequate level to meet the requirements of post-industrial development of the society is an important prerequisite for any company development.

## 6. REFERENCES

- [1] I. Adidzes, **Managing Corporate Lifecycles: how to get and stay at the top**, Santa Barbara: The Adidzes Institute Publishing, 2004.
- [2] A.M. Ahmed, "Virtual integrated performance measurement", **International Journal of Quality & Reliability Management**, Vol.19, No.4, 2002, pp.414–441.
- [3] M. Armstrong, **Handbook of Human Resource Management Practice**, 10 th, London: Kogan Page, 2006.
- [4] I. Boitmane, **Personāla atlase un novērtēšana**, Rīga: SIA Lietišķās informācijas dienests, 2006.
- [5] J. Bratton, J. Gold, **Human resource management: Theory and Practice**, 4th edition, New York: Plagrove MacMillan, 2007.
- [6] K.S. Cameron, R.E. Quinn, **Diagnosing and changing organizational culture**, MA: Addison-Wesley, 1999.
- [7] M.P. Driessen, P.S. Zwart, "The Entrepreneur Scan Measuring Characteristics and Traits of Entrepreneurs", **Entrepreneur Scan Research articles**, 2010, <http://www.entrepreneurscan.co.uk/background/research-articles/> [electronic source]
- [8] D.D. Dubois, W.J. Rothwell, **Competency-Based Human Resource Management**, Palo Alto, CA: Davies-Black, 2004.
- [9] E. Gaile-Sarkane, I. Andersone, "Influence of Factors on Consumer Behavior", **Business and Management** 2008, 5th International Scientific Conference, Vilnius: Technika, 2008, pp.331-337.
- [10] E. Gaile-Sarkane, I. Lapina, "Knowledge Society: Challenges for Higher Education in the Baltic States", **KSI Transaction on Knowledge Society**, Volume 3, 2<sup>nd</sup> International Science Conference "Knowledge Society", Bulgaria: Knowledge Society Institute, 2009. – pp.31-34.
- [11] J. Gido, J. Clements, **Successful Project Management**, 4<sup>th</sup> edition, South-Western College Pub, 2008.
- [12] L. Greiner, "Evolution and Revolution as Organizations Grow," **Harvard Business Review**, Vol. 76, No. 3, 1998, pp. 55-67.
- [13] S. Horton, **The Competency Movement. Competency Management in the Public Sector**, Amsterdam: IOS Press, 2002.
- [14] Ph. Kotler, **Marketing management**, Prentice Hall, 2002.
- [15] L.M. Miller, "Organization Life Cycles: The Creation and Destruction of Wealth", Paper adapted from **Barbarians to Bureaucrats: Corporate Life Cycle Strategies**, <http://www.lmmiller.com/papers-presentations.php> [electronic source]
- [16] G.A. Moore, **Crossing the Chasm**, New York: Harper Business Publishing, 2002.
- [17] I. Mootee **Understanding the Brand Life Cycle Model**, September 10, 2007, <http://www.mootee.typepad.com> [electronic source]
- [18] Quality management systems - Requirements. **ISO 9001 International Standard**, 3<sup>rd</sup> edition, 2000, Reference number ISO 9001:2000(E), p.6
- [19] A. Rauhvargers, "Establishing the European Higher Education Area – progress made in the period of 2005-2007", **Humanities and social Sciences in Latvia**, Vol.3. (52), Riga: LU, 2007, pp.5-35
- [20] E.H. Schein, **Organizational Culture and Leadership**, The Jossey-Bass Business & Management Series, 4th edition, USA: Jossey-Bass, 2010.
- [21] L.M. Spencer, S.M. Spencer, **Competence at work: Models for Superior Performance**, New York: John Wiley & Sons, 1993.
- [22] Tuning Educational Structure in Europe. Pilot Project Final Report, Phase 2, **TUNING Educational Structures in Europe**, 2005, <http://www.tuning.unideusto.org/tuningeu/> [electronic source]

# **Nurturing Opportunity Identification for Business Sophistication in a Cross-disciplinary Study Environment**

**Karine OGANISJANA, Tatjana KOKE**

**Faculty of Education, Psychology and Art, University of Latvia  
Jurmālas gatve 74/76, Rīga, LV 1083, Latvia**

## **ABSTRACT**

Opportunity identification is the key element of the entrepreneurial process; therefore the issue of developing this skill in students is a crucial task in contemporary European education which has recognized entrepreneurship as one of the lifelong learning key competences. The earlier opportunity identification becomes a habitual way of thinking and behavior across a broad range of contexts, the more likely that entrepreneurial disposition will steadily reside in students. In order to nurture opportunity identification in students for making them able to organize sophisticated businesses in the future, certain demands ought to be put forward as well to the teacher – the person who is to promote these qualities in their students. The paper reflects some findings of a research conducted within the frameworks of a workplace learning project for the teachers of one of Riga secondary schools (Latvia). The main goal of the project was to teach the teachers to identify hidden inner links between apparently unrelated things, phenomena and events within 10<sup>th</sup> grade study curriculum and connect them together and create new opportunities. The creation and solution of cross-disciplinary tasks were the means for achieving this goal.

**Keywords:** opportunity identification, cross-disciplinary teaching and learning, the development of students' entrepreneurship.

## **1. INTRODUCTION**

Opportunity identification is widely viewed as one of the key elements of entrepreneurship education; according to Howard Stevenson: “entrepreneurship is the pursuit of opportunities beyond the resources you currently control” [32]. Traditional teaching methods such as lectures, literature reviews and examinations do not activate students' entrepreneurship [11, 30, 14, 13]. It is argued that the most effective way to promote students' entrepreneurial dispositions and mindsets is to “push” students into entrepreneurship through the structuring of learning like an entrepreneurial process [16, 20, 33, 15]. Then students actively interact with the environment and different life situations based on their knowledge, skills, personal needs, desires making their learning vital [18]

and pass through all the stages of the entrepreneurial process started from opportunity identification or creation till its realization into new values [21].

It is argued that the development of students' entrepreneurship shouldn't be considered only in the context of learning economics, management, business, etc, which are study disciplines traditionally connected with entrepreneurship, but as stated by Allan Gibb, entrepreneurship should be taken out of the “locker room of economics”, and based “within a wider interdisciplinary context with a pluralistic and diffused view of society” [11]. As life itself is not mono-disciplinary, the idea of developing entrepreneurship in cross-disciplinary study environment which is maximally approximated to the real life is the prerequisite of entrepreneurship promoting education which is embedded across and within different subjects. Hannon argues that the codification of knowledge into distinct “subjects” creates challenges for cross-disciplinary notions of entrepreneurship as a state of being, or as a process of change or development. When subjects are translated into formalized courses for teaching, they are often “full” of subject “content”. Enterprise or entrepreneurship outcomes largely remain peripheral [12]. Therefore training for entrepreneurship by necessity must actively deal with the multiplicity of becomings, which is life, by sensitive conversations with local situations [15]. For that it is essential to build projects and programs across disciplines [36] and make students members of cross-curricula teams [27].

Therefore the project analyzed in this paper intended to give a boost both to teachers and students for starting certain thinking and behavioral habits – to recognize inner causal links and connections between things and phenomena which couldn't be noticed at first sight or even expected, if based on traditional way of teaching and learning of physics, music, mathematics, biology, history, literature, chemistry, geography, languages, etc. Both quantitative and qualitative data were acquired in the course of the reflection and survey of the teachers and their students in different stages of the project. “The model of opportunity identification and creation in cross-disciplinary teaching and learning environment” was elaborated based on the theoretical analysis of literature and on the outcomes of the quantitative data processing as well as of the content analysis of the text made of the respondents' comments.

## **2. OPPORTUNITY IDENTIFICATION AND CROSS-DISCIPLINARY TEACHING AND LEARNING**

The analysis of different opportunity theories reveals the lack of conceptual clarity and the inconsistency in defining the key constructs of opportunity, as even the origination of opportunities is explained in different ways. Scholars argue that opportunities are:

- recognized [2, 7, 28];
- perceived [31];
- discovered [17, 35];
- created [25, 26];
- identified [5, 6, 19, 22];
- developed [1, 24].

Despite the variety of the conceptual approaches, they are somehow complementary as they lay emphasis on different aspects, stages and factors of opportunity, started from its emergence till its formation and development. However there are certain contradictions. While the proponents of opportunity recognition argue that opportunities “exist out there” and the role of entrepreneurs is to recognize them [7], the opponents of opportunity recognition criticize this approach calling it as misleading, since it implies that opportunities are “out there” waiting to be found [1]; thus it assigns opportunity an instantaneous character [24]. On the opposite to “opportunity recognition” end are those who are for “opportunity creation”. Sarasvathy et al. consider that opportunities do not pre-exist –either to be recognized or to be discovered. Instead they get created as the residual of a process [25]. Gartner et al. also are for the active behavioral approach to creating opportunities; they argue that opportunities are the result of what individuals do, rather than the result of what they see [10]. Thus, as stated by Shackle individuals are the only source of opportunities as through their imagination they can create opportunities from almost nothing by using their mind [26] and acting upon them.

Concerning the “perception of opportunities”, there is a point of view that the procedure of perceiving opportunities must be comprehensive enough to serve as a cognitive objective for the entrepreneur [31]. However, researches show that initial perception of opportunities is often rudimentary, and most opportunities are developed – both before and after venture foundation. This provides evidence for the conceptualization of “opportunity” in terms of “opportunity development” [24]. Long & McMullan describe opportunity identification as a creative structuring process [19].

The authors share the standpoint that “opportunity identification is a more inclusive term that encompasses both potential opportunities already existing in the environment and opportunities that are created by entrepreneurs” [5, p. 367]. So, opportunity identification concerns both external – objective environmental and internal – subjective human factors. Opportunity identification is a cognitive task [29] which is related to education, work experience and entrepreneurial

experience [4]. For entrepreneurs it’s very important to creatively interpret the external environment and relate the opportunities identified in it to their own knowledge corridor [6]. Prior knowledge plays a significant role in the discovery perspective of opportunities [24], as people notice information that is related to their existing knowledge [9]. Moreover, in order to be useful, new information often needs to be complemented with prior knowledge [29]. As prior knowledge of people is inevitably different, according to Stefan Sanz –Velasco that is the main reason why no two individuals perceive exactly the same opportunity [24]; each person’s prior knowledge enables the person, but not others, to recognize certain opportunities [34].

Robert A. Baron has complemented prior knowledge and experience with two more factors: 1) an active search for opportunities and 2) alertness to opportunities, which is the capacity to recognize them when they emerge; these three factors together play a crucial role in opportunity recognition [2].

When working with the teachers of 18 different study disciplines in the cross-disciplinary study environment within the workplace learning project in the above mentioned Riga secondary school, new opportunities were recognized, identified, created and developed by applying procedures from one area of knowledge to another, giving rise to novel associations and these associations enabled them to form the basis of creative ideas [23]. When the teachers were engaged in analyzing problems from one study discipline through others and finding hidden links between apparently unrelated things, phenomena and events and connected them together and created new vision and facets of the reality, it broadened their perception, interpretation, and understanding of environmental forces which according to Dutta and Crossan make the basis for creativity and opportunities not only to be identified but also to be developed and evaluated, that is to be enacted [8].

When a problem from literature was considered through physics, sports, health and chemistry; a problem from biology - through music, handicraft, design and history or a story from mythology – through geography, logics, mathematics and biology in a non-traditional way, it gave rise to the creation of cross-disciplinary problems, which already were new opportunities created by the teachers and aimed at “pushing” studies closer to real life situations and at making all the teaching and learning more attractive and full of discoveries. This approach to the identification and creation of opportunities is justified first of all by a method well known in history and popularized by Thomas Edison; its main idea was to seek to “discover” an invention by combining two items at a time [6]. A few decades later Edward de Bono used “Random entry” technique of lateral thinking which is similar to the above described Edisonian approach to innovation aimed at the creation of new ideas for solving practical problems [3]. However the technique of the identification and creation of new opportunities used in this project had a principal difference from the ones

offered by Edison and de Bono. While looking at a problem of one discipline through the prism of other disciplines, the character of links and interconnections discovered, despite being unexpected and non-traditional, yet were not of “random” but, on the contrary, of “deeply causal” character. A special emphasis was placed on the generation of ideas by seeking links between problems across study disciplines which are classified within different science groups. This served as a ground for the elaboration of “The model of opportunity identification and creation in cross-disciplinary teaching and learning environment” (see fig. 2) which will be considered in the next chapter.

### 3. THE COURSE OF THE RESEARCH

The project was realized and the research was conducted in three main stages:

1. The introduction of the idea and analysis of the essence of cross-disciplinary studies to the teachers who participated in the project and their training;
2. The creation of cross-disciplinary problems by the teachers in groups with their colleagues who teach other disciplines;
3. The solution and analysis of the teachers’ cross-disciplinary problems together with their students.

#### The first stage of the project

It consisted of six phases:

a) While the teachers were presented theory and practice of cross-disciplinarity, special attention was paid to the non-traditional links between parts of the problems based on disciplines from different science groups; it was as well accompanied with the comments on what skills and qualities they may develop.

b) It was analysed and ascertained that ideas for cross-disciplinary problems can be found absolutely everywhere in real life situations and the main problem here is that most people are not used or trained to recognize them.

c) The teachers were offered to create their own ideas for cross-disciplinary problems. For that they were supposed to:

- write down five things, phenomena or events which attracted their attention that day;
- decide in what way these five items could be integrated with the themes of the lessons in study disciplines which they teach;
- relate these themes from their disciplines and the problems behind them to three and more other study disciplines – the farther from theirs the better.

d) The teachers were offered to analyse cross-disciplinary problems which had been created using statistical information taken from the “Worldometers: World statistics updated in real time” [37].

e) The teachers formed three groups and trained to create their own cross-disciplinary problems by using data from the “World statistics updated in real time” [37].

#### The findings of the research conducted in the first stage of the project

The teachers who participated in the project were surveyed with the aim of making a comparative analysis of the advantages and disadvantages of, on the one hand, traditional studies in which the study content is divided into separate disciplines and, on the other hand, of cross-disciplinary studies. The quantitative aspects of the research are shown in table 1.

Table 1. The results of the survey of the teachers on the advantages and disadvantages of the traditional separate-subject studies and the cross-disciplinary studies

Type of studies	+ / -	Aspects	No. of respond. (%)
Traditional separate-subject studies	Advantage	Easy to teach	38
		Easy to perceive for students	19
	Disadvantage	Low potential for creativity	62
		Far from real life context	43
Cross-disciplinary studies	Advantage	Ample opportunities for creativity	67
		Studies are very close to the real life	62
	Disadvantage	Difficult to create cross-disciplinary problems by a single teacher alone	43
		For students may be difficult to comprehend such problems	24

Table 1 shows that despite the difficulties of providing cross-disciplinary studies, it is closer to the real life and has greater potential for creativity.

In the course of the qualitative content analysis of the text made of the teachers’ comments, a few significant aspects of the cross-disciplinary studies which concern opportunity identification emerged out of the teachers’ experience. Having analysed and integrated the results of the quantitative and qualitative researches, the findings of the first stage were finalized as follows.

1. While traditional studies make students concentrate their attention on separate disciplines, cross-disciplinary studies encompass all of them and connect them together as a whole, enabling students to discover new dimensions for thought and activities, perspectives and possibilities for unexpected solutions of real life problems.

2. Cross-disciplinary studies require a special preparation of teachers and a new quality of cooperation among teachers of different study disciplines, as it’s not real for a single teacher to be competent in all areas in order to create valuable cross-disciplinary problems or solve them.

### The second stage of the project

It was aimed at the team work of the teachers who split up into three groups:

Group 1 consisted of six teachers of different languages and a teacher of biology;

Group 2 was formed by teachers of mathematics, handicraft, native language, sports, physics, music and social sciences, one teacher from each discipline.

Group 3 was represented by teachers of mathematics, native language, foreign languages, history and philosophy, economics, geography and dancing.

The authors as the leaders of the above mentioned project, asked the teachers to form mixed groups from teachers of diverse disciplines. However it was the teachers' free choice to get grouped in the way they see it. Each group was to create their own cross-disciplinary problem, based on the topic chosen by them during a two-week time period. In the end of this stage, all the three groups were supposed to demonstrate their problems to the other participants of the project and analyse them together.

### The findings of the research conducted in the second stage of the project

This stage of the project revealed the hindrances which may occur in the process of the active inclusion of cross-disciplinary studies into educational practice.

1. Those teachers, who had worked in school for a long time teaching one or two disciplines, especially from the group of human sciences (group 1), were used to think and act within the frameworks of their study disciplines to such an extent, that for them it caused certain difficulties to "step out" across the boundaries and cooperate with those colleagues who were from other science group.

2. Instead of creating really cross-disciplinary problems in close cooperation with their colleagues, the teachers of group 1 tried to substitute the idea of cross-disciplinary problems by the compilation of independent questions and tasks from the themes which they were teaching in school in that period.

3. While analysing the cross-disciplinary problems, all the participants of the project agreed that the more captivating and many faceted were those problems, which had been created by the groups of teachers of disciplines traditionally recognized as "remote" from each other (groups 2 and 3). This conclusion was confirmed as well in the third stage of the project on the basis of the survey of the students who had been offered to solve these problems and evaluate their studies during the project. As for the teachers, in their reflection on what they had acquired while learning to solve and create cross-disciplinary problems, among the other acquisitions, they emphasized as well that they had become more perceptive towards things, seeing hidden aspects of them which they wouldn't have paid attention to before.

The third stage of the project was the work with the students who solved and analysed the cross-disciplinary problems created in the second stage. It lasted three days, one day for each cross-disciplinary problem created by each group of teachers. The reflection of the students was organized at the end of each day, as the problems were very different in their content, character and the means which were necessary for their solution. The course of the research was filmed for the further analysis and for being used as a methodical material for cross-disciplinary teaching and learning.

### The findings of the third stage of the research

The students' reflection first of all confirmed a high level of positive emotional perception of the cross-disciplinary learning by the students during all the three days. However there was one important tendency which emerged after having processed the quantitative data (see fig. 1 and 2).

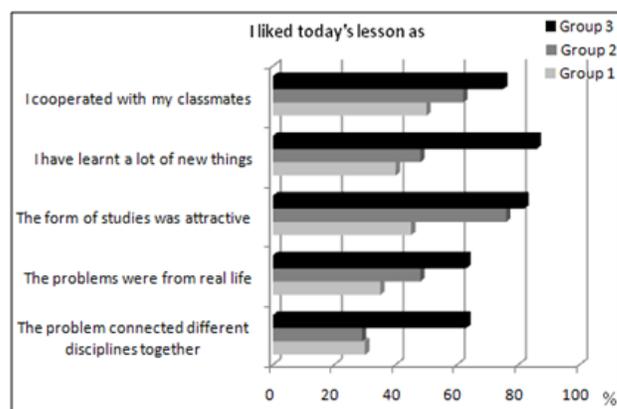


Fig. 1 The students' reflection on why they liked that day's lesson

The diagram in figure 1 shows that the students liked better and evaluated higher the cross-disciplinary lessons provided by those groups of teachers who represent more diversity of disciplines (groups 3 and 2).

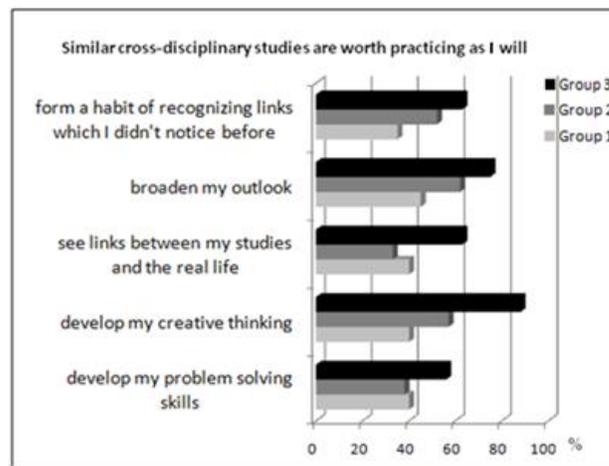


Fig. 2 The students' reflection on why they think that cross-disciplinary studies are valuable

The diagram in figure 2 reveals that in the end of each of the three days the students evaluated that type of cross-disciplinary studies to be a real value as means for developing different skills, including the habit of recognizing links between things and phenomena, which couldn't have been noticed before. However, more value was seen in the lessons delivered by the teachers of groups 2 and 3, who represented more diverse disciplines than in group 1.

Having summarized all the findings, "The model of opportunity identification and creation in cross-disciplinary teaching and learning environment" was elaborated (see fig. 3).

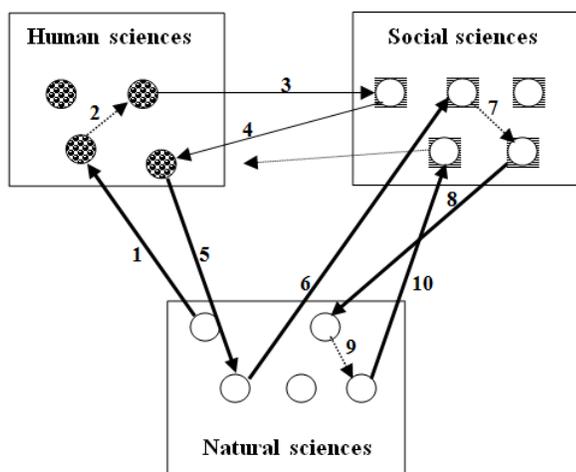


Fig. 3. The model of opportunity identification and creation in cross-disciplinary teaching and learning environment

It shows that the links between study disciplines within one group of sciences (see links 2, 7 and 9 in fig. 3) and between "Human sciences" and "Social sciences" (see links 3 and 4 in fig. 3) are more traditional and apparent; therefore the level of innovativeness of the idea created on these links aren't very high. Meanwhile the links between study disciplines across different science groups, especially between "Natural sciences" and "Social sciences" (see links 6, 8 and 10 in fig. 3) and between "Natural sciences" and "Human sciences" (see links 1 and 5 in fig. 3) traditionally are admitted to be less related and therefore less apparent. This is where the teachers' as well as of their students' ability to connect seemingly unrelated and independent things, phenomena and trends in order to identify and create new opportunities could be developed more successfully. The model doesn't show specifically with which study discipline or which science group the cross-disciplinary studies should start or opportunities should get identified or created. It has a kind of symbolic character to manifest how "intra" and "inter" links are formed and the gaps between seemingly independent areas are overcome; moreover, the bigger the gap the more innovative the idea and opportunity turn out to be.

#### 4. CONCLUSIONS

1. Cross-disciplinary teaching and learning are an appropriate means for developing both teachers' and their students' skills and habits to identify and create opportunities by discovering hidden links between apparently unrelated things, phenomena and events from different study disciplines.
2. In order to provide cross-disciplinary teaching and learning environment, teachers' education should be revised and reorganized and new level of cooperation among teachers of diverse study disciplines should be established.

#### 5. REFERENCES

- [1] Ardichvili, R. Cardozo & S. Ray, "A Theory of Entrepreneurial Opportunity Identification and Development", **Journal of Business Venturing**, Vol. 18, 2003, pp. 105–123.
- [2] R.A. Baron, "Opportunity Recognition as Pattern Recognition: How Entrepreneurs "Connect the Dots" to Identify New Business Opportunities", **Academy of Management Perspectives**, February, 2006, pp. 104–119.
- [3] E. de Bono, **Serious Creativity: Using the Power of Lateral Thinking to Create New Ideas**, Harper Business, 1992.
- [4] P. Davidsson & B. Honig, "The Role of Social and Human Capital among Nascent Entrepreneurs", **Journal of Business Venturing**, Vol. 18, No. 3, 2003, pp. 301–331.
- [5] D.R. DeTienne & G.N. Chandler, "The Role of Gender in Opportunity Identification", **Entrepreneurship Theory and Practice**, May, 2007, pp. 365–386.
- [6] D.R. DeTienne & G.N. Chandler, "Opportunity Identification and Its Role in the Entrepreneurial Classroom: a Pedagogical Approach and Empirical Test", **Academy of Management Learning and Education**, Vol. 3, No. 3, 2004, pp. 242–257.
- [7] P.F. Drucker, "The Discipline of Innovation", **Harvard Business Review**, Vol. 80, No. 8, 1998, pp. 95–102.
- [8] D.K. Dutta & M.M. Crossan, "The Nature of Entrepreneurial Opportunities: Understanding the Process Using the 41 Organizational Learning Framework", **Entrepreneurship: Theory and Practice**, Vol. 29, 2005, pp. 425–49.
- [9] J. Fiet, "The Informational Basis of Entrepreneurial Discovery", **Small Business Economics**, Vol. 8, No. 6, 1996, pp. 419–430.
- [10] W. Gartner, N. Carter and G. Hills, "The Language of Opportunity", in C. Steyaert and D. Hjorth (Eds.), **New Movements in Entrepreneurship**, Edward Elgar, Cheltenham, 2003, pp. 103–124.
- [11] A. Gibb, "Creating Conductive Environments for Learning and Entrepreneurship: Living with, Dealing

- with, Creating and Enjoying Uncertainty and Complexity”, **Industry & Higher Education**, Vol. 19, No. 3, 2002, pp.135–148.
- [12] P. D. Hannon, “Teaching Pigeons to Dance: Sense and Meaning in Entrepreneurship Education”, **Education and Training**, Vol. 48, No. 5, 2006, pp. 296–308.
- [13] P.D. Hannon, L.A. Collins & A.J. Smith, “Exploring Graduate Entrepreneurship: A Collaborative, Co-learned Based Approach for Students, Entrepreneurs and Educators”, **Industry & Higher Education**, Vol. 19, No. 1, 2005, pp. 11–24.
- [14] J. Heinonen & S. Poikkijoki, “An Entrepreneurial-directed Approach to Entrepreneurship Education: Mission Impossible?”, **Journal of Management Development**, Vol. 25, No. 1, 2006, pp. 80–94.
- [15] D. Hjorth & B. Johannisson, “Learning As an Entrepreneurial Process”, in A. Fayolle (Ed.), **Handbook of Research in Entrepreneurship Education**, Cheltenham, UK: Edward. Elgar Publishing, 2007.
- [16] B. Hynes, “Entrepreneurship Education and Training – introducing Entrepreneurship into Non-business Disciplines”, **Journal of European Industrial Training**, Vol. 20, No. 8, 1996, pp. 1–10.
- [17] I. Kirzner, “Entrepreneurial Discovery and the Competitive Market Process: an Austrian Approach”, **Journal of Economic Literature**, Vol. 35, No. 1, 1997, pp. 60–85.
- [18] T. Koke and K. Oganisjana, “Educational Experience as a Path to Active Citizenship, in L. Jögi, E. Przybylska & M. Teresevičienė (Eds.), **Adult Learning for Civil Society. International Perspectives in Adult Education 51**, Bonn -Warsaw: Institute for International Cooperation of the German Adult Education Association, 2005, pp. 39–50.
- [19] W. Long and W. McMullan, “Mapping the Venture Opportunity Identification process”, in J.A. Hornday, F.A. Tardley, J.A. Timmons & K.H. Vesper (Eds.), **Frontiers of Entrepreneurship Research**, Wellesley, MA: Babson College, 1984.
- [20] W. McMullan and W. Long, “Entrepreneurship Education in the Nineties”, **Journal of Business Venturing**, Vol. 2, No. 3, 1987, pp. 261–275.
- [21] K. Oganisjana, “The Development of Students' Enterprise in Study Process” (PhD thesis), Retrieved 6 Febr., 2010. <http://luis.lanet.lv/pls/pub/luj.fprnt?l=1&fn=F1648681065/KarineOganisjana2010.pdf>
- [22] N. Phillips and P. Tracey, “Opportunity Recognition, Entrepreneurial Capabilities and Bricolage: Connecting Institutional Theory and Entrepreneurship in Strategic Organization”, **Strategic Organization**, Vol. 5, No. 3, 2008, pp. 313–320.
- [23] T. Proctor, **The Essence of Management Creativity**, London: Prentice Hall, 1995.
- [24] S.A. Sanz-Velasco, “Opportunity Development as a Learning Process for Entrepreneurs”, **International Journal of Entrepreneurial Behaviour & Research**, Vol. 12, No. 5, 2006, pp. 251–271.
- [25] S. Sarasvathy, N. Dew and S. Venkataraman, “Three Views of Entrepreneurial Opportunity”, in Z. Acs and D. Audretsch (Eds.), **Handbook of Entrepreneurship Research**, Kluwer Academic Publishers, Dordrecht, 2003, pp. 141–160
- [26] G.L.S. Shackle, **Decision, Order and Time in Human Affairs**, Cambridge: Cambridge University Press, 1961.
- [27] G. Shacklock, R. Hattam & J. Smyth, “Enterprise Education and Teachers' Work: Exploring the Links”, **Journal of Education and Work**, Vol. 13, No. 1, 2000, pp. 41–60.
- [28] S. Shane, “Prior Knowledge and the Discovery of Entrepreneurial Opportunities”, **Organization Science**, Vol. 11, No. 4, 2000, pp. 448–469.
- [29] S. Shane & S. Venkataraman, “The Promise of Entrepreneurship as a Field of Research”, **Academy of Management Review**, Vol. 25, No.1, 2000, pp. 217–226.
- [30] O.A. Sogunro, “Efficacy of Role-playing Pedagogy in Training Leaders: Some Reflections”, **Journal of Management Development**, Vol. 23, No. 4, 2004, pp. 355–371.
- [31] H. Stevenson & C. Jarillo, “A Paradigm of Entrepreneurship: Entrepreneurial Management”, **Strategic Management Journal**, Vol. 11, 1990, pp. 17–27.
- [32] H. Stevenson, “The Heart of Entrepreneurship”, **Harvard Business Review**, March-April, 1985, pp. 85–94.
- [33] S. Tan & C. K. Frank Ng, “A Problem-based Learning Approach to Entrepreneurship Education”, **Education and Training**, Vol. 48, No. 6, 2006, pp. 416–428.
- [34] S.Venkataraman, “The Distinctive Domain of Entrepreneurship Research” in J. Katz (Ed.), **Advances in Entrepreneurship, Firm Emergence and Growth**, JAI Press, Greenwich,, Ct, 1997.
- [35] L. von Mises, **Human Action**, New Haven, CT: Yale University Press, 1949.
- [36] K. Wilson, “Entrepreneurship Education in Europe”, in J. Potter (ed.), **Entrepreneurship and Higher Education**, OECD, 2008.
- [37] Worldometers: **World statistics updated in real time**. Retrieved on 16 April, 2011 <http://www.worldometers.info/>

## 6. ACKNOWLEDGEMENT

The authors would like to acknowledge the valuable contribution of the European Commission and express gratitude for having supported the implementation of the project “Raising of Teaching Staff and School Competitiveness in Accordance with EU 2020”.

# Conceptual Models of Risk Assessment in Renewable Energy Production in Latvia

Sandija RIVZA, Irina PILVERE, Peteris RIVZA, Baiba RIVZA  
Faculty of Economics, Latvia University of Agriculture  
Jelgava, LV 3001, Latvia

## ABSTRACT

Risk assessment is an important factor for successful and sustainable entrepreneurship of bioenergy production that has become one of the priorities in energy sector of Latvia. Promotion of the use of renewable energy is included as one of the strategic goals for European Union (EU) and Latvia. As this field of energy production in Latvia is rather new and scantily explored there are many risk factors arising in different stages of production, starting with planning and building of a bioreactor and ending with production and further use and distribution of energy. The present research focuses on risk assessment in renewable energy production form biomass as this kind of energy is seen as a perspective source for renewable energy under the conditions of Latvia. A risk assessment module for renewable energy production made by using the Analytic Network Process (ANP) software is described in the paper.

**Keywords:** Analytic Network Process (ANP), renewable energy, risk assessment, Latvia

## 1. INTRODUCTION

European Union (EU) consistently works on setting up a common energy policy with an important place allocated to the renewable energy production, energy efficiency, sustainable use of resources, energy security and independence. The new Directive on renewable energy (Directive 2009/28/EC of the European Parliament and of the Council) sets ambitious targets for all Member States: the share of energy from renewable sources in EU reaching 20% by the year 2020 (8.5% in the year 2005), and a 10% share of renewable energy specifically in the transport sector [1]. To reach this common goal, each member state has to increase the amount of renewable energy production and exploitation as a source for electricity, heating, cooling, and transportation. In the year 2010, renewable energy composed 37% in the total structure of energy in Latvia, with a target of reaching 40% in the year 2020. Currently, the most of the electricity from renewable resources is made from hydropower plants, but 1% of electricity in Latvia is produced by cogeneration of biomass, which is seen as a perspective source for increasing renewable energy production under the conditions of Latvia. To encourage

the development of cogeneration plants, funding from the EU structural funds, Cohesion fund and European Agriculture Fond for Rural Development is available, and it is planned to attract 74 mill. lats (105 mill. euros) from the government of Latvia and the EU in the following years till 2013.

## 2. THEORIES/METHODS

Even though the concept of risk dates back to the 18<sup>th</sup> century, the awareness of risk and its role in the human society has become topical both theoretically and practically on the turn of the 21<sup>st</sup> century and is linked with the ideas of two sociologists, Anthony Giddens and Ulrich Beck, who have admitted that modern society faces risk growth in comparison with the previous development stage [2, 3]. The authors have introduced a new term *risk society*, describing modern society and its tight interaction with the various threats [4]. The quality of risk evaluation is combined of many factors. In terms of the origin of risk, depending on the aim of its classification, the subgroups of risk might be either all - embracing or very specific. For example, the international risk evaluation centre „RiskMetrics Group” suggests the range of 12 types including market, management, environment, social, accounting, legal, credit risks, etc. [5] yet some risk classifications include only two risk groups – external and manufactured risks [2].

Although researchers of the Latvia University of Agriculture have a certain experience in working with risk determination and assessment issues in various fields of agriculture, veterinary medicine, food science, etc. [6, 7, 8], the field of renewable energy production is rather new and is scantily explored, therefore we should adopt the experience of other countries, for example, USA [9], UK [10] and organizations as United Nations Environment Program [11] In this study, the Analytic Network Process (ANP) was used for risk assessment in renewable energy production as it allows including various factors and criteria - tangible and intangible [12] - that is characteristic to risk assessment.

The ANP is a general theory of relative measurement used to derive composite priority ratio scales from individual ratio scales that represent relative measurements of the influence of elements that interact with respect to control criteria. Through its supermatrix

whose elements are themselves matrices of column priorities, the ANP captures the outcome of dependence and feedback within and between clusters of elements. The Analytic Hierarchy Process (AHP) with its dependence assumptions on clusters and elements is a special case of the ANP. The ANP is a new and an essential phase in decision making, neglected so far because of the linear structures used in traditional approaches and their inability to deal with feedback in order to choose alternatives not simply according to attributes and criteria but also according to their consequences both positive and negative – an essential and so far a missing consideration in decision making [13].

### 3. RESULTS AND DISCUSSION

The present research focuses on risk assessment in production of renewable energy in the process of cogeneration. The risks in this research are divided into 5 groups or clusters: personnel, production, property, environment, and legislative risks. They are evaluated by the ability to use one of three strategies – risk reduction, risk transfer, or risk undertaking strategy. The module allows detecting which is the dominant strategy in the whole risk assessment and which strategy is better for separate risks or risk groups. This module is designed to fit in the risk assessment system for agricultural enterprises made by the researchers of the Latvia University of Agriculture in the year 2009 following the Finnish module of risk assessment [14]. The system generally is a questionnaire for farmers available online. Risk assessment is made from the answers of series of questions about the risk factors in the farm. There are 5 risk groups – the same as used in this particular research – production, property, environment, and legislative risks. Totally the system includes 772 questions, like „Is the noise level measured in the work place?” or „Is unauthorized access to premises averted?”. The choice of the farmer is made between 3 possible answers - „Yes”; „No”; „Partly”, and a possibility to indicate, that the question does not refer to the economic activities of the farm. Each of the questions has a coefficient determined by the experts, which is considered when summing the results. The results of the risk evaluation show the average risk levels in each section, lowest, highest and highly hazardous risks and the total risk level of all sections, additionally indicating the average risk level in Latvia. The mentioned system was suitable for various fields of agriculture (crop farming, dairy, livestock farming, etc.), but it did not include renewable energy production, therefore it is planned to combine these two systems (risk assessment by questionnaire and renewable risk assessment by ANP) in the future.

In the ANP module, each of the 5 groups includes several risks (Table 1). The group of personnel risks consisted of three risks that are connected with the

responsibility, qualification and experience of employees and adherence of work safety on the production site. Production risks are the biggest group in our risk classification as it includes six risks that cover the bioenergy production process starting from the preparation of biomass and its quality to supply management, procession of biomass in the cogeneration process and the further use of electricity and heat. Property risks refer to the security and sustainability of the property that is used in the production. Environment risks mentioned in this assessment are closely linked with the procession phases of the energy production – storage and transportation of biomass, storage of digestate after the production of energy and use of the digestate in the fertilization of fields. The last group is legislative risks; these external risks are the ones that arise from the actions of governmental institutions – Ministry of Economics, Municipalities etc.

Table 1 Classification of accessed risks

Risk group (cluster)	Risk
1.Personnel	1.1.Responsibility of the personnel 1.2. Qualification and experience 1.3. Work safety
2.Production	2.1. Quality of biomass 2.2. Stability of the microbiological processes in the bioreactor 2.3. Regular supply of biomass 2.4. Connection with the state electricity network 2.5. Utilization possibilities of the produced heat and their stability 2.6. Accessibility of service for technical equipment
3.Property	3.1. The outer security of the bioreactor and other production facilities 3.2. Credit risk 3.3. Fire security
4.Environment	4.1. Storage of digestate 4.2. Transportation of biomass 4.3. Storage of biomass 4.4. The use of digestate in the fertilization of fields
5.Legislative	5.1. Changes in energy policy 5.2. Changes in the purchase tariffs

Besides five risk groups or clusters, the current risk assessment module includes 3 alternatives in the same level of risk groups: to reduce risks, to transfer risks, and to undertake risks. These alternatives illustrate the possible choices in risk management. The further analysis with the ANP method in the *Superdecisions* software is done to detect which alternative would be the best choice to each of the risks and risk groups (as an average value of all risks in the group). The Figure 1

shows the connections (influences) between risk groups and alternatives. The arrow from one cluster to another show that all or some elements in the first cluster influence all or some elements in the second cluster, but

the reflexive loops (internal cycles) correspond to mutual influence between the elements in one cluster.

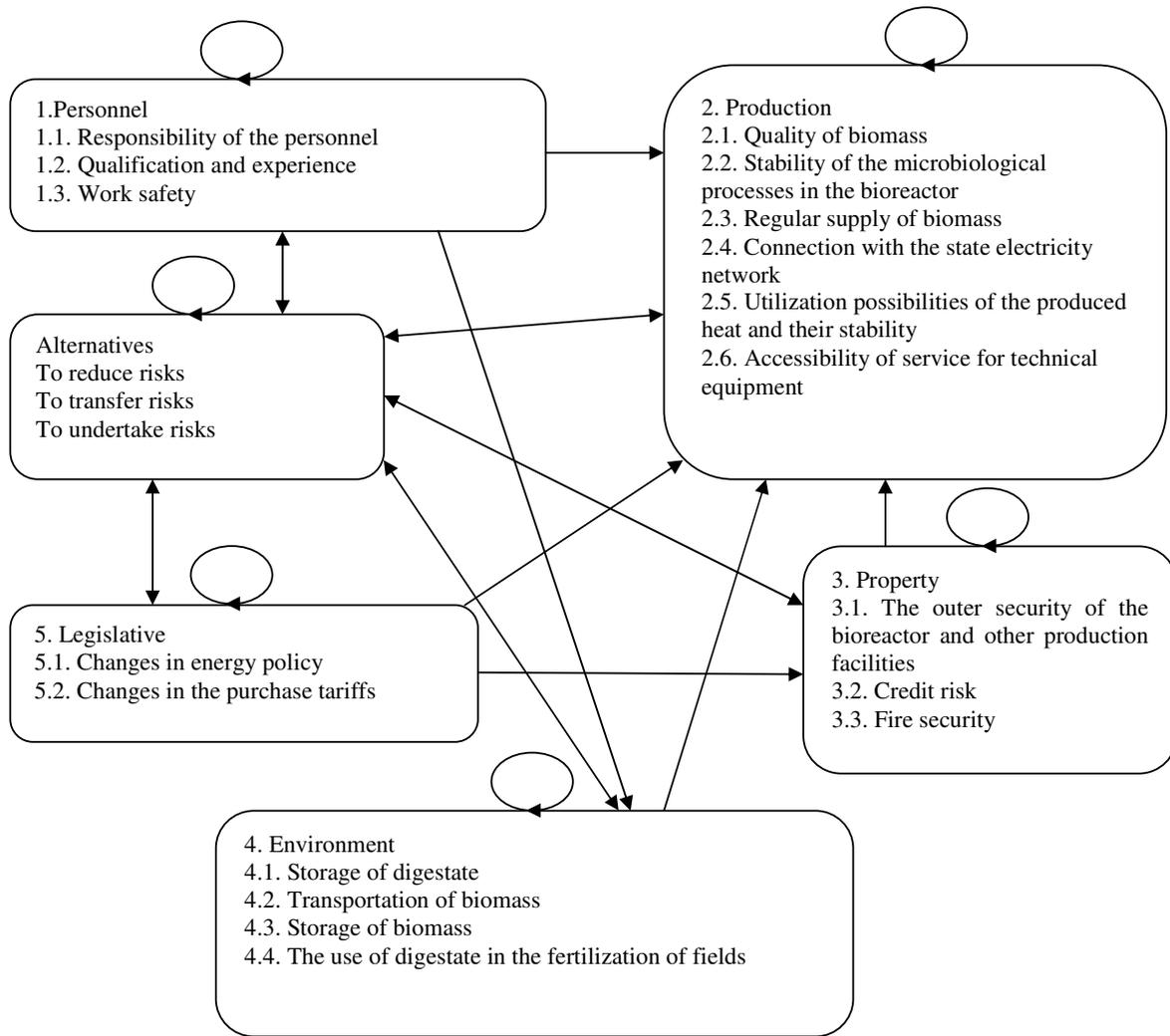


Figure 1 The influence graph describing relations between risk groups

As it can be observed in the Figure 1, Production risks are the ones being the most influenced by all other risk groups – personnel risks (in the aspect of responsibility and qualification of workers and impact of these qualities to the biomass preparation, regular supply and management of the cogeneration processes), legislative risks (in terms of possibilities to connect to state electricity network, and utilize heat), Environment risks (as the need to follow the environment protection regulations in the production process) and property risk (in terms of production site fire-safety and outer security, recoupment/credit risk, also affected by purchase tariffs).

All risk groups and alternatives have the inner dependence – one of the elements in the cluster influences other elements in the same cluster, for example the risk „changes in the energy policy” in the cluster „legislative risks” influence the risk „changes in the purchase tariffs”, because the purchase tariffs are set by the government and therefore influenced by political decisions.

After the determination of influences clusters and nodes of related risks were evaluated. In ANP an intensity of influence is being estimated by the experts with use of pairwise comparison's procedure and the

fundamental ratio scale [15]. Pair comparison technique is the most universal method of measurements, as it can be applied at absence of any scales and standards, in particular at a measurement of intangible attributes. Comparing two objects with respect to a common attribute (criterion, property) the expert estimates a relative preference of one object over another, choosing a suitable estimation from the fundamental scale [16]. In this case the importance or significance of one element over another element is measured.

The results of the performed assessment show (Figure 2) that the alternative that is rated as the best choice for risk management is to reduce risks (average value of all risk groups is 0.21), although the transfer of risk shows a high dispersion (average: 0.176; min: 0.023; max: 0.522) displaying that this alternative is highly suggested for certain risks or risk groups and avoidable or impossible for other.

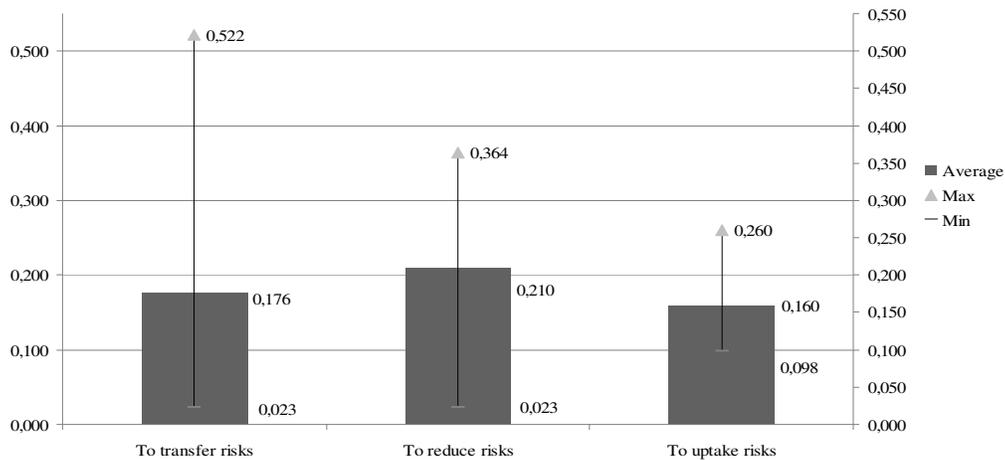


Figure 2 Assessment of alternatives

The Figure 3 allows to identify the suitability of alternatives to risk groups. As it can be seen, legislative risks, according to the assesment (0.209), are not highly manageable in terms of transfer or reduction, one can uptake these risks and pay attention to changes, for

example, follow up the drafting process of new government regulations that includes tarif change, but there are minor possibilities of affecting these processes from the energy producers point of view.

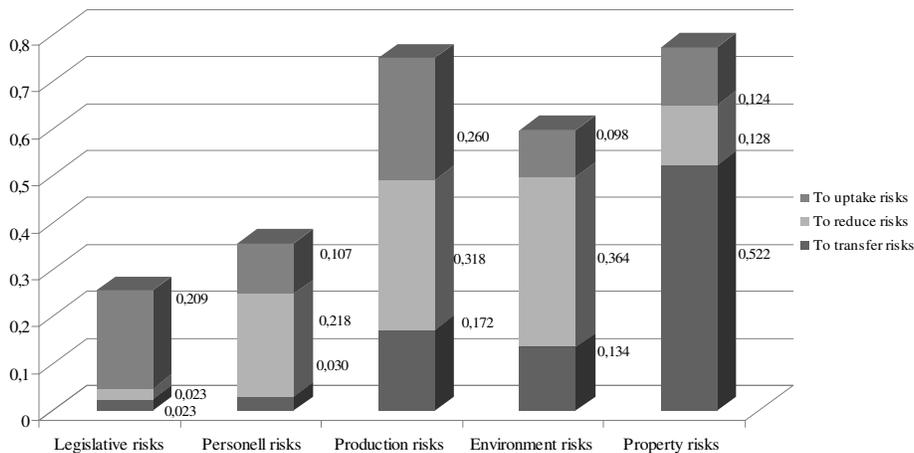


Figure 3 Choice of alternatives within the risk groups

Also in the production risks group there is a high evaluation for the alternative of uptaking risks (2.6), meaning, that not all problems in production process can

be forecasted and reduced, jet „to reduce risks” is the highest evaluated alternative (3.18) in this group, showing the high need of control in the production process. Personell and Environment risks are advised to

be reduced (0.218 and 3.64 respectively), but the property risks in this research are suggested to be transferred (0.522), for example with insurance, yet some of these risks could be reduced or uptaken.

#### 4. CONCLUSIONS

Sector of agriculture production is subjected to various risks therefore comprehensive risk determination and classification is an important precondition to successful and meaningful risk management.

The use of ANP in risk management gives an opportunity to perform the risk assessment by including tangible and intangible factors, and to evaluate various dependencies between risks and alternatives, making it a valuable tool for risk assessment.

The current risk assessment in renewable energy production shows that most of the risks are suggested to be reduced (personnel, production and environment risks) or transferred (property risks in particular), yet there are several risks, mainly in the group of legislative risks, that can be only undertaken i. e. taken into further consideration.

#### 5. REFERENCES

- [1] **Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources** European Parliament, 2009 [Online] Available at: <http://eurlex.europa.eu/lexuriserv/lexuriserv.do?uri=celex:3200910028:en:not>
- [2] A. Giddens, **Risk and Responsibility**. Modern Law Review 62(1); 1999, pp.1-10.
- [3] U. Beck, **Risk Society: Towards a New Modernity**. New Delhi: Sage. 1992.
- [4] P. Caplan, **Risk Revisited**, Pluto Press, London, 2000, pp. 7.
- [5] **RiskMetrics Group**, 2010 [Online] Available at: <http://www.riskmetrics.com>
- [6] P. Rivža, **Riska faktoru izvērtēšana un riska vadība Latvijas lauksaimniecībā**, Latvijas Lauksaimniecības universitātes raksti, LLU, Nr. 11, 2004, pp. 3.-8.
- [7] **Lauksaimniecības un pārtikas risku vadīšana**, monograph, Jelgava, LLU, 2007, pp. 41.-51.
- [8] **Riski lauksaimniecībā un privātajā mežsaimniecībā**, monograph, Jelgava, LLU, 2005, pp. 47.-59.
- [9] G. Rausser, M. Papineau, **Managing R&D Risk in Renewable Energy**. University of California, Berkeley, 2008, [Online] Available at: <http://escholarship.org/uc/item/41j9f6ks#page-1>
- [10] **Sustainable Energy Security: Strategic risks and opportunities for business**, White paper, LLOID`S 360° Risk Insight, 2010, [Online] Available at: [http://www.chathamhouse.org.uk/files/16720\\_0610\\_froggatt\\_lahn.pdf](http://www.chathamhouse.org.uk/files/16720_0610_froggatt_lahn.pdf)
- [11] **Financial Risk Management Instruments for Renewable energy projects**. Summary document. United Nations Environment Programme. Words and Publications, Oxford, UK, 2004, ISBN: 92-807-2445-2
- [12] T.L. Saaty, **Principia Mathematica Decernendi: Mathematical Principles of Decision Making**, RWS Publications, Pittsburgh, PA, USA. 2010.
- [13] T.L. Saaty, **Fundamentals of the Analytic Network Process**, Proceedings of the ISAHP, 1999
- [14] J. Leppälä, M. Murtonen, J. Suutarinen, **Farm-RM – risk management tools for farms**, NJF Report Vol 4, No 7, 2008, pp. 22.-25.
- [15] T.L. Saaty, **Decision making with dependence and feedback. The analytic network process**, Pittsburgh: RWS Publications, 2001.
- [16] Andreichicova O.N., Radyshevskaya T.N. **Application of the Analytic Network Process to a forecasting of oral health**, Proceedings of the ISAHP, 2007.

#### 6. ACKNOWLEDGEMENTS

This publication has been prepared within the framework of the ESF Project „Attraction of human resources to the research of the renewable energy sources”, Contract Nr. 2009/0225/1DP/1.1.1.2.0/09/APIA/VIAA/ 129.

## Marketing Portfolio as a Media for Marketing Assets Interaction to Optimize Marketing Efficiency Measured with Adequate Metric

Aleksandras Vytautas RUTKAUSKAS, Indrė LAPINSKAITĖ-VVOHLFAHRT, Viktorija STASYTYTĖ,  
Laura UTURYTĖ-VRUBLIAUSKIENĖ  
Department of Finance Engineering, Vilnius Gediminas Technical University  
Vilnius, Sauletekio ave. 11, LT-10223, Lithuania

Edvard MICHNEVIČ  
Department of Engineering Mechanics, Vilnius Gediminas Technical University  
Vilnius, Sauletekio ave. 11, LT-10223, Lithuania

### ABSTRACT

A whole set of problems rises as Marketing portfolio stratum (cluster) forms: measuring the impact of marketing on business development, identifying the assets of marketing portfolio, structuring the interaction of marketing assets, selecting the quantitative models of mentioned interactions, generating the information required quantitative description by expert or other means. Finally, finding the optimal solution conditioned by marketing as complex media opportunities, what is usually done by formulating particularly complex task of stochastic programming and creating as often as not quite complex decision methods.

The article deals with the situation which can be described in the following way. It is assumed that in order to increase marketing contribution, a certain amount of capital is additionally distributed in between 4P elements seeking to achieve optimal additional increase of business results triggered by such investment. Herewith it is aimed to combine for investment the efforts with the possibility in order to spread the obtained increase realization between two segments of customers, with different profit possibilities. Additionally, one more problem appears – how to make the choice between the two creditors when it is possible to evaluate debt service imposed the present value probability distribution of possibilities to each of them.

The purpose of situation content is - to find such investment in the 4P mix, sales in A and B segments and loan between the creditor number 1 and 2 proportions which would allow to find the optimum solution, according to a certain composition of the possibility's effectiveness, reliability and subject risk. To find the solutions, Markowitz random field technique proposed by the authors was used.

**Keywords:** marketing portfolio, marketing asset, efficient surface, Markowitz random field.

### 1. INTRODUCTION

The particular prime stage of portfolio concept's germs stratification can be considered H. Markowitz and co-authors development of Modern portfolio theory oriented to the investment portfolio as the media allowing to consider the interactions between investment assets and interactions with micro and macro factors, in order to create the maximum value, measured according to a certain profitability and riskiness composition to the owners of the assets.

Following the investment portfolios stratum, the other strata, combining different assets and goals, also began to develop actively: product line portfolios, corporate strategies portfolios, company business portfolios, growth share portfolios, industry marketing portfolios, industrial purchasing portfolios, strategic resource portfolios, supplier-consumer relationship portfolios, marketing portfolios, etc.

Efficiency measurement principles of investment portfolios stratum of assets interaction had a significant impact on each of mentioned portfolios strata. Investment portfolio efficiency criteria are usually oriented towards the desired composition of profitability and riskiness. It is true that a direct income generation is not typical for a number of strata portfolio of assets, however a possibility of the diptych efficiency-riskiness remains.

Almost all strata have the common features. It is the possibility when measuring both profitability and riskiness the average values are used: profitability average and average riskiness, which is identified with the average standard deviation. And more, the best solution is searched on the generated curve of those parameters, which usually identified as efficiency line.

In turn, in both the investment portfolios strata and in other strata are constantly examined, what should be the utility function of the portfolio owner (holder), to stimulate his interest of efficient portfolio management. One of the major shots constructing the utility function was the Sharpe ratio:

$$C = \frac{e_v - e_{rf}}{Stdev(e)} \quad (1)$$

Where  $e_v$  - the average of portfolio profitability possibilities (distribution);

$e_{rf}$  - the profitability of risk-free asset;

$Stdev(e)$  - the standard deviation of portfolio profitability possibilities.

We will start the conversance with marketing portfolio stratum from Ryals L. *et al.*, 2007, which discussed quite in detail the origination of portfolio category in marketing, as well as specific educational issues of marketing portfolio opportunities. On the next page the issues based on Ryals L. *et al.*, 2007 review will be presented, which is performed in our opinion very competently.

## 2. MARKETING PORTFOLIO AS A MEDIA FOR MARKETING ASSETS INTERACTION TO OPTIMIZE MARKETING EFFICIENCY MEASURED WITH ADEQUATE METRIC

Portfolio (from Wikipedia) literary means “a case for carrying loose papers” (from Latin, the imperative of *portare* “to carry” and the plural of *folium*, meaning “a sheet for writing upon”). There can be found a lot of expressions in literature where “portfolio” is connective word, for example: a type of briefcase; portfolio (government), the post and responsibilities of a head of a government department; career portfolio, an organized presentation of an individual’s education, work samples, and skills; artist’s portfolio, a sample of an artist’s work or a case used to display artwork, photographs etc.; electronic portfolio, a collection of electronic documents; patent portfolio, a collection of patents owned by a single entity; product portfolio (business administration), separation of products by their market share and profits or growth rates portfolio, projects portfolio in project portfolio management; the portfolio of projects in an organization; Atari portfolio, a palmtop computer; [Portfolio.com](http://Portfolio.com), a business magazine; minister without portfolio.

From about 1930 it has also come to mean a “collection of securities or responsibilities held by an individual”.

Further we will be in rely with investment portfolio definition as a “Pool of different investments by which an investor bets to make a profit (or income) while aiming to preserve the invested (principal) amount. These investments are chosen generally on the basis of different risk-reward combinations: from “low risk, low yield” (gilt edged) to “high risk, high yield” (junk bonds) ones; or different types of income streams: steady but fixed, or variable but with a potential for growth” but also try to tend to concept that “investment portfolio is a media for investment assets interaction to optimize investment effect measured with adequate metric”.

When we talk about marketing assets, everything is more complicated, because marketing is a means for product or service to reach the consumer therefore marketing assets often accept the elementary marketing means – advertisement stands, message texts and other; marketing MIX – as the contact points of marketing and business; consumer – as final user of business results; marketing media – as means to unfold the marketing information. That is why we would like to pay some attention to the concepts of marketing assets and marketing metrics.

### Marketing assets

There are some basic definitions what does asset mean:

- A resource with economic value that an individual, corporation or country owns or controls with the expectation that it will provide future benefit [13].
- Something valuable that an entity owns, benefits from, or has use of, in generating income [4].
- In financial accounting, assets are economic resources. Anything tangible or intangible that is capable of being owned or controlled to produce value and that is held to have positive economic value is considered an asset [31].

It is not easy to define what marketing assets are as marketing tactics aren’t black and white like cash and debt. Marketing tactics are very dependent on objectives, customer segments and expectations.

For a marketer, an asset is a tool or a platform, something you can use over and over without using it up. In fact, it’s something that gets better the more you invest [27].

### Marketing metrics

A metric is a measuring system that quantifies a trend, dynamic, or characteristic. In virtually all disciplines, practitioners use metrics to explain phenomena, diagnose causes, share findings, and project the results of future events. Throughout the worlds of science, business, and government, metrics encourage rigor and objectivity. They make it possible to compare observations across regions and time periods. They facilitate understanding and collaboration [2; 11; 14].

As marketers progress in their careers, it becomes increasingly necessary to coordinate their plans with other functional areas. Sales forecasts, budgeting, and estimating returns from proposed marketing initiatives are often the focus of discussions between marketing and finance. For marketers with little exposure to basic finance metrics, a good starting point is to gain a deeper understanding of “rate of return”. “Return” is generally associated with profit, or at least positive cash flow. “Return” also implies that something has left-cash outflow [11]. At a time when firms are cutting costs, it is essential for all functional disciplines within the firm to be financially accountable. This introduces the need for measurement, as without measurement it is impossible to be accountable. For firms to measure the return on marketing, it is essential for them to treat marketing expenditures as an investment [26].

### At the approach of marketing spend optimization

Considering the fact that marketing activity results are not the object of official accountability [23], as there are not enough analytical research to reveal how the influence of marketing creates shareholder value [8; 9; 20], the role of marketing assets goes to different changes of business income caused by the efforts of marketing activity therein the efforts of risk management. Therefore in literature continuously proceed the discussions – could Modern Portfolio Theory (MPT) be analysed to marketing [5] and especially for risk impact assessment.

Financial portfolios use MPT, which deals with problems of risk and return, to make investment allocation decisions. The impact of MPT on business decision-making has been substantial; major capital spending projects, for example, are now routinely assessed for risk as well as return. This message has not yet been taken up by marketing. If marketing calculations take no account of risk, decisions about resources and how to prioritise marketing spending may be sub-optimal [6].

So, can MPT be applied to marketing? Marketing spend allocation decisions can be viewed as portfolio investment decisions [1], whether the portfolio is considered in terms of customers or customer segments [6; 19; 23], products [3] or brands [21]. However, as Devinney, Stewart and Shocker (1985) point out [7], unlike financial portfolios, investment marketing assets is expected to affect the returns from those assets. Thus, MPT would need modification before it could be applied to marketing [5].

Despite ongoing interest in the notion of marketing portfolios and the emergence of portfolio management tools such as the Boston Matrix, Directional Policy Matrix, and StratPort, risk and return has received relatively little consideration in the marketing literature. Previous discussion of the management of marketing

portfolios has tended to focus either on profit maximisation [16; 17] or on customer lifetime value maximisation [18]. An early exception is Kotler (1971), who uses variance of returns as a proxy for risk [15]. However, MPT views risk as depending in large measure on the covariance of its component investments [1]; in other words, diversification reduces portfolio risk. More recently, Srivastava and Reibstein (2004) consider risk in terms of volatility of cash flows, and Dhar and Glazer (2003) have revived the argument for using financial portfolio theory to address marketing portfolios, stressing the importance of understanding risk.

In the Ryals L. *et al.* (2007) the application of MPT to marketing through a model that takes into account risk and return is demonstrated. First, MPT is applied to marketing portfolio made up of customer segments. However, MPT does not apply literally to marketing portfolios since returns on financial portfolios are generally considered to be determined by the market and therefore independent of spend allocation, which is not the case in marketing. Therefore in Ryals L. *et al.* (2007) MPT is applied to a particular conditions in marketing, in which returns are affected by the allocation of marketing spend and allows to optimize risk and returns.

**3. STRAIGHT FORWARD APPLICATION OF MODERN PORTFOLIO THEORY**

In this chapter we will analyse the earlier mentioned marketing portfolio [25] where marketing segments are initiated as marketing assets and which will be approached as direct appliance of MPT.

Analysing the basics of MPT we should note that our proposed portfolio concept – portfolio is a media for the assets interaction to realize holders’ interests measured with adequate metric and supply information for behaviour strategies – become universal to different portfolios.

The authors applying MPT to marketing suggests that optimal marketing portfolios are those marketing portfolios for which:

1. No other combination of customer segments will yield higher returns with the same level of risk; or
2. No other combination of customer segments will yield the same returns with lower risk.

Realistically, many combinations of customer segments portfolios are possible and their risk/return positions can be

plotted. Those that satisfy the conditions of optimality will lie on what is known as the efficient frontier [29].

MPT was developed in the 1950s throughout the early 1970s and was considered an important advance in the mathematical modelling of finance. Nevertheless there remains the set of fundamental results up till now, which maintain their power and perfectly fit for the marketing. Of course since then, many theoretical and practical criticism have been levelled against it.

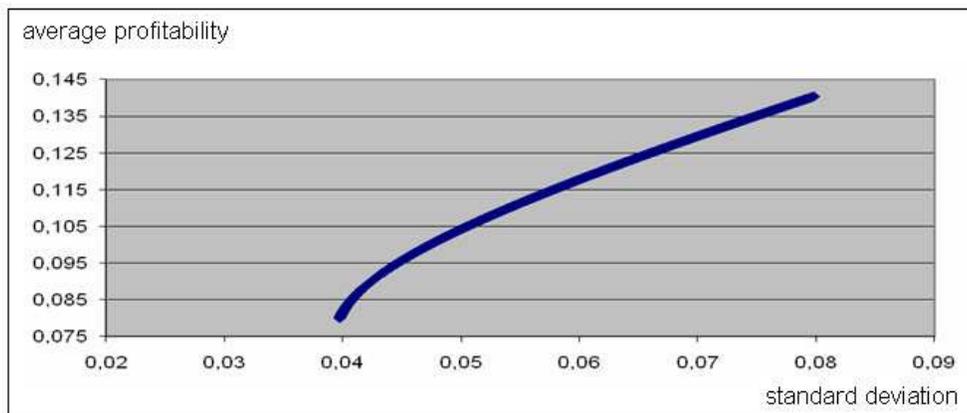
The main features and possibilities of MPT could be understood throughout understanding the concept and main mathematical relations of MPT.

The fundamental concept behind MPT is that the assets in an investment portfolio should not be selected individually, each on their own merits. Rather, it is important to consider how each asset changes in price relative to how every other asset in the portfolio changes in price.

Investing is a tradeoff between risk and expected return. In general, assets with higher expected returns are riskier. For a given amount of risk, MPT describes how to select a portfolio with the highest possible expected return. Or, for a given expected return, MPT explains how to select a portfolio with the lowest possible risk (the targeted expected return cannot be more than the highest-returning available security, of course, unless negative holdings of assets are possible.)

MPT is therefore a form of diversification. Under certain assumptions and for specific quantitative definitions of risk and return, MPT explains how to find the best possible diversification strategy.

Figure 1 is an obvious scheme to solve this problem. Only in the case of two customer segments when in segment A the average of profitability possibilities is 0.08 and the standard deviation is 0.04, and in segment B analogically 0.14 and 0.08, we have here points A and B connecting, and in this case, the coincident possibilities sets of effective portfolio values. Any point in the set of portfolio values uniquely describes the diversification ratio, which is required to obtain this value. Thus, the optimal possibility is strongly related to the diversification ratio or portfolio structure.



**Fig.1** Profit possibilities map (portfolio value set and efficiency line) when sales can be distributed between two customer segments

**4. ADEQUATE PORTFOLIO AS INTENTION TO COMMENSURATE AND INTEGRATE PROFITABILITY, RELIABILITY AND RISK**

Function of fundamental modern (Markowitz) portfolio and its further amplifications [10; 22] is an intention to commensurate investment profitability and risk objectively and to give an opportunity to select a portfolio taking into consideration investor's indifference curve. Efficiency line of portfolio values is fundamental mean of such choice and optimization [28]. However, evaluation of the aimed profitability's reliability and along with general commensuration of profitability, risk, and reliability levels, the essence of which discloses analytically through designing an effective surface in three-dimensional – profitability, risk, reliability – space is of premium and natural

importance for today's investor. Efficient surface, which is formed as an intersection of survival functions of portfolio possibilities values and iso-guaranties, not only contributes for such a commensuration, but also becomes a set of constraints searching for the possibility of the highest profitability for an investor, in other words a criteria invoking his utility function, that depends on profitability, risk, and reliability. Here the word risk is distinguished in order to stress the principal difference between the riskness of investment possibilities' and investor's risk, which depends also on individual features of an investor.

Detailed presentation of adequate portfolio anatomy one could find in Rutkauskas A. *et al.*, 2011 and final picture is presented on Fig. 2.

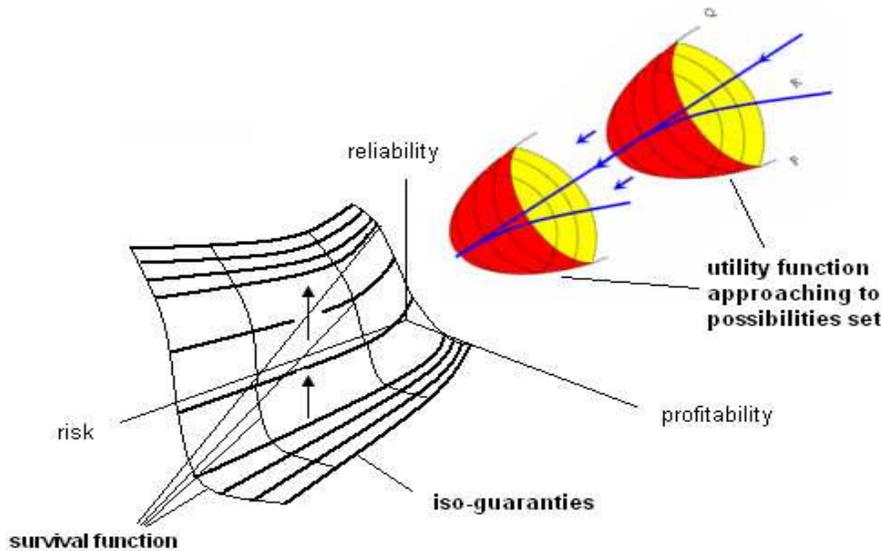


Fig. 2. A scheme of efficient surface [24]

On the efficient surface, i.e. in three-dimensional space, the role of efficiency lines is assigned to iso-guarantees. Here it is possible to analyse the selection of utility possibilities measured in three parameters: profitability, reliability of profitability and risk with the help of three-dimensional utility function.

**Practical application of utility function to the set of possibilities to find an optimal solution**

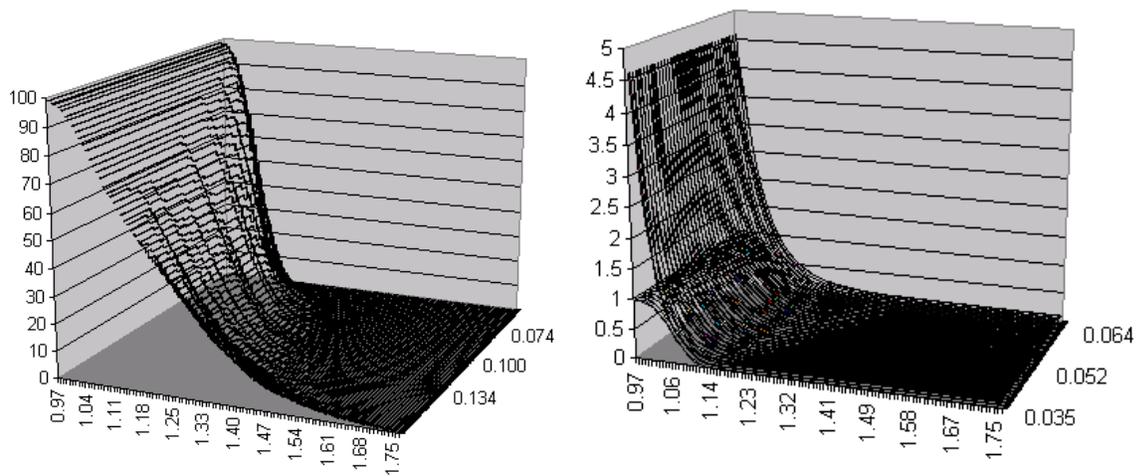
The configurations of possibilities' set (efficiency zone) and utility (objective) function and their inter-position, as well as analytical expression of our applied utility function points out that the magnitude of the possibility, as well as the increase of reliability of possibility both positively influence the growth of utility. However, the analytical expression of the utility function being used provides that the increase of risk negatively influences the growth of utility:

$$U = \exp\left(\frac{p}{r}\right) \cdot g, \tag{2}$$

where  $U$  is the utility level of possibility,  $p$  denotes profitability,  $r$  is the risk and  $g$  – the guarantee.

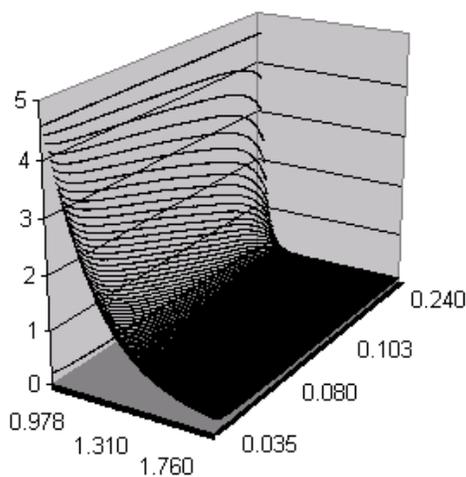
Such specification of utility function and decision-making procedure is analytically meaningful, because it allows to solve a complex stochastic programming task with the help of imitative technologies and graphical decision-making methods.

Fig. 3 exhibits formed set of portfolio values – efficiency surface (a. section), utilities function family (b. section) and mutual position of efficient surface with utility surface under certain utility level (c. section). D section discloses that there is intersection of two continuous and convex to each other surfaces.

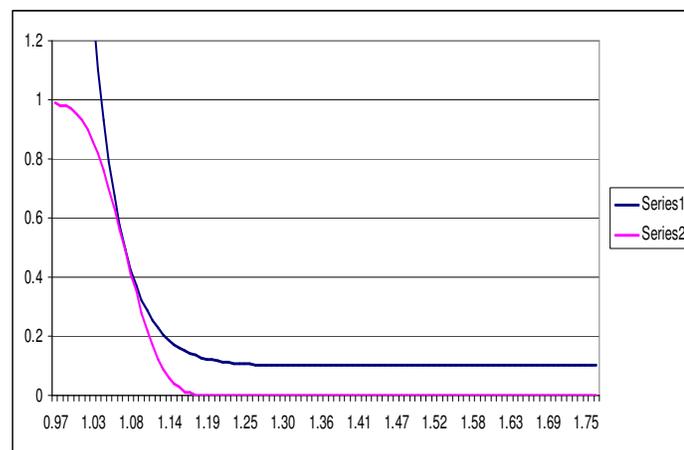


a) Efficient surface

c) Tangency point of the two surfaces



b) Utility surface



d) The section of tangency of efficient surface and utility surface under certain risk level

**Fig. 3.** Anatomy of optimal portfolio investment decision

Indeed, when utility degree of utility function is decreasing, one of the sections of efficient surface, perpendicular to the abscissa axis OX and, passing through certain survival function, is first to touch the utility function itself, along with that touching one of iso-guarantees, while the respective sections of the higher or lower risk levels do not reach their survival functions. In fig. 3 is the case presented right in the beginning of the next chapter.

### 5. MARKETING PORTFOLIO INTEGRATING DIFFERENT CLASSES'S MARKETING ASSETS RETURN POSSIBILITIES TO MAXIMIZE HOLDER'S UTILITY

In this chapter we will integrate the issues covered in chapters 3 and 4. In chapter 4, we analysed the issue how to divide the marginal investment unit between the different elements of the marketing structure -  $P_1, P_2, P_3, P_4$  that, designed total return amount would be the most effective, following the estimates provided by the experts, that an additional return from unitary

investment can be described by Normal probability distributions:  $N_1(1,032; 0,04)$ ,  $N_2(1,11; 0,09)$ ,  $N_3(1,17; 0,2)$ ,  $N_4(1,22;0,28)$ . Graphical view of problem solution is presented in Fig. 3.

Note: the data used in paragraph 2 are taken from scorecard of real situation and the data here obtained in expert way with the description presented by Ginevičius R. *et al.*, this edition [12].

Meanwhile, in chapter 3 the problem was solved how to distribute sales volume between A and B customers' segments hoping to get the most effective according to the return rate and risk return possibility, when the profitability possibility in segment A is described by the Normal distribution -  $N_A$  ( $m = 8\%$ ,  $\sigma = 4\%$ ) and in segment B by Normal distribution -  $N_B$  ( $m = 14\%$ ,  $\sigma = 7\%$ ).

Now we understand the task tackled in presented situation the following way: how to distribute invested marginal unit between the elements of marketing structure and resulted additional product's realization between segments A and B to distribute the way to get the maximum utility measured by the adequate utility function. Frequently is used function  $U = \frac{e \times g}{r}$ , where  $u$  - the extent of utility, and  $e$  - level of profitability,  $g$  - guarantee of profitability and  $r$  - possible risk, can be understood as improved Sharp ratio.

We have chosen an adequate portfolio optimization as decision methodology, which is described in detail in chapter 4 and which here can be explained as follows:

1. Four ways of investing -  $P_1, P_2, P_3, P_4$ , and two received additional product's realization segments A and B define eight "investment and realization" ways, covering all possibilities of situation:  $P_1 \cap A, P_2 \cap A, P_3 \cap A, P_4 \cap A, P_1 \cap B, P_2 \cap B, P_3 \cap B, P_4 \cap B$ .

The return possibilities of each invested unit for every mentioned way can be described with random variable -  $N_j(a_i, \sigma_i) * N_j(a_j, \sigma_j)$ ,  $i = 1, 2, 3, 4$ ;  $j = A, B$ .

2. Selecting all additionally invested capital division proportions between "investment and realization" opportunities  $W_i > 0; W_1 + W_2 + W_3 + W_4 + W_5 + W_6 + W_7 + W_8 = 1$  we obtain effective surface of additionally invested capital earned profitability, where each possibility is set out by return rate, the guarantee of this rate and the risk linked to such possibility.

3. According to the utility functions of the portfolio owner we determine, which possibility is the most effective and what  $W_i \geq 0; \sum_i^8 W_i = 1$  configuration ensures such possibility.

The graphical illustration of decision finding process is in Figure 4; in the right side section a we have the set of portfolio value possibilities, section b – the view were touches that set and utility and section c – selection of the optimal decision possibility.

**Plus debt services expenses**

In the above-described situation, it was accepted to invest a certain capital (defined as the unit) in the 4P mix, borrowed from two banks B1 and B2, with quite adequately evaluable probability distributions of present values of debt service expenses. In this case, profitability distributions are described as follows:

$$N_{B1}(a_1 = 0.05; \sigma = 0.05) \text{ and } N_{B2}(a_2 = -0.06; \sigma = 0.01)$$

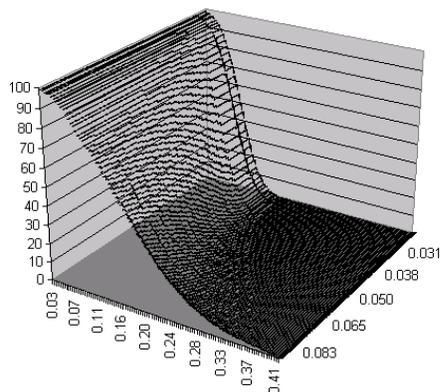
Thus, understanding that the debt service costs would be the factor which will reduce the final net profit of the project, we will seek to integrate the debt service the most efficiently, considering the possibilities to choose between two creditors and, of course, the profit possibilities bringing under their size, reliability and relevance. Decision anatomy and the solution itself are presented in Figure 4. right side is analogous to the left side for gross profit assessment.

Figures 3 and 4 present in full the optimization way. Figure 3 section d shows the point of maximum efficiency with coordinates: efficiency  $e = 1.073$ , reliability  $g = 0.49$ , riskness  $r = 0.042$ . This point is given by the following portfolio structure:  $S_1=0.66; S_2=0.2; S_3=0.04; S_4=0.1$ .

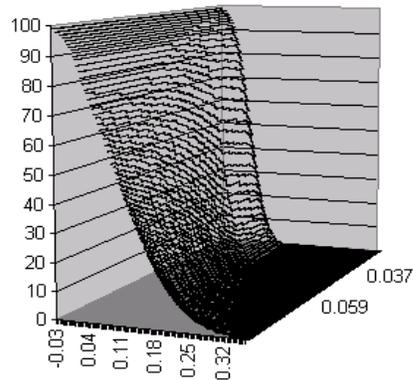
When we are looking for its total profit maximum, then two more elements of the portfolio structure attach to:  $N_1$  and  $N_2$ . Now efficiency  $e = 0.104$ , reliability  $g = 0.53$ , riskness  $r = 0.032$  and the optimal point is given by the new structure:  $S_1=0.2824; S_2=0.0664 S_3=0.2658 S_4=0.1329; S_5=0.0954; S_6=0.0224; S_7=0.0898; S_8=0.0449$ .

Finally, when we combine the conditions of debt optimization between two creditors, the net profit is  $e = 0.1059$ , reliability  $g = 0.46$ , riskness  $r = 0.035$  and the most valuable result is given by the final structure  $S_1=0.6084; S_2=0; S_3=0; S_4=0.3357; S_5=0.0361; S_6=0; S_7=0; S_8=0.0199; S_9=0.944; S_{10}=0.056$ .

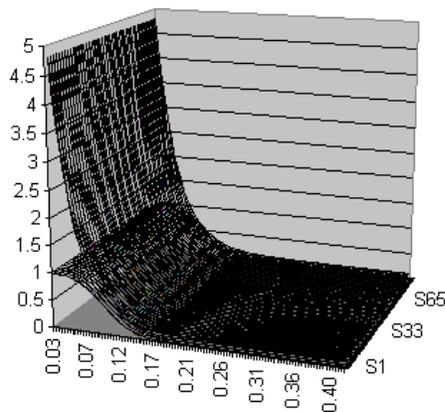
It should be clearly understood, that the next stage should not and cannot repeat the components of the first structure, since sales possibilities do not correlate with investment possibilities and so on.



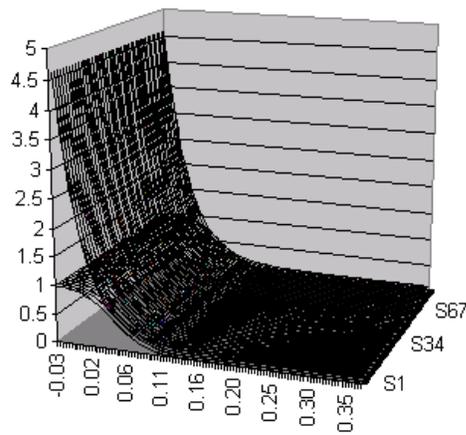
a<sub>1</sub>)



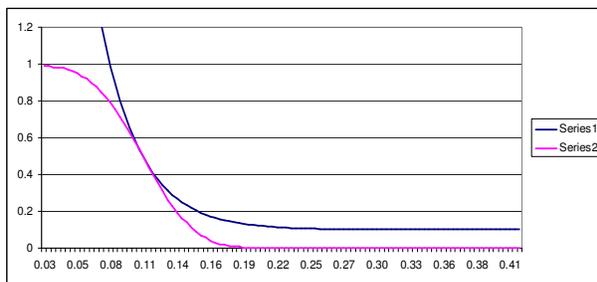
a<sub>2</sub>)



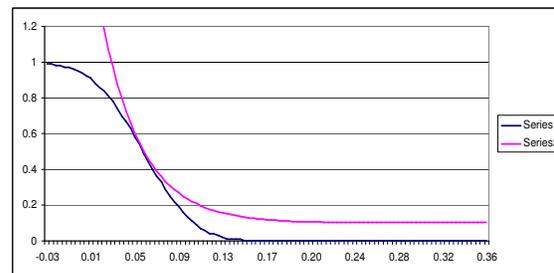
b<sub>1</sub>)



b<sub>2</sub>)



c<sub>1</sub>)



c<sub>2</sub>)

**Fig. 4** The decision search anatomy: the left side – the general profit extent of unitary investment; the right side – the absolute profit extent (clearing debt service expenses). Sections a<sub>1</sub> a<sub>2</sub> – efficient surface; b<sub>1</sub>, b<sub>2</sub> – tangency point of efficient surface and utility surface; c<sub>1</sub>, c<sub>2</sub> – the tangency of efficient surface and utility surface under the certain risk level.

## 6. CONCLUSIONS AND SUGGESTIONS

1. Marketing portfolio is particularly effective tool for both structuring marketing assets, and highlighting the most effective moments of actives' interaction.
2. Integrating, i.e. combining interaction of different marketing asset groups and considering the formation of the resources they need, portfolio management requires particularly complex technique of stochastic process management.
3. Full-rate quantitative description of marketing and business interactions requires to form in principle a new understanding of marketing assets as appropriate evaluation metrics of the impact those assets provide for business.
4. Experimental calculations have shown that Markowitz random field analytical possibilities are sufficient to understand the complex tasks of marketing portfolio management.

## 7. REFERENCES

- [1] P.F. Anderson, "Marketing Investment Analysis", In: Sheth, Jagdish N. (ed.), "**Research in Marketing**", Vol. 4, Chapter 1, 1981, pp. 1-37, Greenwich, CT: JAI Press.
- [2] J. Angelis, E. Pinheiro de Lima, J. Širališova, "Servitised Experiences: Business and Management Implications. **Verslo ir teisės aktualijos [Current Issues of Business and Law]**, Vol. 5, 2010, pp. 6-21.
- [3] R. Bordley, "Determining the Appropriate Depth and Breadth of a Firm's Product Portfolio", **Journal of Marketing Research**, Vol. 40, No.1, 2003, pp. 39-53.
- [4] Business Dictionary, "Asset". Available at: <http://www.businessdictionary.com/definition/asset.html>.
- [5] R.N. Cardozo, Jr. Smith, K. David, "On the Use of Financial Portfolio Theory in Marketing Decisions: A Reply to Devinney, Stewart, and Shocker", **Journal of Marketing**, Vol. 49, No. 4, 1985, pp. 113-115.
- [6] R. Dhar, R. Glazer, "Hedging Customers", **Harvard Business Review**, Vol. 81, No. 5, 2003, pp. 86-92.
- [7] T.M. Devinney, D.W. Stewart, A.O. Shocker, "A Note on the Application of Portfolio Theory: A Comment on Cardozo and Smith", **Journal of Marketing**, Vol. 49, No. 4, 1985, pp. 107-112.
- [8] R. Dobbs, "Managing Value and Performance", **McKinsey Quarterly**, Vol. 3 (Special Issue), 2005, pp. 4-5.
- [9] P. Doyle, "Valuing Marketing's Contribution", **European Management Journal**, Vol. 18, No. 3, 2000, pp. 233-245.
- [10] F.J. Fabozzi, H.M. Markowitz, **The Theory and Practice of Investment Management**, USA: John Wiley & Sons, 2002.
- [11] P.W. Farris, N.T. Bendle, P.E. Pfeifer, D.J. Reibstein, **Marketing Metrics: 50+ Metrics Every Executive Should Master**, Pearson Education, Inc. publishing as Wharton School Publishing, Upper Saddle River, NJ, 2006.
- [12] R. Ginevičius, V. Podvezko, A. Ginevičius, "Determining the Significance of the Criteria Describing Enterprise Marketing", **The 15th World Multi-Conference on Systemics, Cybernetics and Informatics**, July 19–22, 2011, Orlando, Florida, USA. International Institute of Informatics and Systemics. This edition.
- [13] Investopedia, "What Does Asset Mean?" Available at: <http://www.investopedia.com/terms/a/asset.asp>.
- [14] A. Kaziliūnas, "Impacts of Different Factors on the Implementation of Quality Management Systems and Performance Outcomes", **Verslo ir teisės aktualijos [Current Issues of Business and Law]**, Vol. 5, 2010, pp. 75-92.
- [15] Ph. Kotler, **Marketing Decision Making**. New York: Holt, Rinehart and Winston, 1971.
- [16] J.-C. Larreche, V. Srinivasan, "STRATPORT: A Decision Support System for Strategic Planning", **Journal of Marketing**, Vol. 45, (Fall), 1981, pp. 39-52.
- [17] J.-C. Larreche, V. Srinivasan, "STRATPORT: A Model for the Evaluation and Formulation of Business Portfolio Strategies", **Management Science**, Vol. 28, No. 9, 1982, pp. 979-1001.
- [18] K. Lemon, R.T. Rust, V.A. Zeithaml, "What Drives Customer Equity?", **Marketing Management**, Vol. 10, No. 1, 2001, pp. 20-25.
- [19] B. Libai, D. Narayandas, C. Humby, "Towards an Individual Customer Profitability Model: A Segment-Based Approach", **Journal of Service Research**, Vol. 5, No. 1, 2002, pp. 69-76.
- [20] B.A. Lukas, G.J. Whitwell, P. Doyle, "How Can a Shareholder Value Approach Improve Marketing's Strategic Influence?", **Journal of Business Research**, Vol. 58, No. 4, 2005, pp. 414-422.
- [21] M. Petromilli, D. Morrison, M. Million, "Brand Architecture: Building Brand Portfolio Value", **Strategy and Leadership**, Vol. 30, No. 5, 2002, pp. 22-28.
- [22] F.K. Reilly, K.C. Brown, **Investment Analysis and Portfolio Management**, Seventh edition, Thompson: South-Western, 2003.
- [23] R.T. Rust, K.N. Lemon, V.A. Zeithaml, "Where Should the Next Marketing Dollar Go?", **Marketing Management**, Vol. 10, No. 3, (Sept-Oct), 2001, pp. 24-28.
- [24] A.V. Rutkauskas, V. Stasytytė, "Optimal Portfolio Search using Efficient Surface and Three-Dimensional Utility Function", **Technological and Economic Development of Economy**, Vol. 17, No. 2, 2011, pp. 305-326.
- [25] L. Ryals, S. Dias, M. Berger, "Optimising Marketing Spend: Return Maximization and Risk Minimisation in the Marketing Portfolio", **Journal of Marketing Management**, Vol. 23, No. 9-10, 2007, pp. 991-1011.
- [26] S.H. Seggie, E. Cavusgil, S. E. Phelan, "Measurement of Return on Marketing Investment: A Conceptual Framework and the Future of Marketing Metrics", **Industrial Marketing Management**, Vol. 36, 2007, pp. 834-841.
- [27] Seth Godin's blog. "Do Marketers have Assets?" Available at: [http://sethgodin.typepad.com/seths\\_blog/2009/09/do-marketers-have-assets.html](http://sethgodin.typepad.com/seths_blog/2009/09/do-marketers-have-assets.html)
- [28] W.F. Sharpe, "Capital Asset Price: A Theory of Market Equilibrium under Conditions of Risk", **Journal of Finance**, Vol. 19, No. 3, 1964, pp. 425-442.
- [29] W.F. Sharpe, **Investments**, 2<sup>nd</sup> edition, New Jersey: Prentice-Hall, 1981.
- [30] R. Srivastava, D.J. Reibstein, "Metrics for Linking Marketing to Financial Performance", **Working paper submitted to Marketing Science Institute**, October 19<sup>th</sup> 2004, pp. 1-18, Available at: <http://executiveeducation.wharton.upenn.edu/globals/documents/metrics.pdf>
- [31] Wikipedia – The Free Encyclopedia. "Asset". Available at: <http://en.wikipedia.org/wiki/Asset>.

# Markowitz Random Field as a Stand for Investment Analysis and Decision Making

Aleksandras Vytautas RUTKAUSKAS  
Faculty of Business Management, Vilnius Gediminas Technical University  
Vilnius, Sauletekio ave. 11, LT-10223, Lithuania

and

Viktorija STASYTYTĖ  
Faculty of Business Management, Vilnius Gediminas Technical University  
Vilnius, Sauletekio ave. 11, LT-10223, Lithuania

## ABSTRACT

The paper presents analytical discussion about the innovative methodology of dealing with random nature of investment assets. The core element of the methodology is the concept of Markowitz Random Field (MRF), on the basis of which the investment portfolio analysis is moved to three-dimensional space. Applying the proposed procedure, an investor or investment analyst is able to answer the question: how the increases of return and riskness interact on the determined level of investment possibilities' reliability? The graphical procedures of decisions search are fully presented in the paper.

**Keywords:** Random Variable, Markowitz Random Field, Efficient Surface, Stochastic Optimization.

## 1. INTRODUCTION

It is broadly discussed in financial and economic literature that the complex of assets return possibilities can be taken only as a random variable [1, 2, 3, 4, 6, 12].

Additionally, the practical observations of the authors do not contradict the assumption that Gumbel probability distribution function precisely approaches and becomes increasingly dominant in currency and capital markets, as the most adequately defining the emerging market stock returns or exchange rate profitability possibilities.

However, the decisions of professional investment are usually made through the so-called stock or currency portfolios. Therefore it is particularly important to explore what standard features portfolio investment offers to us and what decision-making instrumentation should be sufficient to find sustainable strategies of effective solutions.

## 2. MARKOWITZ RANDOM FIELD (MRF) AS THE INVESTMENT ANALYSIS AND DECISION-MAKING STAND

At first let us in brief recall the main concepts of random processes and random fields. It is worth noticing that further random and stochastic concepts will be used in the same sense.

Stochastic variables or processes in the paper will be perceived as equivalents of the determined variables and processes in stochastic plane, when the analysed indicator is identified by the probability distribution of possibilities.

Each complex reality recognition method has the potential to accumulate stronger powers of cognition if it covers wider fields of research, but, more importantly, it would retain the possibility to refer to a greater amount of theoretically and pragmatically reasoned principles.

Further purpose of this paper is to reveal the fact that our applied method of decision management in financial markets corresponds with broad and powerful canons of financial markets' processes recognition system. The story would go on concepts and structures of stochastic (random) processes and random fields, which are gaining wider applications in various situations of complex financial processes analysis, when these structures and fields assume the features of deep recognition and accumulate the power and universality of many basic recognition principles.

To understand and identify Markowitz Random Field (MRF) in our proposed decisions management system in financial markets it is necessary to recall a few definitions described below.

### Probability space = probability field

By the term *probability space* –  $\{\Omega, \Phi, P\}$  – we call the entirety of three objects: the set of elementary events ( $\Omega$ ),  $\sigma$ -algebra ( $\Phi$ ) of the subsets  $\Omega$  and the probability measure defined for each set  $A \subset \Phi(P(A))$ :  $P(\Omega) = 1, P(\Phi) = 0$ .

**Random process (RP)**

Random process in the probability space  $\{\Omega, \Phi, P\}$  will be called the family of random events  $\xi_t(w)$ , which depends on the scalar parameter  $t \in T$ . It is understandable that T is a subset of real numbers' set.

This subset is called a *determination area* of the random process. The space X, where random process assumes its values, is called a *random process phasic* (transformation from one to another) *space*. That way the classification of a random process emerges: scalar, complex, vectorial.

**Random variable (RV) and random experiment**

Random variable is the value measured in random experiments. The experiment is an observation of event becoming under certain conditions. The experiment is called determined if the result of becoming is known (realized) in advance. If at particular or the same conditions different events may appear that negate each other, such experiment is called random (stochastic) or probabilistic.

**Random variable in probabilistic space**

Random variable in probabilistic space  $\{\Omega, \Phi, P\}$  is called any  $\Phi$ -measurable function defined in space  $\Omega$ .

**Measurable function with regard to algebra  $\Phi$**

Function  $\xi(w)$ , which satisfies the condition  $\{w, \xi(w) \in \Delta\} \in \Phi$  for each interval  $\Delta \subset R$  is measurable regarding to  $\Phi$  or simply  $\Phi$ -measurable.

**Random Field (RF)**

The random function of a few scalar variables is called a random field. More specifically, if we have a probability space  $\{\Omega, \Phi, P\}$  and a subset  $D \subset R^n$  of the n-dimensional space  $R^n$ , a function  $\xi(w; x_1, x_2, \dots, x_n) = \xi(w, X)$ , defined to all  $w \in R$  and  $x \in D$ , is called a random field, to define in set D, if it is measurable according to  $w$  for each set of  $x_1, x_2, \dots, x_n$ .

Often the notion is used that a random field is a random process, determination area of which is a certain continuous set.

If the function's  $\sum(w; x_1, x_2, \dots, x_n)$  set of values is  $R$  – we have a scalar field, if  $R^n$  – then a vectorial field. Attention should be paid to the case when the field is defined by  $DX [O, T]$ , where D is a certain subset of  $R^n$ , and [OT] – the time interval.

**Markowitz random field (MRF)**

In Markowitz random fields' probability space  $\{\Omega, \Phi, P\}$  we will take such family of random variables  $\xi(w)$ , under which the definition area is an interval of continuously changing standard deviation  $\Delta_\sigma : \sigma \in \Delta_\sigma$ , and the points of efficient surface  $\Omega$  form a set of elementary events.

**Event algebra**

The event algebra is such a subset  $\Phi$  of possibilities' set  $\Omega$ , which satisfies the following conditions:

- 1)  $\Omega \in \Phi$ ;
- 2)  $\emptyset \in \Phi$ ;
- 3) If  $\Delta \in \Phi$ , then  $\bar{\Delta} \in \Phi$ ;
- 4) If  $A \in \Phi$  and  $B \in \Phi$ , then  $A \cup B \in \Phi$  and  $A \cap B \in \Phi$ .

Note. If for every infinite set of events belonging to F set  $\{A_1, A_2, \dots, A_n\}$  and  $\bigcup_k A_k \in \Phi$  and  $\bigcap_k A_k \in \Phi$ , then we have a  $\sigma$ -algebra of events.

**3. MARKOWITZ RANDOM FIELD. HOW TO GET THERE?**

Let us invoke the discreet description technique of reality for further discussion. The selected investment scheme can be described by the following characteristics:

- 1) For the analysed  $k$  investment assets the capital returns are described as random values  $\xi_1, \xi_2, \dots, \xi_k$ .
- 2) The objective is to invest a unit capital, along with that analyzing all the capital distribution possibilities among the separate assets. This can be done in the following manner: let us have the so-called *even capital conversion matrix* from the first assets to the following:

$$(W_{ij}) : i = \overline{1, n}, j = \overline{1, k}, W_{ij} \geq 0 \text{ ir } \sum_{j=1}^k W_{ij} = 1. \quad (1)$$

The first row of the matrix ( $w_{1,1}=1, w_{1,2}=0, \dots, w_{1,k}=0$ ), showing the full concentration of the capital in the first asset, is substituted by other rows, revealing transformation of the capital to the subsequent assets until the last row ( $w_{n,1}=0, w_{n,2}=0, \dots, w_{n,k}=1$ ) presents that the whole capital invested was allocated to the last asset. The number of rows  $n$  was identified with number of investment trials, which will be denoted as  $I_1, I_2, \dots, I_n$ .

- 3) The result of every trial can be expressed by the following dependence:

$$\begin{pmatrix} I_1 \\ I_2 \\ \dots \\ I_n \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1k} \\ w_{21} & w_{22} & \dots & w_{2k} \\ \dots & \dots & \dots & \dots \\ w_{1k} & w_{2k} & \dots & w_{nk} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_k \end{pmatrix} \quad (2)$$

or simply

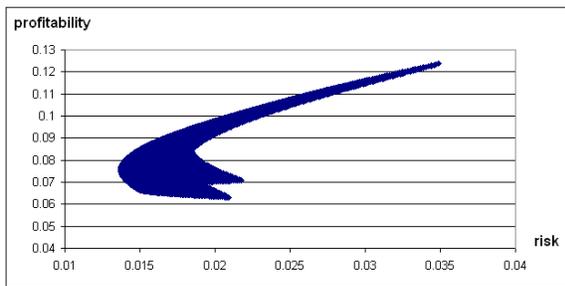
$$I_i = w_{i1}\xi_1 + w_{i2}\xi_2 + \dots + w_{ik}\xi_k \quad (3)$$

for every  $i = \overline{1, n}$ .

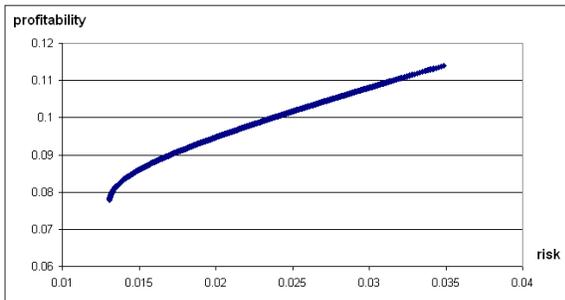
- 4) From the point 3 it is clear that every investment trial is a random value, determined by the stochasticity of assets and the selected investment structure.

Thus, every trial is a random value, which is described by its possibilities' probability distribution. Continuing the discreet reality presentation logics and admitting that every investment trial as a random value can be fully described by  $p$  number of its possibilities, the investment process should be analysed as a behaviour of  $n \cdot p$  possibilities' set. In order to reveal the dependences and consistent patterns of this set, the following procedures should be performed:

1) The initial step – it is the search for Markowitz portfolio, consisting of possible profitability values' set of the  $k$ -assets (we will continue to operate with these assets), measured by possibilities' mean and standard deviation, (Fig. 1, section *a*) and its efficient frontier (Fig. 1, section *b*).



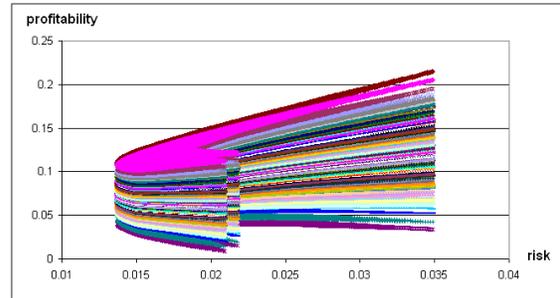
a)The set of “mean – standard deviation” portfolio possibilities



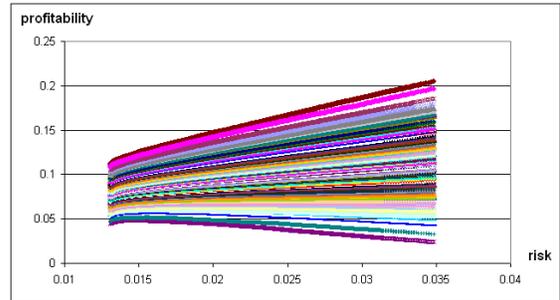
b)The efficient frontier

Fig. 1. Markowitz portfolio of three assets “standard deviation – the mean profitabilities” set of possibilities (section *a*) and efficient frontier (section *b*)

2) In order to reflect not only the mean but also all the possibilities, not only the information provided by the “standard deviation – mean” portfolio values is used, but all the “standard deviation – quintiles (percentiles)” (Fig. 2, section *a*) and the set of its efficient frontiers (Fig. 2, section *b*) is analysed.



a) “Standard deviation – percentiles” sets of profitability possibilities

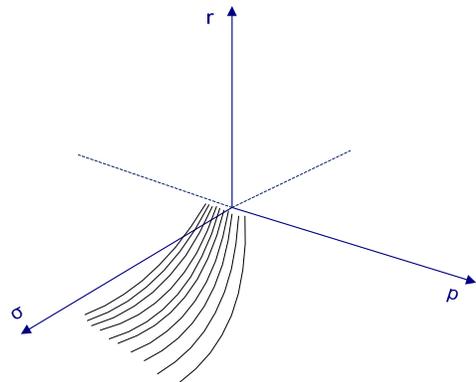


b) The set of efficient frontiers

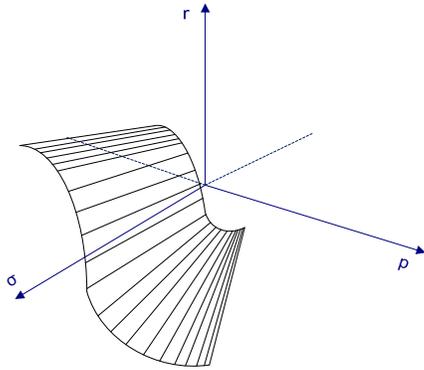
Fig. 2. Markowitz portfolio of three assets “standard deviation – percentiles” sets of profitability possibilities (section *a*) and its efficient frontiers (section *b*)

3) In order to achieve evidence how information is reorganized, conforming to the canons of the formed Markowitz random field (Fig. 4),  $(\sigma, p)$  plane (Fig. 2, section *b*) will be shifted to a horizontal position and will be rotated at a  $180^\circ$  angle (Fig. 3, section *a*), along with that the third coordinate –  $r$  (reliability) – is included (Fig. 3, section *b*).

Note. Fig. 3 presents only schematical view of the performed transformations. In terms of analytical geometry, Fig. 2 is transformed into Fig. 4.



a) Visualization of the initial step of efficient frontiers transformation to the three-dimensional space



b)The lift of efficient frontiers for efficient surface formation

Fig. 3. The sequence of three-dimensional efficient set formation of portfolio return possibilities ( $\sigma$  – standard deviation,  $p$  – value of return,  $r$  – reliability of return).

Now each of the efficient frontiers (Fig. 3, section a) will be lifted to the particular level that describes the corresponding percentile (Fig. 3, section b) and the so-called efficient surface will be formed (Fig. 4) [7] named after the Markowitz efficient frontier [5].

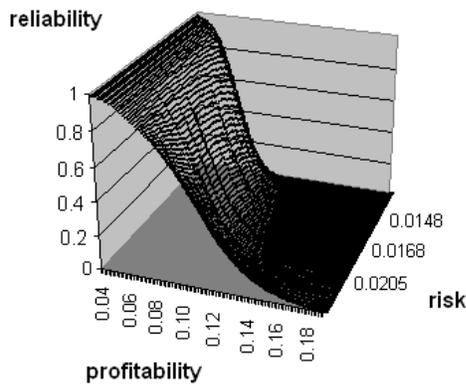


Fig 4. Efficient surface

Fig. 4 illustrates how a change (growth) of portfolio standard deviation changes the investment portfolio probabilities' distribution of possibilities, which in this particular case is a member of the corresponding family of random variables. Its type and parameters can be evaluated using analytical calculations or identified according to the comprised form of survival function [11]. Moreover, a continuously decreasing distance between contiguous standard deviations and fining the percentiles to mili-percentiles and etc., the efficient surface will clearly gain the features of marginal continuous surface, which allows carrying out evaluations that require such continuity.

To talk about the formed structure as a random field we have to remember that for each  $\sigma$  value corresponding to the survival function, there exists a

random variable, and practically every utility function is  $\sigma$ -measurable function.

Efficient surface is the source of exceptionally important and universal information on investment decisions' possibilities:

First, the efficient frontier “standard deviation – quintile”, raised to the respective quintile's level of efficiency, presents how the increases of return and riskness interact on the determined level of investment possibilities' reliability. This is a very useful information, applied in the process of best investment possibility selection.

Second, on every risk level  $\sigma_0$ , particularly – in every intersection point of the plane  $\sigma = \sigma_0$  with efficient surface – the probability distribution of all investment return possibilities on the determined riskness level is revealed. However, the initial graph is composed as a respective survival function  $P\{\xi \geq x\} = 1 - P\{\xi < x\}$  [10]. Thus, in the mentioned section the following composition is present: survival function, accumulated distribution function and density function (Fig. 5).

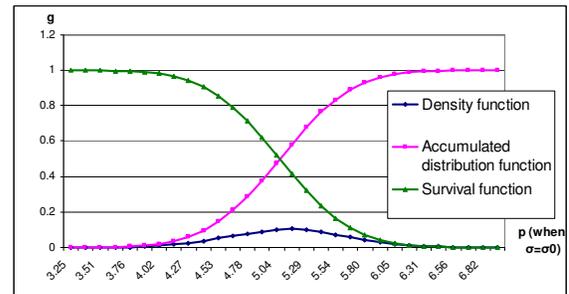


Fig. 5. The view of the plane  $\sigma = \sigma_0$  intersection with efficient surface and the functions derived from that surface

Third, the efficient surface is composed from the efficient points, i.e. the points having the highest returns under determined risk.

Thus, the efficient surface can be treated as a stochastic family of events  $\xi_\sigma(w)$  when  $\sigma$  varies in the interval  $[\sigma_{min}, \sigma_{max}]$ .

#### 4. MARKOWITZ RANDOM FIELD AS A STOCHASTIC OPTIMIZATION TOOL

Since the efficient surface is composed of the “standard deviations – a portfolio of possibilities” sets of efficient frontiers, thus it can be claimed that this way the main function of stochastic optimization is performed – only those possibilities are left that give maximum profitability at the selected risk. Therefore, it is enough to choose the point on the survival function which is determined in the best way by the utility function  $u = u(p; \sigma; r)$  terms.

Fig. 6 shows that the influence of randomness hardly touches the realization of the efficient surface. We can see from the figure that the surface realizations have no significant influence on the decision, i.e. the solution (obtained portfolio structure) practically does not depend on the possible realization it has been searched for.

In turn, the interval of random realizations, when a discrete substitution of a continuous distribution becomes more adequate, i.e. when minimum value of step  $w$  is decreasing, random realizations converge to a quintile line (izoguarantee), connecting quintiles of the same reliability level under different risk levels. There is no doubt that in such a case when survival functions of the random variables are continuous, a continuous surface of survival functions will be formed with all the possibilities to analyse interdependencies and interactions of the three parameters – profitability, reliability (guarantee) and risk.

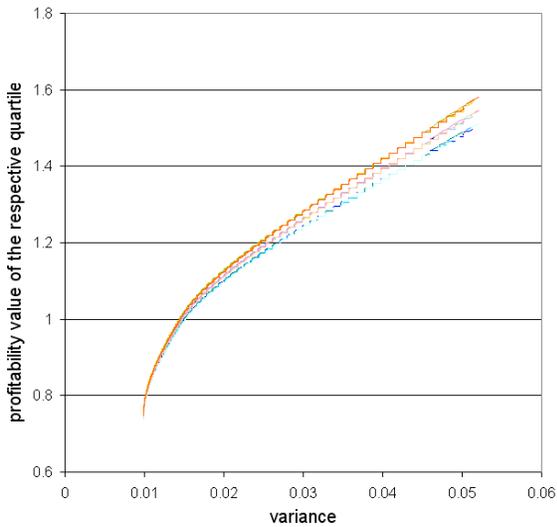
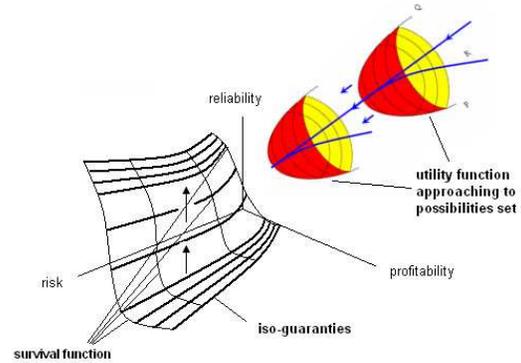
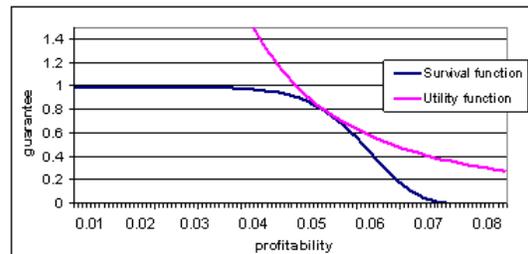
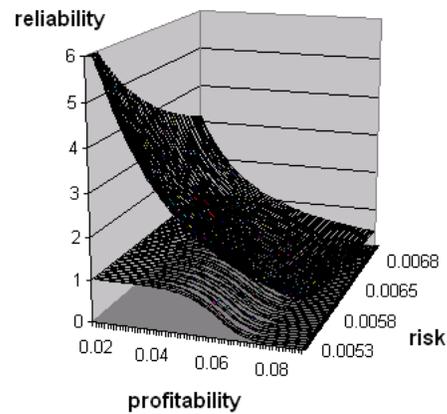


Fig. 6. Dispersions of quartiles' efficient frontiers at different realizations

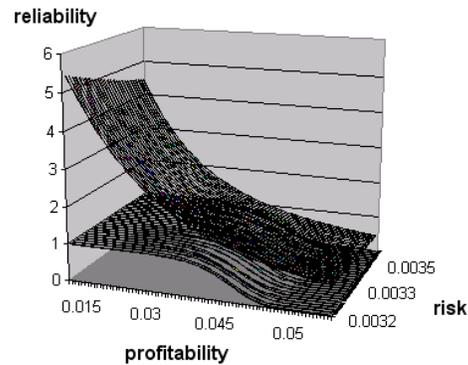
Fig. 7 illustrates how to select the optimal (according to utility function) solution. The cases of three selected capital markets are presented. Section *a* shows the general view of utility function, which at the point when its utility level is sinking, approaches to the possibilities (efficient surface), and at sections *b*, *c* and *d*, representing the decision search in different stock markets and each section comprising of the two graphs, we can see how that particular unique solution is found. The unique solution is because the efficient surface and utility surface are convex regarding to each other.

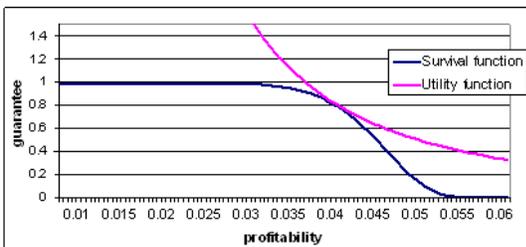


a) The scheme of the best possibility choice [10]

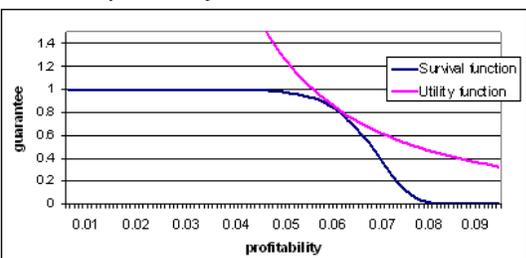
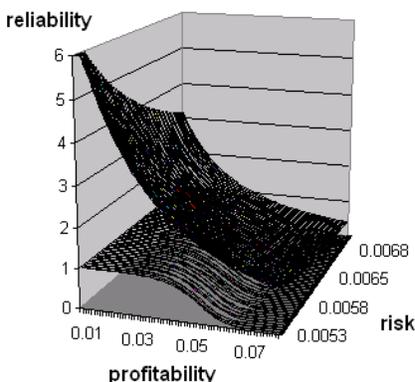


b) The best possibility choice in Australian stock market





c) The best possibility choice in NYSE stock market



d) The best possibility choice in UK stock market

Fig. 7. The final of optimization (section a). The best possibility choice (b, c and d, sections)

Applying a certain forecasting system and from the possible options choosing possibly the best possibility the system's potential in the selected market on the basis of historical data was verified. The experimental results have shown the return above the average.

The particular quantitative results are not presented here due to the shortage of space, however, they are available upon request. Also, the results of the presented methodology application in selected capital and exchange markets can be found in the researches previously published by the authors [7, 8, 9].

### 5. CONCLUSIONS

1. As investment assets' returns possess the main characteristics of the random values and random processes, the MRF appears to be the appropriate analytical tool for the proper analysis of such processes.
2. Applying the canons of MRF, the geometrical representation of optimal investment solution search is

transformed from two-dimensional plane to three-dimensional space (profitability, reliability and riskness) thus allowing to construct an efficient surface.

3. Along with efficient surface, constructing a three-dimensional utility function allows investors to find an optimal solution in the point of intersection of these two convex surfaces. Therefore the proposed methodology in fact corresponds with the main principles of stochastic optimization.

### 6. REFERENCES

- [1] Y. Fang, K.K. Lai, S.-Y. Wang, "Portfolio Rebalancing Model with Transaction Costs Based on Fuzzy Decision Theory", **European Journal of Operational Research**, Vol. 175, 2006, pp. 879-893.
- [2] C.W.J. Granger, Z. Ding, "Some Properties of Absolute Return: An Alternative Measure of Risk", **Annales d'economie et de statistique**, Vol. 40, 1995, pp. 67-91.
- [3] K. Kerstens, A. Mounir, I. Van de Woestyne, "Geometric Representation of the Mean-Variance-Skewness Portfolio Frontier Based upon the Shortage Function", **Document de travail du LEM 2008-21**, UMR 8179, Brussels, 2008, 32 p.
- [4] H.E. Leland, "Optimal Portfolio Management with Transactions Costs and Capital Gains Taxes". **Working Paper RPF-290**, IBER, UC Berkeley, December 1999, Available at SSRN: <http://ssrn.com/abstract=206871> or doi:10.2139/ssrn.206871.
- [5] H. M. Markowitz, "Portfolio Selection", **Journal of Finance**, Vol. 7, No. 1, 1952, pp. 77-91.
- [6] S. Ortobelli, S.T. Rachev, F.J. Fabozzi, "Risk Management and Dynamic Portfolio Selection with Stable Paretian Distributions", **Journal of Empirical Finance**, Vol. 17, 2010, pp. 195-211.
- [7] A.V. Rutkauskas, "Adequate Investment Portfolio Anatomy and Decisions, Applying Imitative Technologies", **Economics**, Vol. 75, 2006, pp. 52-76.
- [8] A.V. Rutkauskas, V. Stasytytė, "The Double Trump Portfolio as the Core of Sustainable Decision Making Strategy in Currency Markets", **The 10th World Multi-Conference on Systemics, Cybernetics and Informatics**, July 16-19, 2006, Orlando, Florida, USA. International Institute of Informatics and Systemics, 2006, pp. 57-62.
- [9] A.V. Rutkauskas, A. Miečinskienė, V. Stasytytė, "Investment Decisions Modelling Along Sustainable Development Concept on Financial Markets", **Technological and Economic Development of Economy**, Vol. 14, No. 3, 2008, pp. 417-427.
- [10] A.V. Rutkauskas, V. Stasytytė, "Effectiveness, Reliability and Subject Risk – Shaping Drivers for the Set of Possibilities and Utility Function when Investment Decision is Made under Uncertainty", **The 6th International Scientific Conference "Business and Management – 2010"**, May 13-14, 2010, Vilnius, Lithuania, pp. 176-183.
- [11] A.V. Rutkauskas, V. Stasytytė, "Optimal Portfolio Search using Efficient Surface and Three-Dimensional Utility Function", **Technological and Economic Development of Economy**, Vol. 17, No. 2, 2011, pp. 305-326.
- [12] A.K. Singh, R. Sahu, Sh. Bharadwaj, "Portfolio Evaluation Using OWA-Heuristic Algorithm and Data Envelopment Analysis", **The Journal of Risk Finance**, Vol. 11, No. 1, 2010, pp. 75-88.

## Multi-Attribute Value Measurement Approach to Assess Business Sustainability

Jelena STANKEVICIENE

Department of Finance Engineering, Vilnius Gediminas Technical University  
Vilnius, Sauletekio ave. 11, LT-10223, Lithuania

### ABSTRACT

The relationship between balanced scorecard (BSC) and company's value as an area of inquiry has sustained interest among researchers over the past decade. More recently, some have challenged the notion differential value among companies through multi-attribute value measurement (MAVM) as it evolves into a ubiquitous sustainability-like set of technologies. This study examines the relationship between the BSC implementation through economic value added (EVA) and its differential business value to firms using analytical hierarchy processing (AHP) method as one of application of multi-attribute value measurement. By drawing upon multi-attribute value theory, this paper develops a process framework to assess business sustainability and overall performance of firms that effectively deploy and use balanced scorecard (BSC). Using data from Lithuanian telecommunication company, the study finds empirical support for the developing an individualized balanced scorecard, combined with EVA, which brings additional value created along with a number of sustainability oriented dimensions. These findings are discussed as optimistic results for developing multi-attribute value theory as a major causal driver of sustainability. The paper has practical relevance for the effectiveness of the balanced scorecard inside a telecommunication company, particularly as it is gaining additional importance due to the fact that the concept brings increasing value for a company.

**Keywords:** multi-attribute value measurement (MAVM), sustainability, balanced scorecard (BSC), analytical hierarchy process (AHP), economic value added (EVA), Simple Additive Weighing (SAW).

### 1. INTRODUCTION

Multiple-attribute value theory (MAVT) can be used to address problems that involve a finite and discrete set of alternative policies that have to be evaluated on the basis of conflicting objectives (Beinat, Nijkamp 1998). For any given objective, one or more different attributes or criteria are used to measure the performance in relation to that objective. These aspects, the impacts of all alternative options for all attributes, which are usually, measured on different measurement scales. MAVT can be used to assess the sustainability of a policy because it allows to simultaneously taking into account indicators that refers to the three dimensions of sustainability: the economic, the social and the economic one. In other words, it can be used to combine information in such a way that it can clarify sustainable development aspects. MAVT can incorporate the following aspects as separate criteria to compare alternative policies: (de-)coupling aspects, adaptability and (ir-)reversibility (Sharifi, Herwijnen 2002, Herwijnen 1999).

Multiple-attribute value theory (MAVT) is a compensatory technique. This means that the method does allow compensation of weak performance of one criterion by a good performance of another criterion. MAVT aggregates the options' performance across all the criteria to form an overall assessment. Well known and often used MAVT methods are Simple Additive Weighing (SAW) and Analytic Hierarchy Process (AHP).

The paper focuses on a complementary system of managerial metrics, which was created; the system gives opportunity to link EVA system to the Balanced Scorecard (BSC) through multi-attribute value measurement (MAVT) and analytical hierarchy processing (AHP) to assess business sustainability. The balanced scorecard was presented as a mechanism for identifying value drivers and drilling down into the operations of the telecommunication company. Afterwards, these complementary frameworks were combined through the MAVT and AHP methodology, to develop a comprehensive measurement system for assessing the overall performance and sustainability of telecommunication company.

### 2. STRATEGIES FOR SUSTAINABLE VALUE CREATION OF ENTERPRISE

The conception of Value Based Management was created in the late 1980s. The new approach put an emphasis on management style that included finance and strategy. From this standpoint it is possible to analyze value drivers and identify corporate strategy that led to better creation of company's value. Due to the reason of integrating strategy and value, Value Based Management is the key point to finding better investment opportunities and following strategic movements. Value Based Management used in practice showed that constant decisions cannot be made and strategy development must be a continuous process. (Clackworthy 2006, Fong *et al.* 2007). Managing for value is a competitiveness advantage of company. Value Based Management involves transforming behavior in a way that encourages employees to think and act like owners (Martin, Petty 2001). Understanding what influences the value of the business is part of Value Based Management. According to Copeland, Koller and Murrin (1996), the organization cannot act directly on value. It has to act on things it can influence, such as customer satisfaction, cost, capital expenditures, and *etc.* The closer look will be taken at analyzing value drivers through financial perspective. Researchers vary as to the number of these value drivers; for instance, five (Ruhl and Cowen 1990), six (Moskowitz 1988), and seven (Rappaport 1998, Mills and Print 1995, Mills *et al.* 1992). Turner (1998) has identified eight value drivers. Creating value of the company, the most accurate value drivers are: weighted average cost of capital (WACC), return on invested capital (ROIC), growth rate (g), and economic value added (EVA). There are several reasons for

ROIC – a ratio for measuring firm's operating profitability that shows the expected rate of return on net new investment – to be a widely used measure:

1. The historical performance of a company can be evaluated;
2. It is component when calculating company's earnings growth rate;
3. It is the leading component in economic value added (EVA);
4. It indicates if company's wealth is being created or destroyed.

Economic Value Added measures true economic profit, or the amount by which the earnings of a project, an operation, or a corporation exceed (or fall short of) the total amount of capital that was originally invested by the company's owners (Bloomsbury business library 2007). Accounting net income is not enough if it doesn't translate into cash returns that exceed the individual investor's own opportunity costs. When a company is able to earn more than it has invested, the business is adding value, whereas, if there are not enough earnings generated, the value is destroyed (Bhalla 2004; Soenen 2005, Fletcher and Smith 2004). At the beginning, company is concerned on future returns, for this reason FCF and EVA are metrics to evaluate future possibilities on investments made and at the business growing stage when returns from investments are generated, that are measured with ROIC as this measure is used for past results analysis and in this way can be compared to expectations if they are fulfilled. In a growing company that is making investments, the goal from value drivers' analysis is to find factors that are making impact to investment returns, as it reduces or increases value. The problem that may lead to wrong interpretation of company's performance is a growth rate, as it reaches its maximum at the beginning of business, while a fast and aggressive expansion policy is being held, and then lowers a lot, when company has fulfilled its business development projects. The higher a firm's operating profit margin and capital turnover, the higher its operating profitability. Researchers (Jie, Xun 2007, Kwan 2006, Lert 2005, Ly *et al.* 2007, Miller *et al.* 2004, Meng, Weerasinghe 2006, Parise, Casher 2003, Roman *et al.* 2006, Yaozhong, Øksendal 2007, Martins, Galdi 2006) are putting an emphasis on value improvement through value drivers. According Hawawini and Viallet (2005) a higher operating profit margin is achieved by:

1. Increasing sales through higher prices and/or higher volume at a higher rate than operating expenses;
2. Reducing operating expenses at a higher rate than sales.

A higher capital turnover is achieved through a better use of the assets required to support the firm's sales activity:

1. A faster inventory turn;
2. A shorter collection period for the firm's receivables;
3. Fewer fixed assets per EUR of sales.

Returns from investments can be measured according the following measures: return on assets (ROA), return on equity (ROE), and cash flow return on investment (CFROI) and other.

### 3. INTEGRATING EVA AND THE BALANCED SCORECARD

Kaplan and Norton (1992) presented the balanced scorecard (BSC) as a performance measurement tool. The balanced scorecard has gained prominence as a way of integrating financial and non-financial performance measures into an

overall control system (Atkison *et al.* 1997, Hoque, James 2000, Malina, Selto 2001, Simons 2000). It appears promising to use the balanced scorecard methodology to integrate environmental and social management with the general management of a firm (Figge *et al.* 2001). For many years the telecommunication industry is a good example of detailed performance measurement system (Pieper 2005). This industry has a specific balanced scorecard implementation, which is not relevant for other industries. Managers of many companies seek to improve profitability in a competitive environment, profitability is needed for increasing value of a company. This is the reason of implementing value-based measures, one of which is Economic Value Added (EVA). Companies should switch beyond narrow metrics to the usage of EVA like a strategic decision tool; this will help in improving of implementation of value based management (VBM). The process should include identified value drivers, integration of budgeting with strategic planning, and development of a sophisticated performance measurement system (Stankeviciene, Sviderske 2010).

The aims of the approach are to satisfy four telecommunication company's requirements regarding balanced scorecard perspectives:

1) *Strategy*: determine strategic value of telecommunication company. This strategic value to the telecommunication organization still remains unanswered and needs further consideration.

2) *People*: the recognition that the success of the telecommunication organization depends on how well its people (telecommunication professionals) are valued.

3) *Process*: apply the telecommunication business model to effectively prove added efficiency and productivity within the telecommunication organization when a remote customer monitoring system is deployed, and to prove that telecommunication service costs may potentially be reduced. A telecommunication business model must encompass all aspects of a telecommunication framework that is directed towards a metric-based strategic plan with measurable relationship across the internal and external telecommunication organization processes - from a top-down management view of how policies, procedures, processes and operations should serve the telecommunication organization to the bottom-up view of how the customer processes are made more efficient further ensuring that IT processes become more cost efficient in the future ongoing support environments (in source, outsource or managed services). There is currently no way of determining the ongoing support costs for the telecommunication organizations.

4) *Technology*: telecommunication company's primary purpose is to provide qualified telecommunication service to customers. The technology of production should also be improved and the cost reduced through technological progress.

These four telecommunication company requirements would further need to be linked to measure systems from a Strategy, People, Process and Technology performance perspective.

Combining EVA and BSC to achieve strategic alignment to enhance value of both tools by combining them in a single application could be an effectively usage of EVA calculation to drive the definition of categories within Balanced Scorecard's financial perspective.

#### 4. THE MAVT AND AHP-BASED VALUATION FRAMEWORK

AHP, as one of the MAVT instruments, is a multi-criteria decision support system, developed by Saaty (1996) and discussed by Podvezko (2009), that allows a decision maker to structure a complex problem in the form of a hierarchy. Generally, the AHP methodology is a method, which splits out a sophisticated case into several components, organizes these components (variables) into a hierarchic order, settles numeric values to subjective judgments on the relative importance of each variable, and synthesizes the judgments in order to identify which variables have the highest priority and should be acted up to influence the outcome of the situation. AHP consolidates personal values and judgments in a logical way.

Some applications of the AHP model have been already made in a plenty of different problems in order to improve decision making process (Hogan, Olson 1999, 2004, 2006; Ishizaka, Lusti 2004; Travana 2004). The most important aspect of the AHP model includes the ability to reconcile sophisticated quantitative and qualitative information in the process of decision making. Furthermore, the simplicity of use and ability to adjust the consistency into the decision-making process bring additional benefits of AHP method.

Hogan, Olson and Sillup (2009) propose that there are four general steps required to implement the AHP. First, the decision maker identifies the criteria and determines their relative importance in achieving the goal and identifies the sub-criteria and determines their relative importance in achieving the related criterion. Second, the decision maker determines the relative importance of the ratings categories for each of the sub-criteria. Third, the alternatives are evaluated in the context of each of the ratings categories. Finally, the results are synthesized to compute the relative contribution of the alternatives in achieving the goal.

The first three steps are the same as in most MCA methods. Step 4 is specific for MAVT. MAVT is based on the assumption that in every decision problem a real value function  $U$  exists that represents the preferences of the decision maker. This function aggregates for each alternative  $a_j$  ( $j = 1 \dots M$ ) the criteria  $c_i$  ( $i = 1 \dots N$ ) that are under consideration by the decision maker.

#### 5. INTEGRATION OF EVA AND BSC USING AHP IN A TELECOMMUNICATION COMPANY

A balanced scorecard needs to be used to realize the full value of a telecommunication company. By Schneiderman (1999) a balanced scorecard fails when company management is not familiar with information from identified relevant performance measures. As noted by Jennings, Graham (1997), the balancing of long-term development with short-term requirements for survival is a particularly important issue for companies – failing to get the budget process aligned with the strategic goals of the enterprise can make achieving this balance harder.

In order to implement a balanced scorecard system to the particular company, some steps of implementation should be fulfilled.

*Strategic Analysis.* The significance of clearly identified objectives is mentioned by Lingle, Schiemann (1996), who warns from so named “fuzzy objectives”, because this is a reason for shortage of implementation and targeted results.

*Strategy Mapping.* The managers carried out a comprehensive strategy mapping of the strategic topic. Strategy mapping, as it is described by Pandey (2005), is a pictorial description of the strategy and its elements. Balanced scorecard designs in telecommunication companies normally include an elaborate process for identifying measures in order to give clear picture of company’s direction towards achievement of its goals (Olve, Roy, Wetter 1999).

The telecommunication company’s vision is to implement management system into the company and be recognized by stakeholders and interested parties as a leading telecommunication company by achieving a superior level of management for employees, consumers and customers. The main strategic plans of telecommunication company are: (1) to maintain existing market share and to invest more in different markets; (2) to develop economically good platform for company’s future; (3) to reach growth of sales and to hold leading position; and (4) to develop and maintain relationships with governmental, regulatory and interested party groups protecting the shareholder’s interest.

**Table 1.** Scorecard for telecommunication company (adopted and improved from Stankeviciene, Sviderske 2010)

	<i>High performance Organization</i>	<i>Systematic Execution &amp; Implementation of system requirements</i>	<i>Stakeholder service excellence</i>	<i>Excellence in financial performance</i>
<b>Objectives/Critical Success Factors</b>	<ul style="list-style-type: none"> <li>- Attract key employees for a company;</li> <li>- Provide health working environment;</li> <li>- Increase team work and spirit of a company.</li> </ul>	<ul style="list-style-type: none"> <li>- Allow for easy storage, access, compilation and dissemination of data, know-how and experience;</li> <li>- Ensure good communication channels.</li> </ul>	<ul style="list-style-type: none"> <li>- Provide feedback and recommendations to internal stakeholders regarding performance and compliance with guidelines and standards to improve the business;</li> <li>- Insure all information is presented in-time, accurately and thoroughly.</li> </ul>	<ul style="list-style-type: none"> <li>- Improvements in operational performance;</li> <li>- Support appropriate targets to achieve business needs;</li> </ul>
<b>Key performance</b>	<ul style="list-style-type: none"> <li>- Employee satisfaction rate;</li> <li>- Training compliance rate;</li> <li>- Successful job rotation.</li> </ul>	<ul style="list-style-type: none"> <li>- Audit recommendation implementation score;</li> <li>- Information sharing score.</li> </ul>	<ul style="list-style-type: none"> <li>- Business improvement rate;</li> <li>- Management satisfaction rate;</li> <li>- External rating score.</li> </ul>	<ul style="list-style-type: none"> <li>- Risk management score;</li> <li>- Expense spending control;</li> <li>- Performance score;</li> <li>- Target achievement rate.</li> </ul>

According to telecommunication company’s strategic plan, there are four main objectives: to increase sales growth, profitability, customer and employees’ satisfaction and their retention. The AHP framework was used by management to

distinguish the relative importance of each of these strategic drivers and develop an index to predict overall telecommunication company performance.

To construct AHP model connecting telecommunication company's EVA mission and objectives with the balanced scorecard. An AHP model must have an exact goal, in case of telecommunication company it is to achieve an EVA target of 1 million EUR. A hierarchy is constructed to link the goal with alternatives. Alternatives can be viewed as strategic options for achieving the goal. Afterwards, the criteria are selected. They represent the measures used to access the performance of financial and non-financial drivers of value in telecommunication company (Ponikvar, Tajnikar, Pušnik 2009). This hierarchy is illustrated in Table 1 and is identified in such way:

1. Goal: Maximize stakeholder and shareholder value by achieving an EVA of 1 million EUR.
2. Strategic Alternatives: increase profitability, grow in sales, satisfy customer needs and satisfy employees' expectations as well as to retain number of customers and employees in a company.
3. Criteria: Each key performance measure is located under each perspectives of balanced scorecard.

To relate each measure with its underlying balanced scorecard perspective is one method how to be sure that each perspective has been taken into account in the process of strategy development and performance management.

The weights of each key performance measures should be selected using AHP model. AHP models need twosome comparisons to evaluate the relative importance of each of the strategic objectives and the importance of each of the criteria in meeting the EVA goal and each strategic objective. Such comparisons are presented by several modes, including verbal, numerical and graphical approaches. The weights are based on management's evaluation of the significance of the performance measure of telecommunication company's unique competitive environment. Customer and sales oriented objectives and measures were higher priorities than financial measures.

It was important to grow sales and to retain its customers in order to enhance telecommunication company's ability to capture competitor's customers.

To use key performance measures for creating an index to predict overall telecommunication company performance. Telecommunication company decided to create three possible alternatives of company's performance, depending on different targets, set by a company. Three scenarios were given and, quantifying the relative effect of each change, an index to predict the company's progress toward its strategic targets was constructed. Evaluated AHP measures can show possible company's problem areas, if the measure does not meet the target; and they also can indicate what company will do well, if the targeted goal is reached.

The maximizing criteria imply that, if their values are growing, the situation is getting better, while for minimizing criteria this means a worsening situation. The integration is achieved by normalization which helps to convert all the criteria values into non-dimensional, i.e. comparable quantities. Quantitative methods quantitatively evaluate each alternative determining the differences in the values obtained for the alternatives considered (Ginevičius 2008). The essence of multi-criteria evaluation can be clearly shown by the so-called Simple Additive Weighting (SAW) method expressed as:

$$S_j = \sum_{i=1}^m \omega_i \tilde{r}_{ij}, \quad (1)$$

where  $S_j$  is the value obtained in multi-criteria evaluation of the  $j$ -th alternative;  $\omega_i$  is the  $i$ -th criterion weight;  $\tilde{r}_{ij}$  is normalized value of the  $i$ -th criterion for the  $j$ -th alternative. As can be seen from the formula (1), the normalized values of the criteria are used to determine the multi-criteria evaluation (Ginevičius, Podvezko 2008; Ginevičius 2007).

If telecommunication company's decide to choose alternative 1 for strategic planning, the overall performance will be accomplished only by 92.13 %. Only one perspective perfectly reached the target, other performance measurement perspectives fell short of its target. By alternative 1 such financial measures as RONA (12.61 % better than targeted), WACC (31.12 %) and Accounts Receivable days (15.42 %) are even better than they were targeted, the better results are also presented by such customer perspective measures: revenue from value-added partnerships (37.44 %) and on-time delivery (0.87 %), only one measure of internal business perspective is twice better than it was targeted – innovation/learning perspective measures – employee retention measure brings better results than targeted (2.91 %). Alternative 1 has only several rates, which are in bad condition, this is an employee suggestion rate (-35.15 %), some decisions should be make to improve this area.

The alternative 2 suggests that 8 measures will bring better performance than it was targeted, they are: WACC (45.98 % better performance), customer retention rate (0.75 %), revenue from value-added partnerships (17.98 %), out-of stock rate (4.83 %), accounts payable days (18.81 %), revenue from new products (17.50 %) and employee training days (38.03 %). So, choosing this alternative, not too many measures will achieve the main target, but the main attraction is that the most weighted measures reach the target and this brings the real value increase to performance index. Some measures indicate poor future results, it means that managers could pay attention to the areas where the results are worse, these areas include employee suggestion increase in a company (-55.47 %), broken packages (-54.83%) and others. The overall performance index of alternative 2 is 106.55 %, it is even better strategic choice than the target.

The last choice is alternative 3 – the performance index of this alternative is 79.34 %, it is a little bit higher than alternative 1, but less than it is targeted. Target will be achieved by RONA (4.76 % greater than targeted), sales growth (23.32 %), accounts payable days (64.64 %), broken packages (30.435 %) and employee retention (7.41 %). The areas, where some changes should be made are concerning revenue from value-added partnership (-48.79 %), revenue from new products (-50.05 %) and employee training days (-22.02 %).

Furthermore, a value cycle of telecommunication company should be created, showing the process of value creation in a company. The first step is to understand and create company's main vision, afterwards, the direction should be made towards second step – strategic mapping. Telecommunication company had decided upon the priorities of balanced scorecard perspectives and strategic objectives. When the most important fields were clarified, the telecommunication company has invested in different business operations, such as people, process, customers and financial operations. Afterward, some performance results appear, and these results should be measured according to the value, which was brought to the telecommunication company after investment. In the case of telecommunication company the value is measured by economic value added (EVA), which is integrated in balanced scorecard

perspectives through AHP methodology. So, the performance is measured and the results are calculated, it means that based on obtained results, company's managers can make some conclusions about assess business sustainability.

This is the next step in the value cycle of a telecommunication company – knowledge, learning and assessing business sustainability. After the analysis of obtained results, managers establish new vision of a telecommunication company and the process circulated again and again.

And, finally, in order to understand the really best strategic choice for telecommunication company's value, Economic Value Added should be calculated according to results of performance. So, taking into account the fact that the main goal was to create a comprehensive measurement system for assessing the overall performance of telecommunication company through combining complementary frameworks (EVA and BSC) through the MAVT and AHP methodology, it can be proved, that the best alternative is second one, because it brings the best results for a telecommunication company, with value and performance increase.

## 6. CONCLUSIONS

1. The balanced scorecard is a mechanism for identifying value drivers and drilling down into the operations of the telecommunication company, whereas economic value added measures the created value for a company. The analytical hierarchy processing methodology helps to combine these two frameworks in order to develop a comprehensive measurement system for assessing the overall performance of telecommunication company.

2. The work has introduced a new framework for a telecommunication company to improve the implementation of value based management by adopting balanced scorecard in order to identify value drivers and develop a quantitative measurement system relating the telecommunication company's objectives of maximizing shareholder value.

3. The combination of EVA and BSC financial perspective has the influence to the remaining three perspectives, because if the goal is to grow sales, by increasing the average value of the product, which will lead to the purchasing behavior changes, the managers will have to institute new activities such as the increase in the quality of the delivery service. To see how the actions are really working in the system the financial outcomes should be calculated accordingly to the EVA methodology. The illustration of the achievement of overall strategy is given in using the Causal Chain example of BSC and EVA combination.

4. AHP valuation framework improves the company's capacity to predict strategy implementation on a real-time basis through increased timeliness and accuracy which gives the better opportunity to improve company's performance and to create value for future.

5. A rational framework which includes multicriteria methods is used. The maximizing and minimizing criteria are expressed in various dimensions and then integrated into one criterion. A rational approach to decision-making requires rules for the aggregation of used criteria into an evaluation score.

7. The value cycle for telecommunication company is a good strategic decision. After selected vision, balanced scorecard and strategic objectives are designed. Another step

is investment to business operations, appealing to the selected objectives, after which is performance and value added measurements. Through this analysis, knowledge and learning appear for the managers. Based on this, they create new vision for a telecommunication company and the value cycle repeats more and more.

## 7. REFERENCES

- [1]. L. G. Aidemark. **The meaning of balanced scorecards in the health care organizations.** Financial Accountability & Management, Vol. 17 No. 1, pp. 23-40, 2001.
- [2]. E. Beinat, P. Nijkamp. **Multicriteria Analysis for Land-Use Management.** Kluwer Academic Publishers: Dordrecht, 1998.
- [3]. V. K. Bhalla. **Creating Wealth Corporate Financial Strategy and Decision Making.** Journal of Management Research, Vol. 4 No. 1, pp. 13-34. 2004.
- [4]. S. Clackworthy. **Value management. Financial Management.** 36 p., 2006.
- [5]. T. Copeland, T. Koller, J. Murrin. **Valuation measuring and managing the value of companies.** New York: John Wiley and Sons, Inc. 550 p., 1996.
- [6]. A. Damodaran. **Investment valuation** [online]. 2002. Available from: <http://pages.stern.nyu.edu/~adamodar/> [Accessed 2008-05-21].
- [7]. M. J. Epsein and A. Rejc. **Identify, Measuring, and Managing Organizational Risks for Improved Performance: Management Accounting Guideline.** Management Accountants of Canada, 42 p., 2005.
- [8]. F. Figge, T. Hahn, S. Schaltegger, M. Wagner. **The Sustainability Balanced Scorecard – a tool for value-oriented sustainability management in strategy focused organizations.** Conference Proceedings of the 2001 Eco-Management and Auditing Conference. ERP Environment: Shipley; pp. 83-90, 2001.
- [9]. H. D. Fletcher, D. B. Smith. **Managing for value: developing a performance measurement system integrating economic value added and the balanced scorecard in strategic planning.** Journal of Business Strategies, Vol. 21 No. 1, pp. 1-17, 2004.
- [10]. P. Fong, M. J. Hills, C. S. Hayles. **Dynamic Knowledge Creation through Value Management Teams.** Journal of Management in Engineering, Vol. 23 No. 1, pp. 40-49. 2007.
- [11]. A. Ginevičius. **Quantitative evaluation of enterprise marketing effectiveness.** (In Lithuanian). Technological and Economic Development of economy, Vol. 13 No. 1, pp. 19-23, 2007.
- [12]. R. Ginevicius, S. Cirba. **Additive Measurement of Market Concentration.** Journal of Business Economics and Management, Vol. 10 No. 3, pp. 191-198, 2009.
- [13]. R. Ginevičius, **Normalization of quantities of various dimensions.** Journal of Business Economics and Management Vol. 9 No. 1, pp. 79-86, 2008.
- [14]. R. Ginevičius, V. Podvezko. **Multicriteria evaluation of Lithuanian Banks from the Perspective of their reliability for clients.** Journal of Business Economics and Management, Vol. 9 No. 4, pp. 257-267, 2008.
- [15]. G. Hawawini, C. Viallet. **Finance for executives: managing for value creation.** Ohio: South – Western Thomson learning. 606 p., 2002.

- [16]. M. Herwijnen. **Multiple-attribute value theory (MAVT)**. [online] Available from internet: [http://www.ivm.vu.nl/en/Images/MCA1\\_tcm53-161527.pdf](http://www.ivm.vu.nl/en/Images/MCA1_tcm53-161527.pdf)
- [17]. K. Hogan, G. Olson. **Evaluating potential acquisitions using the analytic hierarchy process**. *Advances in Mathematical Programming & Financial Planning*, Vol. 5, pp. 3-18, 1999.
- [18]. K. Hogan, G. Olson. **A multi-criteria decision model for portfolio allocation for the individual investor**. *Mathematical Programming: Applications of Management Science*, Vol. 11: pp. 3-16, 2004.
- [19]. K. Hogan, G. Olson, G. Sillup. **Identifying compounds for pharmaceutical companies' development tracks using a multi-criteria decision model**. *Applications of Management Science in Productivity, Finance, and Operations*, Vol. 12: pp. 89-108, 2006.
- [20]. Z. Hoque, W. James. **Linking Balanced Scorecard Measures to Size and Market Factors: Impact on Organizational Performance**. *Journal of Management Accounting Research*, Vol. 12, pp. 1-17, 2000.
- [21]. A. Ishizaka, M. Lusti. **An expert module to improve the consistency of AHP matrices**. *International Transactions in Operational Research*, Vol. 11, pp. 97-105, 2004.
- [22]. P. Jennings, G. Beaver. **The performance and competitive advantage of small firms: a management perspective**. *International Small Business Journal*, Vol. 15 No. 2, pp. 63-75, 1997.
- [23]. X. Jie, Y. Z. Xun. **Mean-variance portfolio selection under partial information**. *SIAM Journal on Control & Optimization*, Vol. 16 No. 4, pp. 986-1006, 2007.
- [24]. R. Kaplan, D. Norton. **Balanced Scorecard: Strategien erfolgreich umsetzen**. Schaffer-Poeschel: Stuttgart, 1997.
- [25]. R. Kaplan, D. Norton. **The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment**, Harvard Business School Press, Boston, MA, 2001.
- [26]. J. H. Lingle, W. A. Schieman. **From Balanced Scorecard to strategic gauges: is measurement worth it**. *Management Review* Vol. 85 No. 3, pp. 56-62, 1996.
- [27]. V. V. Ly, M. Mnif, P. Huyên. **A model of optimal portfolio selection under liquidity risk and price impact**. *Finance & Stochastics*, Vol. 11 No. 1, pp. 51-90, 2007.
- [28]. V. Martinez. **What is the Value of Using Performance Management Systems?** Centre for Business Performance, UK: Cranfield School of Management, 2005.
- [29]. Q. Meng, A. Weerasinghe. **Optimal portfolio selection strategies in the presence of transaction costs**. *International Journal of Theoretical & Applied Finance*, Vol. 9 No. 4, pp. 619-642, 2006.
- [30]. T. W. Miller, R. E. Mathisen, J. P. McAllister. **Estimates of the sensitivities of the value of the firm to profitability, growth, and capital intensity**. *Journal of Accounting & Finance Research*, Vol. 12 No. 7, pp. 20-33, 2004.
- [31]. N. Olve, J. Roy, M. Wetter. **Performance Drivers: A practical guide to using the Balanced Scorecard**. Wiley, UK, 1999.
- [32]. S. H. Pieper. **Reading the right signals: how to strategically manage with scorecards**. *Healthcare Executive*, Vol. 20 No. 3, pp. 9-14, 2005.
- [33]. V. Podvezko. **Application of AHP technique**. *Journal of Business Economics and Management*, Vol. 10 No. 2, pp. 181-189, 2009.
- [34]. N. Ponikvar, M. Tajnikar, K. Pušnik. **Performance ratios for managerial decision-making in a growing firm**. *Journal of Business Economics and Management*, Vol. 10 No. 2, pp. 109-120, 2009.
- [35]. D. Roman, K. Darby-Dowman, G. Mitra. **Portfolio construction based on stochastic dominance and target return distributions**. *Mathematical Programming*, Vol. 108 No. 2-3, pp. 541-569, 2006.
- [36]. V. Rutkauskas, J. Stankevičienė. **Formation of an investment portfolio adequate for stochasticity of profit possibilities**. *Journal of Business Economics and Management*, Vol. IV, No 1, pp. 3-12, 2003.
- [37]. T. L. Saaty. **The analytic hierarchy process**. Pittsburgh, PA: RWS Publications, 1996.
- [38]. M. Schneiderman. **Why Balanced Scorecards fail**. *Journal of Strategic Performance Measurement*, January Special Edition: 6, 1999.
- [39]. R. Simons. **Performance Measurement and Control Systems for Implementing Strategy**. Englewood Cliffs, NJ: Prentice Hall, 2000.
- [40]. J. Stankevičienė. **Real Estate Investment Portfolio Decision Making Under the Risk and Uncertainty**. In: *The 10th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2006: ISAS July 16-19, 2006 Orlando, Florida, USA: proceedings*. Vol I, pp. 51-56, 2006.
- [41]. J. Stankevičienė, I. Jachimavičienė. **A Multi-Objective Project Portfolio Formation Model: Case Study of Lithuanian Transport Sector**. In: *Proceedings of the 6th International Scientific Conference "Business and Management 2010"*. Vilnius, May 13-14, 2010, pp. 230-238. doi 10.3846/bm.2010.032
- [42]. J. Stankevičienė, T. Sviderskė. **Developing A Performance Measurement System Integrating Economic Value Added And The Balanced Scorecard In Pharmaceutical Company**. In: *Proceedings of the 6th International Scientific Conference "Business and Management 2010"*. Vilnius, May 13-14, 2010, pp. 239-247. doi 10.3846/bm.2010.033
- [43]. M. Travana. **A subjective assessment of alternative mission architectures for the human exploration of Mars at NASA using multicriteria decision making**. *Computers & Operational Research*, Vol. 31: pp. 11-47, 2004.
- [44]. J. Weber, U. Schaffer. **Balanced Scorecard and Controlling: Implementierung, Nutzen für Manager und Controller – Erfahrungen in deutschen Unternehmen**. Gabler: Wiesbaden, 2000.
- [45]. H. Yaozhong, B. Øksendal. **Optimal smooth portfolio selection for an insider**. *Journal of Applied Probability*, Vol. 44, pp. 742-752, 2007.
- [46]. S. D. Young, S. F. O'Byrne. **EVA and value-based management**. New York: McGraw-Hill, 2001.

## **Development of Bank Value Model**

**Jelena TITKO**

**Faculty of Engineering Economics and Management, Riga Technical University  
Riga, Latvia**

**and**

**Konstantins KOZLOVSKIS**

**Faculty of Engineering Economics and Management, Riga Technical University  
Riga, Latvia**

### **ABSTRACT**

Finance specialists predict a new wave of M&A activities in banking sector. The integral part of any M&A transaction is a valuation of an entity. Thus, bank valuation is one of the most actual issues in today's financial business. Some valuation specialists consider that a valuation of a financial institution can be undertaken mainly using earnings-based methods, in particular, Discounted Cash Flow (DCF) approach. However, we face some difficulties while using this method for valuation of Latvian commercial banks. One of them is the determination of discount rate, because the models of discount rate calculation were devised by foreign specialists for using in developed countries with mature stock markets. The goal of the paper is to develop a mathematical model as an alternative to existing company's valuation models. As a proxy for a bank's value the market capitalization was used. Taking into account, that shares of only few banks of Baltic States are quoted in the stock exchange, the statistical base was formed from data about European banks.

**Keywords:** Bank Value, Regression Analysis, Financial Ratios.

### **1. INTRODUCTION**

The process of globalization exercises a significant influence on banking business today. Globalization in banking sphere is attended by financial market liberalization. It means the abolition of restrictions for foreign financial institutions' entrance into domestic banking markets that, in turn, exacerbates a competition and activates processes of banking capital consolidation. Competitive pressure has prompted financial institutions pursue diversification strategies, often including mergers and acquisitions (M&A) [18]. As a result, M&A activity is increased worldwide during last decades. In spite of negative repercussions of global financial crisis, recent survey findings indicate a positive outlook for corporate transactions for the nearest future [11, 19].

The valuation of an entity is an integral part of any M&A transaction. Besides, the value-based management skill is one of the main components of successful doing business today. However, the concept of value makes sense only if it is possible to estimate it. Thus, bank valuation is one of the most actual issues in today's financial business.

There are different valuation methodologies that provide an estimate of a company's value. All the methods can be combined into three groups: earning-based methods, assets-based methods and market-based methods [4]. Earning-based methods use the fundamental principle of finance – time value of money [5]. Market-based valuation methods use price or enterprise value multiples, such as price/earnings (P/E) ratio or EBIT multiple [2]. The asset approach to business valuation is based on the principle of substitution: no rational investor will pay more for the business assets than the cost of procuring assets of similar economic utility [6].

The method of Discounted Cash Flow (DCF method) is the most often applied bank valuation method [3, 4]. Using this method, all future cash flows are estimated and discounted to determinate the present value. It based on valuing either a stream of dividends, which is the Dividend Discount Model (DDM), or a stream of free cash flows, which is the Discounted Cash Flow (DCF) method [17].

Using the argument that the only cash flows that a stockholder in a publicly traded firm receives are dividends, equity is valued as the present value of the expected dividends [4]. However, to apply DDM properly, it is necessary to predict values of future dividends using retrospective information. For instance, many of Latvian commercial banks do not pay dividends or pay them irregularly. Today, considering the complex and dynamic financial environment, financial service companies are expected to reinvest all their profit into business activities. Besides, using DDM can lead to the improper valuation of a bank, if, for instance, dividends are paid less, than a bank can afford to pay. However, using Cash Flow to Equity Discount model in Latvia, we

face other challenges, such as discount rate estimation [16].

Due to the limitations in the technical applicability of the DCF, analysts are forced to rely in practice upon valuation multiples and subjective judgments of whether the market price 'feels right' [9].

The goal of the paper is to develop a proxy model for valuation of a commercial bank by the analogy with Z-score models [8]. The factors affecting bank value are represented by financial indices that were selected, using correlation analysis. The degree of correlation between selected variable was quantified by estimating Pearson product-momentum correlation coefficient  $r$ .

Application of multi-factor regression models allows reducing complexity of valuation. Besides, this is a good valuation alternative for such countries as Latvia, where application capabilities of worldwide used methods are limited.

## 2. RESEARCH DESCRIPTION AND EMPIRICAL RESULTS

To achieve the goal of the paper, the following tasks should be accomplished:

- 1) To form the initial data base. In or case the resulting variable ( $y$ ) is a market capitalization (CAP) of a bank. The shares of only few banks in Baltic region are quoted in the stock exchange. This was the reason for using statistical information about European banks' market capitalization;
- 2) To check the degree of correlation between the selected arguments (financial indices –  $x_1, x_2...x_n$ ) and the function (a bank value -  $y$ );
- 3) To form a regression model that describes a relationships between function and the most valuable arguments;
- 4) Using the developed model, to calculate the value of selected banks;
- 5) To analyse the reliability of the results, comparing estimated market values with real values of the banks (market capitalization).

### Selection of the indices for including into the model

Stock price, and as a consequence, market capitalization of a company is influenced by range of factors, such as company's performance results and development plans, trends in economy and attitude of market participants [14].

The first task of our survey was to select a range of indices that can be included into the valuation model. We started with analyzing banks' performance, using financial ration analysis [1, 7].

Based on theory, company's value depends on its ability to generate cash flows from business activities [4, 5, 15]. Thus, it is logically to assume the strong relationships between profitability and company's value. That is why we focused primarily on profitability ratios: return on assets (ROA), return on equity (ROE) and earnings per share (EPS).

Besides, we used the analytical indices of European Central Bank: Cost-to-income ratio (CI) and operating income-to-assets (I/A). We selected Net interest income (NII) ratio for our analysis because of its critical weight in the total bank income. We also analyzed relationships between market capitalization and dividend payout ratio (DPR – dividends per net income) of selected banks. This hypothesis was based on the assertion that company's value depends on investor expectations [13]. Investors, expecting high dividends from a company, increase a demand for its stocks and, consequently, stock price also rises.

The results of the correlation analysis of the market capitalization and selected financial indices of the banks are presented in the Table 1.

The objects of our analysis are European banks, which stocks are quoted on the stock exchanges. The statistical information is provided by financial reports of the selected banks and by information and financial agencies, such as Morningstar, Reuters and Financial Times.

The values of Pearson's correlation coefficient are received, based on processing of statistical information over a period of 2002-2010.

The received results indicate the fact that the correlation coefficients between the selected financial ratios and value of bank market capitalization differ widely among the banks. The cases of strong negative and positive correlation occur simultaneously. Income-to-assets ratio, net interest income and dividend payout ratio have very low correlation coefficients.

The most suitable indices are ROE and Cost-to-income ratio. From the economic point of view ROE should correlate positively with market capitalization, and cost-to-income ratio should have negative correlation with it. It means that an increase in bank's profitability should lead to increase in bank's value. In turn, increase of value of bank's cost should reduce its value. Only few banks (see Table 1) have a negative correlation between ROE and market capitalization, and only one bank has a positive relationship between costs and market capitalization. Thus, in most cases the logical assumption is confirmed by figures. However, the average correlation coefficients are not sufficiently high to include them into the valuation model.

Table 1

Pearson's correlation coefficient between Market Capitalization and financial indices of the banks (I)

Bank name	ROE	CI	I/A	NII	DPR
Banco Comracial Portugues	0,491	0,161	0,509	-0,081	-0,706
Barclays	0,505	-0,266	0,397	-0,259	-0,101
Credit Suisse	0,794	-0,332	-0,131	-0,486	0,534
BNP Paribas	0,549	-0,833	0,111	0,259	0,578
Danske Bank	0,77	-0,569	-0,122	-0,355	0,627
Deutsche Bank	0,418	-0,832	0,485	-0,498	0,004
DnB NOR	0,57	-0,74	-0,515	0,305	-0,574
Dexia	0,689	-0,755	0,562	-0,587	0,716
Erste Group	-0,429	-0,781	-0,226	0,161	-0,574
KBC Groep	0,765	-0,674	0,49	-0,706	0,789
Handelsbanken	0,275	-0,434	-0,074	-0,118	0,157
Lloyds	-0,327	-0,139	-0,305	0,587	-0,718
National Bank of Greece	-0,35	-0,883	0,784	0,196	n/a
Nordea	0,413	-0,794	-0,436	0,473	-0,169
SAMPO OYJ	0,419	-0,865	-0,608	n/a	-0,353
Santander	0,554	-0,73	0,067	0,402	-0,498
Sabadell	0,931	-0,412	-0,871	0,343	n/a
Storebrand	0,351	-0,726	0,36	n/a	0,285
SEB	0,179	-0,353	-0,125	-0,019	0,119
Sydbank	0,622	-0,807	-0,218	n/a	n/a
UBS AG	0,64	-0,503	0,316	0,195	-0,785
Swedbank	0,325	-0,201	0,265	-0,673	0,371
Unicredit	0,424	-0,28	-0,505	-0,182	-0,463
<b>Min</b>	<b>-0,429</b>	<b>-0,883</b>	<b>-0,871</b>	<b>-0,706</b>	<b>-0,785</b>
<b>Max</b>	<b>0,931</b>	<b>0,161</b>	<b>0,784</b>	<b>0,587</b>	<b>0,789</b>
<b>Average</b>	<b>0,416</b>	<b>-0,554</b>	<b>0,009</b>	<b>-0,052</b>	<b>-0,038</b>

We selected indices ROA and EPS as arguments for the linear regression function, because the average correlation between these variables and market capitalization is higher than in previous cases (Fig. 1, Fig. 2).

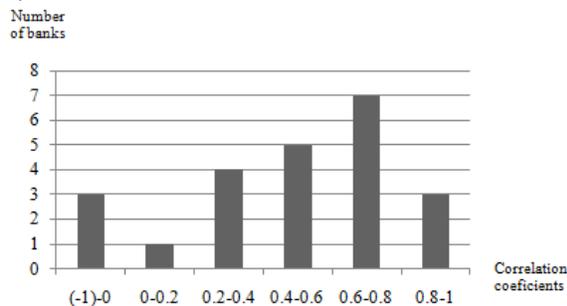


Fig.1. Distribution of the correlation coefficients – relationships between ROA and market capitalization

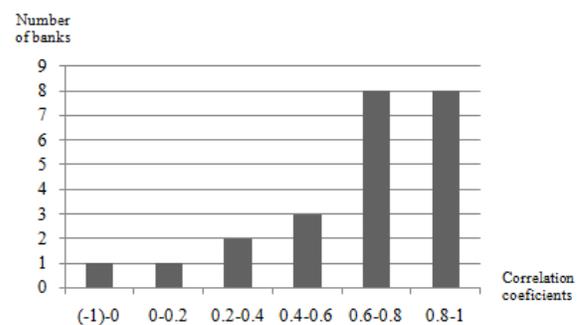


Fig.2. Distribution of the correlation coefficients – relationships between EPS and market capitalization

If we assume the accidental nature of negative correlation between ROA, EPS and market capitalization of some banks and remove them (Table 2), the average values will be equal to 0.552 and 0.676 respectively.

Table 2

Pearson's correlation coefficient between Market Capitalization and financial indices of the banks (II)

Bank name	ROA	EPS
Banco Comercial Portugues	0,756	0,704
Barclays	0,398	0,191
Credit Suisse	0,75	0,837
BNP Paribas	-0,547	0,775
Danske Bank	0,686	0,88
Deutsche Bank	0,884	0,895
DnB NOR ASA	0,553	0,772
Dexia	0,831	0,817
Erste Group	0,833	0,744
KBC Groep NV	0,761	0,805
Handelsbanken	0,458	0,452
Lloyds	-0,373	-0,282
National Bank of Greece	0,425	0,791
Nordea	0,207	0,428
SAMPO OYJ	0,424	0,51
Santander	-0,021	0,817
Sabadell	0,61	0,816
Storebrand ASA	0,647	0,838
SEB	0,213	0,2
Sydbank	0,599	0,795
UBS AG	0,612	0,721
Swedbank	0,345	0,307
Unicredit	0,056	0,786
<b>Min</b>	<b>-0,547</b>	<b>-0,282</b>
<b>Max</b>	<b>0,884</b>	<b>0,895</b>
<b>Average</b>	<b>0,439</b>	<b>0,635</b>

**Development of the bank valuation model**

Thus, we have values of two arguments ( $x_1$  – ROA,  $x_2$  – EPS) and values of function ( $y$  – market capitalization) to construct a linear regression model for bank valuation. To determine the final view of the model (to find the regression coefficients), we need to solve the set of equations that consists of the elements such as Eq. (1):

$$Y_i = a_i x_{1i} + b_i x_{2i} \quad (1)$$

where  $i=1 \dots n$ ;

$n$  – number of selected banks;

$a_i, b_i$  - regression coefficients.

Statistical data that was used for the development of the model are represented in the Table 3.

Table 3

Financial indices of the banks, 2009

Bank name	CAP, EURm	ROA, %	EPS, EUR
Barclays	34997	0,54	0,96
Banco Comercial Portugues	3967	0,24	0,03
BNP Paribas	66215	0,28	5,20
Danske Bank	11081	0,05	0,34
Erste Group Bank	9849	0,45	2,57
Dexia	7861	0,16	0,57
DnB NOR	12273	0,47	0,77
National Bank of Greece	10987	0,25	0,28
Swedbank	6570	-0,58	-1,03
Nordea Bank	294328	0,47	0,60
Sabadell	4650	0,64	0,44
Sampo	9555	16,50	1,14
SEB	9538	0,01	0,06
Handelsbanken	12338	0,80	1,59
Storebrand	2137	0,26	0,25
UBS	38386	-0,16	-0,50
Unicredit	23100	0,25	0,10
Sydbank	1335	0,49	1,57
KBC	10867	-0,76	-7,26
Lloyds	35965	0,28	8,33
Credit Suisse	40795	0,65	3,55
Santander	95043	0,86	1,05
Deutsche Bank	30683	0,33	7,92

Data processing was conducted, using the statistical program eViews as a software tool. To find the optimal equation for our model, the OLS method was applied.

Main statistical indices for the dataset from Table 3 are presented in Fig.3.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ROA	705.5449	1749.138	0.403367	0.6908
EPS	3256.620	1493.416	2.180652	0.0407
R-squared	-0.531591	Mean dependent var		21241.35
Adjusted R-squared	-0.604524	S.D. dependent var		22727.13
S.E. of regression	28788.41	Akaike info criterion		23.45627
Sum squared resid	1.74E+10	Schwarz criterion		23.55501
Log likelihood	-267.7472	Hannan-Quinn criter.		23.48111
Durbin-Watson stat	1.553103			

Fig. 3. Descriptive statistics for dataset analysis

The R-squared ( $R^2$ ) statistic measures the success of the regression in predicting the values of the dependent variable within the sample. In program eViews it can be negative if the regression does not have an intercept, as it is in our case.

The regression coefficients are positive, and it is logically from the economic point of view: profitability ratios positively impact banks' value.

EPS has statistically significant impact on market capitalization with probability more than 95 per cent. However, results of the analysis indicated the problem with ROA index. Based on received results, ROA is not statistically significant ratio. Due to this fact, the estimated values of the selected bank may considerably vary from real values. One of the topics of the future research is to find the more appropriate financial index for the model, which will have the stronger statistical significance.

The percentage gap between the estimated values of the selected banks and the market capitalization values is presented in the Table 4.

Table 4  
Market capitalization and estimated values of the banks

Bank name	CAP, EURm	ROA	EPS	Value, EURm	GAP, %
Barclays	34997	0,54	0,96	3501	-90
Banco Comercial Portugues	3967	0,24	0,03	280	-93
BNP Paribas	66215	0,28	5,20	17132	-74
Danske Bank	11081	0,05	0,34	1130	-90
Erste Group Bank	9849	0,45	2,57	8687	-12
Dexia	7861	0,16	0,57	1969	-75
DnB NOR	12273	0,47	0,77	2846	-77
National Bank of Greece	10987	0,25	0,28	1088	-90
Swedbank	6570	-0,58	-1,03	-3773	-157
Nordea Bank	294328	0,47	0,60	2286	-99
Sabadell	4650	0,64	0,44	1884	-59
Sampo	9555	16,50	1,14	15354	61
SEB	9538	0,01	0,06	186	-98
Handelsbanken	12338	0,80	1,59	5755	-53
Storebrand	2137	0,26	0,25	998	-53
UBS	38386	-0,16	-0,50	-1754	-105
Unicredit	23100	0,25	0,10	502	-98
Sydbank	1335	0,49	1,57	5465	310
KBC	10867	-0,76	-7,26	-24179	-323
Lloyds	35965	0,28	8,33	27338	-24
Credit Suisse	40795	0,65	3,55	12016	-71
Santander	95043	0,86	1,05	4012	-96
Deutsche Bank	30683	0,33	7,92	26025	-15

It is critically important point to be made that the model cannot be applied for banks with negative values of return on assets or earnings per share indices. Taking into

account that regression coefficients are positive, in this case estimated value will be negative.

However, even removing from the dataset the banks with negative ratios, the analysis of estimated value indicates the fact that in the most cases the estimated values are considerably lower than the values of market capitalization (Fig. 4).

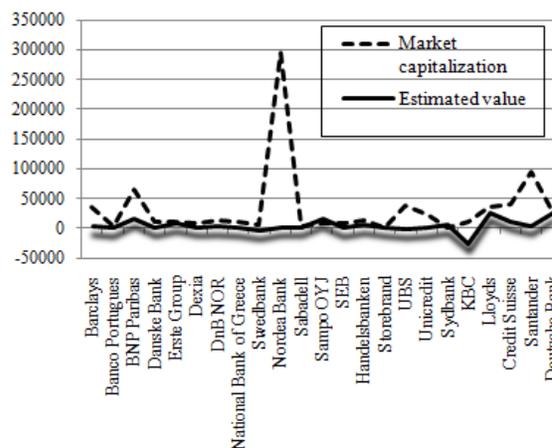


Fig. 4. Real and estimated values of the banks

It can be explained with the fact that, nowadays, intangible assets, in particular the relational capital, amount to over 70 percents of company's value [10]. Relational capital describes an organization's relations with customers, suppliers, investors, co-operation partners and the public [12]. Relational capital keeps customers from abandoning a commercial relationship. Using in valuation model only financial indices, banks will be underestimated.

Besides, it is necessary to check the level of relationships between bank value and other financial indices. As a basis for the selection, it is possible to use the financial ratios from the European Bank's statistical reviews.

Thus, studying the relationships between non-financial indices and banks' market capitalization in order to develop more relevant valuation model seem to be interesting topic for future research.

### 3. CONCLUSIONS

The current research represents an attempt to develop a linear regression model that can be used by external analysts to estimate a value of a bank.

The model was developed based on the analysis of financial indices of the banks, which stocks are publicly quoted.

The estimated values of the selected banks were sufficiently lower than their market capitalization. It points to the fact that such kind of models should involve not only financial ratios, but also non-financial measures, because the intangible assets, such as customer base, have a critical weight in a company's value.

Using the developed model for valuation of Latvian banks, we received absolutely non-adequate results. Thus, the topic of our future research will be creation of specific valuation model for Latvian banks. One of the possible non-financial indices for the model is EPSI rating that is the most popular index to measure customer satisfaction and loyalty in European countries.

#### 4. REFERENCES

- [1] Z. Bodie, R. C. Merton, **Finance**, Upper Saddle River, NJ: Prentice Hall, 2000.
- [2] CFA Institute, **Equity. CFA Program Curriculum**, New York: Pearson Custom Publishing, 2010.
- [3] T. Copeland, T. Koller, J. Murrin, **Valuation: Measuring and Managing the Value of Companies**, New York: John Wiley & Sons, Inc., 1995.
- [4] A. Damodaran, **Tools and Techniques for Determining the Value of Any Asset**, New York: John Wiley & Sons, 2002.
- [5] F.J. Fabozzi, P.P. Drake, **Finance: Capital Markets, Financial Management and Investment Management**, Hoboken, NJ: John Wiley & Sons, 2009.
- [6] J. P. Friedman, N.J.D. Ordway, **Income Property Appraisal and Analysis**, New Jersey: Prentice Hall, 1989.
- [7] L.J. Gitman, **Principles of Managerial Finance**, Boston: Pearson, Addison Wesley, 2006.
- [8] J. Horne, **Financial Management Policy**, Upper Saddle River, NJ: Prentice Hall, 2002.
- [9] S. Imam, R. Barker, C. Clubb, "The Use of Valuation Models by UK Investment Analysts", **European Accounting Review**, Vol. 17, Issue 3, 2008, pp. 503-535.
- [10] R. S. Kaplan, D. P. Norton, **Strategy Maps: Converting Intangible Assets into Tangible Outcomes**, Boston: Harvard Business School Press, 2003.
- [11] A. Kronimus, A. Roos, **M&A: Back to the New Reality: a Survey of European Companies' Mergers and Acquisition Plans for 2011**, The Boston Consulting Group Inc., 2010.
- [12] K. Mertins, W. Wang, M. Will, "InCaS: Intellectual Capital Management in European SME - Its Strategic Relevance and the Importance of its Certification", **The Electronic Journal of Knowledge Management**, Vol.7, 2009, pp. 111 – 122.
- [13] E. Olsen, F. Plaschke, D. Stelter, **Missing Link: Focusing Corporate Strategy on Value Creation**, Boston: Boston Consulting Group, 2008.
- [14] A. Pelane, I. Ukenābele, **10:00-14:00. Vērtspapīru tirgus. Latvijas pieredze 2008, [10:00-14:00. Stock Market. Latvian Experience 2008]**, Riga: Riga Stock Exchange, 2008.
- [15] J.F. Sinkey, **Commercial Bank Financial Management**, Upper Saddle River, NJ: Prentice Hall, 2007.
- [16] J. Titko, N. Lāce, "Valuation of a Commercial Bank in Emerging Markets: Case of Latvia", **VII International Scientific Conference "Management and Engineering' 09"**, June 22-24, 2009, Sozopol, Bulgaria, pp. 364.-374.
- [17] P. Vernimmen, P. Quiry, M. Dallochio, Y. Le Fur, A. Salvi, **Corporate Finance: Theory and Practice**, Hoboken, NJ: John Wiley & Sons, 2009.
- [18] J. O.S. Wilson, B. Casu, C. Girardone, P. Molyneux, Emergind, "Themes in Banking: Recent Literature and Directions for Future Research", **The British Accounting Review**, No. 42, 2010, pp. 153-169.
- [19] S. Wong, "Energy and Financial Sector Deals Heat Up", **Global Finance**, Vol. 25, No. 4, 2011, pp. 87-88.

## An energy efficient routing algorithm (X-Centric routing) for sensor networks

Goktug Atac, Tamer Dag  
 Computer Engineering Department  
 Kadir Has University, Istanbul, Turkey  
 goktugatac@yahoo.com, tamer.dag@khas.edu.tr

**Abstract**-Recent developments in wireless communications and electronics technologies have enabled the progress in low cost sensor networks. Sensor networks differ from traditional networks in several ways, such as the severe energy constraints, redundant low-rate data and many-to-one flows that the sensor networks require. One of the major challenges facing the design of a routing protocol for Wireless Sensor Networks (WSNs) is to find the most reliable path between the sources and the sink node by considering the energy awareness as an essential design parameter. This paper introduces a new routing protocol called as X-Centric routing by considering the above parameters. Under the X-Centric routing, the decision making mechanism depends on the capacity of the sink node by switching between address-centric routing (AC-Routing) and data-centric routing (DC-Routing). The design tradeoffs between energy and communication overhead savings in these routing algorithms have been considered by considering the advantages and performance issues of each routing algorithm.

**Keywords:** Wireless sensor networks; AC-Routing; DC-Routing; X-Centric Routing

### I. INTRODUCTION

Artificial sensors have become an important part of ordinary and industrial part in the second part of the twentieth century. Sensors of a great variety of types have been established in process industries, agriculture, medicine, military services and many other areas. And still, the development of new sensing capabilities is currently proceeding at an unprecedented rate. This technology grows by various factors including a growing concern for the protection of the environment, for the improvement of fossil fuel economy for the enhancement of safety and security [1].

It has been observed that, during the recent years sensor markets in developed countries have been growing at an average rate in excess of %10 per year. As a result, there has been a correspondingly high level of investment in sensor research, and development as the demand has arisen for sensor performance to catch up with the signal processing capabilities. [2]

A general sensor network is made up of multiple detection stations called as sensor nodes. A sensor node is small, lightweight and portable physical quantity. Each sensor node is equipped with the following basic components:

- Transducer – responsible for generating electrical signals based on sensed physical effects and phenomena
- Microcomputer – responsible for processing and storing the sensor output

- Transceiver – responsible for receiving commands from a central computer or transmitting data to that computer in a wireless or hard-wired environment
- Power source – responsible for the power requirement of a sensor node from an electrical utility or from a battery

A sensor network can also be considered as a group of specialized transducers with a communications infrastructure intended to monitor and record conditions at diverse locations. For a sensor network, the most common monitored parameters are temperature, humidity, pressure, wind direction and speed, illumination intensity, vibration intensity, sound intensity, power-line voltage, chemical concentrations, pollutant levels and vital body functions.

A sensor's sensitivity indicates how much the sensor's output changes when the measured quantity changes. For instance, if the mercury in a thermometer moves 1 cm when the temperature changes by 1 °C, the sensitivity is 1 cm/°C. Sensors that measure very small changes must have very high sensitivities. Sensors also have an impact on what they measure; for instance, a room temperature thermometer inserted into a hot cup of liquid cools the liquid while the liquid heats the thermometer. For this reason, sensors need to be designed to have a minimal effect on the parameter that is being measured. Having smaller size sensors often has a positive impact on this aspect and it may introduce other advantages. Technological progress allows an increasing number of sensors to be manufactured on a microscopic scale such as micro sensors using MEMS technology. In most cases, a micro sensor reaches a significantly higher speed and sensitivity compared with macroscopic approaches [2,4].

*Routing* is the act of moving information across an internetwork from a source to a destination. Routing is needed for sensor network because the main idea of sensor network is to transfer information from source to destination so routing becomes the most important element for sensor networks. [3,4,5,6]

In this paper, a new routing algorithm for sensor networks called as the X-Centric routing algorithm for is introduced. The X-Centric routing algorithm can be considered as an approach, which tries to combine address-centric (AC) routing, and data-centric (DC) routing by taking into advantage of the positive aspects of both algorithms. The X-Centric routing algorithm has a decision making process which makes transitions between AC-Routing and DC-routing by considering the sink occupancy levels for desired conditions.

The rest of the paper is organized as follows: Section 2 introduces AC-Routing and DC-Routing used in sensor networks. Section 3 described the proposed X-Centric routing briefly. Section 4 presents some simulation results performed on X-Centric routing by presenting the performance in terms of sink occupancy, total average delay, total packets lost and also total energy saved. The paper concludes with the conclusions made in Section 5.

II. ROUTING ALGORITHMS FOR SENSOR NETWORKS

This section describes the main routing protocols used for sensor networks such as AC-Routing and DC-Routing along with their basic properties.

In order to connect networks together so that they may exchange information and in order to move traffic through these networks efficiently, a method is needed whereby a specific path thus a route is found among the many available nodes (routers, servers, workstations) and routes that connect two or more network together. In other words *Routing* is the act of moving information across an internetwork from a source to a destination. But usually it is not just a route that allows traffic to be exchanged between the users; it is the “best” route between these users.

The term “best” might vary based on the requirements of the application traffic. For instance, while a real-time video conference considers the best route as a route which offers the lowest and most consistent delay, a funds transfer to a bank considers the best route which offers encryption services. [6,7]

For sensor networks the routing protocols can be grouped into address-centric and data-centric routing. Although both routing protocols aim to find the shortest path between the source and destination pairs, the data-centric approach differentiates from the address-centric approach in terms of data aggregation as explained in the following subsections.

A. AC- Routing

In the AC-Routing approach, each source node sends its information separately to the sink. This model is known as the traditional end-to-end routing model. In this protocol, the main idea is to find the shortest path between pairs of addressable end nodes and transmit information along the shortest path.

The example network of Fig. 1 illustrates an example on the behaviour of AC-Routing where there are two source nodes trying to transmit information to the sink. Source 1 is transmitting data “labeled 1” through node A and source 2 is transmitting data “labeled 2” through nodes C and B. These paths are the shortest paths between the individual sources and the sink node. [8,9]

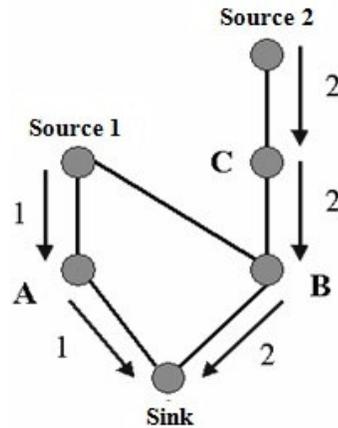


Fig. 1. AC- Routing illustration on a small-scale network

B. DC-Routing

In the DC-Routing approach, the source nodes transmit information to the sink as well. However, in this case, data is aggregated along the path to the destination. Different from the AC-Routing protocol, the nodes on the path of data flow examine the data content and apply some aggregation functions on the transmitted information. The data-centric paradigm results in some positive outcomes in terms of data redundancy, energy consumption and ultimately helping the optimum usage of sensor networks, so that this method becomes an alternative answer to the question of how to transform data in sensor networks.

Fig. 2 illustrates the behaviour of DC-Routing on a small scale network where 2 source nodes are transmitting information towards the sink. In this case, the data originated from the two sources is aggregated at node B, and the combined data “labeled 1+2” is transmitted from node B to the sink. As a result, considerable amount of energy is saved by transmitting less number of packets. [8,9]

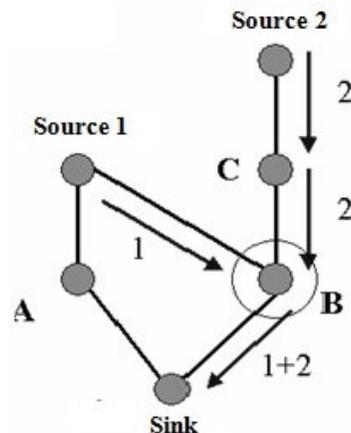


Fig. 2. DC- Routing illustration on a small-scale network

### III. X-CENTRIC ROUTING

Both routing algorithms (AC-Routing and DC-Routing) described in the previous section have the disadvantage of using the occupancy of sink when the observations has done in the parameter of, total average delay, total packet lost, total sink occupancy usage. The main idea of XC routing had born regarding of the disadvantage of using the sink occupancy. Using the occupancy effective gains energy save because of less delay and less packet lost.

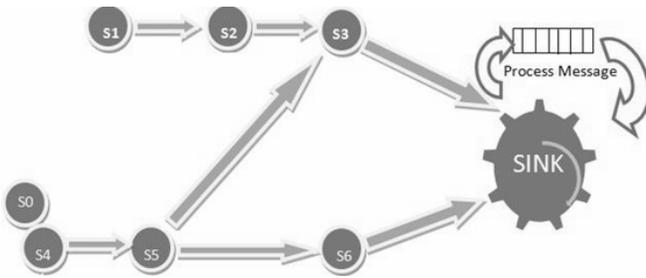


Fig. 3. X-Centric Routing (XC)

In addition, the main problem is to gain a low cost energy saving sensors, the new algorithm is try to determine in which condition it has to use the address centric algorithm or data centric algorithm to make fewer transmission. Because the sink has limited capacity and has a limited energy to process the coming messages to make fewer transmission, sink has to decide the right algorithm to gain this energy save. Figure 3 illustrates the process flow of the message from source to sink, but the main purpose is how the sink process the incoming messages. As we told that it has limited capacity and energy. But in other side it has to respond all incoming messages by processing and forward them to the real destination. We say the sink process the incoming messages, what means? In sensor processing means, first deciding it where to forward, and second controlling the message.

When sink receives a packet or let's say message it doesn't forward it directly to destination, it first try to optimize the packet and if needed consolidate it with other messages. Because the received message can be same so it has to make it separate.

In addition, the sink has limited capacity not all the messages can be received by sink, for this reason delays and energy loose will be obtained in each sensor and sink. Here the routing algorithm effects this condition, next section will observe first, what if sensors choose only address centric routing for routing algorithm in wireless sensor networks and second observe what if sensors choose only data centric algorithm, and last what will happen if the sink decide which routing algorithm to use in different conditions using simulations.

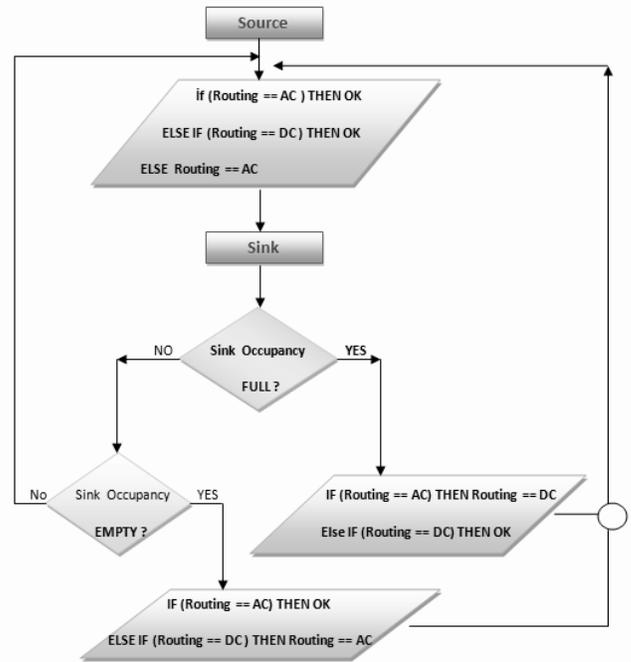


Fig. 4. X-Centric Routing Flow Chart

Fig. 4 presents the flowchart for the X-Centric Routing algorithm as a packet is transmitted from the source to the sink. When a packet is departed from a source, the default routing protocol is the AC-Routing. When the packet reaches the sink, the sink occupancy is checked. If the sink is full, then the routing algorithm is switched into DC-routing. The algorithm will continue as DC-routing until the sink is emptied. Once the sink becomes empty, the routing algorithm will switch back to AC-routing.

### IV. SIMULATION RESULTS FOR X-CENTRIC ROUTING

This section presents the simulations and their results conducted to compare the various characteristics of AC-Routing, DC-Routing and X-Centric routing protocols.

In order to evaluate and compare the performance of these routing protocols, two main parameters have been used:

- Sink Refreshment Time (SR) - the time interval that the sink refreshes itself to process the packets which have arrived and waiting in the sink,
- Periodic Time Intervals (PT) - the periodic time interval that a packet is produced at a source node while the source node collects information.

For the simulations various values for SR and PT have been used to evaluate the response of AC-Routing, DC-Routing and X-Centric Routing to the relative values in between SR and PT. The following values have been used during the simulations:

- 0.4 sec, 1 sec and 1.2 sec for SR,
- 0.2 sec, 1 sec and 1.8 sec for PT,
- In addition, it has been assumed that the maximum occupancy for the sink was 5 units thus capable of holding 5 packets.

By changing the PT and SR values, the following characteristics of sensor networks have been measured for the indicated routing algorithms:

- Sink occupancy
- Number of lost packets
- Number of processed packets
- Average delay

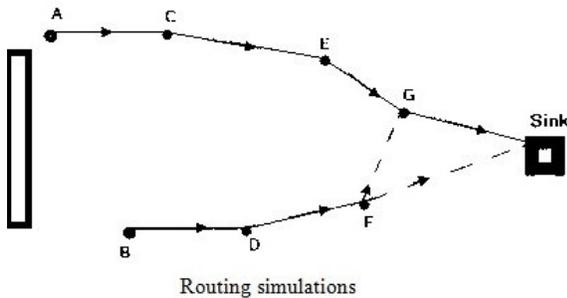


Fig. 5. The example network used for simulations

Fig. 5 shows the network on which the simulations are conducted. It is assumed that two source nodes A and B are transmitting the collected data to the sink node. Under AC-Routing, the path for data flow is A-C-E-G-Sink for source A and B-D-F-Sink for Source B. Under DC-Routing data is aggregated at node G, thus the path for data flow is A-C-E-G-Sink and B-D-F-G-Sink for source A and B respectively.

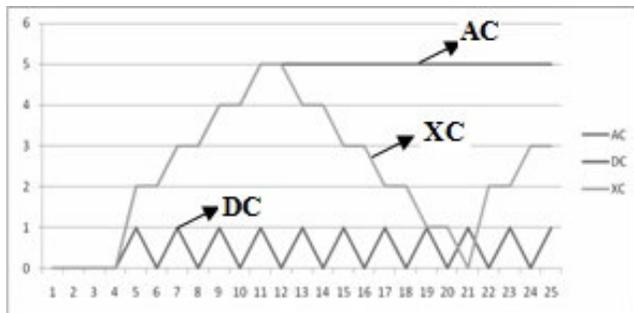


Fig. 6. Sink occupancy for AC-Routing, DC-Routing and X-Centric Routing

Fig. 6 illustrates the sink occupancy for AC-Routing, DC-Routing and X-Centric Routing vs. time when SR=1.2 sec and PT=1.8 sec. In this figure, it can be observed that under the AC-Routing the sink fills up with packets immediately and stays at the maximum occupancy level causing packet losses. On the other hand, under DC-Routing the sink does not fill up as the SR is less than PT and the number of packets arriving at the sink is much less compared to AC-Routing.

Under the XC-Routing, the sink occupancy fluctuates in between minimum and maximum occupancy levels as the algorithm switches between AC-Routing and DC-Routing.

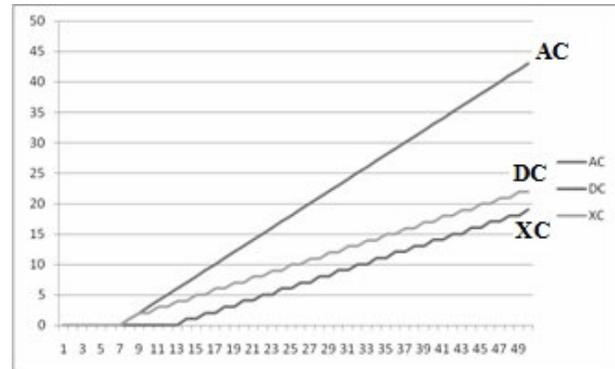


Fig. 7. Number of lost packets under AC-Routing, DC-Routing and X-Centric Routing

Fig. 7 presents the number of lost packets for AC-Routing, DC-Routing and X-Centric Routing vs. time when SR=0.4 sec and PT=0.2 sec. In this case, more traffic arrives to the sink during each SR period as the value of PT is decreased. As expected, the number of lost data for AC-Routing is the most, while for DC-Routing is the least, as the number of packets arriving at the sink is the most for AC-Routing and is the least for DC-Routing because of the data aggregation. The behaviour of X-Centric Routing falls in between the other two because of the decision making mechanism applied under the X-Centric Routing.

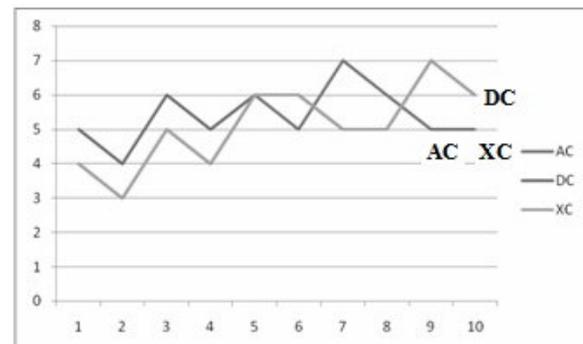


Fig. 8. Average packet delay under AC-Routing, DC-Routing and X-Centric Routing

Fig. 8 illustrates the average packet delays for AC-Routing, DC-Routing and X-Centric Routing vs. the number of packets for P.T equal 0,2 sec and S.R equal to 0,4 sc. Y axis is th avarage delay time (T/P) and X axis of the diagram shows number of packets. The avarage delay value for XC is between AC and DC. But simulation shows that when the network scale increase XC routing has greater performance and delay is decrease.

## V. CONCLUSIONS

In sensor networks the scarcest resource is energy, and one of the most energy-expensive operations is route discovery and data transmission. For this reason, algorithmic research in WSN mostly focuses on the study and design of energy aware algorithm for routing and data transmission from the sensor nodes to the destination.

The goal of this paper is to propose an energy efficient routing algorithm (X-Centric routing) for sensor networks and it had modeled and analyzed the performance of the X -centric routing comparing with AC and DC routing.

The XC routing algorithm is also tries to gain maximum efficiency from a network topology, because it tries not to waste energy. The simulation shows that with the given energy XC routing try to make maximum work.[10]

Day by day, the importance of wireless sensors are increase and current technology try to overcome the demands, but this is not easy enough because to reach the best solution, technology needs new inventions. As the near future technology, for wireless sensor nodes are typically low-cost, low-power, small devices equipped with limited sensing, data processing and wireless communication capabilities, as well as power supply XC routing try to open a new window in routing method and try to gain maximum efficiency with minimum resources. The next step for XC routing, can be integrating this method with other new inventions and try to accomplish other technologies disadvantages. [11]

As the technology needs new inventions XC routing can be a new way to follow and develop.

## REFERENCES

- [1]P. T. Moseley, A. J. Crocker, "Sensor materials", IOP Publishing 1996, British Library Cataloguing-in-publication Data
- [2] Holger Karl, Andreas Willig "Protocols and Architectures for Wireless Sensor Networks", @2005.
- [3] D. Estrin *et al.* "Next Century Challenges: Scalable Coordination in Sensor Networks," MobiCom 1999.
- [4] J. Heidemann *et al.*, "Building Efficient Wireless Sensor Networks with Low-Level Naming," *18th ACM Symposium on Operating Systems Principles*, October 21-24, 2001.
- [5] Bhaskar Krishnamachari, Deborah Estrin, Stephen Wicker "Data-Centric Routing- The Impact of Data Aggregation in Wireless Sensor Networks" *IEEE 2002*
- [6] Modelling Address-Centric and Data-Centric Routing in Wireless Sensor Networks, Bhaskar Krishnamachari, Deborah Estrin, Stephen Wicker , IEEE INFOCOM 2002
- [7] Imad Mahgoub, Mohammad Ilyas, "Sensor Network Protocols", @2006
- [8] Ivan Stojmenovic "Algorithms and Architectures of Sensor Networks" @2005
- [9] Darothea Wagner, Roger Wattenhofer, "Algorithms for Sensor Networks and Ad-Hoc Networks" @2007
- [10] J. Warrior, "Smart Sensor Networks of the Future," *Sensors Magazine*, March 1997.
- [11] A. J. Goldsmith and S. B. Wicker, "Design Challenges for Energy-constrained Ad Hoc Wireless Networks," *to appear in IEEE Communications Magazine*.

# Identity Based Authentication using a Cross Layer Design approach in Wireless Sensor Networks

Harsh Kupwade Patil  
Department of Computer Science  
Southern Methodist University  
Dallas, Texas, USA  
Email: hkupwade@smu.edu

Joseph Camp  
Department of Electrical Engineering  
Southern Methodist University  
Dallas, Texas, USA

Stephen A. Szygenda  
Department of Computer Science  
Southern Methodist University  
Dallas, Texas, USA

**Abstract**—The development of Wireless Sensor Networks (WSN) was originally motivated for military applications such as battlefield surveillance. However, in recent years it has gained popularity in many civilian applications: such as habitat monitoring, healthcare, home automation, traffic control and environmental monitoring. As more low cost, low power and multifunctional sensor nodes are being deployed, security in such sensor networks become one of the prominent issues in WSN. Recent advances in wireless networks did not give the necessary attention to security with regard to device constraints, since they base their design on legacy wireless networks. As more security solutions are being proposed in WSN, there is an increase in the lack of co-ordination between various security measures at different layers, leading to functional redundancy and increased overhead. As WSN scale to a very large number, current malicious node detection schemes will be resource intensive and inefficient. Therefore, new approaches are being sought to efficiently use information from different protocol layers to propose security. They not only focus on layer interactions within a node, but also adapt to changes in the network conditions and adaptively optimize cross-layer interactions across different nodes. It has been shown that pairing based crypto operations are possible on sensor nodes (motes) and a completely new security suite is being developed for WSN using Identity based Cryptography. In this paper, we propose a new cross-layer design approach for WSN using Identity based cryptography. This approach combines cross-layer design principles along with Identity based Cryptography to provide a new set of security solutions, which could be more efficient in storage, computation and energy.

## I. INTRODUCTION

Wireless Sensor Networks (WSN) has been one of the most promising solutions to many applications and security in these networks is crucial to WSN [1]. A typical sensor network consists of computationally limited low cost sensors that can scale from a few hundreds to thousands in number. Besides incorporating the basic vulnerabilities in ad-hoc networks, WSN pose new challenges such as survivability (physical access to adversaries) and lesser computational power than the conventional ad-hoc networks. Many scalable and efficient security solutions have been proposed to improve the energy efficiency in WSN and one approach among them is Elliptic Curve Cryptography (ECC) [2],[3]. Given the limited computational power and the resource constrained nature of wireless sensors, [2] has shown considerable reduction in memory access and computational time in key agreement

by efficiently implementing binary field algorithms such as squaring, multiplication, inversion and modular reduction on Micaz mote. [3] has implemented pairing based signature (BLS-SS [4] and BB-SS[5]) and encryption schemes (BF-IBE [6]) on the MICA family, and has shown that it achieves faster computation and lower memory consumption by choosing super-singular elliptic curve as a pairing group.

However, one of the main commonality with these protocols is that they abide by the traditional layered architecture. Although they may achieve a very high performance in their assigned specific layers, a cross layer approach for digital signature generation and key distribution can jointly optimize and increase the overall network security in WSN. It has been observed that the modular approach to the development of security protocols for individual layers might provide redundant security services and hence consume more energy than required in WSN. In some situations, an over-engineered security design may exhaust all the energy resources, leading to a security service denial of service attack. For example, the non-adaptive nature of the security service can be ineffective in preventing the more sophisticated attacks that look for vulnerabilities at each layer. For example, security provisioning at the network layer can be ineffective since the attacker can spoof the target's MAC address and launch a Denial of Service (DoS) attack. In this paper, we look at a new cross-layer approach to authentication using pairing based cryptography.

## II. RELATED WORK

### 1) Identity based cryptography:

a) *Overview:* In 1984, Shamir introduced Identity Based Cryptography (IBC) [16]. He could construct an Identity Based Signature (IBS) scheme using the existing RSA function [17], but was unsuccessful in constructing an Identity Based Encryption (IBE) scheme and it remained a long standing problem for almost a decade. In 2001, Boneh and Franklin came up with an independent solution using the concept of bilinear maps [6]. It led to a new era of research where many identity based digital signature schemes were proposed using bilinear maps. At the same time, Cocks proposed an Identity Based Encryption (IBE) scheme using quadratic residuosity [7]. However, Cocks's scheme is limited in its applicability

to WSN due to the generation of long ciphers and slower performance as it is based on ternary quadratic form.

In IBC, an end user can choose an arbitrary string to be his identity. This identity is used in generating the public key. Therefore, they do not need digital certificates from Certificate Authorities (CA) to verify digital signatures. A trusted third party Private Key Generator (PKG) is responsible to distribute private keys corresponding to their identity (Figure 1).

b) *Key Distribution in IBE*: The general key distribution problem refers to the onerous task of distributing secret keys between communicating parties to provide security properties such as authentication or confidentiality or both. But with the advent of private key distribution, new problems arise such as the inherent key escrow problem and the need for secure distribution of private keys [22].

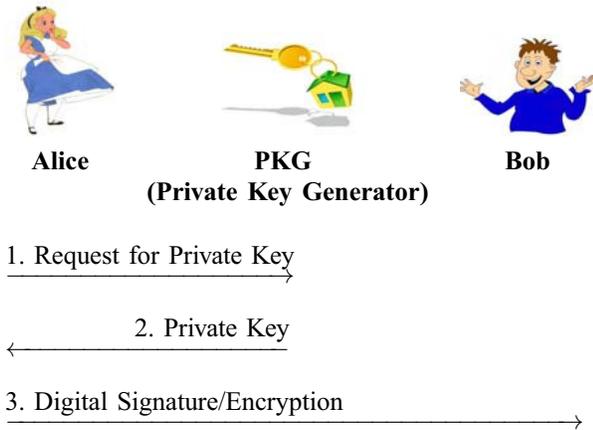


Figure 1: Key distribution in Identity Based System

Byoungcheon Lee *et. al* came up with a secure scheme for private key distribution in IBC. We briefly review [22].

Let  $H_1, H_2$  and  $H_3$  be three hash functions such that

$$H_1 : \{0, 1\}^l \rightarrow G_1 \tag{1}$$

where  $l$  is the length of the plain text.

$$H_2 : \{0, 1\}^l \times G_2 \rightarrow Z_q \tag{2}$$

$Z_q = Z/qZ$  denotes integers mod  $q$  where  $q$  is a large prime. Therefore  $Z_q$  denotes the group  $\{0, 1, 2, \dots, q-1\}$  and  $Z_q^* = Z \setminus \{0\}$ .

$$H_3 : G_2 \rightarrow Z_q^* \tag{3}$$

The PKG specifies two groups  $G_1$  and  $G_2$  of order  $q$ , where  $G_1$  is an additive group and  $G_2$  is a multiplicative group. Let  $e$  be a bilinear map such that  $e : G_1 \times G_1 \rightarrow G_2$  with the following properties

- Bilinear: Let  $(x_1, x_2$  and  $y) \in G_1$ . Then

$$e(x_1 + x_2, y) = e(x_1, y).e(x_2, y) \tag{4}$$

- Non-degenerate: There exists  $x \in G_1$  and  $y \in G_1$  such that

$$e(x, y) \neq 1 \tag{5}$$

In fact  $G_1$  is a point subgroup on an elliptic curve over a finite field and  $G_2$  is a subgroup of a cyclic group of a larger finite field. The pairings are derived from the Weil, Tate or  $\eta_T$  pairing [6], [8]. The PKG chooses a private key  $s_0 \in Z_q^*$  and computes the master public key

$$P_0 = s_0P \text{ where } P \in G_1 \tag{6}$$

The security of the master public key is dependent on the elliptic curve discrete log problem [18]. The PKG publishes the description of the groups  $G_1$  and  $G_2$ , public key  $P_0$ , hash functions  $(H_1, H_2$  and  $H_3)$ , the bilinear map  $e$  and the group element  $P$ . Alice and Bob choose their secrets to compute their blinding factors. Alice with identity  $ID_A$  chooses a random secret  $x \in Z_q^*$  and computes a blinding factor  $X = xP$ . An eavesdropper will not be able to generate the private key since he has no knowledge of the secret  $x$ . She then requests the PKG to issue a partial private key by sending  $X$  and  $ID_A$ . The PKG would use some pre-shared credentials to verify the authenticity of an end user's identity.

The PKG validates Alice's identity and computes the public key of Alice as

$$Q_{ID_A} = H_1(ID_A) \tag{7}$$

It computes a blinded partial private key as

$$Q_{bl_A} = H_3[e(s_0X, P_0)]s_0Q_{ID_A} \tag{8}$$

It then generates a signature  $Sig(Q_{bl_A})$  for providing integrity protection.

$$Sig(Q_{bl_A}) = soQ_{bl_A} \tag{9}$$

It sends  $Sig(Q_{bl_A})$  and  $Q_{bl_A}$  to Alice.

Alice verifies the signature using the below mentioned formula,

$$e(Sig(Q_{bl_A}), P) \stackrel{?}{=} e(Q_{bl_A}, P_0) \tag{10}$$

and finally retrieves her private key  $D_{ID_A}$  by un-blinding  $Q_{bl_A}$  as follows

$$D_{ID_A} = \frac{Q_{bl_A}}{H_3[e(P_0, P_0)^x]} \tag{11}$$

Hence there is a secure key exchange between the PKG and the end users. Identity based authentication can be easily extended to a two level hierarchical domain environment as shown in figure 2.

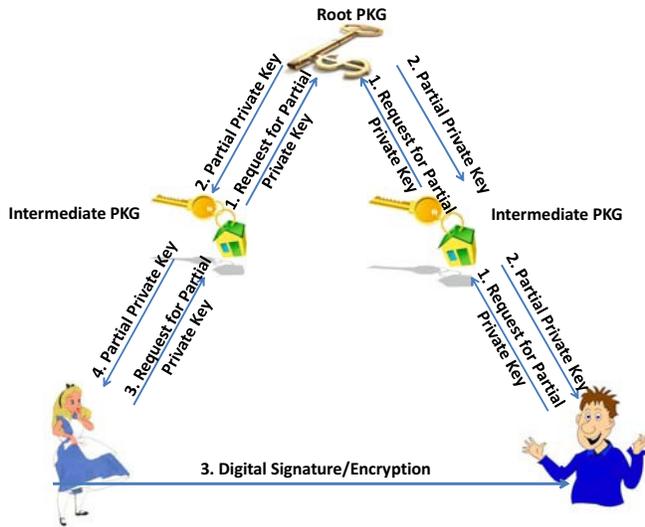


Figure 2: Identity based authentication in a two level hierarchical environment

Alice would then use Gentry and Silverberg's algorithm to generate a digital signature or Chow *et al's* algorithm to generate a signcrypt message [25],[26]. Signcrypt messages can only be verified by the intended recipient as he would use his own private key along with sender's public key to validate the ciphertext. Concerning the key escrow problem, one of the promising approaches is in using threshold key cryptography in which one full piece of secret information is derived from a set of secret shares [19],[20]. [6] directly apply the techniques of threshold key cryptography to their IBE systems. The master secret  $s_0 \in Z_q^*$  is used to generate the private key  $D_{ID} = s_0 Q_{ID}$ , where  $Q_{ID}$  is derived from the user's public key identity. This private key can be easily distributed in a  $t$ -out-of- $n$  fashion by giving each of the  $n$ -PKGs one share  $s_i$  of the secret  $s$ . When generating the private key each of the  $t$  chosen PKGs simply respond with  $D_{ID_{priv}}^{(i)} = s_i Q_{ID}$ . The user can then construct  $D_{ID_{priv}}$  as  $D_{priv} = \sum \lambda_i D_{ID_{priv}}^{(i)}$  where  $\lambda_i$  are the appropriate Lagrange coefficients.

*c) Accountable Authority Identity based Encryption (A-IBE):* A-IBE is a variant of the IBE scheme having an exponential or super-polynomial number of possible decryption keys available corresponding to every identity. Each decryption key  $d_{ID}$  for an identity will belong to a unique decryption key family (denoted by the number  $nF$ ). Roughly speaking, in the definitions of security we will require that: given a decryption key belonging to a family, it should be intractable to find a decryption key belonging to a different family (although it may be possible to find another decryption key belonging to the same family). The end user gets the decryption key corresponding to his identity from the PKG using a secure key generation protocol [22]. The protocol allows the user to obtain a single decryption key  $d_{ID}$  for his identity without letting the PKG know which key he obtained. If the PKG turns malicious and generates a decryption key  $d'_{ID}$  for that

specific identity, with all but negligible probability, it will be different from the key, which the end user obtained. Therefore the generation of a key pair  $(d_{ID}, d'_{ID})$  is a cryptographic proof of malicious behavior of the PKG as only one key per identity should be used in circulation.

*d) Key Distribution in WSN:* The traditional key distribution mechanism based on Public Key Infrastructure (PKI) would be impractical in WSN due to the increase in consumption of energy [15]. A more pragmatic approach to key distribution in WSN is by relying on pre-key distribution. One of the approaches to pre-key distribution is by using a single pre-key distribution for the entire sensor network. This approach would be inefficient if any of the sensor nodes is captured by an adversary and would lead to the compromise of the entire network. To avoid the compromise of the entire network, the other approach is to setup a pair-wise key sharing between every two sensor nodes. Therefore every node will have  $n-1$  keys and the entire network will have  $\frac{n(n-1)}{2}$  keys. One of the major disadvantages with this approach is that the entire network has to be re-keyed if new nodes are added into the network or with the deletion of existing nodes. In addition, storage of  $n-1$  keys in every sensor node will be resource intensive.

Eschenauer and Gligor proposed a new centralized key management scheme called as the basic random key scheme [9]. It consists of three phases.

#### Phase I: Key Pre-distribution

Each sensor stores keys in the form of a key-ring. The ring consists of randomly chosen  $k$  keys from a large pool of keys. Each key in the key-ring has a key identifier which is stored in the sensor. Further, each sensor node  $i$  would share a pairwise key  $K$  with its corresponding base station. The scheme is weakened if the base station within the domain were compromised.

#### Phase II: Shared key discovery

Each sensor discovers its neighbors who are in the communication domain by simply broadcasting a list of identifiers for its corresponding keys in an un-encrypted fashion. However, a malicious neighboring node can easily eavesdrop on the key sharing pattern among other sensors within the communication domain. They propose a challenge response mechanism for secure sharing of keys among nodes, which is computationally intensive.

*Phase III: Path key establishment:* A path key is established at the end of the shared key phase between the two communicating nodes. The topology for the sensor array as seen by the network layer (routing layer) is set by the end of phase II. However, it would hamper routing as a path key between nodes has to be established between end-to-end communicating parties. Revocation is inefficient, as the base station has to send a message to all the nodes within the communication domain to revoke the compromised key.

### III. IDENTITY BASED AUTHENTICATION USING CROSS LAYER DOMAIN.

In 1997, Zheng showed that by combining authentication with encryption into a single primitive, it was possible to achieve significant savings on both communication and computational overhead [14]. Implementing signcryption schemes in WSN would be advantageous as the end users verify the sender's identity as well as decrypt messages at the same time. Below we briefly review two important signcryption schemes and its applicability in WSN.

#### A. Lynn's algorithm [24]

Let

$$H_4 : \{0, 1\}^* \times \{0, 1\}^* \rightarrow Z_q \quad (12)$$

$$H_5 : Z_q \times G_2 \rightarrow \{0, 1\}^* \quad (13)$$

$$H_6 : \{0, 1\}^l \rightarrow \{0, 1\}^* \quad (14)$$

Alice would pick a random secret  $k \in Z_q^*$  and compute

$$q = H_4(k, m) \text{ where } m \text{ is the message} \quad (15)$$

and

$$w = e(D_{ID_A}, Q_{ID_B}) \quad (16)$$

Alice would then send the signcrypted message  $\prec U, V, W \succ$  to Bob where

$$\prec U, V, W \succ = \prec q, En_{H_5[q, w]}(k), En_{H_6[k]}(m) \succ \quad (17)$$

$En_{(Key)}$  refers to encryption using AES algorithm [29].

Bob would decrypt the message  $m$  as shown below

$$e(Q_{ID_A}, D_{ID_B}) = w \quad (18)$$

$$Dn_{H_5[U, w]}(V) = k \quad (19)$$

$$Dn_{H_6[k]}(W) = m \quad (20)$$

#### B. Chow et al's algorithm [26]

Let us consider a WSN network consisting of a sink node (node with abundant resources) along with a few cluster nodes. The sensor nodes send sensed data to their respective cluster nodes, which may aggregate the data and further send it to the sink node.

Subsequently, let domain one be controlled by the root PKG (Sink Node) and domain two be controlled by the cluster nodes.

Let the Root PKG's master private key be  $M_s \in Z_q^*$  and the master public key be  $Q_0 = M_s P$  where  $P \in G_1$ . Let  $s_1 \in Z_q^*$  then

$$Q_1 = s_1 P \quad (21)$$

where  $Q_1$  is a public parameter generated by one of the cluster nodes.

$$H_7 : G_2 \rightarrow \{0, 1\}^* \quad (22)$$

Alice would generate the signature

$$Si g_c = D_{ID_A} + k P_M \quad (23)$$

where  $P_M = H_1(m)$

and compute

$$g = e(Q_0, k Q_{ID_2}) \quad (24)$$

She would generate the signcrypted message as shown below,

$$\prec \overbrace{k Q_{ID_B}}^{U_1}, \overbrace{En_{H_7[g]}(m || Si g_c || ID_B)}^V, \overbrace{Q_0}^{W_1}, \overbrace{Q_1}^{W_2}, \overbrace{Q_M}^{W_3} \succ \quad (25)$$

where  $Q_M = k P$ .

Alice would send  $\prec U_1, U_2, V, W_1, W_2, W_3 \succ$  to Bob.

Bob would compute

$$\frac{e(W_3, D_{ID_B})}{e(Q_2, U_1)} = g \quad (26)$$

and then decrypt  $V$  as shown below,

$$Dn_{H_7[g]}(V) = (m || Si g_c || ID_B) \quad (27)$$

He would then verify the signature as shown below,

$$\frac{e(P, Si g_c)}{e(W_2, Q_{ID_A})} \stackrel{?}{=} e(W_1, Q_{ID_1}) e(W_3, P_M)$$

In case of authentication, we could apply Hess's Identity based signature scheme in WSN [23]

The use of Identity field should not be restricted to the upper 3 layers (Application, Network and Data Link) and should also take the physical layer parameters as identities for authentication. Figure 3 describes the possible identity fields in different layers.

Layer	Identity Header
Application	Device Name
Network	IP address
Data Link	MAC address
Physical	Channel frequency response

Figure 3: Identity field at different layers

The conventional approach to assigning security to the upper layers would leave a gap between the upper and lower layers (physical layer). An unsecured physical layer is susceptible to a broad range of security attacks. Therefore the physical layer properties in a wireless medium have to be leveraged to address the security threats. A legitimate user can be differentiated from a rogue end user by measuring the channel frequency response and hypothesis testing [32].

On the contrary, solely depending on the physical layer is not the best approach to providing security. The traditional security schemes in the physical layer (Frequency Hopping Spread Spectrum or Direct Sequence Spread Spectrum) could be rendered insecure, if the intermediate adversary along the active routing path has prior knowledge of the hopping pattern or the spreading code. In such cases an adaptive, location and energy efficient cross layer design approach should be adopted to provide a holistic security solution.

We suggest that the identity field should not be fixed to a specific layer and should opt for a cross layer approach, wherein the end user decides the identity depending on the type of authentication mechanism (hop-by-hop or end to end). Figure 4 shows end-to-end vs hop-by-hop authentication.

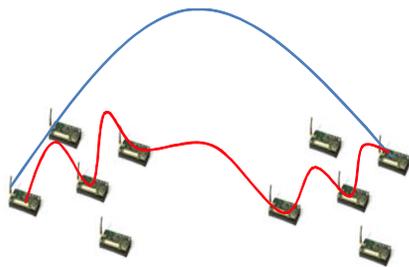


Figure 4: End-to-end Vs hop-by-hop authentication

In case the end user decides for end-to-end authentication in a network and if one of the intermediate nodes is compromised, the usual tendency of the end user is to wait till the timer expires and retransmits the digitally signed message via the same route. If the compromised sensor decides to drop all packets and distribute spurious routing tables to its neighbors, it will lead to a total compromise of the entire network. In such situation an adaptive authentication scheme would be beneficial as it would detect the malicious intermediate node by adapting from end-to-end authentication to hop-by-hop authentication. In addition, it would also increase the efficiency of the routing scheme of the entire network, as it will help un-compromised intermediate nodes to re-route the packets in an intelligent manner. If an end user would want to perform end-to-end authentication and send confidential information in WSN, a signcryption scheme would be more suitable as end users will be able to verify digital signatures and decrypt data in one operation. This scheme would fit the WSN environment as it would save the bandwidth and the battery life of the communicating sensors.

Additionally, the use of signcryption schemes will be beneficial in location and energy aware routing. Let us consider a WSN network, where the sink node has an inbuilt Global Positioning System (GPS) chip and is able to locate its position. Subsequently, all the other nodes are able to locate their position with reference to the sink node. Hence, each node is aware of its immediate neighbor's location and as well as the sink node. If the network adapts to a hop-by-hop signcryption scheme, each node could send its location

co-ordinates and its energy levels in a signcrypted fashion, leading to intelligent location and energy aware routing.

The security of the network could be strengthened if the network and data-link identities could be concatenated using Unique Universal IDs (UUIDs) [31]. Application level identities can also be concatenated using UUIDs.

Concerning ID based key distribution schemes in WSN, a PKG could be a single entity who could be able to distribute private keys or it could be a group of sensor nodes (in a distributed architecture) acting as a PKG and the security would be based on threshold key cryptography. Although Accountable IBE is a promising approach in a centralized architecture, where the PKG could be held accountable in case it leaks the end user's key for an identity, its applicability in the WSN is restricted and would be totally inefficient in an ad-hoc environment. Key revocation is also possible if a collected group of sensor nodes vote against a sensor node, and if the number of votes against the sensor node exceeds a specific threshold, the node will be revoked [30]. However, this scheme would be inefficient if all the nodes that vote were compromised and could vote against an honest node; thereby compromising the entire network. The ease of developing short new identities at any layer offers huge operational advantages in resource constrained environments such as WSN. Paradoxically, the problem of key escrow is disadvantages with its applicability to public domains, but would be advantages in case of military applications in which an ad-hoc group moderator would want to monitor traffic between communicating parties.

#### IV. CONCLUSION

As WSN are expected to proliferate in large numbers in the coming decade, security would play a vital role in the smooth running of these networks. However, these networks pose a unique challenge within security domain due to the traditional layered security scheme. In this paper, we make an early attempt to combine the cross layer approach with the identity based authentication schemes using pairing based cryptography to open a new paradigm of security solutions. As more and more security proposals are being proposed in WSN using pairing based cryptography, to maximize efficiency in computation and energy consumption, a fusion with the cross layer approach will help in providing better routing schemes, revocation mechanisms, immediate detection of sophisticated cross layer attacks and network survivability. The lightweight nature of TinyPbc (Pairing based cryptography library for WSN) would take less than 10% of RAM and about 20% of ROM on a Micaz mote, which would leave plenty of space to run reasonably heavy applications such as the identity based signature/signcryption schemes. In addition, we show that ID based authentication schemes are not only computationally efficient, but can also be used in making intelligent location and energy aware routing decisions.

#### REFERENCES

- [1] X. Du and H. Chen, "Security in wireless sensor networks," IEEE Wireless Communications, vol. 15, issue 4, pp 60-66, Aug. 2008

- [2] D. F. Aranha, R. Dahab, J. López, L. B. Oliveira, "Efficient implementation of elliptic curve cryptography in wireless sensors" *Advances in Mathematics of Communications*, vol. 4, issue 2, pp. 169-187, 2010.
- [3] X. Xiong, D. S. Wong and X. Deng, "TinyPairing: A Fast and Lightweight Pairing-based Cryptographic Library for Wireless Sensor Networks," in *IEEE Wireless Communications and Networking Conference*, pp 1-6, April 2010.
- [4] D. Boneh, H. Shacham, and B. Lynn, "Short signatures from the Weil pairing," in *Advances in Cryptology – ASIACRYPT 2001*, pp. 514-532, 2001.
- [5] D. Boneh and X. Boyen, "Short signatures without random oracles and the SDH assumption in bilinear groups," *Journal of Cryptology*, Springer-Verlag, vol. 21(2), pp.149–177, 2008
- [6] D. Boneh and M. Franklin, "Identity-based encryption from the Weil pairing," in *Advances in Cryptology – CRYPTO 2001*, LNCS 2139, pp.213-229, 2001.
- [7] C. Cocks, "An Identity Based Encryption Scheme Based on Quadratic Residues, Cryptography" in *Coding - Institute of Mathematics and Its Applications International Conference on Cryptography and Coding - Proceedings of IMA 2001*, LNCS 2260, pp 360-363, Springer-Verlag, 2001.
- [8] S. D. Galbraith, K. Harrison and D. Soldera, "Implementing the Tate Pairing," in *Lecture Notes in Computer Science (LNCS)*, vol. 2369, pp 69-86, Springer Berlin, 2002
- [9] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in *CCS 02: Proceedings of the 9th ACM conference on computer and communication security*, pp 41-47, 2002.
- [10] R. Ramanathan and J. Redi, "A Brief Overview of Ad Hoc Networks: Challenges and Directions", *IEEE Communication. Mag.*, vol 40, no 5, pp 20-22, 2002.
- [11] F. Ye, H. Luo, S. Lu and L. Zhang, "Statistical En-route Filtering of Injected False Data in Sensor Networks" *IEEE Journal in Selected Areas in Communication*, pp 839-850, 2005.
- [12] M. Xiao, X. Wang and G. Yang, "Cross-Layer Design for the Security of Wireless Sensor Networks," in *Proceedings of the 6th World Congress on Intelligent Control and Automation*, vol 1, pp 104-108, 2006.
- [13] D.J. Malan, M. Welsh and M.D Smith "A Public-Key Infrastructure for Key Distribution in TinyOS Based on Elliptic Curve Cryptography" in *IEEE SENCN*, pp 71-80, 2004.
- [14] Y. Zheng, "Signcryption and Its Applications in Efficient Public Key Solutions," in *Proceedings of Information Security Workshop*, 1997, LNCS, vol. 1397, Springer-Verlag, pp. 291-312, 1998.
- [15] R. Housley, W. Ford, W. Polk and D. Solo, "Internet X.509 Public Key Infrastructure Certificate and CRL Profile," IETF RFC 2459..
- [16] A. Shamir, "Identity-based Cryptosystems and Signature Schemes," in *Advances in Cryptology: Proceedings of CRYPTO 84*, Lecture Notes in Computer Science, vol. 196, Springer-Verlag, pp. 47-53, 1984.
- [17] R. L. Rivest, A. Shamir, and L. M. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM*, vol. 21, pp. 120-126, ACM NY, 1978.
- [18] N.P. Smart, "The Discrete Logarithm Problem on Elliptic Curves of Trace One," *Journal of Cryptology*, vol. 12, Springer New York, pp. 193-196, 1999.
- [19] P. Gemmell, "An introduction to threshold cryptography", in *Crypto-Bytes*, a technical newsletter of RSA Laboratories, Vol. 2, No. 7, 1997.
- [20] R. Gennaro, S. Jarecki, H. Krawczyk, T. Rabin, "Secure Distributed Key Generation for Discrete-Log Based Cryptosystems", *Advances in Cryptology - Eurocrypt '99*, Lecture Notes in Computer Science, Vol. 1592, Springer-Verlag, pp. 295-310, 1999.
- [21] V. Goyal, S. Lu, A. Sahai and B. Waters, "Black-Box Accountable Authority Identity Based Encryption", in *ACM Conference on Computer and Communications Security*, pages 89–98. ACM, 2008.
- [22] B. Lee, C. Boyd, E. Dawson, K. Kim, J. Yang and S. Yoo, "Secure Key Issuing in ID-based Cryptography," in *Conferences in Research and Practice in Information Technology*, 2004, vol. 32, pp. 69-74.
- [23] F. Hess, "Efficient Identity Based Signature Schemes based on Pairings," in *Selected Areas in Cryptography: 9th Annual International Workshop*, 2002, LNCS, vol. 2595, Springer-Verlag, pp. 310-324, 2003
- [24] B. Lynn, "Authenticated Identity-Based Encryption," available at <http://eprint.iacr.org/2002/072/>.
- [25] A. Gentry and A. Silverberg, "Hierarchical ID-Based Cryptography," in *Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security*, 2002, LNCS, vol. 2501, Springer-Verlag, pp. 548 - 566, 2002.
- [26] S. Chow, T. Yuen, L. Hui and S. Yiu, "Signcryption in Hierarchical Identity Based Cryptosystem," *Security and Privacy in the Age of Ubiquitous Computing, International Federation for Information Processing*, vol. 181, pp. 443-457, Springer Boston, 2005
- [27] Secure Hash Standard available at <http://csrc.nist.gov/publications/fips/fips180-2/fips180-2withchangenotice.pdf>.
- [28] S. Josefsson, "The Base16, Base32, and Base64 Data Encodings," IETF RFC 3548, July 2003.
- [29] National Institute of Standards and Technology (NIST). FIPS-197: Advanced Encryption Standard, November 2001, available at <http://www.itl.nist.gov/fipspubs/>.
- [30] H. Chan, V. Gligor, A. Perrig, and G. Muralidharan, "On the distribution and revocation of cryptographic keys in sensor networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 2, no. 3, pp 233-247, July-Sept. 2005
- [31] P. Leach, M. Mealling and R. Salz, "A Universally Unique Identifier (UUID) URN Namespace," IETF RFC 4122, July 2005.
- [32] L. Xiao, L. J. Greenstein, N. B. Mandayam and W. Trappe, "Using the Physical Layer for Wireless Authentication in Time-Variant Channels" in *IEEE Transactions on Wireless Communications*, vol. 7, No. 7, pp. 2571-2579, July 2008
- [33] D. F. Aranha, "Tiny Pairing-based Cryptography Library," available at <http://code.google.com/p/relic-toolkit/>
- [34] D. Galindo, R. Roman, and J. Lopez, "A killer application for pairing: Authenticated key establishment in underwater wireless sensor networks," in *Proc. of the 7th International conference on Cryptology and Network Security (CANS 2008)*, pp 120-132, 2008

# Data security in transportation solutions

Tomas ZELINKA

Faculty of Transportation Sciences, Czech Technical University of Prague  
110 00 Praha 1, Czech Republic

and

Zdenek LOKAJ

Faculty of Transportation Sciences, Czech Technical University of Prague  
110 00 Praha 1, Czech Republic

## ABSTRACT

Intelligent Transport Services expect availability of the secure seamless communications solutions typically covering widely spread areas. Different ITS solutions require different levels of telecommunications service quality. These parameters have to correspond with ITS service performance parameters required by specific service. Even though quite extensive range of wireless data services with reasonable coverage are provided by public telecommunications providers, practically all services are provided with no guaranteed data service quality and security. ITS requirements can be in most parameters better met if multi-path communications systems are applied where typically used public solution is combined with both other public as well as private ones where and when it is needed. Such solution requires implementation of relevant flexible system architecture supported by the efficient decision processes.

This paper is concentrated the telecommunications security issues relevant to the ITS wide area networking. Expected level of security varies in dependence on relevant ITS service requirements. It is well known fact that data volumes transferred both in private data vehicle on board networks as well as between vehicles and infrastructure or other vehicles progressively grow. Such trend leads to increase of the of the fatal problems appearance probability if security of the wide area networks is not relevantly treated. Probability of hazardous appearances additionally principally grows if vehicles networks are integrated in the wide area networks. That is reason why relevant communications security treatment becomes crucial part of the ITS solution. Besides of available “off shelf” security tools we present solution based on non-public universal identifier with dynamical extension (time and position dependency as a autonomous variables) and data selection according to actor role or category.

Presented results were obtained within projects e-Ident<sup>1</sup>, DOTEK<sup>2</sup> and SRATVU<sup>3</sup>.

## 1 INTRODUCTION

The processes in the ITS architecture are defined by chaining system components through the information links – see Fig. 1. The system component carries the implicit system function (like F1, F2, F3). The terminator (e.g. driver, consignee, emergency

vehicle) is often the initiator and also the terminator of the selected process.

The chains of functions (processes) are mapped in physical subsystems or modules. Second process is defined e.g. by chaining the functions G1, G2 and G3 and information flows between functions specify the communication links between subsystems or modules. If time, performance or other constrains are assigned to different functions and information links, the result of presented analysis is represented by table of system requirements assigned to each physical subsystem (module) and physical communication link between subsystems.

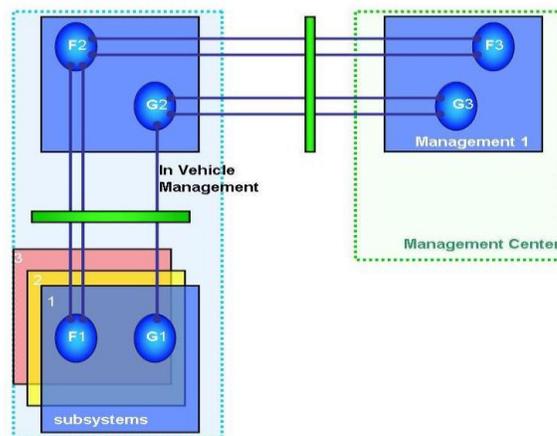


Fig. 1 ITS architecture

It is feasible to consider creation of several subsystem classes of the modular sub-system structure. In this case addition of appropriate optional module can extend or improve system parameters. The same principles are applied in the communication solution design. Such decomposition also simplifies analysis as well as synthesis of the systems where security parameters are accepted as the critical criteria.

## 2 TELEMATIC SUB-SYSTEM REQUIREMENTS

The methodology for the definition and measurement of following individual system parameters is being developed in frame of the ITS architecture and it is described in [1] - [5]. Individual system parameters – performance indicators - were accepted in frame of the ITS architecture:

- Reliability - the ability to perform required function under given conditions for a given time interval.
- Availability - the ability to perform required function at the initialization of the intended operation.
- Integrity - the ability to provide timely and valid alerts to the user when a system must not be used for the intended operation.

<sup>1</sup>e-Ident – Electronic identification systems within transport process – grant 2A-2TP1/108.of Ministry of Industry and Trade of the Czech Republic (MPO),

<sup>2</sup>DOTEK – Communication module for transport telematic applications –grant 2A-2TP1/105 of MPO,

<sup>3</sup>SRATVU – “System Requirements and Architecture of the universal Telematic Vehicle Unit” is grant 2A-1TP1/138 of MPO,

- Continuity - the ability to perform required function without non-scheduled interruption during the intended operation.
- Accuracy - the degree of conformance between a platform's true parameter and its estimated value, etc.
- Safety - ability of the system to ensure that no material damage nor of human life losses will occur in cases of non-standard events.

Decomposition of system parameters enables application of the follow-up analysis of telematic chains according to the various criteria (optimization of the information transfer between a mobile unit and processing centre, maximum use of the existing information and telecommunication infrastructure, etc.). It is obvious that quantification of requirements on relevant telecommunication solutions within telematic chains plays one of key roles in this process.

Mobility of the communication solution represents one of the crucial system properties namely in context of specific demand on availability as well as security of the solution.

Following communications performance indicators quantify communications service quality (see e.g. [6] - [10]):

- Availability – (Service Activation Time, Mean Time to Restore (MTTR), Mean Time Between Failure (MTBF) and VC availability),
- Delay is an accumulative parameter and it is effected by either interfaces rates, frame size or load/congestion of all in line active nodes (switches),
- Packet/Frames Loss (as a tool which not direct mean network failure),
- Security.

Performance indicators applied for such communications applications must be transformable into telematic performance indicators structure and vice versa. Indicators transformability simplifies system synthesis. Additive impact of the communications performance indicators vector  $tc_i$  on the vector of telematics performance indicators  $\Delta tmi$  can be expressed as  $\Delta tmi = TM \cdot tc_i$ , where TM represents transformation matrix. It is valid, however, only under condition that probability levels of all studied phenomena are on the same level and all performance indicators are expressed exclusively by parameters with the same physical dimension – typically in time or in time convertible variable (see e.g. [7]). Transformation matrix construction is dependent on the detailed communication solution and its integration into telematic system. Probability of each phenomena appearance in context of other processes is not deeply evaluated in the introductory period, when specific structure of transformation matrix is identified. In [7] - [10] are presented details of proposed iterative method.

### 3 COMMUNICATIONS SOLUTION STRUCTURE

Figure 2 presents telecommunications chain diagram, originally applied within the pilot project at Airport Prague (see e.g. [7]). We accepted this structure as typical architecture of ITS telematic solutions. On Board Units (OBU), GNSS Sensing System (SS) and set of Wireless Units (WL) are installed in the moving object. SS applies now exclusively GPS (Global Positioning System with no SLA publicly available), but there is expected launch of the European Galileo GNSS services as well as the second generation of the GPS services with guaranteed quality of service. OBU represents not only control but also display and human communication services and WL<sub>i</sub>

represents i-th cellular technology of the wireless complex solution. Terrestrial communication part consist of set of mobile cellular Base Stations (BS<sub>ij</sub>) (i-th bases station of the j-th cellular system) integrated by the terrestrial network based on L3/L2 switches/nodes (TN<sub>i</sub>) interconnected with Servers (S<sub>i</sub>). E2E (End to End) service is provided based on IP protocol, L2 switching is Ethernet protocol based.

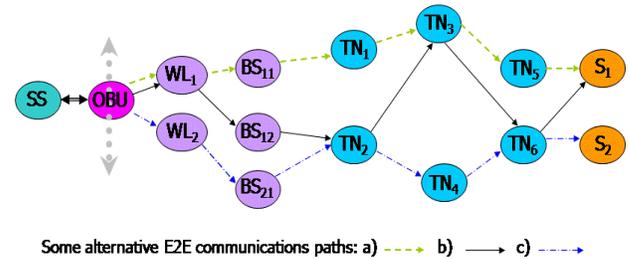


Fig. 2. Telematic telecommunication scheme in chain diagram

One preferred access wireless technology would be accepted (if possible) as the core solution to be combined with alternative solutions when and where it is needed. Principles of procedures supporting selection of the best possible communications solution quantified both by performance indicators and some other parameters e.g. like service cost, company policy as well. ISO TC204, WG16.1 “Communications Air-interface for Long and Medium range” (CALM) group presented their complex approach to resolve described procedures – see e.g. [11] or [12]. Complexity of the ISO approach offers solution with transparent RM OSI compatible architecture, however, such approach also represents highly demanding implementation phase requiring most probably some additional years to reach the market in reasonable pricing.

The IEEE 802.21 presents handover in heterogeneous networks standard known as Media-Independent Handovers (MIH) – see [18]. This standard is designed to enable mobile users to use full advantage of overlapping and diverse of access networks. IEEE 802.21-2008 provides properties that meet the requirements of effective heterogeneous handovers. It allows transparent service continuity during handovers by specifying mechanisms to gather and distribute information from various link types. The collected information comprises timely and consistent notifications about changes in link conditions and available access networks. Scope of IEEE 802.21-2008 is restricted to access technology-independent handovers and additional activities in this area are on the way. Handover decision and target assessment constitute a multiphase process where the assistance of IEEE 802.21 is essential. However, the actual handover execution is outside the scope of the IEEE 802.21 standard.

Authors of this paper recently introduced easily implementable alternative solution applicable namely for compact solutions like On Board Units (OBU) where all telecommunications technologies units are integrated into one compact system. This alternative is adoptable in much shorter time horizon if compared with system based on complex ISO CALM approach or IEEE 802.21 standard. Author's research team goal is to enable its solution for implementations in time period before solutions based on accepted CALM or 802.21 standards are commercially available in reasonable pricing. Authors adopted L3 “intelligent” routing which allows fast implementation namely in compact units like vehicle OBUs. It is based dominantly on the SW package system integration with minimal or no additional requirements on HW specific support. Results of the research are summarized e.g. in [31].

#### 4 DATA SECURITY

Security performance indicator see e.g. [15] describes ability of the system to ensure that no material damage or losses of human life will occur in cases of non-standard events (e.g. fake transaction). It means that system detects the forgery on a defined level of probability.

$$P(|W_i - W_{m,i}| \leq \varepsilon) \geq \gamma. \quad (1)$$

This equation describes that the difference between desired or non-standard risk situation  $W_i$  and real situations of risk or abnormal  $W_{m,i}$  does not exceed the level of probability  $\gamma$ . “Car to Infrastructure” (C2I) and “Car to Car” (C2C) communication as well as vehicles on board data communication via Controlled Area Network (CAN) bus are areas with progressive growth of transferred data volumes. If private on board network solution is not connected to any communication channel than it remains reasonably secure and no additional security treatment is typically needed and implemented. However, vehicle private data network security and integrity can be violated in a moment when this network is connected to any other device or network. It is absolutely necessary to take in account that most of CAN based vehicles are minimally equipped with interface for diagnostics purposes, anyhow. Above that interconnection of the CAN bus to the C2C or C2I communications structures becomes “trendy” namely due to on network representative data availability applicable for services like car identity or car units integrity or functionality remote identification. However, data security in such applications represents sensitive issue to be carefully studied and treated.

There are many in vehicle systems interconnected via CAN which can be attacked by hackers with potential of even fatal consequences. Reliable and secure identification of both partners for remote communication represents between others one of important security tools to prevent unauthorized exchange of any data. It must be combined with other security tools like encryption or more effective tunneling. Authentication of two actors for mutual communication based on identifier like VIN code or OBU-ID, however, is not acceptable as sufficient tool and extended approach must be applied.

Second security aspect which follows authentication is data privacy and actors authorization to data content knowledge. Authors’ approach is based on selective data transmission according to actor role/category. Security approach is covered in two steps – reliable and secure authentication and the only relevant to actor’s rights data exchange (data which can be provide to defined actor). These tools must be combined with other available security tools.

##### 4.1 Unique identifier

Presented approach is based on usage of Universal Identifier of Vehicle (UIV) is generated as set of all important partial vehicle identifiers where each of them describes non-changeable part of the car detailed identification. There are some examples of identifiers:

- VIN
- Nr. of axels
- Emission class
- Weight

- Year of manufacture

The UIV represents set of partial identifiers extended by unique non-public part generated from agreed data by standard cryptography algorithm (e.g. AES or SHA-2) to prevent possibility of UIV algorithm identification in case set of identifiers is for any reason known to the hacker. Check part at the end of identifier is connected for fast check of identifier validity (like validity check of credit card number). The example of UIV is on the Fig. 3.

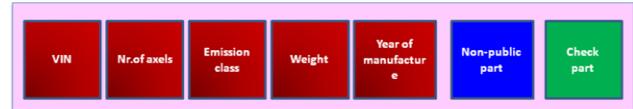


Fig.3. Example of unique identifier

It is not necessary to take care of UIV uniqueness because this functionality is ensured by unique VIN code. Advantage of such approach is that complex information about vehicle integrated in the UIV can be used for different telematic applications. Threat of sensitive data abuse is prevented by data selection availability to user in dependence on service class assignment to each one. System allows to use the only that parts of identifier which is dedicated to identified service class – like emergency, public and commercial services.

##### 4.2 Communication and security identification

As described above due to high sensitivity on data privacy exchanged between vehicle and service infrastructure VID must be reasonably protected against potential hackers attacks. Three categories of telematic system security in ITS are provided:

- Identifier and data security in vehicle (Vehicle environment),
- Identifier and data security for data transmission (wireless environment),
- Identifier and data security in receiver part (server area).

In this paper the only wireless environment part will be discussed.

The communication channel might be secured by standard cryptography methods, however, insufficient protection might be broken and hacker’s successful attack can misuse transferred data with fatal consequences. Proposed approach to data security yields in dynamical component extension (time and position dependency) and symmetric or asymmetric encryption, which is chosen depending on application.

For Point to Point (P2P) communication typically symmetric encryption is applied - e.g. like radar control on the highway. In case of Point to Multipoint (P2M) communications namely if large number of active terminals are served asymmetric cryptography can be efficiently used, as well. Such approach is relevant in case communication between systems of different owners with significant difficulties to manage relevant keys distribution.

In this solution the identifier is concatenated by actual time, current GNSS coordinates (i.e. exclusively in direction from by GNSS equipped vehicle to infrastructure) and finally by the user ID. Identifier is then encrypted by either asymmetric or symmetric cryptographic algorithm. Examples described on the Figure 4.

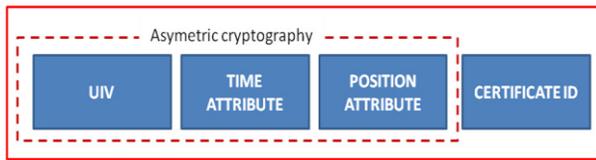


Fig. 4. Dynamic version of the identifier

Encryption of the UIV is described as follows:

$$M_1 = E_K (UIV \parallel T_i \parallel P_i), \quad (2)$$

where

- UIV = Universal Identifier of the Vehicle,
- $E_K$  = asymmetric encryption with public key K of the end receiver,
- $T_i$  = clock state in time of message generation,
- $P_i$  = position in time of message generation,
- $UIV \parallel T_i \parallel P_i$  = identifier with link to current time and position.

After receiving the request by system central system, the message  $M_1$  is decrypted and UIV is read in „static form“ - received time  $T_i$  and  $P_i$  are checked for validity. It means, that the message is not older than  $n$  seconds and the message has been sent from area with maximum of  $m$  meters tolerated difference. Data message with identifier in dynamic format is not impacted by this process and this approach doesn't influence usage of the other security tools.

The goal of this approach is to highly secure data against attacks mainly like eavesdropping and usage of the data for forgery.

### 5 SERVICE CATEGORIES

Proposed approach covers categorization of the telematic services. Each category has defined set of data allowed to user application. Because the unique identifier includes complex information about vehicle there must be special tool implemented on both sides (sender and receiver) which process incoming identifier and transfers and publish the only relevant data to user. On Figure 5 this component is described as an “Interface”. This component also covers “dynamisation” of the message content as it was already described above.

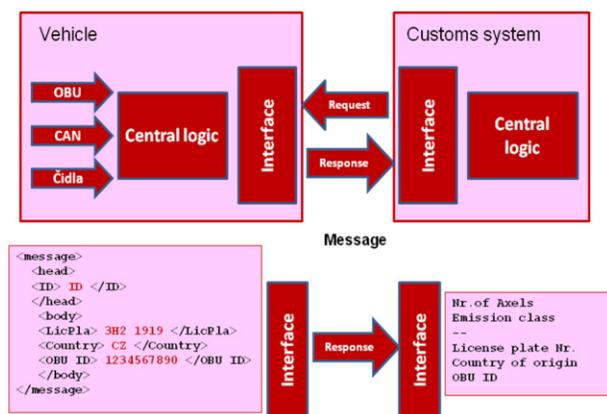


Fig. 5. Public service example – customs administration

Three service categories were defined:

- Security services – e.g. emergency, fire dept., police,
- Public services (public authorities) – e.g. customs,
- Commercial services.

Example on Figure 5 describes public services support dedicated for public institutions. Set of available data is identified by the unique identifier. Hand reader operated by customs officer generates request for identification and sends it to the vehicle unit - encrypted message contains user Public Encryption Key (PEK). Vehicle unit processes the request and sends relevant service category data according to the rights of customs administrations. Category is identified by PEK. User requires for example emission class, number of axles, license plate number, country of origin and OBU ID. Even though the other “public class” data are included in a sent UIV, the interface component splits the unique identifier and the only relevant data are publish, i.e. in this case just emission class, number of axles, license plate number, country of origin and OBU ID. Remaining data from the identifier are suppressed and are kept unreadable for the system.

### CONCLUSION

Due to complexity of ITS services requiring typically mobile services wide area coverage and selectable classes of services we focused our effort on wireless access solution designed as seamless combination of more independent access solutions of the same or alternative technology. Before analysis of the seamless handover/handoff based solution was designed and implemented the detailed analysis of the most frequently used communications solutions was done. Authors decided to design quickly and easily implementable alternative solution to complex solution based either on family of ISO standards known as CALM or IEEE 802.21. Solution adopts software based L3 routing effective in implementations in compact units like vehicles OBU.

“Car to Infrastructure” (C2I) and “Car to Car” (C2C) communication as well as vehicles on board data communication via Controlled Area Network (CAN) bus are areas with progressive growth of transferred data volumes. If private on board network solution is not connected to any communication channel than it remains reasonably secure and no additional security treatment is typically needed and implemented. However, vehicle private data network security and integrity can be violated in a moment when this network is connected via OBU to any other device or network.

CAN network with OBU interconnect will be provided more and more frequently namely due to representative data available on the CAN vehicle network. This data are applicable for telematics services either like vehicle as whole or its parts identity remote identification or vehicle units (like engine) functionality record remote processing. However, data security in such applications represents sensitive issue which has been carefully studied and treated.

Security of telecommunications channel can be resolved with wide range of available security tools. Strongly required reliable and secure authentication of the actors and data privacy (content) are, however, usually not sufficiently resolved by standardly available products. This paper presents solution based on non-public universal identifier with dynamical extension (time and position dependency as autonomous variables where relevant) as well as provided data selection according to actor role or category.

These tools have been implemented in laboratory OBU and they have been under procedure of the complex tests. The first results evaluations will be presented in occasion of WMSCI 2011 conference.

#### REFERENCES

- [1] M. Svitek, **Architecture of ITS Systems and Services in the Czech Republic**, International Conference Smart Moving 2005, Birmingham 2005, England.
- [2] M. Svitek, **Intelligent Transport Systems - Architecture, Design methodology and Practical Implementation**, Key-note lesson, 5th WSEAS/IASME Int. Conf. on Systems Theory and Scientific Computation, Malta 2005.
- [3] M. Svitek, T. Zelinka, **Communications Tools for Intelligent Transport Systems**, Proceedings of 10th WSEAS International Conference on Communications, pp 519 – 522, Athens 2006, ISSN 1790-5117, ISBN 960-8457-47-5.
- [4] M. Svitek, T. Zelinka, T., **Communications Solutions for ITS Telematic Subsystems**, WSEAS Transactions on Business and Economics, Issue 4 (2006), Vol. 3, pp 361 – 367, Athens 2006, ISSN 1109-9526,
- [5] M. Svitek, T. Zelinka, **Telecommunications solutions for ITS**. Towards Common Engineering & Technology for Land, Maritime, air and Space Transportation – ITCT 2006, CNISF, Paris 2006.
- [6] M. Svitek, T. Zelinka, **Communication solution for GPS based airport service vehicles navigation**, EATIS'97 ACM-DL Proceedings, Faro (Portugal) 2007, ISBN #978-1-59593-598-4.
- [7] T. Zelinka, M. Svitek, **Communication solution for Vehicles Navigation on the Airport territory**, Proceedings of the 2007 IEEE Intelligent Vehicle Symposium, Istanbul, Turkey, pp 528–534, IEEE Catalogue number 07TH8947, ISBN 1-4244-1068-1.
- [8] M. Svitek, T. Zelinka, **Communications Environment for Telematic Subsystems**, Proceedings of 11-th World Multi-Conference on Systemics, Cybernetics and Informatics, Volume II, pp 362-367, IIS/IFSR, Orlando, FL, USA, ISBN-10: 1-934272-16-7, ISBN-13: 978-1-934272-16-9
- [9] M. Svitek, T. Zelinka, **Communications Challenges of the Airport Over-ground Traffic Management**, Proceedings of the 11th WSEAS International Multi-conference CSCC, Volume – Advances in Communications, pp. 228 – 234, Agion Nikolaos, Crete Island, Greece, ISSN 1790-5117, ISBN 978-969-8457-91-1.
- [10] T. Zelinka, M. Svitek, **Communications Scheme for Airport Service Vehicles Navigation**, Proceedings of International Conference TRANSTEC Prague, Czech Technical University, Faculty of Transport Science and University of California, Santa Barbara, Praha 2007, pp. 160 – 166, ISBN 978-80-01-03782-9
- [11] B. Williams, **CALM handbook V1.0**. Document ISO TC204 WG.16.1 CALM, 2004.
- [12] N. Wall, **CALM - why ITS needs it**, ITSS 6 (September), 2006
- [13] T. Zelinka, M. Svitek, **CALM - Telecommunication Environment for Transport Telematics**, Technology & Prosperity, 2006, Vol. XI, special edition (11/06), ISSN 1213-7162.
- [14] K. Yang, J. Wittgreffe, M. Azmoodeh: **Policy-Based Model-Driven Creation of Adaptive Services in Wireless Environments**. IEEE Vehicular technology Magazine, September 2007, pp. 14-20.
- [15] M. Svitek, **Dynamical Systems with Reduced Dimensionality**, Neural Network World edition, II ASCR and CTU FTS, Praha 2006, ISBN:80-903298-6-1, EAN: 978-80-903298-6-7.
- [16] A. Dempster, N. Laird, D. Rubin, **Maximum likelihood from incomplete data via EM algorithm**. J. Royal Stat. Soc. 39, 1977, pp 1-38.
- [17] T. Zelinka, M. Svitek, **Communication Scheme of Airport Over-ground Traffic Navigation System**. Proceedings of the International Symposium on Communications and Information Technologies - ISCIT 2007. IEEE Sydney, 2007, pp 329 - 334. IEEE Catalogue No. 07EX1682(C), ISBN 1-4244-977-2, Library of Congress 2007920360.
- [18] M. Svitek, T. Zelinka, **Monitoring of Transport Means on Airport Surface**. Proceedings of 7-th International Conference on Transport Systems Telematics, Katovice-Ustron, 2007, pp. 285 – 292, ISBN 978-83-917156-7-3.
- [19] M. Svitek, T. Zelinka, **Monitoring of Transport Means on Airport Surface**. Advances in Transport Systems Telematics, Monograph edited by Jerzy Mikulski, Selesian University of Technology, Katowice, pp. 285 – 292, ISBN 978-83-917156-6-6.
- [20] T. Zelinka, M. Svitek, **Decision processes in telematic multi-path communications access systems**. International Journal of Communications, North Atlantic University Network NOUN, Issue 2, Volume 1, 2007, pp.11 – 16.
- [21] M. Svitek, T. Zelinka, **Communications multi-path access decision scheme**. Neural Network World, ICS AS CR and CTU, FTS, Praha, No. 6., 2008, pp 3 - 14, 2008, ISSN 1210 0552,
- [22] M. Svitek, T. Zelinka, **Decision processes in Communications Multi-path Access Systems applied within ITS**. Transactions on Transport Science, MTCR, Praha, No. 1, 2008, pp 3-12, ISSN 1802-971X,
- [23] T. Zelinka, M. Svitek, **Identification of Communication Solution designated for Transport Telematic Applications**. WSEAS Transactions on Communications, Issue 2, Volume 7, Athens, 2008, pp 114 – 122, ISSN: 1109-2742.
- [24] T. Zelinka, M. Svitek, **Multi-path communications access decision scheme**. Proceedings of the 12-th World Multi-Conference on Systemics, Cybernetics and Informatics, Volume III, pp 3233-237, IIS/IFSR, Orlando, FL, USA, ISBN-10: 1-934272-32-7, ISBN-13: 978-1-934272-33-6.
- [25] T. Zelinka, M. Svitek, **Adaptive communications solutions in complex transport telematics systems**. Proceedings of the 11th WSEAS International Multiconference CSCC 2008, Volume – New Aspects of Communication, pp. 206 – 212, Heraklion, Greece, ISSN 1790-5117, ISBN 978-960-6766-84-8.
- [26] T. Zelinka, M. Svitek, **Adaptive communications solutions in complex transport telematics systems**. Monograph on Computers and Simulation in Modern Science-Volume II, WSEAS Press, Athens 2009, pp. 234 -241, ISBN 978-960-474-032-1.

[27] T. Zelinka, M. Svitek, **Adaptive Wireless Access Environment in Transport Solutions**. Proceedings of, 13-th World Multi-Conference on Systemics, Cybernetics and Informatics, Volume IV, pp 310 - 315, IIS/IFSR, Orlando, FL, ISBN 978-1-934272-62-6.

[28] T. Zelinka, M. Svitek, M. Vosatka, **Adaptive Approach to Management of the Multi-path Wireless Solutions**. Proceedings of the Symposium Recent Advance in Data Network, Communications, Computers, WSEAS Press, Morgan State University, Baltimore, 2009, pp. 161 – 168, ISBN 978-960-474-134-2.

[29] T. Zelinka, M. Svitek, M. Srotyr, M. Vosatka, **Adaptive multi-path Telecommunications Solutions for ITS**.

Proceedings of, 14-th World Multi-Conference on Systemics, Cybernetics and Informatics, Volume I, pp 89 - 94, IIS/IFSR, Orlando, FL, USA, ISBN 978-1-934272-98-5.

[30] T. Zelinka, M. Svitek, Z. Lokaj, **Adaptive Decision Processes the Multi-path Wireless Access Solutions Implementable on the IP Routing layer**. EATIS'10 Proceedings, Panama City (Panama), 2010, ISBN 978-958-44-7280-9

[31] T. Zelinka, M. Svitek, M. Srotyr, M. Vosatka, **Adaptive Multi-path Telecommunications Solutions for ITS**, Journal of Systemics, Cybernetics and Informatics Volume 9, No. 1, pp. 14 – 20, Orlando, 2011, ISSN: 1690-4524.

# Cloud Service Feature driven Security Policies for Virtualized Infrastructures

Ramaswamy Chandramouli  
National Institute of Standards & Technology  
Gaithersburg, MD, USA  
mouli@nist.gov

## ABSTRACT

With the increasing maturity of various cloud service delivery models ( Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS)) and deployment models (Private, Community, Public, Hybrid), the security risk profile of each cloud service configuration is coming into focus. In this paper, we take up the example of a Public Infrastructure as a Service (IaaS) cloud provider who provides computing services through a data center with a virtualized infrastructure. In order to provide the needed security assurance for its IaaS cloud offering, the cloud provider needs to implement various security measures as part of the infrastructure configuration. A precursor to developing security measures is a comprehensive security policy. Now these policies are dependent upon the set of service features that the IaaS cloud provider provides as part of its offering as well as internal administrative capabilities needed to support those features. The focus of this paper is to illustrate an approach for derivation of appropriate security policies based on the security goals of functions associated with internal administration capabilities and cloud service features.

Keywords - Cloud computing, Infrastructure as a Service, Public cloud, Security Policy, Virtualization.

## 1. INTRODUCTION

The security risk profile of a cloud service is dictated by the following factors. We call the variables associated with these factors are Cloud Service Orchestration variables.

- Cloud Delivery Model (Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS)) [1]
- Cloud Deployment Model (Private, Hybrid, Community, Public)
- The functional capabilities offered by the Cloud Service Provider (we call this as Service Feature Set in this paper) and
- The Technologies used to enable the functional capabilities (e.g., Virtualization, Federation Protocols etc)

A given cloud service, therefore, is defined in terms of the values of the applicable cloud service orchestration variables from the above factors. Out of the above four factors, once the cloud delivery model and cloud deployment model is chosen, the detailed security risk analysis and the security threats are entirely dictated by the functional capabilities offered by the cloud service provider and the underlying technology used in the IT infrastructure used by the cloud service provider to enable those functional capabilities that constitute the service package.

The cloud service that is the focus of this paper has Infrastructure as a Service (IaaS) as the cloud delivery model and Public Cloud as the Cloud deployment model. The main cloud technology we consider is virtualization [2]. We also consider a comprehensive set of functional capabilities that can be provided based on the current state of virtualization technology and market orientation.

It must be mentioned that not every IaaS cloud offering has to be enabled only using virtualization technologies. The economic viability of an IaaS is in large part enhanced due to abstraction of resources (computing, storage and network) and virtualization technology is the most prevalent one used today for providing that abstraction.

Now, based on the description of the our reference cloud service, it should be clear that the cloud service provider scenario we are referring to is an IaaS cloud provider who provides storage and computing services through a data center with a virtualized infrastructure. The use of a virtualized infrastructure by an IaaS provider makes the most economic sense because of the following:

- IaaS cloud provider is able to utilize the entire set of computing, storage and network equipment in an optimum fashion with high level of utilization.
- IaaS cloud provider is able to quickly ramp up the resources with minimal re-configuration whenever the demand from the existing cloud subscribers increases or new subscribers join the service. [3]
- IaaS is able to offer diversity in terms of computing power/throughput (through custom configuration of Virtual Machines) as well as platform diversity (e.g., different operating systems).

Once the IaaS cloud service provider takes care of the adequacy of the resources (computing, storage and networking capabilities), the next two critical issues are:

- Service Management
- Security (Adequate protection for cloud subscriber's data and application) [4]

In order to offer a robust and secure IaaS cloud offering, the cloud provider has to have the following:

- A set of administrative capabilities relating to the virtualized infrastructure in its data center.
- A set of features that is part of the cloud service that the cloud provider offers which the cloud subscriber is able to invoke based on its subscription package.

In order that the internal virtualized infrastructure of the IaaS cloud provider is secure and the service orchestration/configuration maintains its integrity, there needs to be a set of security policies associated with each internal administrative capability as well as with a cloud service feature. The purpose of this paper is to enumerate these internal administrative capabilities (IAC) and cloud service features (CSF), analyze the security goal associated with each and then provide a statement or scope of security policies that should go with each administrative capability or cloud service feature. The enumeration is based on the author's analysis of the current state of virtualization technology deployed in commercial IaaS cloud service offerings as well as common subscriber-facing service features currently available in those offerings.

The need for strong security measures and hence a comprehensive set of security policies is much more critical for protecting virtualized infrastructures owned/operated by cloud providers compared to virtualized infrastructures deployed as part of the enterprise IT architecture [5]. This is due to the fact that enterprise users who are cloud service subscribers have shifted their trust from the IT resources in their data centers to IT resources under the control of cloud providers. This trust is needed for satisfying the internal audit requirements as well as public regulatory compliance needs for enterprise IT users.

Before we provide the organization of this paper, a note regarding certain naming conventions used in this paper is in order. The following are the conventions used:

- IAC-Fx – denotes function Fx (x being the running sequence) associated with Internal Administration Capability
- IAC-SPx – denotes security policy SPx associated with Internal Administrative Capability
- CSF-Fx – denotes function Fx associated with Cloud Service Feature
- CSF-SPx – denotes security policy SPx associated with Cloud Service Feature

The organization of the rest of the paper is as follows. In sections 2 through 5 we look at the administrative capabilities (IAC-F1 through IAC-F8) (both technical and procedural) required for an IaaS cloud provider to provide security for its virtualized infrastructure and derive the associated security policies (IAC-SP1 through IAC-SP8). In sections 6 through 9, we describe the various features that an IaaS cloud provider could offer as part of its service package (CSF-F1 through CSF-F15) and the security policies associated with each of those service features (CSF-SP1 through CSF-SP15). Section 10 provides the conclusions and benefits.

## 2. FEATURES AND POLICIES RELATING TO PROTECTION OF VIRTUALIZED HOSTS

A physical host on which is mounted a Virtual Machine Monitor (VMM) or Hypervisor is called a Virtualized Host. A Virtualized host is capable of supporting multiple Virtual Machines (VMs) each with its own operating system (called a Guest O/S).

A virtualized host is the basic building block of the virtualized infrastructure of an IaaS cloud service provider. It is also the foundational resource offered in the service since cloud subscribers define their service entities (i.e., virtual machines or

VMs) on it. Hence adequate protection for all virtualized hosts is critical for the security of the entire IaaS cloud service. With this as the driving security goal, we have to look at the set of administrative capabilities or functions that are needed to realize it.

In a datacenter set up for a public IaaS cloud service, the cloud provider organization needs to provide direct access to virtualized hosts only for a limited set of users who need to manage and/or monitor the virtualized infrastructure elements such as the Virtual Machines, Virtual Security Appliances and Virtual Networks.

This may require the following functions:

*IAC-F1: The users accessing a virtualized host should be authenticated using a robust authentication mechanism*

*IAC-F2: The means of user and network accesses to the virtualized hosts should be tightly controlled.*

*IAC-F3: To perform management functions on a virtualized host (such as monitoring of inter-VM traffic), a virtual network must be setup inside a virtualized host (hypervisor specifically).*

The security policies associated with the above functions based on the state of the technology are as follows:

*IAC-SP1: Authentication policies should call not only for robustness in some selected virtualized hosts but also some uniformity and consistency across all virtualized hosts. This may require that authentication of users into individual virtualized hosts be integrated with a directory service (e.g., Active Directory) using a secure protocol (e.g., Kerberos) so as to define and enforce consistent authentication policies.*

*IAC-SP2: Access policies relating to user and network access to virtualized hosts should cover designation of users/secure channels for remote access (e.g., SSH), designation of client nodes (e.g., TCP Wrappers) and designation of allowable in-bound and out-bound communication ports (e.g., through firewall rules)[6].*

*IAC-SP3: The virtual network in the hypervisor should be configured such that there is a dedicated management network. This is often accomplished by having a dedicated virtual switch inside a hypervisor connect to a dedicated physical network adapter/network interface card on the host.*

## 3. FEATURES AND POLICIES RELATING TO CREATION & STORAGE OF VIRTUAL MACHINE IMAGES

Many of the IaaS cloud providers offer a set of virtual machine (VM) images for use by their subscribers to create and launch VM instances. The following are the administrative capabilities required if the IaaS cloud provider offers this feature:

*IAC-F4: The cloud providers should have tools to create secure VM images.*

*IAC-F5: The cloud provider should provide a means to securely store the VM images in a Image Repository*

*IAC-F6: The cloud provider should maintain the security status of VM images in the Image Repository*

The corresponding security policies required to support the above capabilities are given below:

*IAC-SP4: The IaaS cloud provider should develop security baselines for various types of VM Images it wants to offer to cloud subscribers. These baselines should cover Guest OS versions, configuration values and anti-malware software.*

*IAC-SP5: The IaaS cloud provider should ensure that only authorized administrators can access, create, store or replace VM images in the image repository.*

*IAC-SP6: The IaaS cloud provider should have policies for periodically scanning and updating the VM images through operations such as patch application, configuration changes or updating anti-malware signatures.*

#### 4. FEATURES AND POLICIES RELATING TO MANAGEMENT OF VIRTUALIZED INFRASTRUCTURE

Having looked at the administrative capabilities and policies governing secure access to virtualized hosts, the next step is configuration of privileges and permissions needed for the management of the entire virtualized infrastructure. The management functions that can be performed from the individual virtualized hosts are limited to those that pertain to virtual machines residing on that host. In most practical infrastructures owned by the IaaS cloud provider, a scalable and efficient management process takes place through a centralized management server [6] that has visibility into multiple virtualized hosts.

Some of the functions related to management of entire virtualized infrastructure are stated below:

*IAC-F7: The cloud provider should have tools and capabilities for performing key management functions on the virtualized infrastructure such as: (a) Creation of a cluster consisting of a group of virtualized hosts and (b) Balancing the workload among the hosts within a cluster by having the capability to perform live VM migration across the hosts within the cluster.*

The policies associated with the above privilege management functions are:

*IAC-SP7: (a) The cloud provider should limit the capability to perform management functions on the virtualized infrastructure to a selected set of qualified administrators (b) The permissions should be granular and their assignment to authorized administrators should be based on principles such as Least Privilege, Separation of Duty etc and (c) The integrity and efficiency of permission assignment process should be improved by encapsulating related set of permissions into a role with meaningful names and using the roles as the basis for allocation to administrative users.*

#### 5. FEATURES AND POLICIES RELATING TO OPERATIONAL PROCESSES AND PROCEDURES

Cloud subscribers may be hosting data on an IaaS cloud provider's infrastructure whose creation, protection and dissemination are subject to public regulations such as SoX, HIPAA, GLBA and PCI-DSS. When these categories of data are hosted in their own data centers, the trust in systems, processes and procedures that are needed to ensure compliance rest with the employees and management of the enterprise. However when this data is resident in cloud provider infrastructure, the trusted has to be shifted to the employees and management of the cloud provider.

Hence the cloud provider has to perform the following:

*IAC-F8: The IaaS cloud provider should create trust in the cloud subscriber with respect to personnel and procedures for secure operation of its infrastructure.*

In order to provide this trust, the cloud provider should have the following policy in place:

*IAC-SP8: The cloud provider should have a process in place to perform minimal background verification of the personnel that will be managing its cloud service related assets and resources. Also it should have in place processes and procedures for Administrator accountability (through granular logging mechanisms), Incident Notification, Response and Remediation.*

#### 6. FEATURES & POLICIES RELATING TO SUBSCRIPTION HANDLING

All of the administrative capabilities and the governing policies we have looked at so far are purely internal to the IaaS cloud provider. Let us now look at the features and associated security policies that are part of the cloud provider's service offering and hence subscriber-facing.

Every cloud provider provides a web based interface for registration of new/potential cloud subscribers.

*CSR-F1: Before a new subscriber is enrolled in the service, the IaaS cloud provider should perform the following: Before a new user is formally registered and allowed to open accounts, the cloud provider should have a mechanism to perform some minimal background checks (e.g., credit history check). The threat due to a rogue subscriber is real and the "Abuse and Nefarious Use of Cloud Computing" as a threat is outlined in a study by Cloud Security Alliance [7].*

In order that the cloud provider's integrity of operations is not threatened due to a rogue subscriber, the following policy is called for:

*CSR-SPI: The cloud provider should perform some minimal background check on every new subscriber. Additionally there should be a means of monitoring a subscriber's network traffic to ensure that a rogue subscriber is not using the cloud provider's computing, storage and network resources to host malicious programs such as botnets, Trojan horses etc.*

#### 7. FEATURES & POLICIES RELATING TO VIRTUAL MACHINE CREATION AND OPERATION

Recall that the IaaS cloud provider provides network, computing and storage resources. The unit of computing resources that the cloud provider provides to a cloud subscriber is the Virtual Machine. Baseline features in VM resource offering by an IaaS cloud provider includes the ability for the subscriber to specify the O/S and the applications to be run on the subscribed VM and the capability to launch that VM with the required profile and subsequently stop, monitor and re-start that VM. The following is a set of features and associated policies related to VM creation and operation.

##### 7.1 Authenticity of VM Images

*CSR-F2: The cloud provider offers pre-defined VM Images or Templates. These templates may contain just a O/S (e.g., with a hardened version of O/S) or they may be application-specific (e.g., a webserver with a Linux O/S and Apache Web Server).*

The security policy that will provide the authenticity of the image is as follows:

*CSR-SP2: All VM Images offered by cloud provider should be digitally signed objects so as to provide assurance to the cloud subscriber that they were created by the cloud provider or its authorized agent and have not been tampered with.*

## 7.2. Security Assurance for VM Images

*CSR-F3: The cloud provider would like to offer secure VM Images. A secure VM Image [8] contains in addition to the O/S and/or necessary application, all security tools typically found in a physical server such as Firewall, IDS/IPS, Anti-Malware [9], Anti-Virus along with applicable patches for the version of O/S.*

The policy that will enable this service feature is:

*CSR-SP3: The IaaS cloud service provider should define and publish baseline security configurations that it has used for creating all types of VM images it offers.*

## 7.3. Launching Customized VM Images

*CSR-F4: The cloud provider offers tools to customize a VM Image from a pre-defined Image or tools to build a customized VM Image from scratch.*

*CSR-F5: The cloud provider offers tools for the subscriber to launch either a pre-defined VM Image or a customized VM Image (called as Public Image and Private Image respectively by some Cloud providers). Once launched, a VM Image becomes a VM instance. Generally it will be possible to launch multiple instances using a particular VM Image (pre-defined or customized).*

An associated policy is:

*CSR-SP4/5: For all VMs launched in the cloud provider infrastructure, a unique ID should be provided. This unique ID is necessary to monitor the security status [10], [11] (as well as performance) of the VM instance. Further, the cloud provider should provide a secure channel (e.g., Virtual Private Network (VPN)) to communicate and perform all tasks (both user-level and administrative) on those VM instances.*

## 7.4 Design of VM Profiles and allocation of VMs to minimize impact of Denial of Service Attacks

*CSR-F6: To enable cloud subscribers to build a VM that meets their application needs, cloud providers provides a pre-defined types of VMs with designated resource profiles. A resource profile is some combination of CPU and Memory resources that will be available to the VM instance that is going to be launched and hence is an indicator of the type of performance that can be expected. For example if the subscriber needs a VM for a compute-intensive application, he/she may choose VM type whose profile has a relatively higher CPU resource as opposed to Memory resource. On the other hand if the subscriber is going to deploy the VM instance for voluminous data crunching such as database processing, he/she may choose a VM type with a profile with relatively large memory resource compared to CPU resource so as to support the large throughput.*

The associated policy is:

*CSR-SP6: The cloud provider should have VM deployment/allocation policies that assign the type of VMs that should be run on various virtualized hosts based on the resource capacity of the latter. This policy is needed to minimize the impact of any Denial of Service attack launched*

*from any VM from incapacitating the entire physical host on which the VM resides.*

## 7.5 Re-starting of Dormant VMs

*CSR-F7: The IaaS cloud provider provides the capability to pause/suspend and later on re-start VMs for the cloud subscriber.*

The policy that should govern this operation is:

*CSR-SP7: The cloud provider should have a tool to enable the cloud subscriber to do the following: (a) Verify that the dormant VM being re-started conforms to a security baseline and (b) Has the ability to generate an alert with the exact violation (e.g., security patch for Guest O/S is out of date)*

The above policy is critical because of the following reasons:

1. If the VM has been dormant for a long time, the security patch for the O/S and or an application (e.g., Web server or DBMS) would have gone out of date
2. The rules for the Virtual Firewall in the VM would have to be changed following re-configuration of Virtual LANs

We are aware of the fact that some IaaS cloud service providers do not want to take responsibility for VM instances launched by cloud subscribers and they make it part of their SLAs. However, we feel that ensuring that VM instances running in your infrastructure has some minimal security is important since there is the possibility of coexistence of VMs from two different cloud subscribers on the same physical host and the threat of inter-VM attacks due to the VM from one of the subscribers being a rogue (or a victim being used as a launch pad for originating an attack).

## 7.6 Publication of VM Representation formats

Cloud subscribers would like to migrate their VM instances from one cloud provider to another and from their own internal infrastructure to that of a cloud provider.

To enable this, the cloud provider should provide the following:

*CSR-F8: The cloud provider provides information about the formats used for representation of VM Images.*

The following policy should go with the above feature.

*CSR-SP8: Cloud providers should publish the representational formats for the VMs running in their Virtualized infrastructure. The representational formats may vary with the type of virtualization product (technology) used by the cloud provider. The representational formats could be one of the following:*

- Proprietary Format
- Open and Interoperable Format (e.g., OVF)
- Open and Interoperable Format with some extensions

It may be argued that publication of the representational format for the VM images that run in the cloud provider infrastructure has a bearing on portability and interoperability of VMs and has nothing to do with the security. However, we feel that for VMs to operate with the same security protection, it is necessary for cloud providers to publish the representational format for VMs.

## 7.7 Migrating a VM from one Physical Host to Another

Many IaaS cloud providers would like to create the concept of clusters (groups of physical hosts) for better management of virtual machines and for efficient utilization of resources within their data center. With the help of clusters, the cloud provider can move virtual machines across the physical hosts in case of failures or planned maintenance of those hosts. Also the

migration of VMs can be used in situations where workloads need to be balanced across physical host systems.

The description of the service and its associated policies are:

*CSR-F9: The cloud provider has tools and capabilities to migrate VMs across (virtualized) physical hosts within a cluster in its infrastructure*

*CSR-SP9: IaaS cloud providers with capabilities to migrate VMs across physical hosts should have policies for the following: (a) The integrity of the VM is not jeopardized during the migration (e.g., through tampering of VM-related files-containing configuration & storage details) and (b) The security policy data associated with VM is also migrated (e.g., the sensitivity level of the VLAN in which the VM can be connected).*

### 7.8 Porting Applications out of a VM or De-provisioning a VM

Cloud subscribers may port an application running in a cloud provider VM to their own internal infrastructure or to another VM in another cloud provider. Similarly they could de-provision an entire VM instance. In these instances, it is possible that the ported application is a mission critical one that has a bearing on the business competitiveness or safety of the cloud subscriber organization. Alternately, the data handled by the ported application may have privacy implication. Hence it is necessary from a security perspective that the cloud provider has well articulated policies for destruction of files related to the ported application and the data associated with it. (Even if the application files and data are moved instead of being copied, copies of files relating to application and data may still be residing in the backup volumes of the cloud service provider).

*CSR-F10: The cloud provider provides the capability for the subscriber to de-provision VMs*

*CSR-SP10: The cloud provider has a clear set of procedures for:*

- *destruction of system files (O/S and Configuration) relating to de-provisioned VMs*
- *destruction of files relating to applications ported out of the cloud provider infrastructure*

## 8. FEATURES AND POLICIES RELATING TO SUBSCRIBER APPLICATION-LEVEL PROTECTION

A cloud subscriber to an IaaS cloud service may install and run several applications on the set of VMs that he/she has created in the provider's infrastructure. The subscriber may demand separation of the traffic between the applications based on the following logic:

- Separation of traffic between different tiers of the application. In this scenario, the subscriber will dedicate separate VMs for each tier of the application such as Web Server, Application Server, Database Server etc.
- Separation of traffic between different environments. In this scenario, the subscriber needs logical separation of traffic between VMs belonging to different application environment entities such as Production, Development, Quality Assurance etc.
- Separation of traffic between applications with different sensitivity levels. In this scenario, the subscriber needs separation of traffic between VMs running applications of

different sensitivity levels (e.g., an application processing credit card numbers Vs an application that provides access to product documentation)

In addition to separation of traffic pertaining to subscriber application, the cloud provider may need to create a separate virtual sub network for traffic relating to management of physical (virtualized) host as well as for access to storage resources under a Storage Area Network (SAN) or any other Network Attached Storage (NAS).

To enable the above the cloud provider may provide the following feature:

*CSR-F11: The cloud provider may provide capabilities to segment one physical LAN into a number of logical LANs or Virtual LANs (VLANs) to provide for separation of traffic emanating from different VMs on the same physical host [13].*

The following policy should be in place for cloud providers offering this VLAN segmentation feature.

*CSR-SP11: The policies for VLAN segmentation should consist of the following: (a) VLANs should be configured such that there is separation of traffic relating to provider's infrastructure related functions and subscriber application functions. The provider's infrastructure related functions include management of (virtualized) physical host and access to network attached storage resources (b) The configuration of virtual subnets relating to subscriber application functions should conform to the logical separation required by subscribers (based on application tiers, application sensitivity levels, application environments (production, development) etc)*

In many cloud provider infrastructures, VLAN segmentation can span several physical (virtualized hosts) whenever there exists the concept of cluster containing groups of virtualized hosts. The same policies stated above should apply to that scenario as well.

In some instances, the cloud provider can enable segmentation of traffic within a virtualized host using the following feature:

*CSR-F12: A cloud provider may install and run a virtual security appliance that runs in a specialized security-hardened VM and provides firewall (and in some instance IDS/IPS) functions.*

The following policies should be in place for offering this security appliance feature.

*CSR-SP12: The security appliance configuration policy should consist of the following: (a) Security Appliance should have visibility into traffic emanating from all VMs in the physical host. (b) The Security Appliance's firewall should be configured to restrict inter-VM traffic based on the requirements of the cloud subscriber.*

### 8.1 Features and Policies Relating to Service Management API

Cloud provider provides Interfaces and APIs to cloud subscribers for performing several operations. Since the cloud subscriber may be hosting several applications on the VMs, he/she may need to provision several types of information needed for application services such as authentication and authorization.

The following are some features relating to these functions:

*CSR-F13: The cloud provider provides API for the following functions:*

- *Uploading Authentication Credentials*

- Provisioning of User Information (e.g., Account Names and Privileges)

The security policy related to this API offering is:

*CSR-SP13: The cloud provider should ensure that the interfaces and APIs they provide to cloud subscribers can only be used for the intended functions and cannot be used to launch attacks (including Denial of Service (DOS)) on the cloud provider infrastructure.*

## 9. FEATURES AND POLICIES RELATING TO SUBSCRIBER MONITORING/MANAGEMENT CAPABILITIES

### 9.1 Subscriber Monitoring & Limited Management

As part of IaaS cloud offering, cloud providers provide monitoring (performance, security, usage etc) and some limited management capabilities to the cloud subscribers.

The capability relating to this service feature is described as follows:

*CSR-F14: The Cloud provider may have installed several management tools for the virtualized infrastructure (individual virtualized hosts and cluster of hosts) for monitoring (performance and security states), metering the usage of resources and may provide access to some of these tools to the cloud subscribers.*

The due diligence policy associated with the above is as follows:

*CSR-SP14: The cloud provider should ensure that all management interfaces to hypervisor and other components of the virtualized infrastructure are connected to a dedicated VLAN meant only for management function and no other form of user interaction. Further access to the management capabilities must be provided to subscribers and cloud provider only through a mechanism that provides strong authentication (preferably multi-factor) and a robust role-based authorization.*

## 11. REFERENCES

- [1] P.Mell, T.Grace, "A NIST Definition of Cloud Computing, NIST SP 800-145, <http://csrc.nist.gov/publications/PubsDrafts.html#SP-800-145>
- [2] P.Barham et al., "Xen and the Art of Virtualization," **Proc. 19th ACM Symp. Operating System Principles**, ACM Press, 2003, pp. 164-177.
- [3] R.Ghosh et al., "Performability Analysis for Infrastructure-as-a-service cloud: an interacting stochastic model approach," in **IBM Research Report, RC 25006**, 2010.
- [4] D.Binning, "Top Five Cloud Computing Security Issues, **Computer Weekly**, April 2009, <http://www.computerweekly.com/Articles/2010/01/12/235782/Top-five-cloud-computing-security-issues.htm>
- [5] T.Ormandy, "An Empirical study into the Security Exposure to Hosts of Hostile Virtualized Environments", 2007, <http://tavisio.decsystem.org/virtsec.pdf>
- [6] S. Lowe, **Mastering VMware vSphere 4**. Wiley Publishing Inc, 2009
- [7] Top Threats to Cloud Computing V 1.0, Cloud Security Alliance, March 2010, <http://www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf>

### 9.2 Routing of Security Alerts

The provider capability and the associated configuration policy are listed below:

*CSR-F15: Cloud provider has installed several security tools such as IDS/IPS, Malware and Virus detectors either on individual VMs or in a security appliances residing in a security-hardened VM on the host.*

*CSR-SP15: Security Alerts generated by the security monitoring tools should be automatically directed to the concerned subscriber. To enable implementation of this policy, the cloud provider should be able to generate a unique ID for each VM that a subscriber has launched and being able to associate that unique ID with the name and contact information for the subscriber.*

## 10. CONCLUSIONS AND BENEFITS

In this paper, we have provided a comprehensive set of policies needed for secure operation of a virtualized infrastructure owned by a public IaaS Cloud provider. We looked at the typical set of service features offered by most IaaS cloud providers and derived security policies associated with each them based on the current state of virtualization technologies available to implement those features. The catalog of security policies derived in this paper can be of benefit to both an IaaS cloud provider and an IaaS cloud subscriber in the following ways:

- The cloud subscriber may look at the set of features advertised by the IaaS cloud provider in its service package and then verify whether the associated security policies can be adhered to by the provider.
- The IaaS cloud provider can extract the generic security policies enumerated in this paper for various administrative and subscriber-facing functions and customize it for its own infrastructure and service package.

- [8] J. Wei et al., "Managing Security of Virtual Machine Images in a Cloud Environment," **Proc. ACM Cloud Computing Security Workshop**, Nov 2009, Chicago, IL, USA.
- [9] S.King et al., "Implementing Malware with Virtual Machines," **Proc. IEEE Symposium on Security and Privacy**, Berkeley, CA, USA, May 2006.
- [10] M. Jensen et al., "Technical security issues in cloud computing," **Proc. IEEE International Conference on Cloud computing**, Bangalore, India, Sept 21-25, 2009.
- [11] B.R.Kandukuri et al., "Cloud Security Issues," **Proc. IEEE International Conference on Cloud computing**, Bangalore, India, Sept 21-25, 2009.
- [12] Open Virtualization Format Specification – v 1.1.0, DMTF, January 2010, [http://www.dmtf.org/sites/default/files/standards/documents/DSP0243\\_1.1.0.pdf](http://www.dmtf.org/sites/default/files/standards/documents/DSP0243_1.1.0.pdf)
- [13] A. Bowcom, "New Directions in Virtualization Security: How Segmentation Can Strengthen Your Security Posture" <http://www2.reflexsystems.com/l/19/ns-Virtualization-Security-pdf/JIA0J>
- [14] A. Antonopoulos, J. Burke, "Practicing Virtual Security", <http://www2.reflexsystems.com/l/19/tualization-Best-Practices-pdf/ID401>

# Network Intrusion Detection Using an Evolutionary Fuzzy Rule-Based System

Terrence P. Fries

Department of Computer Science  
Indiana University of Pennsylvania  
Indiana, PA USA  
[t.fries@iup.edu](mailto:t.fries@iup.edu)

**Abstract**—The proliferation of computer networks has brought network security to the forefront. It has become imperative to devise new methods for network intrusion detection. Current methods are frequently unable to detect polymorphic or novel attack modes. In addition, the enormous volume of network traffic makes it difficult to monitor and evaluate all features of communication packets on the network. This research uses evolutionary learning to select key traffic features to be used in intrusion detection. An evolutionary algorithm is then used to learn a set of fuzzy inference rules based on the selected features. The resulting system is able to adapt to dynamic network conditions and identify morphing and unknown attack modes. Tests show a very high success rate in identifying attacks with minimal false positives.

**Keywords**—*evolutionary learning; network security; network intrusion detection; fuzzy logic*

## I. INTRODUCTION

The widespread use of computer networks and the movement of users toward mobility have made networks more vulnerable to intrusions and malicious attacks. As a result, it has become increasingly important to develop more reliable network security safeguards. Intrusion detection systems (IDS) provide network administrators with an automated method of identifying abnormal behavior in network traffic and blocking a wide variety of attacks which may appear in many forms. An IDS must be capable of identifying network intrusions, including malicious attacks, unauthorized access, and other anomalous behaviors.

Intrusion detection is classified into one of two distinct categories based on the type of attack: misuse detection or anomaly detection. Misuse attacks frequently attempt to exploit weaknesses in the system or application software. AN IDS that uses misuse detection techniques attempts to match particular features in a packet with known attack modes stored in a database. Packet Features which typical are used to identify misuse attacks include source address, destination address, source and destination ports, and keywords of the packet. Although misuse detection systems have a low rate of false positives, they lack the robustness needed to identify attack patterns not explicitly defined in their rules or new types of attacks. Such systems require frequent updates of new attack mode to remain effective.

Anomaly detection is able to detect network traffic that deviates from normal behavior. Current network activity is compared against known normal behavior using statistical

and artificial intelligence techniques. These techniques allow the IDS to detect unknown or novel attacks with prior knowledge. Unfortunately, these systems are difficult to implement due to the uncertainty of which packet features are relevant to identifying an attack. Additionally, anomaly detection systems suffer from a higher false positive rate [1, 2, 3].

The majority of IDS use some form of supervised or unsupervised pattern-recognition techniques to construct meta-classifiers which are then used for intrusion detection. IDS methodologies include statistical models, immune system approaches, protocol verification, file and taint checking, neural networks, whitelisting, state transition analysis, expression matching, dedicated languages, genetic algorithms, and burglar alarms [2, 4]. Most existing IDS using these techniques function satisfactorily only in well-defined environments. They are incapable of identifying complex or unknown behaviors and unable to adapt to dynamic environments. Some IDS use machine-learning techniques that allow them to be automatically retrained. However, labeled data needed for retaining is difficult to acquire because it is not practical to manually classify training data due to the enormous volume of network traffic data.

In light of the limitations in current IDS, there is a need for a reactive technique with the capability of identifying previous unknown attack modes and of operating in a dynamic environment. A problem arises because network data often contains few intrusion instances (positive events), making training difficult. A poorly trained system has the potential to either miss attacks or generate many false positives. It can be costly to investigate the false positives that result from the large volume of network transactions. In addition, attackers frequently use new intrusion techniques or variants of existing ones.

Recent attempts have been made to use evolutionary algorithms to address the problems associated with the vast amount of network traffic, the dynamic environment, morphing attack modes, unknown attack modes, and the need for robustness. The self-learning aspect of evolutionary computation makes it particularly appropriate in addressing these concerns.

The use of fuzzy logic, evolutionary computation, and a hybrid of the two have been recently proposed to resolve the inherent uncertainty in intrusion detection and the existence of unknown attack modes. Fuzzy logic is appropriate for

intrusion detection because the combination of quantitative features such as number of connection involves a level of uncertainty or fuzziness. In addition, there may be a level of uncertainty associate with the identification of an attack [5]. It has been shown that fuzzy logic provides better accuracy during the decision phase when uncertainty is present. Deterministic techniques are not flexible enough in applications that involve multiple features in a decision [6]. Evolutionary techniques such as genetic algorithms (GA) offer the ability to overcome the limitations of many current IDS systems. GA properties include robustness to noise, self-learning capability, and the ability to initially build rules without the need for *a priori* knowledge. These inherent properties of GAs make them particularly suitable for intrusion detection systems. Genetic algorithms have been proven to be capable of providing near optimal solutions for NP-complete problems. GAs have additional advantages over traditional techniques for intrusion detection. They are intrinsically parallel because they generate multiple offspring that explore the solution space in multiple directions simultaneously. Parallelism makes them well-suited where solution space is extremely large. The adaptability of GAs allows the system to be easily retrained to evolve new rules [1, 2, 7].

This paper presents a new approach to creating a fuzzy rule-based IDS using evolutionary learning. Section II presents a dataset which is commonly used to test the performance of intrusion detection systems. Section III discusses current intrusion detection systems using evolutionary learning techniques. The new evolutionary fuzzy rule-based intrusion detection system is presented in Section IV and the test results and comparisons with existing approaches are provided in Section V. A summary and future research directions are presented in Section VI.

## II. INTRUSION DETECTION DATASET

When comparing the performance of various IDSs, it is essential to use a common dataset. Without the use of a common set of network traffic data, variations in the dataset can significantly affect the measured success and failure rates. All of the methods discussed in Section III were tested using the Knowledge Discovery and Data Mining (KDD) Cup 1999 Dataset [8]. This dataset has become a standard for testing IDSs.

The KDD Cup 99 dataset was derived from the 1998 DARPA Intrusion Detection Evaluation Program held by MIT Lincoln Labs. Raw TCP/IP dump data was gathered in a simulated military network environment in which a typical U.S. Air Force LAN which was subjected to simulated attacks. Approximately 4 GB of compressed TCP dump data from 7 weeks of network traffic comprise about 5 million connection records. The dataset includes 41 various quantitative and qualitative features which were extracted for each TCP/IP connection. The features are grouped into 3 categories: basic features of individual TCP connections, content features within a connection, and traffic features computed using a two-second time window. The features associated with each of the three categories are shown in Tables I – III.

The dataset simulates 24 types of attacks in the training data. An additional 14 types in the test data provide the opportunity to test the ability of an IDS to handle unknown modes of attack. The simulated attacks cover all four of the most common categories of attack:

- *Denial of service* (DoS). Attacks in this category attempt to overload the utilization of some computing or memory resource, rendering it too busy to handle legitimate requests. These attacks may be initiated by flooding a system with communications, abusing legitimate resources, targeting implementation bugs, or exploiting the system’s configuration.
- *User to root* (U2R). A U2R attacker begins by accessing a normal user account and then exploits vulnerabilities to gain unauthorized access to the root. One of the most common U2R attacks produces buffer overflows.

TABLE I. FEATURES OF INDIVIDUAL TCP CONNECTIONS [10]

Feature Name	Description
duration	length of connections (in sec.)
protocol_type	type of protocol
service	network service on destination
src_bytes	number of data bytes from source to destination
dst_bytes	number of data bytes from destination to source
flag	status of connection: normal or error
land	1 if connection is from/to same port
wrong_fragment	number of “wrong” fragments
urgent	number of urgent packets

TABLE II. CONTENT FEATURES WITHIN A CONNECTION [10]

Feature Name	Description
hot	number of “hot” indicators
num_failed_logins	number of failed login attempts
logged_in	1 if successfully logged in
num_compromised	number of “compromised” conditions
root_shell	1 if root_shell is compromised
su_attempted	1 if “su root” command attempted
num_root	number of root accesses
num_file_creations	number of file creation operations
num_shells	number of shell prompts
num_access_files	number of operations on access control files
num_outbound_cmds	number of outbound commands in an ftp session
is_hot_login	1 if login belongs to hot list
is_guest_login	1 if login is a guest

TABLE III. TRAFFIC FEATURES USING A 2-SECOND WINDOW [10]

Feature Name	Description
count	number of connections to same host in past 2 seconds
error_rate	% of connections with SYN errors
reror_rate	% of connections with REJ errors
same_srv_rate	% of connections to same service
diff_svr_rate	% of connections to different services
srv_count	number of connections to same service in past 2 seconds
srv_error_rate	% of connections with SYN errors
srv_reror_rate	% of connections with REJ errors
srv_diff_host_rate	% of connections to different hosts

- *Remote to user* (R2L). An R2L attacker sends packets to a machine, and then exploits vulnerabilities in that machine to gain local access as a user. Such unauthorized access may result from password guessing. These type of attacks are often the most difficult to detect because packets are most similar to those from legitimate users.
- *Probing* (PROBE). In probing, the attacker scans the network to gather information or to find known vulnerabilities. Attacks of this nature include port scanning.

### III. BACKGROUND ON IDS EVOLUTIONARY LEARNING

A number of intrusion detection techniques using genetic algorithms have been proposed due to the GA's capabilities. These approaches focus on selecting key features of network traffic that reduce the complexity of identifying attackers and generating a relatively small set of rules to identify attackers. Specific techniques discussed below include genetic clustering, optimizing a set of rules to identify attackers, and creating a fuzzy inference system.

#### A. Genetic and Fuzzy Clustering

Attempts have been made to use both genetic and fuzzy clustering in intrusion detection systems. The Intrusion Detection Based on Genetic Clustering (IDBGC) algorithm provides a two-stage approach that establishes clusters of network traffic features and then detects intruders by classifying traffic into one of the clusters [2]. The first stage of the algorithm establishes a set of original clusters using the nearest neighbor method. It groups similar instances into a cluster and then filters noisy data based on similarity or dissimilarity metrics. In the second stage, the original clusters are combined using a GA to obtain a near optimal result. Using the KDD Cup 1999 dataset for testing, the average detection rate was close to 60% while the average false positive rate was only 0.4%. The failure to detect 40% of attacks is obviously unacceptable in a network.

One fuzzy clustering method combines fuzzy clustering with the Dempster-Shafer theory to resolve the inherent uncertainty in the IDS [9]. It then uses the k-nearest neighbor technique to identify attacks. It has been tested using the KDD Cup 99 dataset with mixed success. However, the method has only been tested on U2R and R2L attacks. Another technique using fuzzy c-means to accomplish fuzzy clustering has been relatively successful [10]. When tested on the KDD Cup 99 dataset, the method provided a 91% detection rate with 1.8% false positives. This performance is not acceptable because it allows 9% of attacks to go undetected. While the false positive rate appears low, it is significant when one considers the enormous amount of network traffic involved.

#### B. Optimizing a Set of Fuzzy Rules

Bankovic et al. [1] have proposed a method that creates a set of rules to identify network intrusions such as anomalous behaviors, unauthorized access, and malicious attacks. The first step uses a dimension reduction technique, principal component analysis (PCA). PCA, also known as Karhunen-

Loève transform, extracts a subset of features that preserve the most relevant information by identifying a few orthogonal linear combinations of the original variables with largest variance. The 41 features in the KDD Cup 1999 dataset were reduced to only three:

<i>duration</i>	number of seconds of the connection
<i>src_bytes</i>	number of data bytes from source to destination
<i>error_rate</i>	percentage of connections with "SYN" errors

The second step uses a genetic algorithm to create a set of rules to classify behavior as normal or abnormal. Every selected feature represents one gene in chromosome resulting in the chromosome structure. Each chromosome represents a rule for intrusion detection in an *if-then* clause format. The conditional part of rule (antecedent) is composed of the features connected by an *and* function of a particular type. The consequent of each rule is a confirmation of an intrusion. For example, the chromosome shown in Fig. 1 corresponds to the rule:

**if** (*duration* = "1" **and** *src\_bytes* = "0"  
**and** *error\_rate* = "50") **then** *intrusion*

Although the type of the attack is not of great importance in intrusion detection, it is important for forensics in order to recover from an attack. To accomplish this, the confirmation in the rule consequent can be replaced by an identification of the attack type. The expanded chromosome shown in Fig. 2 corresponds to the rule:

**if** (*duration* = "1" **and** *src\_bytes* = "0"  
**and** *error\_rate* = "50")  
**then** (*attack\_name* = "portsweep")

An initial population of chromosomes is randomly created and then a GA is used to optimize the set of rules. To determine the fitness of each rule, the following fitness function is used:

$$fitness = \frac{\alpha}{A} - \frac{\beta}{B}$$

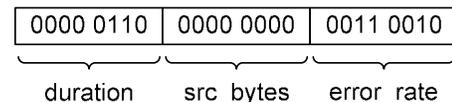


Figure 1. Chromosome example.

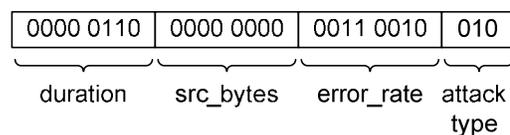


Figure 2. Chromosome with attack type added.

where  $\alpha$  is the number of correctly identified attacks,  $A$  is the total number of attacks in the training dataset,  $\beta$  is the number of normal connections incorrectly characterized as attacks (false positives), and  $B$  is the total number of normal connections in the training dataset. The resulting fitness values are in the range  $[-1, 1]$ .

When tested using the KDD Cup 1999 dataset, this method produced a 94% attack detection rate with no false positives, that is, normal connections were classified 100% correctly. While a 94% attack detection rate may initially appear to be a significant improvement, a 6% failure rate over millions of packets allows far too many attackers to remain undetected. It has been determined that failure to identify a large percentage of attackers is because the method reduces the feature set too much to be practical.

### C. Fuzzy Inference Systems

The use of a GA for generating intrusion identification rules can be expanded to create a set of fuzzy rules that comprise a fuzzy inference system [11, 12, 13, 14, 15, 16]. The resulting fuzzy inference system is capable of detecting intrusive behaviors in computer networks. The fuzzy classifier system uses fuzzy if-then rules that are similar to ones in the prior example except that the antecedent features are weight with fuzzy linguistic variables and the consequent classification is qualified with a certainty factor. A typical rule,  $R_i$  is:

**if**  $x_1$  is  $A_{i1}$  **and** ... **and**  $x_n$  is  $A_{in}$   
**then** Class  $C_i$  with  $CF = CF_i$

where  $R_i$  is the label of the  $i$  th rule

$x_1, \dots, x_n$  are the attributes/features

$A_{i1}, \dots, A_{in}$  are antecedent fuzzy sets

$C_i$  is the consequent class

$CF_i$  is the certainty factor of the rule  $R_i$

A fuzzy set, or linguistic variable, is defined by a membership function which describes the degree of membership in the set [17]. The membership function provides a mapping to any real value in the range 0 to 1, inclusive. The use of fuzzy sets accommodates values that possess a degree of inaccuracy. The fuzzy linguistic variables for the IDS rules are represented by the triangular membership functions shown in Fig. 3.

After an initial set of fuzzy rules are randomly created, a GA is then used to optimize the set of fuzzy rules using the training set from the KDD Cup 1999 dataset. When optimized rules were tested using the KDD Cup 1999 dataset, they method produced a 99% attack detection rate with less than 4% false positives. Although the detection rate is commendable, the existence of 4% false positives is not practical. Due to the large volume of network transactions, it would be costly to investigate all of the false positives.

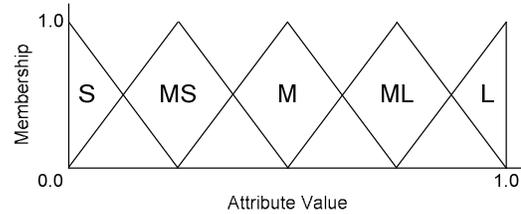


Figure 3. Membership functions of linguistic variables (S: small, MS: medium small, M: medium, ML: medium large, L: large).

## IV. EVOLUTIONARY LEARNING APPROACH TO CREATING A FUZZY RULE-BASED IDS

All of the existing methods which utilize fuzzy logic, genetic algorithms, or a combination of the two suffer from either a poor detection rate or a high false positive rate. This section presents an approach to developing a fuzzy inference IDS using evolutionary learning which overcomes both problems by developing a set fuzzy rules which identify intruders. The algorithm for constructing the IDS is comprised of two layers as shown in Fig. 4: feature reduction followed by fuzzy rules optimization. First, a genetic algorithm is used to establish an optimal subset of the packet features which are necessary to identify intrusions. Second, a evolutionary learning is used to optimize a set of fuzzy rules.

### A. Feature Subset Selection

The reduction of the number of key features is essential to improve performance in terms of learning time, classification accuracy, and comprehensibility of the learned rules [18]. The limitations of traditional dimension reduction techniques based on statistics and genetic clustering has been demonstrated [2]. Clearly, an algorithm is needed that is capable of producing a more robust subset of features.

In this approach, a genetic algorithm is used for feature subset selection with each chromosome corresponding to a candidate feature subset. Each chromosome is encoded as a string of 0's and 1's with the number of bits equal to the total number of features. Each bit represents a particular feature. If the bit is a '1', it indicates the attribute is to be used for training, while a '0' indicates the attribute is not to be used. A GA determines the optimal set of features to be used for training the rule set.

The GA fitness function is based on the accuracy of the subset indicated by the chromosome and the number of attributes in the chromosome. The accuracy is the ratio of correct identifications to the total number of packets in the test set. Since each additional feature adds complexity and computational time to the examination of each packet, the cost is the number of features used in the subset. The fitness of subset  $s$  is:

$$\text{fitness}(s) = w_{\text{accuracy}} * \text{accuracy}(s) + w_{\text{cost}} * \text{cost}(s)$$

where  $w_{\text{accuracy}}$  and  $w_{\text{cost}}$  are the respective weights of the accuracy and cost.

B. Optimization of a Fuzzy Rule Set

Although the fuzzy inference system previously described only has a 4% rate of false positives, this is a significant number when the massive quantity of packets received by a typical network is considered. This relatively large number of false positives undermines the system’s usefulness. The problem lies in the use of triangular fuzzy sets for the antecedents which limits the ability to match the rules. To gain a wide range at a high membership, the tails of the triangular fuzzy set may have to extend far into unacceptable ranges, thus allowing a rule to be matched when the packet is not an actual threat. Conversely, if a narrow triangular fuzzy set is used to minimize the extent of the tails, the rule is limited and may not be matched, thus allowing a threat to go unidentified. The only alternative is the inclusion of additional rules to cover all possible ranges of the antecedent. This would bloat the rule base and requires additional computation in the IDS.

To eliminate this problem, this research uses trapezoidal fuzzy sets for the antecedents. Trapezoidal fuzzy sets are defined by the tuple  $(a, b, c, d)$  which describe the key points as shown in Fig. 4. They have been chosen because they allow full membership over any range in the universe of discourse and the range of the right and left tails can be adjusted, thus, providing great flexibility in the weighting.

The fuzzy rule-based IDS uses fuzzy if-then rules that are similar to ones already discussed. The unique part of this algorithm which provides more reliable intruder detection is the use of trapezoidal fuzzy sets for the antecedent features weights,  $A_{ij}$ , and for the certainty factor,  $CF_i$ , corresponding to the consequent. A typical rule,  $R_i$ , is:

**if**  $x_1$  is  $A_{i1}$  **and** ... **and**  $x_n$  is  $A_{in}$   
**then** Class  $C_i$  with  $CF = CF_i$

where

- $R_i$  is the label of the  $i$  th rule
- $x_1, \dots, x_n$  are the relevant features
- $A_{i1}, \dots, A_{in}$  are trapezoidal fuzzy sets
- $C_i$  is the consequent class (intrusion category)
- $CF_i$  is the trapezoidal fuzzy certainty factor of the rule  $R_i$

Another limitation of existing fuzzy inference engines is use of predefined fuzzy sets. To address this limitation, the proposed method uses a generalized trapezoidal fuzzy set whose parameters  $(a, b, c, d)$  are determined by the genetic algorithm at the same time it optimizes the set of fuzzy

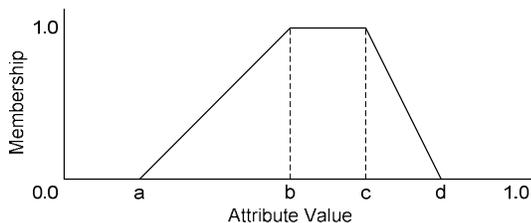


Figure 4. Trapezoidal fuzzy set.

rules. Because the chromosome length affects the convergence time, the fuzzy set parameters  $(a, b, c, d)$  are each limited to one of eight values in the range 0 to 7. This requires only 3 bits to encode each in each gene corresponding to a single feature. This requires only 12 bits for each gene representing feature  $x_j$  as shown in Fig. 5. Thus, the chromosome for rule  $R_i$  requires a 12-bit gene for each feature,  $x_j$ , and one 4-bit gene for the consequent certainty factor,  $CF_i$ , as shown in Fig. 6.

The set of fuzzy rules that evolve from the algorithm provide an intrusion detection system. The rules identify each packet as valid or as an intruder. If the intruder is detected, the rule will provide a classification of the intrusion in the consequent  $C_i$  with a certainty of  $CF_i$ . The system is adaptable in a dynamic environment. When new intruders are found, the system can be retrained offline using the genetic algorithms. The IDS can then be updated with the new set of fuzzy rules to ensure that the new intruders will be detected.

V. TEST RESULTS

The fuzzy-genetic intrusion detection approach was tested using the KDD Cup 1999 Dataset [8]. Evolutionary optimization was used to perform feature subset selection and optimization of the fuzzy rule set as in the algorithm discussed in the previous section. Training was conducted using a 10% training subset of the KDD dataset. The training data contained approximately 500,000 connection records.

Feature subset selection reduced the original 41 features to a subset of 8 features:

<i>duration</i>	number of seconds of the connection
<i>num_access_files</i>	number of operations on access control files
<i>num_failed_logins</i>	number failed login attempts
<i>root_shell</i>	1 if root shell is obtained; 0 otherwise
<i>same_srv_rate</i>	percent of connections to same service
<i>serror_rate</i>	percent of connections with “SYN” errors
<i>src_bytes</i>	number of data bytes from source to destination
<i>srv_count</i>	number of connections to same service in past 2 seconds

The evolutionary learning chromosome required eight 12-bit genes, one for each feature, and one 4-bit gene for the

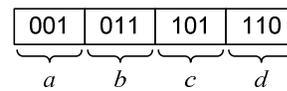


Figure 5. Gene for feature  $x_j$ .

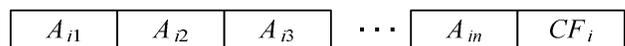


Figure 6. Chromosome for rule  $R_i$ .

certainty factor. This resulted in 100-bit chromosomes in the rule optimization step.

Training was done using 10% of the TCP records in the KDD dataset. Once trained, the new IDS was tested using the full 5 million connection records in the KDD dataset. The full dataset contains 14 types of intrusion attacks not present in the training data. The new IDS successfully identified almost all of the intrusion attacks with a 99.6% success rate and only 0.2% false positives.

Table IV provides a comparison of the fuzzy-genetic IDS with the other methods described in Section III of this paper. It is impossible to compare with the other methods which were not tested with the standard KDD dataset because such testing would encounter an entirely different set of unknown intruders. The proposed fuzzy-genetic intrusion detection system had the best intrusion detection rate of those tested by identifying 99.6% of all attacks. The 0.2% rate of false positives was deemed to be a success. Only the rule optimization method had a lower rate of false positives, however, it had a lower detection rate making it less effective, and the difference in false positive rates is negligible. In addition, the new system was able to correctly identify each of the 14 types of intrusions not in the training data. This demonstrates the robustness of the new IDS.

## VI. CONCLUSION

This paper has presented a fuzzy rule-based approach to network intrusion detection using evolutionary learning to reduce the number of features needed and then to optimize the fuzzy rules. The fuzzy approach was essential in addressing the uncertain nature of mutating and unknown attacks. The two-stage process of developing the IDS consisted of using genetic algorithms to optimize feature selection and the fuzzy rule set. Evolutionary optimization has several advantages over traditional methods: unsupervised learning, robustness, ability to find a near optimal solution in large dimensional problem spaces, and intrinsic parallel operation. Testing has shown that the proposed method provides superior performance when compared to other fuzzy rule-based and GA-based algorithms. The method provides a higher detection rate while reducing the rate of false positives. This method overcomes the poor detection rate of genetic clustering, the costly and impractical 6% rate of failed detections for rule optimization, and the unacceptable 6% rate of false positives for the fuzzy inference system. The method demonstrates

TABLE IV. PERFORMANCE COMPARISON OF IDS METHODS

IDS Method	Intrusion Detection Rate	False Positives
Genetic Clustering	60%	0.4%
Fuzzy Clustering	91%	1.8%
Rule Optimization	94%	0%
Fuzzy Inference System	98%	4%
Proposed Evolutionary Learning IDS	99.6%	0.2%

improved robustness in comparison to other techniques by correctly identifying previously unknown attack behaviors. In addition, the IDS is adaptable by easily retraining it for newly discovered modes of attack.

## REFERENCES

- [1] Z. Bankovic, D. Stepanovic, S. Bojanic, and O. Nieto-Taladriz, "Improving Network Security Using Genetic Algorithm Approach," *Computers and Electrical Engineering*, vol. 33, September-November 2007, pp. 438-451.
- [2] Y. Liu, K. Chen, X. Liao, and W. Zhang, "A genetic clustering method for intrusion detection," *Pattern Recognition*, vol. 37, May 2004, pp. 927-942.
- [3] J. E. Dickerson and J. A. Dickerson, "Fuzzy network profiling for intrusion detection," *Proc. 19th International Conf. North American Fuzzy Information Processing Society*, July 2000, pp. 301-306.
- [4] T. Verwoerd and R. Hunt, "Intrusion detection techniques and approaches," *Computer Communications* vol. 25, September 2002, pp. 1356-1365.
- [5] G. Florez, S. A. Bridges, and R. B. Vaughn, "An improved algorithm for fuzzy data mining for intrusion detection," *Proc. 200221<sup>st</sup> International Conf. North American Fuzzy Information Processing Society*, August 2002, pp. 457-462.
- [6] J. H. Bellec and T. M. Kechadi, "Curfes: clustering using fuzzy representative event selection for fault recognition problem in telecommunication networks," *Proc. ACM Conf. Information and Knowledge Management*, 2007, pp. 55-62.
- [7] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd ed. New York: Springer, 1996.
- [8] University of California, Irvine, "KDD99 Cup 1999 Dataset," 1999. [Online]. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>. [Accessed: October 15, 2009].
- [9] T. S. Chou and K. K. Yen, "Fuzzy belief k-nearest neighbors anomaly detection of user to root and remote to local attacks," *IEEE Information Assurance and Security Wksp*, June 2007, pp. 207-213.
- [10] Q. P. Zeng and S. X. Wu, "A fuzzy clustering approach for intrusion detection," *International Conf. on Web Information Systems and Mining*, December 2009, pp. 728-732.
- [11] M. S. Abadeh, J. Habibi, Z. Barzegar, M. A. Sergi, "Parallel Genetic Local Search Algorithm for Intrusion Detection in Computer Networks," *Engineering Applications of Artificial Intelligence*, vol. 20, December 2007, pp. 1058-1069.
- [12] M. S. Abadeh, J. Habibi, and C. Lucas, "Intrusion Detection Using a Fuzzy Genetics-Based Learning Algorithm," *J. Network and Computer Applications*, vol. 30, January 2007, pp. 414-428.
- [13] A. T. Haghighat, M. Esmaili, A. Saremi, and V. R. Mousavi, "Intrusion Detection via Fuzzy-Genetic Algorithm Combination with Evolutionary Algorithms," *Proc. 6th IEEE/ACIS Conference on Computer and Information Sciences*, 2007.
- [14] T. Ozyer, R. Alhajj and K. Barker, "Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule pre-screening," *Journal of Network and Computer Applications*, vol. 30, January 2007, pp. 99-113.
- [15] T. P. Fries, "A fuzzy-genetic approach to network intrusion detection," *Proc. 2008 Genetic and Evolutionary Computation Conference*, Atlanta, GA, July 2008.
- [16] A. N. Toosi and M. Kahani, "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers," *Computer Communications*, July 2007, pp. 2201-2212.
- [17] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, March 1965, pp. 338-353.
- [18] G. Helmer, J. S. K. Wong, V., Honavar, V., and L. Miller, "Automated Discovery of Concise Predictive Rules for Intrusion Detection," *J. System Software*, vol. 60, February 2002, pp. 165-175.

# Number of product terms in reliability formula of weighted- $k$ -out-of- $n$ system by SDP method

Yoichi Higashiyama<sup>1</sup> and Ventsi Rumchev<sup>2</sup>

<sup>1</sup> Department of Electrical and Electronic Engineering  
 Ehime University  
 Matsuyama, 790-8577 Japan  
 e-mails: mountain@dpc.ehime-u.ac.jp

<sup>2</sup> Department of Mathematics and Statistics  
 Curtin University of Technology  
 P.O. Box U1987, Perth, WA 6845 Australia  
 e-mail: V.Rumchev@curtin.edu.au

## Abstract

A weighted version of the  $k$ -out-of- $n$  system is considered. In the disjoint products version of reliability analysis of weighted- $k$ -out-of- $n$  systems, it is necessary to determine the order in which the weight of components is to be considered. The  $k$ -out-of- $n$ :G(F) system consists of  $n$  components; each component has its own probability and positive integer weight such that the system is operational (failed) if and only if the total weight of some operational (failure) components is at least  $k$ . A method to compute the reliability formulas using SDP (Sum of Disjoint Products) method has been already published. The proposed method expresses the system reliability in fewer reliability formulas than those already published. The purpose of this paper is to point out that the number of product terms in the reliability formula of weighted- $k$ -out-of- $n$  system is influenced by the order of the minimal set of weighted components.

*Keywords:* Weighted- $k$ -out-of- $n$  system; Reliability formulas; SDP method; Weight of components.

## 1. Introduction

The weighted- $k$ -out-of- $n$ :G(F) system consists of  $n$  components, each of which has its own probability and positive integer weight (total system weight =  $w$ ), such that the system is operational (failed) if and only if the total weight of some operational (failure) components is at least  $k$  [6]. The reliability of the weighted- $k$ -out-of- $n$ :G system is the component of the unreliability of a weighted- $(w-k+1)$ -out-of- $n$ :F system. Without loss of generality, we discuss the weighted- $k$ -out-of- $n$ :G system only. The original  $k$ -out-of- $n$ :G system is a special case of the weighted- $k$ -out-of- $n$ :G system wherein the weight of each component is 1. The system model was extended to a two-stage weighted model with components in common [7]. Recently, several different aspects of related problems were investigated [4], [5].

One of the questions that arise when using recursive disjoint products algorithms for reliability of the weighted- $k$ -out-of- $n$  system is the order in which the weight of components should be considered [3]. The system was introduced by Wu and Chen in 1994 [1]. They proposed  $O(n \cdot k)$  algorithm to

compute the exact system reliability. However, their algorithm does not take any account of the order of components. The number of product terms in their reliability formula is strongly influenced by the order of components.

Higashiyama has pointed out the advantages of an alternative order in the method based on the weight of components [2]. Three types of orders are studied in [2]: (1) random order [1], (2) ascending order, and (3) descending order. In ascending order, the components are arranged so that the lower weight has a lower component number. That means that the component order is equivalent to the order of the weight of components in the system. This order is also called best order. For example, if the weight of component  $i$  is less than the weight of component  $j$ , then the component number  $i$  must be lower than the number  $j$ . The descending order is opposite of the ascending order and is also called worst order. The best order method reduces the computing cost and data processing effort required to generate an optimal factored formula [2].

The method proposed in [2] dramatically reduced the computing cost and data processing effort. However, a lot of reliability formulas unused in later steps are automatically derived in the method. This paper gives an efficient algorithm to generate the reliability formulas only to be used in later steps.

A relatively new but popular method for obtaining a reliability formula for coherent system is the Abraham SDP (Sum of Disjoint Products) method [9]. Higashiyama et al. have applied the SDP method to the weighted- $k$ -out-of- $n$ :G systems. The algorithm described in [10] gives the disjoint product terms.

Section 2 describes the notation & assumptions. Section 3 shows an  $O(n \cdot k)$  algorithm by Wu-Chen for the reliability of the weighted- $k$ -out-of- $n$ :G system. Section 4 shows a revised algorithm by Higashiyama to generate a factored reliability formula. Section 5 gives Higashiyama's recent algorithm to reduce the number of computing steps. Section 6 proposes new algorithm using SDP method to derive the reliability formula and the variation of number of product

terms in the reliability formula obtained by SDP method.

## 2. Model

Notation

- $n$  number of components in a system.
  - $k$  minimal total weight of all operational (failed) components which makes the system operational (failed).
  - $w_i$  weight of component  $i$ .
  - $p_i$  operational probability of component  $i$ .
  - $q_i \square 1.0 - p_i$ , failed probability of component  $i$ .
  - $R, B, W$  [random, best, worst] case in which the components of the system are ordered [randomly, the lower weight one has lower number, the higher weight one has lower number].
  - $R_\Omega(i, j)$  reliability formula of the weighted- $j$ -out-of- $i$ :  $G_\Omega$  for  $\Omega = R, B, W$  case.
  - $R_\Omega^N(i, j)$  reliability formula of  $R_\Omega(i, j)$  for  $\Omega = R, B, W$  case in the new method which generates only the reliability formulas that are used in later steps.
  - $M_\Omega(i, j)$  binary random value indicating the state of  $R_\Omega^N(i, j)$  for  $\Omega = R, B, W$ .
- Assumptions
- A. Each component and the system has binary states, i.e., either operational or failed.
  - B. The components and system are non-repairable.
  - C. All components are statistically independent.
  - D. Sensing and switching mechanisms are perfect.
  - E. Each component has a known positive integer weight.
  - F. Operational probability of each component is known.
  - G. The system is operational if and only if the total weight of operational components is at least  $k$ .

### 3. Wu-Chen (random case) [1]

Wu and Chen [1] have presented an  $O(n \cdot k)$  algorithm to evaluate the reliability of the weighted- $k$ -out-of- $n$ : $G_R$  system.

To derive  $R_R(i, j)$ , the algorithm needs to construct the table with  $R_R(i, j)$ , for  $i = 0, 1, 2, \dots, n$ , and  $j = 0, 1, 2, \dots, k$ . Initially,

$$R_R(i, 0) = 1.0, \text{ for } i = 0, 1, 2, \dots, n; \quad (1)$$

$$R_R(0, j) = 0.0, \text{ for } j = 1, 2, \dots, k. \quad (2)$$

Furthermore, if  $j < 0$ , it is obvious that for any  $i$ :

$$R_R(i, j) = 1.0 \quad (3)$$

For  $i = 1, 2, \dots, n$ , and  $j = 1, 2, \dots, k$ , their algorithm generates each  $R_R(i, j)$ ,

$$R_R(i, j) = \begin{cases} p_i \cdot R(i-1, j-w_i) + q_i \cdot R(i-1, j), & \text{if } j-w_i \geq 0; \\ p_i + q_i \cdot R(i-1, j), & \text{otherwise.} \end{cases} \quad (4)$$

Now the algorithm for computing  $R(n, k)$  is:

1. Using equation (1) and equation (2), construct row 1 and column 1 in the  $R_R(i, j)$  table.
2. Using equation (4), construct row 2, row 3, ..., row  $(n+1)$  in that order. Hence,  $R_R(n, k)$  is eventually derived.

Because the size of the  $R_R(i, j)$  table is  $(n+1) \cdot (k+1)$ , the size of the sequential algorithm needs  $O(n \cdot k)$  running time.

This method has a disadvantage in that the number of terms depends on the order of components. Hereafter it is referred

to as random order method.

Consider a weighted-5-out-of-3: $G_R$  system with weights:  $w_1 = 2$ ,  $w_2 = 6$ , and  $w_3 = 4$ .

By equation (1), get column #1 wherein,

$$R_R(0, 0) = R_R(1, 0) = R_R(2, 0) = R_R(3, 0) = 1.0 \quad (5)$$

and by equation (2), get row #1 wherein,

$$R_R(0, 1) = R_R(0, 2) = R_R(0, 3) = R_R(0, 4) = R_R(0, 5) = 0.0 \quad (6)$$

Therefore, by equation (4) rows #2, #3, and #4 are derived as follows:

Row #2:

$$\left. \begin{aligned} R_R(1, 1) &= p_1 \cdot R_R(0, -1) + q_1 \cdot R_R(0, 1) = p_1 \\ R_R(1, 2) &= p_1 \cdot R_R(0, 0) + q_1 \cdot R_R(0, 2) = p_1 \\ R_R(1, 3) &= p_1 \cdot R_R(0, 1) + q_1 \cdot R_R(0, 3) = 0.0 \\ R_R(1, 4) &= p_1 \cdot R_R(0, 2) + q_1 \cdot R_R(0, 4) = 0.0 \\ R_R(1, 5) &= p_1 \cdot R_R(0, 3) + q_1 \cdot R_R(0, 5) = 0.0 \end{aligned} \right\} \quad (7)$$

Row #3:

$$\left. \begin{aligned} R_R(2, 1) &= p_2 \cdot R_R(1, -5) + q_2 \cdot R_R(1, 1) = p_2 + q_2 p_1 \\ R_R(2, 2) &= p_2 \cdot R_R(1, -4) + q_2 \cdot R_R(1, 2) = p_2 + q_2 p_1 \\ R_R(2, 3) &= p_2 \cdot R_R(1, -3) + q_2 \cdot R_R(1, 3) = p_2 \\ R_R(2, 4) &= p_2 \cdot R_R(1, -2) + q_2 \cdot R_R(1, 4) = p_2 \\ R_R(2, 5) &= p_2 \cdot R_R(1, -1) + q_2 \cdot R_R(1, 5) = p_2 \end{aligned} \right\} \quad (8)$$

Row #4:

$$\left. \begin{aligned} R_R(3, 1) &= p_3 \cdot R_R(2, -3) + q_3 \cdot R_R(2, 1) \\ &= p_3 + q_3 \cdot (p_2 + q_2 p_1) = p_3 + q_3 p_2 + q_3 q_2 p_1 \\ R_R(3, 2) &= p_3 \cdot R_R(2, -2) + q_3 \cdot R_R(2, 2) \\ &= p_3 + q_3 \cdot (p_2 + q_2 p_1) = p_3 + q_3 p_2 + q_3 q_2 p_1 \\ R_R(3, 3) &= p_3 \cdot R_R(2, -1) + q_3 \cdot R_R(2, 3) = p_3 + q_3 p_2 \\ R_R(3, 4) &= p_3 \cdot R_R(2, 0) + q_3 \cdot R_R(2, 4) = p_3 + q_3 p_2 \\ R_R(3, 5) &= p_3 \cdot R_R(2, 1) + q_3 \cdot R_R(2, 5) \\ &= p_3 \cdot (p_2 + q_2 p_1) + q_3 p_2 = p_3 p_2 + p_3 q_2 p_1 + q_3 p_2 \end{aligned} \right\} \quad (9)$$

## 4. Higashiyama method -1[2]

### 4.1 Best case

This section presents the best order of components so that the lower weight component has lower component number. After reordering of the components, the same procedure as in [1] can be used to compute the system reliability. Hereafter it is referred to as best order method.

Therefore, consider the reliability formula for the reordered weighted-5-out-of-3: $G_B$  system with weights:  $w_1 = 2$ ,  $w_2 = 4$ , and  $w_3 = 6$ .

By equation (1), get column #1 wherein,

$$R_B(0, 0) = R_B(1, 0) = R_B(2, 0) = R_B(3, 0) = 1.0 \quad (10)$$

and by equation (2), get row #1 wherein,

$$R_B(0, 1) = R_B(0, 2) = R_B(0, 3) = R_B(0, 4) = R_B(0, 5) = 0.0 \quad (11)$$

Therefore, by equation (4) rows #2, #3, and #4 are derived as follows:

Row #2:

$$\left. \begin{aligned} R_B(1, 1) &= p_1 \cdot R_B(0, -1) + q_1 \cdot R_B(0, 1) = p_1 \\ R_B(1, 2) &= p_1 \cdot R_B(0, 0) + q_1 \cdot R_B(0, 2) = p_1 \\ R_B(1, 3) &= p_1 \cdot R_B(0, 1) + q_1 \cdot R_B(0, 3) = 0.0 \\ R_B(1, 4) &= p_1 \cdot R_B(0, 2) + q_1 \cdot R_B(0, 4) = 0.0 \\ R_B(1, 5) &= p_1 \cdot R_B(0, 3) + q_1 \cdot R_B(0, 5) = 0.0 \end{aligned} \right\} \quad (12)$$

Row #3:

$$\left. \begin{aligned} R_B(2,1) &= p_2 \cdot R_B(1,-3) + q_2 \cdot R_B(1,1) = p_2 + q_2 p_1 \\ R_B(2,2) &= p_2 \cdot R_B(1,-2) + q_2 \cdot R_B(1,2) = p_2 + q_2 p_1 \\ R_B(2,3) &= p_2 \cdot R_B(1,-1) + q_2 \cdot R_B(1,3) = p_2 \\ R_B(2,4) &= p_2 \cdot R_B(1,0) + q_2 \cdot R_B(1,4) = p_2 \\ R_B(2,5) &= p_2 \cdot R_B(1,1) + q_2 \cdot R_B(1,5) = p_2 p_1 \end{aligned} \right\} \quad (13)$$

Row #4:

$$R_B(3,5) = p_3 \cdot R_B(2,-1) + q_3 \cdot R_B(2,5) = p_3 + q_3 p_2 p_1 \quad (14)$$

The final result  $R_B(3,5)$  is only generated from reliabilities  $R_B(2,-1)$  and  $R_B(2,5)$ , so it is not necessary to calculate  $R_B(3,1)$ ,  $R_B(3,2)$ , ...,  $R_B(3,4)$ .

#### 4.2 Worst case

This section presents the worst order of components so that the higher weight one has lower component number. After reordering of the components, the same procedure as in [1] can be used to compute the system reliability. Hereafter it is referred to as worst order method.

Consider the reliability formula for the reordered weighted-5-out-of-3:  $G_w$  system with weights:  $w_1 = 6$ ,  $w_2 = 4$ , and  $w_3 = 2$ .

By equation (1), get column #1 wherein,

$$R_w(0,0) = R_w(1,0) = R_w(2,0) = R_w(3,0) = 1.0 \quad (15)$$

and by equation (2), get row #1 wherein,

$$R_w(0,1) = R_w(0,2) = R_w(0,3) = R_w(0,4) = R_w(0,5) = 0.0 \quad (16)$$

Therefore, by equation (4) rows #2, #3, and #4 are derived as follows:

Row #2:

$$\left. \begin{aligned} R_w(1,1) &= p_1 \cdot R_w(0,-5) + q_1 \cdot R_w(0,1) = p_1 \\ R_w(1,2) &= p_1 \cdot R_w(0,-4) + q_1 \cdot R_w(0,2) = p_1 \\ R_w(1,3) &= p_1 \cdot R_w(0,-3) + q_1 \cdot R_w(0,3) = p_1 \\ R_w(1,4) &= p_1 \cdot R_w(0,-2) + q_1 \cdot R_w(0,4) = p_1 \\ R_w(1,5) &= p_1 \cdot R_w(0,-1) + q_1 \cdot R_w(0,5) = p_1 \end{aligned} \right\} \quad (17)$$

Row #3:

$$\left. \begin{aligned} R_w(2,1) &= p_2 \cdot R_w(1,-3) + q_2 \cdot R_w(1,1) = p_1 + q_2 p_1 \\ R_w(2,2) &= p_2 \cdot R_w(1,-2) + q_2 \cdot R_w(1,2) = p_2 + q_2 p_1 \\ R_w(2,3) &= p_2 \cdot R_w(1,-1) + q_2 \cdot R_w(1,3) = p_2 + q_2 p_1 \\ R_w(2,4) &= p_2 \cdot R_w(1,0) + q_2 \cdot R_w(1,4) = p_2 + q_2 p_1 \\ R_w(2,5) &= p_2 \cdot R_w(1,1) + q_2 \cdot R_w(1,5) = p_2 p_1 + q_2 p_1 \end{aligned} \right\} \quad (18)$$

Row #4:

$$\begin{aligned} R_w(3,5) &= p_3 \cdot R_w(2,3) + q_3 \cdot R_w(2,5) \\ &= p_3 \cdot (p_2 + q_2 p_1) + q_3 \cdot (p_2 p_1 + q_2 p_1) \\ &= p_3 p_2 + p_3 q_2 p_1 + q_3 p_2 p_1 + q_3 q_2 p_1 \end{aligned} \quad (19)$$

In the same manner to best case, the final result  $R_w(3,5)$  is only generated from reliabilities  $R_w(2,3)$  and  $R_w(2,5)$ , so it is not necessary to calculate  $R_w(3,1)$ ,  $R_w(3,2)$ , ...,  $R_w(3,4)$ .

#### 4.3. Comparisons between three results

A. Using the component numbers in the weighted-5-out-of-3:  $G_B$  system,  $R_B(3,5)$  (interchange component numbers 2 and 3) and (interchange component numbers 1 and 3) can be rewritten as, respectively;

$$\begin{aligned} R_B(3,5) &= p_3 p_2 + q_3 p_2 p_1 + p_3 q_2 = p_3 \cdot (p_2 + q_2) + q_3 p_2 p_1 \\ &= p_3 + q_3 p_2 p_1 = R_B(3,5) \end{aligned} \quad (20)$$

$$\begin{aligned} R_w(3,5) &= p_2 p_1 + p_3 q_2 p_1 + p_3 p_2 q_1 + p_3 q_2 q_1 \\ &= q_3 p_2 p_1 + p_3 p_2 p_1 + p_3 q_2 p_1 + p_3 p_2 q_1 + p_3 q_2 q_1 \\ &= p_3 \cdot \{(p_2 + q_2) \cdot p_1 + (p_2 + q_2) \cdot q_1\} + q_3 p_2 p_1 \\ &= p_3 + q_3 p_2 p_1 = R_B(3,5) \end{aligned} \quad (21)$$

- B. Best order method generates only 2 product terms and 4 variables, and requires 1 addition (+ -operator) and 2 multiplications ( $\times$ -operator).
- C. Random order method generates 3 product terms and 7 variables, and requires 2 additions and 4 multiplications.
- D. Worst order method generates 4 product terms and 11 variables, and requires 3 additions and 7 multiplications.

### 5. Higashiyama method-2 [8]

The method proposed in [2] dramatically reduced the computing cost and data processing effort. However, a lot of reliability formulas unused in later step are automatically derived in the method. For Example, Best case in the section 4.1 derives  $R_B(3,5)$  as a final result. The final result is only derived from three reliability formulas,  $R_B(2,5)$ ,  $R_B(1,1)$ , and  $R_B(1,5)$ . Each of formulas without three ones are not used to generate the final result, then these formulas do not need to generate the final result. This section gives an efficient algorithm to generate the formulas only to be used in later steps.

#### 5.1 Algorithm

The **Algorithm: Generate reliability formulas only used in later steps** is based on the definition of the system structure function, which is given in *Notation* of Section 2. **Step 1** generates the matrix,  $M$ , ( $i, j$ ) position of which corresponds to a reliability formula,  $R_B(i, j)$ . Each digit 1 of  $M$  means the formula to be derived. Each digit 0 of  $M$  means the formula not to be derived. The format of the algorithm makes it easy to implement in a high-level programming language like Fortran, Pascal, or C.

**Algorithm: Generate reliability formulas only used in later steps**

**input:**  $n, k, w_1 \sim w_n, p_1 \sim p_n$ ;

**common:**  $n, k, w_1 \sim w_n, p_1 \sim p_n, M, R; q_i = 1.0 - p_i$ ;

**Step 1**

**initial clear:**  $M[1 \leq i \leq n, 1 \leq j \leq k] := 0$ ;

$M[n, k] := M[n-1, k] := 1$ ;

**if**  $k - w_n > 0$  **then**  $M[n-1, k - w_n] := 1$ ; **end if**

**for**  $i := n-1$  **step** -1 **until** 2 **do**

**for**  $j := 1$  **until**  $k$  **do**

**if**  $M[i, j] = 1$  **then**  $M[i-1, j] := 1$ ;

**if**  $j - w_i > 0$  **then**  $M[i-1, j - w_i] := 1$ ;

**end if; end if; end for; end for;**

**Step 2**

**initial clear:**  $R[0 \leq i \leq n, j \leq 0] := 1.0$ ;  $R[0, 1 \leq j \leq k] := 0.0$ ;

**for**  $i := 1$  **until**  $n$  **do**

**for**  $j := 1$  **until**  $k$  **do**

**if**  $M[i, j] = 1$

**then**  $R[i, j] := p_i \cdot R[i-1, j - w_i] + q_i \cdot R[i-1, j]$ ;

**end if; end for; end for;**

**Return**

## 5.2 Examples

Consider the weighted-5-out-of-3:G system with weights:  $w_1 = 2$ ,  $w_2 = 6$ , and  $w_3 = 4$ . For each case  $(R, B, W)$ , the **Algorithm** generates the reliability formulas below for each case about the example system. The proposed method only derives the reliability formulas to get the final result, then each of formula numbers corresponds to the formula number to be added in the section 3 and 4.

### 5.2.1 Random case

After executing of **Step 1** in the **Algorithm**, the matrix,  $M_R[\cdot]$ , is:

$$M_R = \begin{bmatrix} 10001 \\ 10001 \\ 00001 \end{bmatrix}$$

By virtue of  $M_R$ , **Step 2** generates the reliability formulas as follows:

$$R_R^N(1,1) = p_1 \cdot R_R^N(0,-1) + q_1 \cdot R_R^N(0,1) = p_1$$

1<sup>st</sup> row in equation (7)

$$R_R^N(1,5) = p_1 \cdot R_R^N(0,3) + q_1 \cdot R_R^N(0,5) = 0.0$$

5<sup>th</sup> row in equation (7)

$$R_R^N(2,1) = p_2 \cdot R_R^N(1,-5) + q_2 \cdot R_R^N(1,1) = p_2 + q_2 p_1$$

1<sup>st</sup> row in equation (8)

$$R_R^N(2,5) = p_2 \cdot R_R^N(1,-1) + q_2 \cdot R_R^N(1,5) = p_2$$

5<sup>th</sup> row in equation (8)

Finally the algorithm derives the final result as follows:

$$\begin{aligned} R_R^N(3,5) &= p_3 \cdot R_R^N(2,1) + q_3 \cdot R_R^N(2,5) \\ &= p_3 \cdot (p_2 + q_2 p_1) + q_3 p_2 = p_3 p_2 + p_3 q_2 p_1 + q_3 p_2 \end{aligned}$$

5<sup>th</sup> row in equation (9)

### 5.2.2 Best case

After executing of **Step 1**, the matrix is:

$$M_B = \begin{bmatrix} 10001 \\ 00001 \\ 00001 \end{bmatrix}$$

By virtue of  $M_B$ , **Step 2** derives the formulas as follows:

$$R_B^N(1,1) = p_1 \cdot R_B^N(0,-1) + q_1 \cdot R_B^N(0,1) = p_1$$

1<sup>st</sup> row in equation (12)

$$R_B^N(1,5) = p_1 \cdot R_B^N(0,3) + q_1 \cdot R_B^N(0,5) = 0.0$$

5<sup>th</sup> row in equation (12)

$$R_B^N(2,5) = p_2 \cdot R_B^N(1,1) + q_2 \cdot R_B^N(1,5) = p_2 p_1$$

5<sup>th</sup> row in equation (13)

$$R_B^N(3,5) = p_3 \cdot R_B^N(2,-1) + q_3 \cdot R_B^N(2,5) = p_3 + q_3 p_2 p_1 \quad (14)$$

### 5.2.3 Worst case

The  $M_W$  and  $R_W^N$  are derived as follows:

$$M_W = \begin{bmatrix} 10101 \\ 00101 \\ 00001 \end{bmatrix}$$

$$R_W^N(1,1) = p_1 \cdot R_W(0,-5) + q_1 \cdot R_W(0,1) = p_1$$

1<sup>st</sup> row in equation (17)

$$R_W^N(1,3) = p_1 \cdot R_W^N(0,-3) + q_1 \cdot R_W^N(0,3) = p_1$$

3<sup>rd</sup> row in equation (17)

$$R_W^N(1,5) = p_1 \cdot R_W^N(0,-1) + q_1 \cdot R_W^N(0,5) = p_1$$

5<sup>th</sup> row in equation (17)

$$R_W^N(2,3) = p_2 \cdot R_W^N(1,-1) + q_2 \cdot R_W^N(1,3) = p_2 + q_2 p_1$$

3<sup>rd</sup> row in equation (18)

$$R_W^N(2,5) = p_2 \cdot R_W^N(1,1) + q_2 \cdot R_W^N(1,5) = p_2 p_1 + q_2 p_1$$

5<sup>th</sup> row in equation (18)

$$\begin{aligned} R_W^N(3,5) &= p_3 \cdot R_W^N(2,3) + q_3 \cdot R_W^N(2,5) \\ &= p_3 \cdot (p_2 + q_2 p_1) + q_3 \cdot (p_2 p_1 + q_2 p_1) \\ &= p_3 p_2 + p_3 q_2 p_1 + q_3 p_2 p_1 + q_3 q_2 p_1 \end{aligned} \quad (19)$$

### 5.2.3 Comparisons

The proposed algorithm can generate three types of the final reliability formula above,  $R_R^N(3,5)$ ,  $R_B^N(3,5)$ , or  $R_W^N(3,5)$ , for each case.

- A. For the random case, the proposed algorithm needs 5 reliability formulas to get the final reliability formula and 6 reliability formulas are omitted.
- B. For the best case, the proposed algorithm needs 4 reliability formulas to get the final reliability formula and 7 reliability formulas are omitted.
- C. For the worst case, the proposed algorithm needs 6 reliability formulas to get the final reliability formula and 5 reliability formulas are omitted.

## 6 New algorithm using SDP method

### 6.1 SDP method

In the reliability problem of weighted- $k$ -out-of- $n$ :G(F) system, the system is operational (failed) if the total weight of some operational components is at least  $k$ . Therefore, the reliability formula of weighted system would be calculated if all the minimal sets were evaluated. Let  $A_i$  denote the product of Boolean variables corresponding to  $i$ -th minimal set. Then the reliability formula is derived as:

$$F = A_1 + A_2 + \dots + A_m \quad (22)$$

where  $m$  is the number of minimal sets. If the minimal set denotes the operational set,  $A_i$  is the product of Boolean variables,  $x$ 's, corresponding to  $i$ -th operational set; otherwise,  $A_i$  is the product of Boolean variables,  $\bar{x}$ 's, corresponding to  $i$ -th failed set.

If the terms of the reliability formula are disjoint, then the reliability formula and the numerical formula are one-to-one identical with one another. As is well known, some of  $A_i$  ( $i = 1, 2, \dots, m$ ) in equation (19) are not disjoint each other. This means that if one substitutes the numerical values into the reliability formula, the system reliability can not be computed.

To make  $A_i$ 's disjoint,  $F$  is transformed as:

$$F = A_i + \bar{A}_1 A_2 + \bar{A}_1 \bar{A}_2 A_3 + \dots + \bar{A}_1 \bar{A}_2 \dots \bar{A}_{j-1} A_j + \dots + \bar{A}_1 \bar{A}_2 \dots \bar{A}_{m-1} A_m \quad (23)$$

Consider the following  $i$ -th function  $F_i$  in equation (23):

$$F_i = \bar{A}_1 \bar{A}_2 \dots \bar{A}_j \dots \bar{A}_{i-1} A_i \quad (24)$$

Let  $B_{ji}$  be the set of variables which exist in  $A_j$  and which do not exist in  $A_i$ , and  $A_{ji}$  be the product of  $x_k$ 's,  $x_k \in B_{ji}$ .

Then equation (24) can be rewritten as:

$$F_i = \bar{A}_{1i} \bar{A}_{2i} \dots \bar{A}_{ji} \dots \bar{A}_{i-1,i} A_i \quad (25)$$

Furthermore, if any two product terms  $A_{ki}$ ,  $A_{ji}$  in equation (23) is satisfied with  $B_{ki} \subset B_{ji}$ , then  $\bar{A}_{ji}$  is dropped from equation (25). Let  $A'_{ji}$  ( $j=1,2,\dots,b$ ) be the un-dropped product terms. Then  $F_i$  can be rewritten as:

$$F_i = \bar{A}'_{1i} \bar{A}'_{2i} \dots \bar{A}_{bi} A_i \quad (26)$$

Let  $A$  has fixed 2-valued indicators  $x_1, x_2, \dots, x_k$ , then

$$\bar{A} = \bar{x}_1 + x_1 \bar{x}_2 + x_1 x_2 \bar{x}_3 + \dots + x_1 x_2 \dots x_{k-1} \bar{x}_k \quad (27)$$

Referring to the formula of equation (27), equation (26) is transformed into sum of disjoint terms.

## 6.2 Basic idea

In this section we will briefly expose, by means of an example, the basic idea in the method to be proposed.

### 6.2.1 Random case of SDP method

Consider the random case of weighted-5-out-of-3 system with weights  $w_1 = 2$ ,  $w_2 = 6$ ,  $w_3 = 4$ . Let  $S_i$  be a minimal set of weighted components whose total weight is more than or equal to  $k$ .  $S_i$  can be generated in ascending manner of components (random order) as follows:

$$S_1 = \{w_1, w_3\}$$

$$S_2 = \{w_2\}$$

Let  $x_i$  be Boolean variable of  $i$ -th component,  $w_i$ , in the weighted- $k$ -out-of- $n$  system. The product of Boolean variables,  $A_i$ , corresponds to the  $i$ -th set,  $S_i$ , as follows:

$$A_1 = x_1 x_3$$

$$A_2 = x_2$$

Then the reliability formula,  $F$ , is derived as:

$$F = A_1 + A_2 = x_1 x_3 + x_2 \quad (28)$$

And using equations (23) – (27), the reliability formula,  $F$ , is transformed by Boolean algebra as flows:

$$F = A_1 + \bar{A}_1 A_2 = x_1 x_3 + (\bar{x}_1 + x_1 \bar{x}_3) x_2 \quad (29)$$

The number of indicators and terms in equation (29) is equal to the number of probabilities and terms in 5<sup>th</sup> row of equation (9). And the final result is independent of the order of weighted components.

Next we consider another example, weighted-7-out-of-4 system with weights,  $w_1 = 2$ ,  $w_2 = 7$ ,  $w_3 = 5$ ,  $w_4 = 3$ .

There are 3 minimal sets as follows:

$$S_1 = \{w_1, w_3\}$$

$$S_2 = \{w_2\}$$

$$S_3 = \{w_3, w_4\}$$

Then a product of Boolean variables is correspond to minimal sets, respectively:

$$A_1 = x_1 x_3$$

$$A_2 = x_2$$

$$A_3 = x_3 x_4$$

The reliability formula,  $F$ , is derived as:

$$F = A_1 + A_2 + A_3 = A_1 + \bar{A}_1 A_2 + \bar{A}_1 \bar{A}_2 A_3 = F_1 + F_2 + F_3 \quad (30)$$

Consider  $F_2$  and  $F_3$  in equation (30):

$$F_2 = (\bar{x}_1 \bar{x}_3) \cdot x_2 = (\bar{x}_1 + x_1 \bar{x}_3) \cdot x_2 \quad (31)$$

$$F_3 = (\bar{x}_1 \bar{x}_3)(\bar{x}_2) \cdot x_3 x_4 = \bar{x}_1 \bar{x}_2 \cdot x_3 x_4 \quad (32)$$

Therefore, the final result,  $F$  is

$$F = A_1 + A_2 + A_3 = A_1 + F_2 + F_3 = x_1 x_3 + (\bar{x}_1 + x_1 \bar{x}_3) x_2 + \bar{x}_1 \bar{x}_2 x_3 x_4 \quad (33)$$

Equation (33) has 4 product terms. We call this order as random case of SDP method.

### 6.2.2 Best case of SDP method

Consider the best case of weighted-5-out-of-3 system with weights  $w_1 = 2$ ,  $w_2 = 6$ ,  $w_3 = 4$  again. The minimal set are ordered so that the smallest on is first, That is, in order of the number of elements in the events (best order) as follows:

$$S_1 = \{w_2\}$$

$$S_2 = \{w_1, w_3\}$$

Then a product of Boolean variables is corresponding to a minimal set, respectively.

$$A_1 = x_2$$

$$A_2 = x_1 x_3$$

Then the reliability formula,  $F$ , is derived as:

$$F = A_1 + A_2 = x_2 + x_1 x_3 \quad (34)$$

And  $F$  is transformed by Boolean algebra as flows:

$$F = A_1 + \bar{A}_1 A_2 = x_2 + \bar{x}_2 x_1 x_3 \quad (35)$$

The number of indicators and terms in equation (35) is equal to the number of probabilities and terms in equation (14). And the final result is independent of the order of weighted components, but not independent of the order of  $A_i$ 's.

Next we also consider another example, weighted-7-out-of-4 system with weights,  $w_1 = 2$ ,  $w_2 = 7$ ,  $w_3 = 5$ ,  $w_4 = 3$ .

There are 3 minimal sets as follows:

$$S_1 = \{w_2\}$$

$$S_2 = \{w_1, w_3\}$$

$$S_3 = \{w_3, w_4\}$$

Then a product of Boolean variables is correspond to minimal sets, respectively:

$$A_1 = x_2$$

$$A_2 = x_1 x_3$$

$$A_3 = x_3 x_4$$

The reliability formula,  $F$ , is derived as:

$$F = A_1 + A_2 + A_3 = A_1 + \bar{A}_1 A_2 + \bar{A}_1 \bar{A}_2 A_3 = F_1 + F_2 + F_3 \quad (36)$$

Consider  $F_2$  and  $F_3$  in equation (36):

$$F_2 = (\bar{x}_2) \cdot x_1 x_3 = \bar{x}_2 \cdot x_1 x_3 \quad (37)$$

$$F_3 = (\bar{x}_2)(\bar{x}_1 \bar{x}_3) \cdot x_3 x_4 = \bar{x}_1 \bar{x}_2 \cdot x_3 x_4 \quad (38)$$

Therefore, the final result,  $F$  is

$$F = A_1 + A_2 + A_3 = A_1 + F_2 + F_3 = x_2 + \bar{x}_2 x_1 x_3 + \bar{x}_1 \bar{x}_2 x_3 x_4 \quad (39)$$

Equation (39) has 3 product terms. The order of minimal sets strongly influences the number of product terms in the reliability formula obtained by SDP method. We call this order as best case of SDP method.

### 6.2.3 Worst case of SDP method

Consider the worst case of weighted-5-out-of-3 system with weights  $w_1 = 2$ ,  $w_2 = 6$ ,  $w_3 = 4$  newly. The minimal sets are ordered so that the largest one is first, that is, in the order of the number of elements in the events. This case is in reverse order of best case (worst order).

$$S_1 = \{w_1, w_3\}$$

$$S_2 = \{w_2\}$$

Then a product of Boolean variables is corresponding to a minimal set, respectively.

$$A_1 = x_1x_3$$

$$A_2 = x_2$$

Then the reliability formula,  $F$ , is derived as:

$$F = A_1 + A_2 = x_1x_3 + x_2 \quad (40)$$

And  $F$  is transformed by Boolean algebra as flows:

$$F = A_1 + \bar{A}_1A_2 = x_1x_3 + (\bar{x}_1 + x_1\bar{x}_3)x_2 \quad (41)$$

The number of indicators and terms in equation (41) is equal to the number of probabilities and terms in 5<sup>th</sup> row of equation (9) and equation (29). And the final result is also independent of the order of weighted components.

Next we also consider another example, weighted-7-out-of-4 system with weights,  $w_1 = 2$ ,  $w_2 = 7$ ,  $w_3 = 5$ ,  $w_4 = 3$ .

There are 3 minimal sets as follows:

$$S_1 = \{w_1, w_3\}$$

$$S_2 = \{w_3, w_4\}$$

$$S_3 = \{w_2\}$$

Then a product of Boolean variables is correspond to minimal sets, respectively:

$$A_1 = x_1x_3$$

$$A_2 = x_3x_4$$

$$A_3 = x_2$$

The reliability formula,  $F$ , is derived as:

$$F = A_1 + A_2 + A_3 = A_1 + \bar{A}_1A_2 + \bar{A}_1\bar{A}_2A_3 = F_1 + F_2 + F_3 \quad (42)$$

Consider  $F_2$  and  $F_3$  in equation (42):

$$F_2 = (\bar{x}_1\bar{x}_3) \cdot x_3x_4 = \bar{x}_1 \cdot x_3x_4 \quad (43)$$

$$F_3 = (\bar{x}_1\bar{x}_3)(\bar{x}_3\bar{x}_4) \cdot x_2 = (\bar{x}_1\bar{x}_3 + \bar{x}_3x_1 + \bar{x}_1\bar{x}_4x_3) \cdot x_2 \quad (44)$$

Therefore, the final result,  $F$  is

$$F = A_1 + A_2 + A_3 = A_1 + F_2 + F_3 \quad (45)$$

$$= x_1x_3 + \bar{x}_1x_3x_4 + (\bar{x}_1\bar{x}_3 + \bar{x}_3x_1 + \bar{x}_1\bar{x}_4x_3) \cdot x_2$$

Equation (45) has 5 product terms. The order of minimal sets strongly influences the number of product terms in the reliability formula obtained by SDP method. We call this order as worst case of SDP method.

## 7. Conclusions

In the old version of the reliability analysis of weighted- $k$ -out-of- $n$  systems, it was necessary to deter-

mine the order in which the weight of components is to be considered. Higashiyama et al. have designed an algorithm using to compute the reliability formula. This paper has investigated the variation of number of product terms in the reliability formula obtained by SDP method. Each final result is independent of the number of weighted components, but not independent of the order of product terms which are generated from minimal sets of weighted components whose total weight is more than or equal to  $k$ .

## References

- [1] J.S. Wu and R.J. Chen, "An algorithm for computing the reliability of weighted- $k$ -out-of- $n$  systems", *IEEE Transactions on Reliability*, Vol.43, no.2, pp327-329, 1990.
- [2] Y. Higashiyama, "A factored reliability formula for weighted- $k$ -out-of- $n$  system", *Asia-Pacific Journal of Operational Research*, Vol.18, pp61-66, 2001.
- [3] J.A. Buzacott, "The ordering of terms in cut-based recursive disjoint products", *IEEE Transactions on Reliability*, Vol.R-32, no. 5, pp472-474, 1983.
- [4] G. Levitin and A.Lisnianski, "Reliability optimization for weighted voting system", *Reliability Engineering and System Safety*, vol. 71, no. 2, pp.131-138, 2001.
- [5] M. Xie and H. Pham, "Modeling the reliability of threshold weighted voting systems", *Reliability Engineering and System Safety*, vol. 87, no. 1, pp.53-63, 2005.
- [6] W. Kuo and M. J. Zuo, "Optimal Reliability Modeling", Chapter 7, John Wiley & Son, pp.258-264, 2003.
- [7] Y. Chen and Q. Yang, "Reliability of two-stage weighted-k-out-of-n systems with components in common", *IEEE Transactions on Reliability*, vol. 54, no. 3, pp.431-440, 2005.
- [8] Y. Higashiyama and V. Rumchev, "A method for reducing the number of reliability formulae in reliability expression of weighted-k-out-of-n system", *International Journal of Performance Engineering*, vol. 4, no. 1, pp.19-29, 2008.
- [9] J. Abraham, "An improved algorithm for network reliability", *IEEE Transactions on Reliability*, vol. R-28, no. 1, pp.58-61, 1979.

## **A New Three Dimensional Bivalent Hypercube Description, Analysis, and Prospects for Research**

**Jeremy Horne**  
**Avenida Moreras 131**  
**San Felipe, Baja California C.P. 21850**  
[mindsurgeon@hotmail.com](mailto:mindsurgeon@hotmail.com)

**Keywords: 3-D Hypercube, Logic, Pattern Classification, Binary Systems, Innate Order**

### **Abstract**

A three dimensional hypercube representing all of the 4,096 dyadic computations in a standard bivalent system has been created. It has been constructed from the 16 functions arrayed in a table of functional completeness that can compute a dyadic relationship. Each component of the dyad is an operator as well as a function, such as “implication” being a result, as well as an operation. Every function in the hypercube has been color keyed to enhance the display of emerging patterns. At the minimum, the hypercube is a limited “multiplication table” or table of dyadic computations and values that shorten the time to do operations that normally would take longer using conventional truth table methods. It also can serve as a theorem prover and creator. With the hypercube comes a complete system without the need for axioms. The main significance of the 3-D hypercube at this point is that it is the most fundamental way of displaying all dyadic computations in binary space, thus serving as a way of normalizing the rendition of uninterpreted, or raw, binary space. The hypercube is a dimensionless entity, a standard by which in binary spaces can be measured, analogous to a meter stick.

### **Introduction**

A three dimensional hypercube representing all of the 4,096 dyadic computations in bivalent systems has been created. There are 16 functions that can compute a dyadic relationship, each component of the dyad, as well as its operator, being a function. These functions are arrayed in a table of functional completeness that reflects a binary counting from 0000 to 1111. Each function in the hypercube has been color keyed as an aid to make any patterns more visible. The hypercube is a new canonization of three-dimensional binary space. At the minimum, the hypercube and the canonization underlying it serve as a “multiplication table” or table of computations and values that shorten the time to do operations that normally would take longer using conventional truth table methods. There are other uses, such as various hypercubes consisting of binary spaces used to compute optimal communications paths ( “hamming distances” ). The main significance of the hypercube at this point is a description of the most fundamental three-dimension space in the binary world and a standard by which there can be a classification and analysis of patterns in binary space, be it randomly generated or from known process. Patterns, or displays of regularity may be produced by a regular process. Patterns emerge from deep innate structures in the universe. The hypercube is a structure created from a known process, and gauging a pattern generated from reputedly random processes against it may be a way of understanding randomness. Currently the hypercube is being presented here for research purposes.

### **Construction of binary logical space and functional notation**

Zeros and ones and permutations of those as successive quantities present themselves as ordered logical space. These semantics are in keeping with a fundamental aspect of mathematics discussed by Giuseppe Peano in 1898 concerning postulates describing ordering based upon increasing quantity. Peano's Postulates lack a critical postulate concerning a definitive association between succession and

increasing quantity by a regular increment. Yet, the notion of succession, or ordering, exists. Mathematics and logic are co-joined by order based on succession marked by increase, or mounting quantity of binary space, and there is a philosophy underpinning it [1].

Our semantics reflects that philosophy of order. The four rows of permutations of existent relationships yield a 16-column space, the Table of Functional Completeness (TFC). It is called “complete”, as all possibilities, or permutations, of 0s and 1s appear for the placeholders p and q. This is generated by the same method as with the above tables – serially and in ascending order (binary counting) in the same manner as the previous tables and in this case from 0000 to 1111, every column being vertically read. Columns are headed by an “f” with subscripts ranging from 0 through 15, each designating a particular function. In computer language, bytes consist of eight bits, and half a byte is a nybble. The TFC consists of 16 columns of nybbles; a function is a nybble. The notation is consistent with that presented by Irving Copi in his *Symbolic Logic* [2]. However, his functions are discussed only in terms of completeness of the binary system. Nothing is written about the nature, philosophy, or the use of the functions as discussed in this paper.

While the TFC includes the p and q generators, or placeholders, they could be omitted, leaving the functions. Philosophically, it can be said that process (a function) is object (result of computation), and object is process. (Notice, also, that in keeping with our ontological commitment of having only two existents, 0 and 1, that stripped of the letters and the function designators, all that remains in the TFC are those 0s and 1s, or only bits) We will see more of this shortly, where a function is an operator, as well as a result of a computation. This makes the logical space an entirely closed and complete space. One function is always the result of two other functions being computed via an operator function. The TFC showing all the permutations of relationships between existents as functions is the following:

p	q	f <sub>0</sub>	f <sub>1</sub>	f <sub>2</sub>	f <sub>3</sub>	f <sub>4</sub>	f <sub>5</sub>	f <sub>6</sub>	f <sub>7</sub>	f <sub>8</sub>	f <sub>9</sub>	f <sub>10</sub>	f <sub>11</sub>	f <sub>12</sub>	f <sub>13</sub>	f <sub>14</sub>	f <sub>15</sub>
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

**Table 1 - Table of Functional Completeness (TFC)**

For example, referring to simple existents, row 3 for function 4 (f<sub>4</sub>), is read, “function 4 relates p=1 to q=0 to yield 0.”

With more than two variables, or existents, the TFC is expanded to 2<sup>n</sup> rows of permuted values, where n=number of variables and (2<sup>n</sup>)<sup>n</sup> columns in the table. With three variables the display of zeros and ones is simply doubled, where “0” becomes “00”, and so forth. With four variables, it becomes “0000”.

From the TFC those teaching propositional logic create notational standards for expressing relationships between two existents. Four functions – nybbles - normally are taught: and (&), or (∨), equivalence (≡), and implication (⊃) - but there are 16 functions, and each can be given an operator symbol, viz:

<b>f<sub>0</sub></b>	X - Contradiction
<b>f<sub>1</sub></b>	&, and, conjunction
<b>f<sub>2</sub></b>	>, p is greater than q
<b>f<sub>3</sub></b>	1>, 1 precedes, or, simply "p"
<b>f<sub>4</sub></b>	<, p is less than q
<b>f<sub>5</sub></b>	>1, 1 follows (or simply "q")
<b>f<sub>6</sub></b>	≠, p or q is true (1) but not both (XOR);exclusive "or"
<b>f<sub>7</sub></b>	∨, p or q is true or both are true; inclusive "or", disjunction
<b>f<sub>8</sub></b>	NOR, neither p nor q or both is/are true
<b>f<sub>9</sub></b>	≡, p is equivalent to q in truth value
<b>f<sub>10</sub></b>	>0, 0 follows (or simply "not q")
<b>f<sub>11</sub></b>	⊂, ← q contains p
<b>f<sub>12</sub></b>	0>, 0 precedes (or simply "not p")
<b>f<sub>13</sub></b>	⊃, or → p contains q (often called "IMP") – <b>defines deduction</b>
<b>f<sub>14</sub></b>	NAND, not both p and q are true
<b>f<sub>15</sub></b>	T, tautology

**Table 2 - Functions, Symbols, and Their Names**

Negation (~) is a unary operation, and the TFC implicitly defines it, with functions f<sub>8</sub> through f<sub>15</sub> being opposite or “mirroring”, reflections of f<sub>0</sub> through f<sub>7</sub>. (Although functional completeness is discussed commonly, names often are not given to the functions or standardized. Operator symbols are not standardized; “&” is the same as “^”, “.”, and “and”.)

### Truth tables and the new canonization

Truth tables are in functional form, e.g., 0011 (third column in the TFC) is f<sub>3</sub>. A calculation like f<sub>10</sub>(f<sub>12</sub>, f<sub>4</sub>) → f<sub>11</sub>) follows the same procedure as with the standard four operators. For example, f<sub>10</sub> means:

<b>p</b>	<b>q</b>	<b>p &gt;0 q</b>
0	0	1
0	1	0
1	0	1
1	1	0

**Table 3 - Truth Table for f<sub>10</sub>**

The resulting value is 1 when 0 is the second value in the relationship. Otherwise, the result is 0. Any function operating over p - (f<sub>3</sub>), and q - (f<sub>5</sub>), or the four permutations of 0 and 1, will yield itself. So, the function defines itself, as in f<sub>10</sub>(f<sub>3</sub>, f<sub>5</sub>) → f<sub>10</sub> and f<sub>9</sub>(f<sub>3</sub>, f<sub>5</sub>) → f<sub>9</sub>. The “→” will be used interchangeably with “⊃” for typographical convenience. Now, replacing the f<sub>12</sub> and f<sub>4</sub> values for the ones in p and q, p = 1100 and q = 0100, respectively, we have the following:

<b>p</b>	<b>q</b>	<b>p &gt;0 q</b>
1	0	1
1	1	0
0	0	1
0	0	1

**Table 4 - Truth Table for  $f_{10}(f_{12}, f_4) \rightarrow f_{11}$**

To illustrate the rapidity of space saving, the new truth table canonization is:

<b>p</b>	<b>q</b>	<b>p &gt;0 q</b>
$f_3$	$f_5$	$f_{11}$

**Table 5 - New Truth Table Canonization**

The syntax for dyadic computations is  $f_n(f_x, f_y) \rightarrow f_p$ , where  $f_n$  represents a binary operator, such as  $f_7$ , or 0111. The  $f_x$  and  $f_y$  represent the operands, and the  $f_p$  is the result of the computation. N-adic computations take the form  $f_q(f_n(f_x, f_y) \rightarrow f_p) \rightarrow f_r \dots f^*$ , with the  $f_q$  and  $f_n$  being operators. The  $f_p$  and  $f_r$  are computational results. Again, any function, as in  $f_r$  and  $f^*$ , can be an operator, or an operand, depending upon its placement in the syntax. Such is one of the factors making binary logical space closed, each function being in dialectical relationship with the others (one in terms of the others), where it can serve in an opposite capacity – operator or operand. Process as an operator becomes the object of an operator (a result of a computation), and object as a function becomes a process. For example,  $f_{13}$  is the material implication operator, but it also can be the result of a computation.

Evaluation is standard, working from the innermost parentheses to the outermost. It is optional whether to re-iterate the  $f_3$  and  $f_5$  underneath the formula being evaluated, as these values already exist in the permutation table. They have been left in to demonstrate that many computations can be done simply by inspecting the function.

To appreciate the space saving nature of the canonization, we have a standard truth table, such as:

<b>p</b>	<b>q</b>	<b>r</b>	<b>s</b>	<b>(p &amp; q)</b>	<b>(r → s)</b>	<b>(p → r)</b>
0	0	0	0	0	1	1
0	0	0	1	0	0	1
0	0	1	0	0	1	1
0	0	1	1	0	0	1
0	1	0	0	0	1	1
0	1	0	1	0	0	1
0	1	1	0	1	1	0
0	1	1	1	1	0	0
1	0	0	0	0	1	1
1	0	0	1	0	0	1
1	0	1	0	0	1	1
1	0	1	1	0	0	1
1	1	0	0	1	1	0
1	1	0	1	1	0	0
1	1	1	0	1	1	0
1	1	1	1	1	0	0

**Table 6 - Standard four variable table**

rendered as:

				1		3		1		2		1		
<b>p</b>	<b>q</b>	<b>r</b>	<b>s</b>	<b>(p &amp; q)</b>	<b>→</b>	<b>[(r ≡ s)</b>	<b>v</b>	<b>(p → r)]</b>						
				f <sub>1</sub>				f <sub>9</sub>		f <sub>7</sub>		f <sub>13</sub>		
f <sub>0</sub>	f <sub>0</sub>	f <sub>3</sub>	f <sub>5</sub>	f <sub>0</sub>	f <sub>0</sub>	f <sub>0</sub>	f <sub>15</sub>	f <sub>3</sub>	f <sub>9</sub>	f <sub>5</sub>	f <sub>15</sub>	f <sub>0</sub>	f <sub>15</sub>	f <sub>3</sub>
f <sub>0</sub>	f <sub>15</sub>	f <sub>3</sub>	f <sub>5</sub>	f <sub>0</sub>	f <sub>0</sub>	f <sub>15</sub>	f <sub>15</sub>	f <sub>3</sub>	f <sub>9</sub>	f <sub>5</sub>	f <sub>15</sub>	f <sub>0</sub>	f <sub>15</sub>	f <sub>3</sub>
f <sub>15</sub>	f <sub>0</sub>	f <sub>3</sub>	f <sub>5</sub>	f <sub>15</sub>	f <sub>0</sub>	f <sub>0</sub>	f <sub>15</sub>	f <sub>3</sub>	f <sub>9</sub>	f <sub>5</sub>	f <sub>11</sub>	f <sub>15</sub>	f <sub>3</sub>	f <sub>3</sub>
f <sub>15</sub>	f <sub>15</sub>	f <sub>3</sub>	f <sub>5</sub>	f <sub>15</sub>	f <sub>15</sub>	f <sub>15</sub>	f <sub>11</sub>	f <sub>3</sub>	f <sub>9</sub>	f <sub>5</sub>	f <sub>11</sub>	f <sub>15</sub>	f <sub>3</sub>	f <sub>3</sub>

**Table 7** - Four variable table in only terms of functional notation

The canonization explanation is published in *The Journal for Systemics and Informatics* [3].

### Composition of the hypercube

The hypercube represents the 4,096 permutations of dyadic (two place) computations of the sixteen functions in the Table of Functional Completeness (TFC), i.e.  $f_n(f_x, f_y)$ , where  $f_n$  is a selected operator and the ordered pair  $x, y$  as operands. The hypercube is to computational completeness for dyadic relationships as the TFC is to the permutations of 0s and 1s as a four place number. The first is three-dimensional, the second two-dimensional. The hypercube contains the smallest volume that can be occupied in Euclidean space. This fact ensures that the resulting permutation space is optimal. There are 16 plates in the hypercube, each corresponding to one of the 16 functions. Each plate displays a Cartesian coordinate form of a particular function operating over the 16 functions, including itself and shows the complete permutation of computations for a function. There are  $16^2$  computations, in each plate or 256 results. In reading the hypercube one starts from the top left of each plate, reads downward and then across the top to arrive at an answer. A number of function pairs are not commutable, i.e., yield the same result if the functions are switched. Thus, in Plate  $f_6$ , for  $f_6(f_9, f_{12})$  to get  $f_5$ , read down the left-hand most column to  $f_9$  and then across to the column headed by  $f_{12}$  in the manner of a distance chart on a highway map to get the  $f_5$ . The same plate shows  $f_6(f_8, f_{11}) = f_3$ . The plates are like TFC generation but with ascending functions. Two samples appear in the *Appendix*.

### Computational significance of the hypercube

The hypercube acts as a multiplication table for doing dyadic computations in binary space. Rather than displaying a full truth table one can do a chained calculation, such as  $f_{13}(f_7(f_4, f_8), f_1(f_{12}, f_3)), f_9)$  simply by starting with the innermost parentheses, as in standard logical calculations, and working to the outermost function,  $f_{13}$ . In this case, using the hypercube, the final result is:

- $f_7(f_4, f_8) \rightarrow f_{12}$   $f_7$  plate – first set of innermost parentheses
- $f_1(f_{12}, f_3) \rightarrow f_0$   $f_1$  plate – second set of innermost parentheses
- $f_7(f_{12}, f_0) \rightarrow f_{12}$   $f_7$  plate –  $f_7$  function operating over the results of the previous two calculations
- $f_{13}(f_{12}, f_9) \rightarrow f_{11}$   $f_{13}$  plate – final calculation.

## **Towards pattern recognition in binary space with the hypercube**

A pattern is a display of repetition. A pattern can indicate of an ordered process that is emerging and we should be able to determine the nature of the process generating it and predict how the display will appear in the future. Patterns display order, and it is order through the lens of the hypercube as an observing and measuring device that we hope to see what produces that order in a binary space. Binary space may be able to display patterns resulting from proofs and demonstrations, machine language programs, Turing machine programs, and cellular automata, among others. Each is an ordered process, hence potentially being able to exhibit a pattern in the hypercube. Much of the following is speculative but offers a foundation for future research in ascertaining the efficacy of the hypercube as a way of detecting and analyzing regularity in binary spaces.

There are two types of spaces, one purposeful, where we know what generated it, and the second, one resulting from an unknown generator. The question about emerging patterns can be answered in a straight-forward way. Create a number of theorems, or deductive arguments, and generate a truth table to exhibit an arrangement of 0s and 1s, some or all of which may exhibit a regularity. Theorems are structures, some of which are of thinking itself. They may represent types of abstraction. Certain modes of thinking present certain types of binary spaces. It may be the case that two or more theorems present the same arrangement. Each arrangement can be classified, as Wolfram has done with his cellular automata [4]. Once a “dictionary” of pattern types resulting from these theorems is produced, we can use it to compare regularity generated randomly or from unknown processes. Theorems can be created by the methods of proof, conditional proof, short truth tables, or by the hypercube. We see may a pattern in the hypercube or that the cube can analyze it. If so, then, we will be on the way of discovering the origin of patterns in a reputedly random space.

## **Tracing the origin of a valid relationship using the hypercube**

We ask what ultimate two functions will give tautology ( $f_{15}$ ) as a result for a theorem's corresponding conditional, that is,  $f_n(f_p, f_q) \rightarrow f_{15}$ . We go to Plate  $f_{13}$  and look inside the plate for the  $f_{15}$ s. Each  $f_{15}$  is the result of one function, or conjunct of them implying another to give a valid relationship, or theorem. Then we determine what two functions  $f_{13}$  operate over to give the  $f_{15}$ , and there are many. We see  $f_{15}(f_0, f_0)$  as the first. Remember, read the intersection of the first row and the first column to see the  $f_0$ , just like a distance finder between two cities on a road map. Thus,  $f_{13}(f_0, f_0) \rightarrow f_{15}$ . Similarly, we see all across the  $f_0$  row a series of  $f_{15}$ s, i.e.,  $f_{13}(f_0, f_n) \rightarrow f_{15}$  or  $f_{13}(f_n, f_n) \rightarrow f_{15}$ , for that matter,  $n$  being any function. For the  $f_1$  row, every other function yields  $f_{15}$ , as in  $f_{13}(f_1, f_1) \rightarrow f_{15}$  and  $f_{13}(f_1, f_3) \rightarrow f_{15}$ . Another example is  $f_{13}(f_6, f_{14}) \rightarrow f_{15}$ . Now, we ask what functions can produce each of the functions used by  $f_{13}$  to produce  $f_{15}$ . Such an exercise is rather extensive, but as has been demonstrated, the hypercube shortens this research considerably, as only a look-up is required. For example, we have  $f_{13}(f_8, f_{12}) \rightarrow f_{15}$  (in fact  $f_8 - f_{15}$ ). Then, how we can get an  $f_8$ ? Just about any of the hypercube plates will tell us quickly. Taking one at random, let us say  $f_4$ , we can locate many, such as  $f_4(f_7, f_{8-15}) \rightarrow f_8$ . Plate  $f_9$ , as another example, shows that  $f_9(f_{13}, f_{10}) \rightarrow f_8$ . Hence, the task of determining exactly what computational result gave rise to a function is not possible, because of their being at least two paths to the same result. One should realize, however, that theorems can be created by such backtracking in the hypercube, just so long as one preserves the relationships set forth in the corresponding conditional. Here, in addition to the the long and short truth tables method, the hypercube serves as a useful shorthand tool for creating theorems. It also can act as a theorem prover. Simply backtrack by inspection from a result to see if what reputedly gave rise to it actually did. A worthy research project would be to create a computer program to generate theorems from the hypercube, classify them according to emerging patterns or methods of generation, and create the dictionary alluded to earlier.

## The Hypercube as foundational for pattern analysis

To date, there has been no method to find exactly all the functions generating a function or a space, save for a nearest neighbor analysis of cellular automata using a Turing-style tape of 0s and 1s [5]. Contrary to this, the analyzable spaces here are not necessarily produced by using a cellular automaton or neighborhood space method. The 0 and 1 values may be inserted from anywhere by any method, thus adding force to the term “raw space”.

We are not trying to say what a pattern represents, such as a face or language character but merely that a regularity exists, and we should ask what gave rise to it. The whole theory underlying this paper is that all binary patterns have their origin in some place in the hypercube, as the hypercube is the most fundamental 3-D building block of binary logic, given what the above discussion about theorem generation illustrates. From that origin, there is built from various functional relations expressed by the cube the emerging pattern. Patterns don't emerge from nowhere. There is an origin and the progression is orderly. For example, in examining the hypercube closely, it may be asked whether certain groupings, such as the assemblage of  $f_{15S}$  on plate  $f_{11}$  constitute a core, or a “seed” for “gliders”, the image so popular in discussing cellular automata [6]. Aside from any reference to order emerging from chaos in the binary world or whether there are Lorentz attractors, there are already in each of the sixteen plates emerging patterns of color, as with Plate  $f_7$ , where, with  $f_7$ , there is a string of  $f_7S$ , followed by eight  $f_{15S}$ , both vertically and horizontally, suggesting boundary conditions of some type. The diagonals in each plate show a functional counting, one diagonal ascending, the other descending. By coupling all the plates together, a three-dimensional view may reveal more. Already can be found emerging grouping, or clustering, of functions, as in Plate  $f_4$ , where  $f_2$  seems to congregate in groups of three at various places along the top of the diagonal. The hypercube also may contain patterns classified in the dictionary of theorems mentioned above.

We also question whether there is a family of patterns or an algebra of spaces. For example, there is  $f_{13}(f_n, f_{15}) \rightarrow f_{15}$ ,  $f_{13}(f_n, f_n) \rightarrow f_{15}$ , and  $f_{13}(f_0, f_n) \rightarrow f_{15}$ , where  $n$  is any function, and from this we may say that  $f_7[f_{13}(f_n, f_{15}), f_{13}(f_0, f_n)] \rightarrow f_{15}$ , and list equivalent relationships with appropriate substitutions. Theorems are algebraic expressions.

When we see a space for the first time, it can be designated as a “raw binary space”, meaning that if we discern any regularity of the values, we may not know what generated them. Let us say a pattern has been generated “randomly”. One may argue that there is “emergence”, but of what? Patterns displayed resulting from allegedly autopoietic, or self-organizing processes present a challenge of determining what the organizing principle is. For any regularity in binary space, we should ask not only what generated it but what its significance is. In Wolfram's automata, what do any of the patterns generated by the automata mean, save for the design being associated with a rule? Rules are specific thought patterns, but is there a more general observation about thinking we can make? Later, we will touch on this subject in discussing a correlation between binary spaces and activity in brain structures. We won't analyze patterns (as many ways exist for doing so [7]) but suggest a way of normalizing a space and propose the hypercube as a measuring device for the raw spaces. To work with patterns in binary space, one can consider a way of making the hypercube a measurement and classification device. Raw spaces should be normalized, such as padding sets of bits less than a nybble (four bits) either before or after with 0s, as spaces can be only of complete functions. One compares the normalized space using the “dictionary of theorems”, mentioned above. A second means of classifying spaces is with the 3-D hypercube; what in that “raw space” looks like the regularity inside any of the plates in the hypercube? The hypercube is a baseline against which to measure deviation of any set of blocks of binary space.

## A system without axioms

To this point we have not identified any axioms but yet have been able to generate theorems by the definition of deduction, truth table methods, and corresponding conditional rule. All are based on the rule that a statement must be derivable from the previous, and that derivation can be found through inspection by using the hypercube. This means that in terms of zeros and ones, the “and”, as well as “material implication” operator/function a deductive relationship holds. Validation that a statement follows from the previous is done by inspection of the hypercube. This means that our system is without axioms, or naturally and intuitively deductive.

All of the functions emerge from juxtaposition of numbers in an ascending fashion. The table of logical space is set, and the functions are only naming devices for particular sets of 0s and 1s in that space. It is not necessary to use axioms to derive the functions. Only an ordering principle is needed, coupled with the primitives, definitions, and so forth. It is proposed that Peano's Postulates form the basis of such a system, plus a postulate that asserts mounting quantity based succession. A start would be focusing on the concept of number based on a fundamental Cartesian cut.

## Philosophical significance of the binary structures

The two-valued, or binary, system is foundational in deduction, as it uses the lowest number of variables possible to construct a system of relations. The simple observation, other than of the whole, is of two. This observation stems from an extreme division of any object in the three dimensional universe; ultimately, it will be reduced to the smallest of the smallest, or Planck volume in terms of not-Planck volume, or vacuum space. Of course this sub-quantum world must be apprehended in terms of the whole in order to place matters in proper perspective. The three dimensional world is syntactically binary [8].

This author's assertion is that logic is a language that describes innate order in the 3-D universe and that it is the basis upon which mathematics rests. Logic is discovered, rather than invented; "...a machinery for the combination of yes-no or true-false elements does not have to be invented. It already exists [9]." Jean Piaget asserts, "There exist outline structures which are precursors of logical structures,... It is not inconceivable that a general theory of structures will...be worked out, which will permit the comparative analysis of structures characterizing the outline structures to the logical structures ... [10]". This idea is not new, as it extends as far back as 5000 years ago in South Asia [11], and with the Chinese 4,000 years ago and the I Ching. Leibniz [12], the first modern scientist to formalize the binary arithmetic in 1703 wrote of it.

When one attempts Cartesian subdivision at the quantum level, the world of uncertainty is met, and one cannot measure position, except statistically. Yet, computationally, we can go to the Planck scale, where all symmetries are broken. To divide something beyond a Planck volume would require more energy than exists in the universe. What exists in terms of not-Planck volume is vacuum space, which is penetrated periodically with energy fluctuations. At the Planck scale, nothing is discrete, so one has to identify what makes a description of reality binary [13]. We certainly do not know what is at the Planck scale level, as this degree of granularity is theoretical and merely computational. At the Planck scale, the very nature of the binary world is transformed, where bivalency transforms to a four dimensional world and a four-valued system [14]. This system is the superset of the three-dimensional bivalent system.

The unfolding structure of the most basic logic in the three dimensional world - binary relationships - comes from natural ordering [15]. Logical space is generated in an ascending fashion and ultimately contains all the relationships possible in this world. Recall, everything is reducible to Planck volume

and non-Planck volume, a duality. From the singular and planar logical spaces comes the three dimensional hypercube.

## Future direction in research

Much of what follows is speculative, but it was stated at the outset in this paper that the hypercube is being presented for research purposes. Aside from lofty considerations of the quantum world [16], the hypercube has more prosaic applications such as an algebra of spaces. Consider commutativity as error checking device. Commutativity is symmetric, where  $f_c(fp, fq) = f_c(fq, fp)$ . This goes for  $C=0, 1, 6, 7, 8, 9, 14,$  and  $15$ . Functional computation may be expressed in algebraic form, such as the simple example in the  $f_5$  plate, where  $f_5(fn, fp) \rightarrow fp$ , and where  $fn$  and  $fp$  represent any two distinct functions. Similarly,  $f_3(fn, fp) \rightarrow fn$  exists for the  $f_3$  plate. For plate  $f_8(f_{15}, fq) = f_0$  and  $f_8(fp, f_{15}) = f_0$ , so  $f_8(f_{15}, fq) = f_8(fp, f_{15})$ . Numerous and more complicated relations may be developed, but such is work for further research.

We have seen where the new canonization with the functional notation can result in an algebra of functions to generate inference and equivalence rules. For example, *modus ponens* is  $p \supset q, p, \therefore q$ . In our canonization, this is  $f_{13}$ (a conjunct of functions resulting from others)  $\rightarrow f_c$  (a conclusion, or derived function) where  $f_{13}$  must result in  $f_{15}$ , or tautology. Research might produce a computer program to generate not only acceptable rules, but these might be used to help produce an algebra of spaces. Already, we have seen where the hypercube has aided us in finding theorems using the corresponding conditional and a computer program can be written to do this.

We said above that each of the 16 functions is recursive, i.e., the outputs forward fed as inputs into the function cause the function to reappear [17]. Thus, each function acts as a self-maintaining, or homeostatic, automaton. Of course, all binary spaces are composed of one or more of these functions, or partial functions (less than a nybble). Starting with a set of set of formulas demarcating an initial space, it would be interesting to see how that space evolves until it repeats itself. No entity at whatever level is static, so tracing the dynamism of an initial state of binary functions would give an insight to pattern generation and possibly shed light on how basins of attraction form.

As a longer term project, one may map each function to a sound or color to see what patterns may emerge. Newton, following an idea by the ancient Greeks, suggested that there may be a correlation between color and sound [18]. Correlating sound to color is not novel these days [19]. In various computer programs designed to play CDs, such as Windows Media, one can view colored patterns emerge when playing music.

Consciousness studies can be explored with processes applied to binary spaces. According to Tononi, consciousness is integrated information, "...the amount of information generated by a complex of elements, above and beyond the information generated by its parts." [20]. Consciousness arises from the condition of neural systems, and these can be represented in a binary manner, i.e., on-off switches, or as Tononi refers says, "photodiodes" [21]. Of course, to represent anything approaching what people think is consciousness would involve enormous complexity, as Tononi admits, but his serves a model for research. Perhaps the the 3-D hypercube developed in this paper could be overlaid on to the binary space generated by Tononi's model, much in the same manner as discussed earlier with respect to binary spaces in general. The theorems and their corresponding patterns generated by the hypercube might have neural correlates and such would involve Tononi's research. This world is just beginning.

## Conclusion and outlook

The most basic binary logical space is generated from a single square to two squares, one containing a value and the other a second value. The permutations of this two-squared space yield four permutations of the two values. These, in turn, produce the sixteen basic functions displayed in the Table of Functional Completeness (TFC). From the TFC is developed the three-dimensional hypercube. Functions also are results of computations and vice versa. A new canonization has been presented that allows for a simplified way of computing dyadic relationships, as well as traditional truth tables. The hypercube is color coded to help display the relationships of functions to each other and identify patterns. At the outset the hypercube can be used as a look-up table to yield the results of any dyadic computation involving any of the 16 functions.

The hypercube allows for more rapid and simplified dyadic computations in bivalent space, but also may enable enhanced methods for computing hamming distances. There are indications that Lorenz attractors may exist within the hypercube, these possibly indicating seeds from which order is generated from what was thought previously to be chaos. Something (a pattern) doesn't come from nothing. The universe at the third dimension has innate order, described by binary structures, the hypercube being one. While there is evidence of randomness (inability to predict), such as Brownian movement and pi ( $\pi$ ), there is an innate order in the universe, and chaos contains encoded order that can be untangled by logical analysis.

Numerous research areas stem from the development of the basic hypercube, all centering on pattern analysis of structures expressed in binary space. Once an undefined binary space is mapped onto the ordered one, the same analytical process of pattern recognition can be applied, thus leading to a uniform way of looking at reality in many of its diverse but reducible forms.

## References (all web links accessed 27 March 2011)

- [1] J. Horne, "The General Theory and Method of Binary Logical Operations", **Journal of Applied Science and Computations**, 3 December 2000.
- [2] I. Copi, **Symbolic Logic**, 5<sup>th</sup> ed., New York: Macmillan, 1979, pp. 220-222.
- [3] J. Horne, "Recursion Of Binary Space As a Foundation Of Repeatable Programs", **The Journal of Informatics and Systemics**, Vol. 4, No. 5, 2006.
- [4] S. Wolfram, **A New Kind of Science**, <http://wolframscience.com/nksonline/toc.html>.
- [5] A. Wuensche, "The Ghost in the Machine: Basins of Attraction of Random Boolean Networks, University of Sussex at Brighton: **Cognitive Science Research Papers**, CSRP 281, 1993, [https://docs.google.com/viewer?url=http://www.informatics.sussex.ac.uk/users/andywu/downloads/papers/ghost\\_in\\_machine.pdf&embedded=true&chrome=true](https://docs.google.com/viewer?url=http://www.informatics.sussex.ac.uk/users/andywu/downloads/papers/ghost_in_machine.pdf&embedded=true&chrome=true) .
- [6] Wolfram, *passim*.
- [7] Hamann, "Definition and Behavior of Langton's Ant in Three Dimensions", , <http://www.springerlink.com/content/m75033714447106g/>, [http://videlectures.net/kdd09\\_ye\\_mdpbmf/](http://videlectures.net/kdd09_ye_mdpbmf/), and <http://www.exploringbinary.com/what-a-binary->

[counter-looks-and-sounds-like/](#) .

[8] J. Horne, “Is Reality Digital or Analog?”, 2011 **FXQI Essay**,  
<http://www.fqxi.org/community/forum/topic/862> .

[9] C.W. Misner , K.S. Thorne, J.A. Wheeler **Gravitation**, New York: W.H. Freeman and Company, 1973.

[10] J. Piaget, **Logic and Psychology**, New York: Basic Books, Inc.; 1958.

[11] **Rig Veda**, Hymn CXXIX – **Creation**, Book the Tenth – **Oxford Companion to Philosophy**,  
[http://www.answers.com/topic/ancient-philosophy#Indian\\_philosophy](http://www.answers.com/topic/ancient-philosophy#Indian_philosophy).

[12] Leibniz, **Explication de l'Arithmétique Binaire**, [https://docs.google.com/viewer?url=http://ads.ccsd.cnrs.fr/docs/00/10/47/81/PDF/p85\\_89\\_vol3483m.pdf&embedded=true&chrome=true](https://docs.google.com/viewer?url=http://ads.ccsd.cnrs.fr/docs/00/10/47/81/PDF/p85_89_vol3483m.pdf&embedded=true&chrome=true), 1703.

[13] J. Horne, “Is Reality Digital or Analog?”, 2011 **FXQI Essay**,  
<http://www.fqxi.org/community/forum/topic/862> .

[14] A. Stern, A., **Matrix Logic**, Amsterdam: North-Holland, 1988.

[15] J. Horne, **Logic as the Language of Innate Order in Consciousness**,. *Informatica*, 1997; 21: 675-682.

[16] D. Rapoport, **Surmounting the Cartesian Cut Through Philosophy, Physics, Logic, Cybernetics, and Geometry: Self-reference, Torsion, the Klein Bottle, the Time Operator, Multivalued Logics and Quantum Mechanics**, Found Phys, 22 August 2009, A. Stern, A., **Matrix Logic**, Amsterdam: North-Holland, 1988.

[17] J. Horne, “Recursion of Logical Operators and Regeneration of Discrete Binary Space”, **Informatica**, 2000;24:275-279.

[18] Newton I. **Opticks, or, a Treatise of the Reflections, Refractions, Inflections, and Colours of Light**. London: Printed for William Innys at the West-End of St. Paul's, 1730.

[19] For example, see <http://www.jacmuse.com/artisticconcepts/newpage14.htm>,  
<http://www.harpcenter.com/page/SWHC/PROD/BestMeditCD/3212R>.

[20] G.B. Tononi, “Consciousness as Integrated Information: a Provisional Manifesto,;Bull. 215: 216-242, 2008.

[21] Ibid.

The author wishes to thank Asia Flood, a professional artist located in Gold Canyon, AZ for her invaluable assistance in color coding the hypercube. Without her work, it would have been difficult, if not impossible to discern the colored patterns.

## Appendix

### Sample Plate of The Three-dimensional Hypercube

$f_1$  - AND – conjunction  $p$  &  $q$

$f_1$	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$
$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$	$f_0$
$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$
$f_2$	$f_0$	$f_0$	$f_2$	$f_2$	$f_0$	$f_0$	$f_2$	$f_2$	$f_0$	$f_0$	$f_2$	$f_2$	$f_0$	$f_0$	$f_2$	$f_2$
$f_3$	$f_0$	$f_1$	$f_2$	$f_3$	$f_0$	$f_1$	$f_2$	$f_3$	$f_0$	$f_1$	$f_2$	$f_3$	$f_0$	$f_1$	$f_2$	$f_3$
$f_4$	$f_0$	$f_0$	$f_0$	$f_0$	$f_4$	$f_4$	$f_4$	$f_4$	$f_0$	$f_0$	$f_0$	$f_0$	$f_4$	$f_4$	$f_4$	$f_4$
$f_5$	$f_0$	$f_1$	$f_0$	$f_1$	$f_4$	$f_5$	$f_4$	$f_5$	$f_0$	$f_1$	$f_0$	$f_1$	$f_4$	$f_5$	$f_4$	$f_5$
$f_6$	$f_0$	$f_0$	$f_2$	$f_2$	$f_4$	$f_4$	$f_6$	$f_6$	$f_0$	$f_0$	$f_2$	$f_2$	$f_4$	$f_4$	$f_6$	$f_6$
$f_7$	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$
$f_8$	$f_0$	$f_8$	$f_0$	$f_8$	$f_8$	$f_8$	$f_8$	$f_8$	$f_8$	$f_8$						
$f_9$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_1$	$f_0$	$f_9$	$f_8$	$f_9$	$f_8$	$f_9$	$f_8$	$f_9$
$f_{10}$	$f_0$	$f_0$	$f_2$	$f_2$	$f_0$	$f_0$	$f_2$	$f_2$	$f_8$	$f_8$	$f_{10}$	$f_{10}$	$f_8$	$f_8$	$f_{10}$	$f_{10}$
$f_{11}$	$f_0$	$f_1$	$f_2$	$f_3$	$f_0$	$f_1$	$f_2$	$f_3$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_8$	$f_9$	$f_{10}$	$f_{11}$
$f_{12}$	$f_0$	$f_0$	$f_0$	$f_0$	$f_4$	$f_4$	$f_4$	$f_4$	$f_8$	$f_8$	$f_8$	$f_8$	$f_{12}$	$f_{12}$	$f_{12}$	$f_{12}$
$f_{13}$	$f_0$	$f_1$	$f_0$	$f_1$	$f_4$	$f_5$	$f_4$	$f_5$	$f_8$	$f_9$	$f_8$	$f_9$	$f_{12}$	$f_{13}$	$f_{12}$	$f_{13}$
$f_{14}$	$f_0$	$f_0$	$f_2$	$f_2$	$f_4$	$f_4$	$f_6$	$f_6$	$f_8$	$f_8$	$f_{10}$	$f_{10}$	$f_{12}$	$f_{12}$	$f_{14}$	$f_{14}$
$f_{15}$	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$

$f_{13}$  -  $\supset$ ,  $p$  contains  $q$ ,  $p \supset q$ , – defines deduction

$f_{13}$	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$
$f_0$	$f_{15}$															
$f_1$	$f_{14}$	$f_{15}$														
$f_2$	$f_{13}$	$f_{13}$	$f_{15}$	$f_{15}$												
$f_3$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$												
$f_4$	$f_{11}$	$f_{11}$	$f_{11}$	$f_{11}$	$f_{15}$	$f_{15}$	$f_{15}$	$f_{15}$	$f_{11}$	$f_{11}$	$f_{11}$	$f_{11}$	$f_{15}$	$f_{15}$	$f_{15}$	$f_{15}$
$f_5$	$f_{10}$	$f_{11}$	$f_{10}$	$f_{11}$	$f_{14}$	$f_{15}$	$f_{14}$	$f_{15}$	$f_{10}$	$f_{11}$	$f_{10}$	$f_{11}$	$f_{14}$	$f_{15}$	$f_{14}$	$f_{15}$
$f_6$	$f_9$	$f_9$	$f_{11}$	$f_{11}$	$f_{13}$	$f_{13}$	$f_{15}$	$f_{15}$	$f_9$	$f_9$	$f_{11}$	$f_{11}$	$f_{13}$	$f_{13}$	$f_{15}$	$f_{15}$
$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$
$f_8$	$f_7$	$f_{15}$														
$f_9$	$f_6$	$f_7$	$f_6$	$f_7$	$f_6$	$f_7$	$f_6$	$f_7$	$f_{14}$	$f_{15}$	$f_{14}$	$f_{15}$	$f_{14}$	$f_{15}$	$f_{14}$	$f_{15}$
$f_{10}$	$f_5$	$f_5$	$f_7$	$f_7$	$f_5$	$f_5$	$f_7$	$f_7$	$f_{13}$	$f_{13}$	$f_{15}$	$f_{15}$	$f_{13}$	$f_{13}$	$f_{15}$	$f_{15}$
$f_{11}$	$f_4$	$f_5$	$f_6$	$f_7$	$f_4$	$f_5$	$f_6$	$f_7$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$
$f_{12}$	$f_3$	$f_3$	$f_3$	$f_3$	$f_7$	$f_7$	$f_7$	$f_7$	$f_{11}$	$f_{11}$	$f_{11}$	$f_{11}$	$f_{15}$	$f_{15}$	$f_{15}$	$f_{15}$
$f_{13}$	$f_2$	$f_3$	$f_2$	$f_3$	$f_6$	$f_7$	$f_6$	$f_7$	$f_{10}$	$f_{11}$	$f_{10}$	$f_{11}$	$f_{14}$	$f_{15}$	$f_{14}$	$f_{15}$
$f_{14}$	$f_1$	$f_1$	$f_3$	$f_3$	$f_5$	$f_5$	$f_7$	$f_7$	$f_9$	$f_9$	$f_{11}$	$f_{11}$	$f_{13}$	$f_{13}$	$f_{15}$	$f_{15}$
$f_{15}$	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$

## Optimum Node Deployment of the Wireless Sensor Network System

Yong Kuk Park, Min Goo Lee, Kyung Kwon Jung, June Jae Yoo  
U-Embedded Convergence Research Center, Korea Technology Institute  
Seongnam-si, 463-816, Korea

and

Hyeong-Seok Kim  
School of Electrical and Electronics Engineering, Chung-Ang University  
Seoul, 156-756, Korea

### ABSTRACT

In this paper, a fuzzy c-means clustering algorithm is proposed to determine the optimum deployment of sensor nodes. It is for a given application space to improve energy efficiency and reduce cost. We performed simulation for building area to find minimum number and optimum location of sensor nodes.

**Keywords:** Sensor Nodes Deployment, Fuzzy C-means algorithm

### 1. INTRODUCTION

One of the fundamental problems in a wireless sensor network system is the efficient deployment of sensor nodes. Since it is not easy to replace batteries of sensor nodes, efficient energy management is one of the most important tasks in the wireless sensor network system. To solve this problem, various methods using efficient MAC protocol [1] and routing protocol have been tried, but these methods tend to be application-specific. However, if the required number and the location of sensor nodes for the given area, coverage [2], are known in advance, the number of unnecessary nodes can be reduced in application, resulting in improved energy efficiency at a reduced cost.

In this paper, a method to determine optimum node deployment of a sensor network is proposed for communication between nodes in a limited area. In this method, the minimum number and the optimum location of sensor nodes are derived from a simulation based on the fuzzy c-means clustering algorithm [3]. The validity of this method is analyzed through experiments with actual deployment of sensor nodes based on the simulation, and measurement of success rate of communication between the nodes. We assume that the communication range and the sensing range of each sensor node are identical.

### 2. WIRESS CHANNEL MODEL

In the literature, there are numerous experimental and theoretical studies of indoor propagation. These models tend to focus on a particular characteristic like temporal fading or inter-floor losses. This paper aims at developing an indoor propagation model from measurements taken using 2.4 GHz band wireless sensor nodes.

In both indoor and outdoor environments the average large scale path loss for an arbitrary Transmitter-Receiver (T-R) separation is expressed as a function of distance by using a path loss exponent,  $n$ . This value of  $n$  depends on the propagation environment, i.e., type of construction material, architecture, and location within a building. Lowering the value of  $n$  lowers the signal loss. For example, in free space,  $n$  is equal to 2, and when obstructions are present,  $n$  will have a larger value [4].

Random shadowing effects occurring over a large number of measurement locations which have the same T-R separation, but different levels of clutter on the propagation path is referred to as Log-Normal Distribution. This phenomenon is referred to as log-normal shadowing. Variations in environmental clutter at different locations having the same T-R separation is not accounted for by the log-distance path loss model alone. This leads to measured signals which are vastly different than the average value predicted by using the log-distance path loss model.

The log-normal path loss model is a radio propagation model that predicts the path loss a signal encounters inside a building or densely populated areas over distance.

To account for these variations, the average path loss  $PL(d)$  for a transmitter and receiver with separation  $d$ .

$$PL(\text{dB}) = PL(d_0) + 10n \log\left(\frac{d}{d_0}\right) + X_\sigma \quad (1)$$

Where  $X$  is a zero-mean Gaussian distributed random variable with standard deviation  $\sigma$ . The reference distance  $d_0$ , the path loss exponent  $n$ , and the standard deviation  $\sigma$ , statistically describe the path loss

model for an arbitrary location having a T-R separation. This model can be used in computer simulation to provide received power levels for random locations in communication system design and analysis.

The channel modeling experimental set-up is displayed in Figure 1. A TIP50 node [5] is used as transmitter and a spectrum analyzer (MS2668C, Anritsu) is used as receiver, and both of them are placed 0.5 m over the floor.

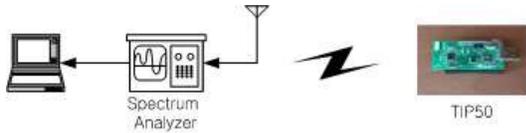


Figure 1. Measurement of received power.

Several positions are chosen for the transmitter, and for each of them the receiver is separated from the transmitter. The steps between the receiver positions and the transmitter are: From 0.5 to 10 meters every 0.5 meters. Centered at each step, for all these positions a 10 times time averaging is performed (10 snapshots per step). The experiments are conducted several times, and environment similar results are obtained in different runs. Figure 2 shows the results according to the methodology explained above. It can be seen that the received power fits a exponential line using the method of least square errors.

The log-normal path loss model parameters obtained through the received RF power are shown in Table 1.

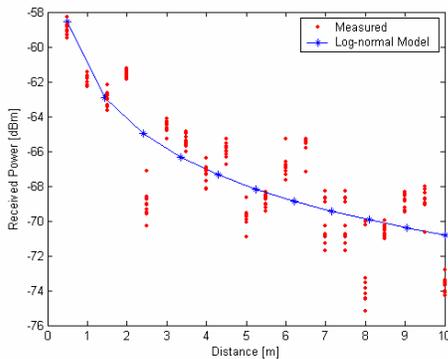


Figure 2. RF received power and channel modeling results.

Table 1. Log-normal path loss model parameters

$PL(d_0)$ [dBm]	$n$	$X\sigma$
61.903	0.942	0.697~28.912

A frame is received correctly if all its bits are received correctly, hence, for a frame of length  $f_b$  the probability of successfully receiving a packet is:

$$P = (1 - P_e)^{f_b} \quad (2)$$

Where,  $P_e$  is the probability of bit error.  $P_e$  depends on the modulation used in CC2420. The CC2420 corresponds to O-QPSK.

$$P_e = Q\left(\sqrt{2 \cdot \frac{E_b}{N_0}}\right) \quad (3)$$

Where,  $Q(\cdot)$  is error function,  $E_b/N_0$  is the energy per bit to noise power spectral density ratio [6].

The packet reception rate (PRR) is shown in Figure 3.

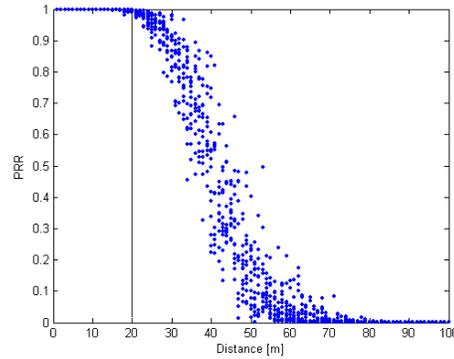


Figure 3. Packet reception rate

At this time, the communication range of each node was defined as a circle of 20m radius.

### 3. DEPLOYMENT OF SENSOR NODES

#### Sensor Node Density

During the design phase of wireless sensor networks, the designer knows the number of sensor nodes which are deployed in a given space in either random or deterministic fashion.

A circular radio range with radius  $r$  is considered in this paper. A grid-based deployment is introduced as a good deployment in wireless sensor networks, especially for the coverage performance. Figure 4 shows a grid deployment of  $M$ -by- $N$  grid points hosts a sensor.

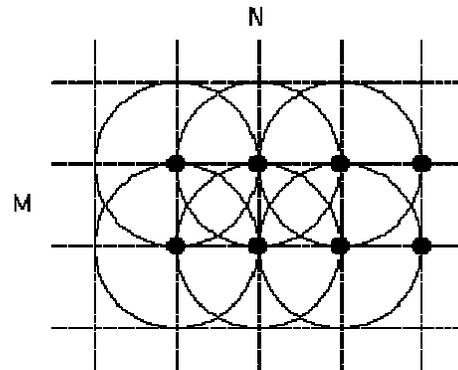


Figure 4. Sensor nodes of grid structure.

The total area of circular boundary,  $S$ , can be calculated Equation (4).

$$S = \pi r^2 + (N-1)\left(\frac{\pi}{3} + \frac{\sqrt{3}}{2}\right)r^2 + (M-1)\left(N\left(\frac{3\sqrt{3}}{2}-1\right) + \frac{\pi}{3} - \sqrt{3}\right)r^2 \geq A \quad (4)$$

Where  $A$  is area of given space.

In the Equation (4), total number of sensor nodes,  $MN$ , term can be written as Equation (5).

$$MN \geq \frac{A - \left(\left(\frac{\pi}{3} - \sqrt{3}\right)M + \left(\frac{\pi}{3} - \sqrt{3} + 1\right)N\right)r^2}{\left(\frac{3\sqrt{3}}{2} - 1\right)r^2} \quad (5)$$

If the given area  $A$  is bigger than square of radio range  $\rho$ , the sensor node density  $\rho$  is derived as follows:

$$\rho = \frac{MN}{A} \geq \frac{1}{\left(\frac{3\sqrt{3}}{2} - 1\right)r^2} \quad (6)$$

According to radio range and building area, the sensor node density is shown in Figure 5.

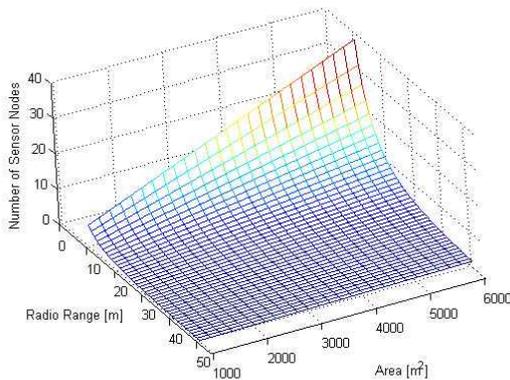


Figure 5. Sensor node density.

### Fuzzy C-means Algorithm

Fuzzy clustering plays an important role in solving problems in the areas of pattern recognition and fuzzy model identification. A variety of fuzzy clustering methods have been proposed and most of them are based upon distance criteria [6]. One widely used algorithm is the fuzzy c-means (FCM) algorithm. It uses reciprocal distance to compute fuzzy weights. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is

used in pattern recognition [3].

FCM clustering is used mainly in dealing large data sets such as genome analysis or image processing. Because there are infinite possibilities in placing the sensor nodes, we applied FCM clustering, which uses membership information to achieve accurate clustering and to find optimum node location. The object function  $J_m$  is given as follows.

$$J_m(u, c) = \sum_{k=1}^K \sum_{i=1}^N (u_{ki})^m d^2(x_i, c_k) \quad (7)$$

$$\sum_{k=1}^K u_{ki} = 1 \quad \text{for all } i=1, \dots, n \quad (8)$$

Here,  $K$  is the number of cluster, and  $N$  is the number of deployable nodes.  $u_{ki}$  is a value between 0 and 1 referring to the probability for the  $i$  th data of the node at  $x_i$  to belong to the  $k$  th cluster.  $m \in [1, \infty)$  is an exponential weight variable indicating the fuzziness of the membership function.  $m = 2$  is generally used to perform the clustering.  $d_2(x_i, c_k)$  is a distance from the center of the cluster  $c_k$  to the deployable node location  $x_i$ . A probabilistic constraint that the sum of the membership for each cluster should be 1 is imposed.

## 4. Simulations

Simulation was performed with L-shape of plane model using Matlab. The L-shaped plane for simulation is shown by Figure 6. The area is 4,800m<sup>2</sup> and number of sensor nodes according to radio range is shown in Figure 7.

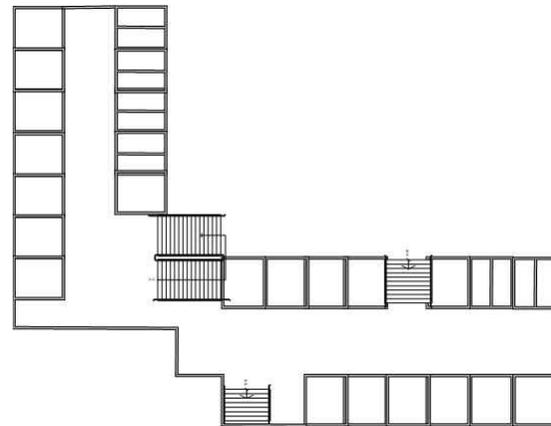


Figure 6. Plane figure.

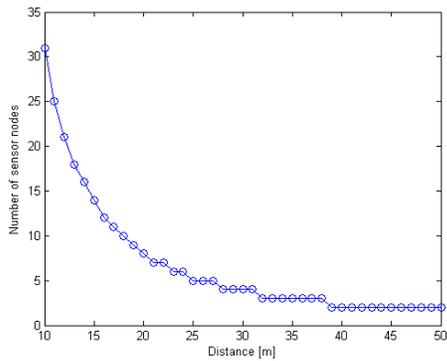


Figure 7. Radio range vs. number of sensor nodes.

Simulation results are shown in Figure 8. In Figure 8, small points indicate randomly generated deployable points of the nodes. Large points indicate optimum node deployment derived through the FCM clustering of small points. Finally, large circles indicate the communication distance of each node. Nodes capable of communicating with each other were connected with a solid line.

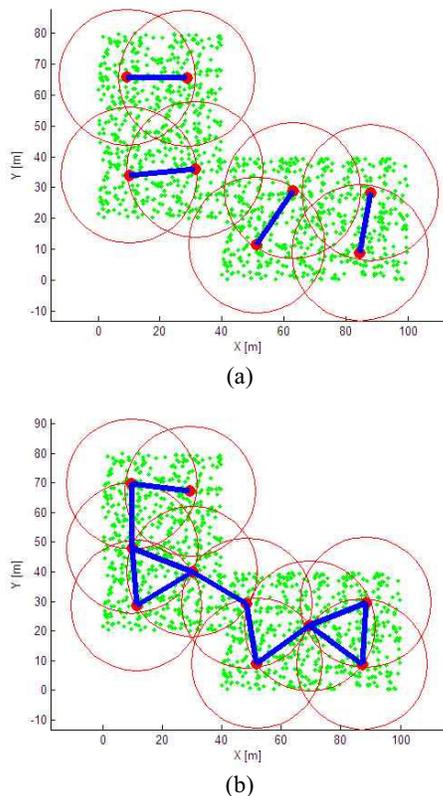


Figure 8. Simulation results. (a) MN=8, (b) MN=10.

#### 4. CONCLUSIONS

In this paper, we propose an optimal placement of sensor nodes with 2.4GHz wireless channel characteristics. The proposed method determines optimal transmission range based on log-normal path loss model, and optimal number of sensor nodes calculating the density of sensor nodes. For the lossless data transmission, we search the optimal locations with fuzzy c-means algorithm, and number of sensor nodes. We demonstrate that optimal transmission range is 20m, and optimal number of sensor nodes is 8 or higher. We performed simulations on the searching for optimal locations and confirmed the link condition of sensor nodes.

#### 5. ACKNOWLEDGMENT

The project related to this paper is performed by the Ministry of Knowledge Economy and the Korea Evaluation Institute of Industrial Technology. [10035491]

#### 6. REFERENCES

- [1] I. Demirkol, C. Ersoy, F. Alagoz, "MAC protocols for wireless sensor networks: a survey," *IEEE Communications Magazine*, Volume 44, Issue 4, pp. 115 – 121, 2006.
- [2] S. Meguerdichian, F. Koushanfar, M. Potkonjak, M. B. Srivastava, "Coverage problems in wireless ad-hoc sensor networks," *INFOCOM*, vol. 3, pp. 1380 – 1387, 2001.
- [3] J. Bezdek. "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. on Patt. Anal. and Mach. Intell*, vol. PAMI-2, pp. 1-8, 1980.
- [4] Bernard Sklar, *Digital Communications: Fundamentals and Applications*, Prentice Hall, 2001.
- [5] Maxfor homepage, <http://www.maxfor.co.kr/>
- [6] Leon W. Couch, *Digital and Analog Communication Systems*, Prentice Hall, 2001.

# Using Motion Levels Of Detail in the Fast Multipole Method for Simulation of Large Particle Systems

Seyed Naser RAZAVI

Department of computer engineering, Iran University of Science and Technology  
Tehran, Iran

Nicolas GAUD

Multi-agent systems and applications group, Laboratoire Systèmes et Transports, UTBM  
Belfort, France

Abderrafiâa KOUKAM

Multi-agent systems and applications group, Laboratoire Systèmes et Transports, UTBM  
Belfort, France

Naser MOZAYANI

Department of computer engineering, Iran University of Science and Technology  
Tehran, Iran

[razavi@iust.ac.ir](mailto:razavi@iust.ac.ir), [nicolas.gaud@utbm.fr](mailto:nicolas.gaud@utbm.fr), [abder.koukam@utbm.fr](mailto:abder.koukam@utbm.fr), [mozayani@iust.ac.ir](mailto:mozayani@iust.ac.ir)

## ABSTRACT

This article introduces a novel approach to increase the performances of  $N$ -body simulations. In an  $N$ -body simulation, we wish to evaluate all pairwise interactions between  $N$  bodies or particles. The direct computation of all pairwise interactions requires  $O(N^2)$  time, which is clearly prohibitive for a very large  $N$ . Our approach combines the Fast Multipole Method (FMM) coming from computational physics with motion levels of detail from computer graphics. The main goal is to speed up the execution of the  $N$ -body simulations while controlling the precision of the associated approximation, a natural trade-off between accuracy and efficiency common in the field of simulation. At each simulation cycle, the motion levels of detail are generated automatically and the appropriate ones are chosen adaptively to reduce computational costs. The new approach follows the overall structure of the FMM. However, clusters are approximated using their Center of Mass (CoM) in force computations. A similarity measure is used to decide which clusters can be approximated without any significant loss in the accuracy of the simulation. The proposed approach is tested for Coulombic system, in which  $N$  charges induce potentials to each other. The preliminary results show a significant complexity reduction without any remarkable loss in the visual appearance of the simulation, indicating the potential use of the proposed model in the simulation of a wide range of  $N$ -Body systems.

**Key words** – *Dynamics Simplification, Fast Multipole Method, Motion levels of Detail, Multi-Agent Based Simulation*

## 1. INTRODUCTION

Nowadays, Thanks to the increasing power of modern computers, simulation plays a very important role in the study of large complex systems and has attracted many researchers from a growing number of diverse areas including sociology,

biology, physics, chemistry, ecology, economy, etc. Simulation of a large complex system often requires computing all the pairwise interactions between particles in the system. Assuming  $N$  particles, direct methods require  $O(N^2)$  computations which is clearly prohibitive for large  $N$ . The challenge of efficiently carrying out the related calculations is generally known as the  $N$ -body problem.

The Fast Multipole Method (FMM), proposed by Rokhlin and Greengard [1,2,3], can be regarded as one of the most successful attempts to reduce the  $O(N^2)$  time complexity required in the  $N$ -body simulations. Also, it has been identified as one the ten most important algorithmic contributions in the past century [4]. Depending on the distribution of particles in the computational domain, FMM evaluates the pairwise interactions in  $O(N \log N)$  or even  $O(N)$  which is a remarkable improvement over the  $O(N^2)$  time required by direct methods.

The FMM is based on the idea that well-separated or far-away groups of particles can be considered as one particle. So, it imposes a hierarchical spatial partitioning structure on the computational domain to determine the near-field and the far-field for each particle or group of particles. Based on our recent work [5], we think that this hierarchical structure can be exploited to further reduce the overall computational cost. As the "no free lunch" theory suggests, this complexity reduction introduces a new source of inaccuracy in the simulation which is tolerable in many applications where we are more interested in the overall behavior of the system as a whole instead of the individual behavior of the constituent particles. The new approach presented here combines the FMM with motion levels of detail from computer graphics to improve simulation speed in the  $N$ -body simulations while controlling the desired level of accuracy.

In fact, the FMM algorithm clusters particles into hierarchical groups based on their relative positions in the domain, but it deals with each particle in the same group separately,

computing a different potential and force for each particle in the group resulting in different behaviors for different particles in the same group. However, particle systems generally exhibit a high level spatial coherence, meaning nearby particles behave in approximately the same way. Therefore, the main idea is to compute only one approximated behavior for a group and then making all particles in the group to follow the same behavior. This is achieved by replacing particles inside a cluster by one weighted particle and approximating the motion models of those particles in the force computations during FMM.

To perform automatic dynamics simplification, a mechanism is needed to generate a hierarchy of motion models and also to switch between different levels in the hierarchy. Here, we have decided to use the same *hierarchical physically-based subdivision* scheme in the FMM. This way we can take advantage of the approximate computations used in the FMM to reduce the complexity beside our automatic dynamics simplification. However, during the force computations, some parts of the FMM tree (quadtree in 2D and octree in 3D) are pruned and replaced by only one node containing a single weighted particle. Particles which are located in the pruned areas follow the same behavior computed for the weighted particle. This way, a lot of computations can be saved.

The pruning described above can result in a huge amount of inaccuracy. Thus, to keep the error below a desired level specified by the user, a mechanism is used to determine which parts of the tree can be pruned safely without any remarkable loss in the visual appearance of the simulation. This important decision that specifies which parts of the tree can be pruned is made at each simulation cycle. Therefore a specific region in the computational domain may be simulated in a higher resolution in the current cycle and it may be simulated in a lower resolution at later cycles because of the change in the distribution of its particles. It is very important to note that this transition between different levels of detail should be very smooth so that it does not cause a noticeable change in the flow of the simulation.

Therefore, at each simulation step the tree is updated based on changing simulation requirements as defined by the user or by the nature of simulation. Appropriate motion levels of detail are adaptively generated based on this subdivision given the requirements for different regions of the simulation and a desired execution time for each step. We have implemented a prototype system that can automatically generate simplified motion models, select appropriate models, and switch between them seamlessly. Then, the proposed framework is applied to the dynamic simulation of a large ensemble of particles in a Coulombic system which is a common example for  $N$ -body systems. The preliminary results are very promising, indicating a significant reduction in the complexity of the simulation while maintaining its correctness, and thus the potential to generalize the framework to the dynamic simulation of more systems.

Although there are existing simulation acceleration techniques such as [6,7,8,9], there is no known algorithm for the *automatic dynamics simplification* of complex physical or biological systems especially when all the pairwise interactions between particles are needed. In this paper, we limit the scope of our investigation to particle systems as a first step toward the design of automatic dynamics simplification. Particle systems are commonly used in computer graphics, physically based modeling, and animation of natural phenomena and group behavior. Additionally, they are often the building blocks of highly complex dynamical systems. Therefore, we hope the results of this paper can lead to the generalization of dynamics

simplification for a broader class of dynamical systems (e.g. pedestrians and crowds simulation).

The rest of this paper is organized as follows: section 0 gives a brief review of some related works regarding simulation acceleration techniques and motion levels of detail. Section 0 presents a quick review of the FMM which is used to solve our target application. Section 4 describes the proposed framework, including the automatic generation of motion levels of detail and the mechanism used to select appropriate models and to switch between them adaptively. Some experimental results are given in section 5. Finally, section 6 concludes with some future research guidelines and perspectives.

## 2. LITERATURE REVIEW

In this section, various attempts related to level of details done in the field of interactive computer graphics are summarized. To achieve *dynamic realism* in computer graphics, often the polygonal geometry of small or distant portions of the model is simplified to reduce the rendering cost without a significant loss in the visual content of the scene. Therefore, the goal of polygonal simplification in rendering is to reduce the complexity of a polygonal model to a level that can be rendered at interactive rates. In an offline preprocessing step, multiple versions of each object are created at progressively coarser *levels of detail*, or LODs. Once the LODs have been created and stored for every object in the model, complexity can be regulated at run-time by choosing for each frame which LOD will represent each object. As an object grows more and more distant, the system switches to coarser and coarser LODs. This type of simplification is called geometrical simplification or graphical levels of detail. It is important to distinct it from simulation levels of detail.

Simulation levels of detail or motion levels of detail techniques have been proposed for reducing the computational cost of the dynamics simulation of the character motion. Motion models can be generated from pre-recorded motion sequences, procedurally approaches, kinematics, or based on dynamics computation. Some of the earlier human motion models in computer animation exploited this concept implicitly by using procedurally generated motion, simplified dynamics and control algorithms, off-line motion mapping, or motion play-back [10,11,12,13,14].

Carlson and Hodgins applied simulation LOD techniques to a graphical environment populated with multiple, physically simulated one-legged robots [6]. They demonstrated that a group of characters that dynamically switches between LODs can sufficiently replicate the performance of a fully simulated group. Although in this work, the generation of simulation LODs, switching and selection are designed by hand but their experimental results are indicative of the potential of automatic simplification of general dynamical systems.

By using the space-time constraint dynamics formulation, Popovic and Witkin introduced a motion transformation technique that preserves the essential properties of animated character motion with drastically less number of degrees of freedom [9]. Multon, et al. suggested a series of simplified walking models for mobilizing on complex terrain, as well as how and when the transition takes place [15].

Other types of simulation acceleration techniques have also been investigated to reduce the total computational simulation costs for a large, complex dynamical system. For example, Chenney *et al.* proposed view-dependent culling of dynamic

systems to speed up the computation of dynamics by ignoring what is not visible to the viewer [7], similar to view culling. As another example, Faloutsos *et al.* developed a system that accomplishes complex tasks by evaluating multiple controllers automatically and selecting an appropriate sequence [16].

Additional graphics researchers are investigating ways to simplify physical simulations to reduce computation costs. Grzeszczuk *et al.* developed a technique that uses neural network approximations to emulate physically simulated characters' equations of motion [8]. After the neural network is trained to sufficiently model the original system, it can produce motions more efficiently than a full dynamic simulation. O'Sullivan and Dingliana investigate the opportunities to replace simulated particles with simplified counterparts when imperceptible to the viewer based on the idea that inaccurate dynamics are less noticeable in the peripheral vision and when the movements are complex [17]. O'Brien *et al.* describe a method to automatically simplify particle simulations through clustering into spatially localized groups [18].

It is very important to notice that the practitioners of interactive computer graphics are concerned less with accurately simulating the physics of a system than with finding better and faster ways to approximate the results of such a simulation. This is very different from a situation in which accuracy of simulation is paramount. However, with careful attention, it is possible to apply the same ideas to these kinds of applications to reduce the time complexity while preserving accuracy. This is a major feature distinguishing this work from those in the field of computer graphics. Therefore, the work presented in this paper may be considered as a link connecting the rich body of literature on real-time computer graphics with those in  $N$ -body simulations.

### 3. REVIEW OF THE FMM

Given  $N$  source densities  $\{\phi_i\}$  located at  $N$  source points  $\{x_i\}$ , we wish to compute the potential  $\{q_i\}$  at  $N$  target points  $\{y_i\}$  induced by a kernel  $G$  using the following relation:

$$q_j = \sum_{i=1}^N G(x_i, y_j) \phi(x_i), j = 1, \dots, N.$$

As mentioned earlier, direct implementation of this summation result in an  $O(N^2)$  algorithm, while for a large class of kernels, FMM computes the same interactions in  $O(N)$  time. However, FMM is an approximate algorithm, in the sense that the summation is not computed exactly. But, the good news is that the error can be bounded from above and the constant factor in the time complexity of the FMM is directly related to the accuracy of the approximation.

In our implementation, we have used the single layer Laplacian kernel  $\phi$  which arises from solving the Coulombic system of charged particles. Given a point charge of unit strength at point  $x_i$  in the complex plane, then for any point  $y_j$  also in the complex plane with  $x_i \neq y_j$ , the potential at point  $y_j$  due to the charge of  $x_i$  is given by  $G(x_i, y_j) = -\log\|y_j - x_i\|$ . In the FMM context it is convenient to use  $G(x, y) = \text{Re}(\log(z_y - z_x))$  where  $z_x$  and  $z_y$  are complex numbers corresponding to  $x$  (source) and  $y$  (target) points on the plane. The main idea of FMM is to represent the potentials of a set of source densities using the *multipole expansion* and *local expansion* at places far away from these sources. Assuming that

the source densities are located inside a disk centered at  $z_c$  with radius  $r$ , then for all  $z$  outside the disk with radius  $R$  ( $R > r$ ), the potential at  $z$  from the source densities can be represented using a set of coefficients  $\{a_k, 0 \leq k \leq p\}$  where

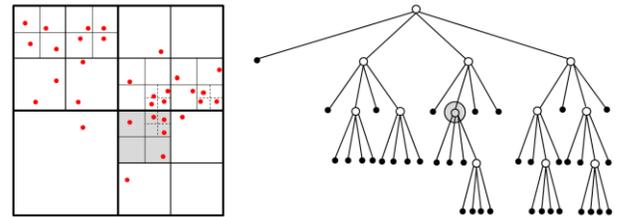
$$q(z) = a_0 \log(z - z_c) + \sum_{k=1}^p \frac{a_k}{(z - z_c)^k} + O\left(\frac{r^p}{R^p}\right) \quad (1)$$

This is called multipole expansion and represents the contribution from the source densities inside the disk centered at  $z_c$  with radius  $r$  on the far-field of that disk. Similarly, the contribution from source points in the far-field on any target point  $z$  inside that disk is called local expansion and can be represented again using a set of coefficients  $\{c_k, 0 \leq k \leq p\}$  where

$$q(z) = \sum_{k=1}^p c_k (z - z_c)^k + O\left(\frac{r^p}{R^p}\right) \quad (2)$$

In both expansions, the *truncation* number  $p$  is usually a small constant determining from the desired accuracy of the result. For the definitions of the coefficients  $\{a_k\}$  and  $\{c_k\}$ , the reader may refer to [2].

The above representations are used by FMM in a recursive manner. This is achieved by a hierarchical partitioning of the computational domain. At the first step of the FMM, which is called *space partitioning*, the computational domain, a box large enough to contain all source and target points, is hierarchically partitioned into a tree structure (a quadtree in 2D or an octree in 3D). The tree is constructed so that the leaves contain no more than a prespecified number of points, say  $s$ . This parameter is usually called the *clustering size*. Figure 1 shows an example distribution of particles and the corresponding quadtree in 2D.



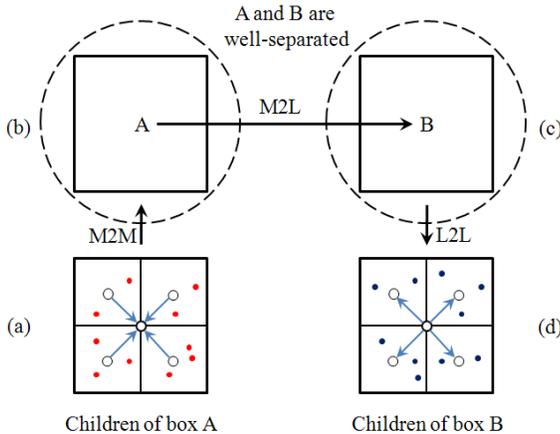
**Figure 1** A 2D particle distribution (left) and its corresponding quadtree (right).

After the tree construction, FMM performs two different passes of the tree with different goals: one *upward pass* and one *downward pass*. During the upward pass, the tree is traversed from bottom to the top in a postorder manner. The goal of the upward pass is to compute the multipole expansion for each box in the tree. For each leaf box, the multipole expansion is computed directly from its source densities using (1), while the multipole expansions for nonleaf boxes are computed using an operator which is called M2M translation (see Figure 2b).

During the downward pass, the tree is traversed in preorder to compute the local expansions. For each box  $B$ , the local expansion is the sum of two parts: first, the local-to-local transformation collects the local expansion of  $B$ 's parent (the contribution from the sources in all boxes which are not adjacent to  $B$ 's parent) as shown in Figure 2d, and second, the multipole-to-local transformation collects the multipole

expansions of the boxes which are the children of the neighbors of  $B$ 's parent but are not adjacent to  $B$  (these boxes compose the *interaction list* of  $B$ ) as shown in Figure 2c. Therefore, the some of these two parts encodes all the contribution from the sources in the boxes which are not adjacent to  $B$  itself.

At the final step of the FMM, which is called *final summation*, for each box the near interaction evaluated by iterating over all the source points in the neighborhood of the target box is combined with the far interaction which is evaluated using local expansion at this box to obtain the potential of each target point inside that box.



**Figure 2** The Fast Multipole Algorithm mechanisms: (a) after spatial decomposition, (b) the child boxes use the Multipole to Multipole translation to shift their multipole expansion to the center of the parent box, (c) using the Multipole to Local translation, well-separated boxes interact by creating a local expansion at the center of box  $B$  due to box  $A$ , (d) the children of box  $B$  feel the potential of box  $A$  by using the Local to Local translation to shift the parent's local expansion.

In addition to these expansions (multipole and local) that can be used for an efficient evaluation, FMM uses three different types of translation operators to translate between these expansions. The combined effect of these expansions and translations make an  $O(N)$  complexity algorithm possible. These translation operators are listed here:

**M2M**: the *multipole to multipole* translation transforms the multipole expansions of a box's children to its own multipole expansion.

**M2L**: the *multipole to local* translation transforms the multipole expansions of a box to the local expansion of another non-adjacent box.

**L2L**: the *local to local* translation transforms the local expansion of a box's parent to its own local expansion.

For the derivation of these translations and a detailed discussion on complexity analysis, error bounds and implementation details, the reader may refer to [2] and [19].

#### 4. INTEGRATING MOTION LEVELS OF DETAIL INTO THE FMM

This section presents the new method for automatic dynamics simplification with application in the simulation of large dynamical systems. As the new method follows the overall structure of the FMM, the focus of this section is mainly on the

differences rather than the similarities. The discussion starts by presenting the details about computing the *Center of Mass* (CoM) particle for each box or group of boxes, and then follows by introducing the criteria used in the new algorithm to determine those parts of the tree that can be pruned to simplify calculations.

##### 4.1 Approximating motion models

For each box (a leaf box or an internal box in the FMM tree), dynamics simplification is achieved by approximating its particles dynamics using the particle dynamics of its center of mass. Given a box  $B$  consisting of  $k$  particles, the corresponding approximated motion model is computed using the following procedure:

- 1) Compute the position  $P_{COM}$  and velocity  $V_{COM}$  of the center of mass particle using:

$$P_{COM} = \frac{\sum_{i=1}^k m_i P_i}{\sum_{i=1}^k m_i}$$

$$V_{COM} = \frac{\sum_{i=1}^k m_i V_i}{\sum_{i=1}^k m_i}$$

Where  $m_i$ ,  $P_i$  and  $V_i$  representing the charge, position and velocity of the  $i$ th particle inside  $B$ . Using the standard computations in the FMM, first compute the potential of the CoM particle for box  $B$  and then update its velocity and position accordingly.

- 2) Apply the same results computed for the COM particle to all particles inside  $B$ .

##### 4.2 Pruning the FMM tree

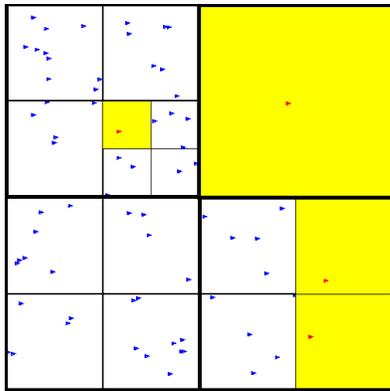
After the construction of FMM tree, some parts of the tree are pruned to speed up the simulation process. This pruning is performed based on the past behaviors of the particles in each box. For a given box, if the behaviors of the particles are more or less the same, then the particles inside that box can be replaced with the CoM particle of that box. So, a mechanism is needed to decide for which boxes the dynamics computation can be simplified. Here, these boxes are called *simplified boxes*. Applying a good mechanism is a key factor determining the overall success or failure of the automatic dynamics simplification.

In this work, two similarity measures are used to determine the candidate boxes for simplification: the *speed ratio* and the *velocity vector angle ratio*. These computations differ for a leaf box and an internal box:

- If  $B$  is a leaf box, these measures are computed from the velocities of its particles. That is, the relative speed of its particles and also their relative angle should be within certain ratios.
- Otherwise, if  $B$  has children, the similarity measures for  $B$  are computed by considering the COM particles of its children. Again the relative speed of the COM particles and their relative angle are constrained to be within certain ratios.

The above computations are done during the upward pass of the FMM algorithm. If box  $B$  satisfies these two conditions, then all of its particles are replaced by its weighted center of mass particle which is computed directly from its particles or indirectly from the center of mass of its children. After this simplification, the new algorithm continues as the classical

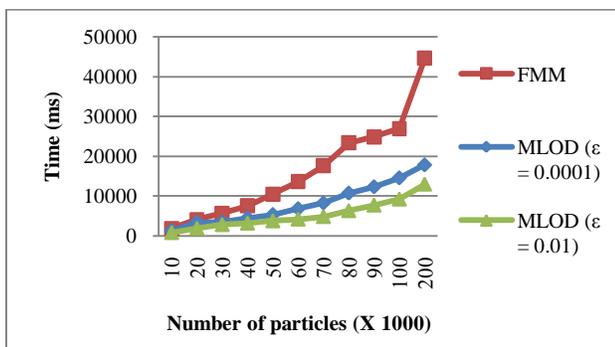
FMM. However, when the potentials have been computed for every particle (a real particle or a weighted center of mass particle), the computed potentials for a CoM particle are applied to all of the particles inside the corresponding simplified box. A snapshot of the developed system for 200 particles is given in Figure 3. Because of dynamics simplification, the number of particles is reduced from 200 to 63 in this system resulting in a considerable reduction in the execution time.



**Figure 3** A snapshot of the developed system with 200 agents. Each colored box represents a box whose particles dynamics are simplified using dynamics of its CoM particle (the red particle in it).

### 4.3 Real-time simulation

During the simulation, it is possible to monitor the execution time of the last few steps and compare it to the target execution time defined by the user. If the execution speed is too slow, the system performs more simplifications by enlarging the *maximum cluster size*. This is achieved by allowing greater boxes (those that are located at higher levels in the hierarchical tree) to be merged. The inverse occurs when the speed is more than necessary. This is achieved by introducing a new parameter called *simplification level*. For a given simplification level, only boxes below that level are allowed to be merged. A small value for this parameter means that boxes at higher levels are allowed to be merged and a large value means that only small boxes at lower levels are allowed to be merged. Therefore, each time we need a faster execution, the simplification level is decreased and this allows larger boxes in higher levels to be merged. Increasing this parameter does the inverse.



**Figure 4** Comparing execution time of the original FMM to the new algorithm.

In summary, the proposed framework has several parameters which enable us to control both the error and the execution time of the simulation. These parameters are: the speed ratio and the velocity vector angle ratio, the maximum number of particles in each box and the maximum cluster size or the simplification level. The set of these parameters gives us the potential to apply our framework for real-time applications as well. Future works are already planned to test this real-time capability of the proposed framework.

## 5. EXPERIMENTAL RESULTS

To test the proposed method, we have developed a prototype system in java with the ability to generate motion levels of details, select the appropriate ones and switch between them. This section presents the results of the new method tested on the particles in a Coulombic system, which is a common example for *N*-body systems. All experiments in this section were conducted on a desktop computer configured with one 2.5GHz Quad-Core Intel Pentium processors and 4GB RAM running a Windows 7 Professional x32 Edition. Our application was developed in Java and the Java VM used to execute the tests was configured with 1GB memory.

### 5.1 Simulation of a large dynamical system

The goal of this experiment is to compare the execution time of the proposed method to the original FMM, while trying to keep the error below the user-specified level. For each experiment, the appropriate values for similarity measures are determined by try and error. Also, all the results are computed by averaging the execution times over 100 consecutive runs. To keep the results comparable, we have fixed the truncation number *p* to 10 and the clustering size *s* to 100 in all experiments. The results are shown in Figure 4.

As it can be seen in Figure 4, the experimental results confirm the theoretical expectations, i.e., more computational time can be saved if the algorithm is allowed to be run less accurately. Of course, it is possible to achieve more efficiency by relaxing accuracy requirements more. This trade-off between accuracy and efficiency enables us to cope with very large and complex systems which would be intractable otherwise.

For a better understanding of how adding motion levels of detail can affect efficiency, we can introduce the *effective number of particles* for the proposed method. For a given amount of execution time, this parameter specifies that how much particles can be processed by the two methods. For example, consider the case in which number of particles is equal to 200000 for the new method. In this case, the execution time is about 13 seconds which is equal to the required time by the original FMM when there are approximately 60000 particles. Therefore, for 200000 particles, the effective number of particles is 60000 which is less than one-third of the original size. Table 1 reports the effective number of particles in the new FMM algorithm for various numbers of particles. All the given numbers in this table are computed approximately using interpolation.

**Table 1** Effective number of particles in the proposed method

Particle No. (× 1000)	Effective Number of Particles (× 1000)	
	$\epsilon = 10^{-2}$	$\epsilon = 10^{-4}$
10	3.5	5
20	10	15
30	12	17
40	15	21.5
50	16	32
60	20	35
70	23	44
80	33	50
90	40	52
100	49	56
200	60	72

## 6. CONCLUSIONS

This paper presents a new framework for simulation of large particle systems by combining the well-known fast multipole method from computational physics with the concept of motion levels of detail from computer graphics. The main goal is to reduce the computational complexity of the FMM by adding automatic dynamics simplification to it and hence to speed up the simulation of large  $N$ -body. Dynamics simplification is achieved by first replacing all particles inside a box with one weighted center of mass particle and then applying the potential computed for the CoM particle to all particles in that box.

This recursive dynamics simplification can be performed for any box at any level. This provides us a hierarchy of approximated motion models which are updated at each simulation cycle. These motion levels of detail enable us to adjust the level of simulation dynamically and hence to trade accuracy for efficiency. The results given in section 5 are very promising indicating the potential use of the method to more dynamical systems. However, this work is a first step toward the design of a general multilevel simulation framework and can be extended in several ways:

- Investigating other possible methods or heuristics which can be served as the similarity measure.
- Designing new experiments to gain a better understanding of performance of the system and giving a more complete complexity and error analysis by studying the effects of main parameters such as truncation number on the complexity and the error of the new method.
- Extending the proposed method for real-time situations (see section 4.3).

## REFERENCES

[1] V. Rokhlin, "Rapid solution of integral equations of classical potential theory," *Journal of computational physics*, pp. 187-207, 1983.

[2] Leslie Greengard and Vladimir Rokhlin, "A fast algorithm for particle simulations," *Journal of Computational Physics*, vol. 73, no. 2, pp. 325-348, 1987.

[3] Leslie Greengard and Vladimir Rokhlin, "Rapid evaluation of potential fields in three dimensions," in *Lecture notes in*

*mathematics*, vol. 1360, Berlin, 1988, pp. 121-141.

[4] J. Dongarra and F. Sullivan, "The top ten algorithms of the century," vol. 2, no. 1, pp. 22-23, 2000.

[5] N. Razavi, N. Gaud, N. Mozayani, and A. Koukam, "Multi-agent based simulations using fast multipole method: application to large scale simulations of flocking dynamical systems," *Artificial Intelligence Review*, vol. 35, no. 1, pp. 53-72, 2011.

[6] D. Carlson and J. Hodgins, "Simulation levels of detail for real-time animation," in *Graphic Interface 1997*, 1997.

[7] S. Cheney and D. Forsyth, "View-dependent culling of dynamic systems in virtual environments," in *ACM Symposium on Interactive 3D Graphics*, 1997.

[8] R. Grzeszczuk, D. Terzopoulos, and G. Hinton, "Neuroanimator: Fast neural network emulation and control of physics-based models," in *SIGGRAPH*, 1998, pp. 9-29.

[9] Z. Popovic and A. Witkin, "Physically based motion transformation," in *SIGGRAPH 1999*, 1999, pp. 11-20.

[10] A. Brudlerlin and T. Calvert, "Goal-directed, dynamic animation of human walking," in *Computer Graphics (Proc. of SIGGRAPH'89)*, 1989, pp. 233-242.

[11] A. Brudlerlin and T. Calvert, "Knowledge-driven, interactive animation of human running," in *Graphics Interface*, 1996, pp. 213-221.

[12] M. Girard and A. Maciejewski, "Computational modeling for computer animation of legged figures," in *Computer Graphics (Proc. of SIGGRAPH)*, vol. 19, 1985, pp. 263-270.

[13] J. Granieri, J. Crabtree, and N. Badler, "Production and playback of human figure motion for 3d virtual environments," in *VRAIS*, 1995, pp. 127-135.

[14] k. Perlin, "Real time responsive animation with personality," in *IEEE Transactions on Visualization and Computer Graphics*, 1995, pp. 5-15.

[15] F. Multon, B. Valton, B. Jouin, and R. Cozot, "Motion levels of detail for real-time virtual worlds," in *ASTC-VR'99*, 1999.

[16] P. Faloutsos, M. van de Panne, and D. Terzopoulos, "Composable controllers for physics-based character animation," in *SIGGRAPH 2001*, 2001, pp. 251-260.

[17] C. O'Sullivan and J. Dingliana, "Collisions and perception," in *ACM Transactions on Graphics*, 2001.

[18] D. O'Brien, S. Fisher, and M. Lin, "Automatic simplification of particle system dynamics," in *Computer Animation*, p. 2001.

[19] H Cheng, Leslie Greengard, and Vladimir Rokhlin, "A fast adaptive multipole algorithm in three dimensions," *Journal of Computational Physics*, vol. 155, pp. 468-498, 1999.

# Statistical Quality Control of Microarray Gene Expression Data

Shen LU

University of Arkansas at Little Rock, Department of Computer Science,  
Donaghey College of Engineering and Information Technology, Little Rock, AR 72204 USA

and

Richard S. SEGALL

Arkansas State University, Department of Computer & Information Technology,  
College of Business, State University, AR 72467-0130 USA

## ABSTRACT

This paper is about how to control the quality of microarray expression data. Since gene-expression microarrays have become almost as widely used as measurement tools in biological research, we survey microarray experimental data to see possibilities and problems to control microarray expression data. We use both variable measure and attribute measure to visualize microarray expression data. According to the attribute data's structure, we use control charts to visualize fold change and t-test attributes in order to find the root causes. Then, we build data mining prediction models to evaluate the output. According to the accuracy of the prediction model, we can prove control charts can effectively visualize root causes.

**Keywords:** statistical quality control, microarray, information product, control charts, root causes

## 1. INTRODUCTION

In microarray experiments, we compare patterns of expression across multiple samples hybridized to a particular array in order to discover new genes that, as drug targets, are responsible for a disease. For gene discovery and prediction, the quality of the data set usually affects the result of the experiment.

Microarray technology has found its applications in recent years in many fields of life science. Generally speaking, all the data analysis behind these applications can be characterized into two major categories: discovery and prediction. Discovery is to discover new knowledge, new genes involved in a pathway; prediction is to create predictive models to be used in such areas as toxicology and disease diagnosis. Fundamental to both discovery and prediction is the selection of genes that are differentially expressed (up or down) when comparing the samples of your interest to the control group.

Lu and Segall [11, 12] present preliminary research on medical record linkage and entity resolution methods as applied to bioinformatics. Segall [14] presented a chapter on data mining of microarray databases for biotechnology. Segall [14] performed data visualization and data mining of microarray databases for continuous numerical-valued Abalone fish data and discrete nominal-valued mushroom

data using evolutionary algorithms specifically for neural networks and genetic algorithms. Segall [15] performed data mining of microarray databases for human lung cancer. Segall [16], Segall [17], and Segall [18] performed data mining of microarray databases of Leukemia cells using single SOM. This paper extends the methodology used in authors' previous research for microarray data analysis to those using statistical quality control techniques.

## 2. PHASE I: MEASURING THE QUALITY ATTRIBUTES

For two-color microarray experiments, as shown in Figure 1, one must decide what the most appropriate comparison is to be made with each array hybridization. The simplest comparisons can be separated into four general classes, such as direct comparison, reference design, balanced block design and loop design. In many ways, direct comparisons are the simplest conceptually; they are used when two distinct classes of experimental samples are to be compared, such as a treated sample and its untreated control. On each array, representatives of the two classes are paired and co-hybridized together such that the relative expression levels are measured directly on each array. The choice of appropriate pairing depends on the experimental question under study. For example, one can pair diseased and normal tissue from the same patient or randomly select animals from mutual and wild-type groups. The strategy to collect data for any given case is influenced by a wide range of factors, including the availability of samples, the quantity of RNA that can be obtained, the size of the study, and the logistical constraints in the laboratory.

For each gene, the process begins with defining an expression vector that represents its location in expression space. In this view of gene expression, each hybridization represents a separate distinct axis in space, and the  $\log_2(\text{ratio})$  measured for that gene in that particular hybridization represents its geometric coordinate. In this way, expression data can be represented in m-dimensional expression space, where m is the number of hybridizations and where each gene expression vector is represented as a single point in that space. It should be noted that one could use a similar approach to representing each hybridization assay using a sample vector consisting of the expression

values for each gene; these define a sample space whose dimension is equal to the number of genes assayed in each array.

We collect Microarray experimental data and to see possibilities and problems about whether the data are sufficient and can be used to generate, evaluate, and improve the cancer-related prediction model and about whether the data can be used to select the proper pre-processing and modeling techniques. Several different data sets are considered. According to Babu [1], a microarray is typically a glass slide onto which DNA molecules are placed as spots. A microarray may contain thousands of spots and each spot may contain a few million copies of identical DNA molecules that uniquely correspond to a gene as shown in part A of Figure 1 from Babu [1].

For liver cancer [3], there are 17,400 genes and 179 samples, for lung cancer [2], there are 12,600 genes and 245 samples, for NIH cancer dataset [3], 12,196 genes and 240 samples, for prostate cancer [10], there are 26,260 genes and 103 samples. Most of Microarray data have a small size of samples in which the number of genes is large. Obviously, in comparison with the number of genes, we can make such a conclusion that most of Microarray experiments can neither supply enough samples to do statistical analysis, nor generate a prediction model. A wide range of methods for microarray data analysis have evolved, ranging from simple fold-change approaches to testing for differential expression, to many computationally demanding and complex techniques. In this paper, in order to control the quality of Microarray experimental data, we generate such a process that we collect Microarray experimental data, check the quality of the data, remove noise, and build a prediction model to evaluate the output.

Other than expression ratio value, we also use fold change and t-test as attribute measures. Fold change and t-test can be used to identify best distinguish genes between the sample classes.

The student's t-test can be used to test whether a difference is significant, which is an assessment of signal-to-noise ratio for the particular gene in question.

$$t = (\text{signal} / \text{noise}) =$$

$$(\text{difference between groups} / \text{variability of groups})$$

A large value for the t statistics indicates that the populations representing measurements of a gene for condition A and B are well separated. It can be used to estimate how likely that a gene is differentially expressed between conditions.

Fold change is a mathematical operation describing how much two variables differs. It is the ratio of the final value and initial value (B/A), if the final value is larger.

For t-test, a p-value is normally calculated to quantify the significance. And the most common interpretation for a p-value of 0.05 is that there is a 5% probability that the observed difference in expression may simply due to

chance, not independent. Calculating a fold change is straightforward, although one does have to decide which of the three methods to use to calculate an average (arithmetic, geometric, harmonic).

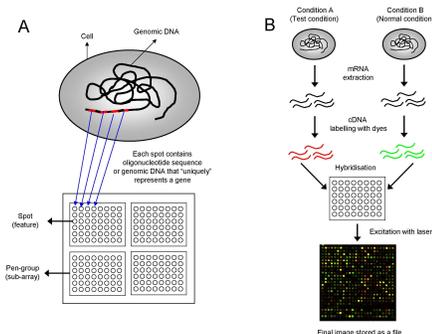


Figure 1. Illustration of a microarray that may contain thousands of “spots” of genomic data [1]

### 3. PHASE II: ANALYZE THE INFORMATION PRODUCT

Statistical process for microarray expression data includes the following steps:

1. Pre-processing: because of experimental errors, some values of expression data are missing. We use K Nearest Neighbor (KNN) algorithm to automatically impute missing values first.
2. Sample selection: since microarray expression data set is not very big, we can use total data for any experiments and applications. However, regarding to the different number of treated samples and untreated samples, we randomly generate data sets in which both treated and untreated classes have the same number of samples.
3. Feature selection: even if data mining analysis can be performed, it is still extremely useful to reduce the data set to those genes that are best distinguished between the sample classes.

Before statistical analysis of microarray expression data, we have to decide the attribute data which can be used to analyze root causes. For an information system, there are two types of data: one is variable data; the other is attribute data. Variable data can be measured, primarily continuous in nature. Attribute data are observed to be either present or absent, conforming or non-conforming. The effectiveness of charts depends on the attribute data's structure. The problem is how to choose attribute data. The critical is if the categories of non-conforming are sufficiently focused, so that there is likely to be only one assignable cause per category. For microarray expression data, since the difference and independence of difference samples are significant, we use fold change and t-test to measure the quality of the sample. For microarray expression data, since it belongs to binomial distribution and each gene is tested in the same number of experiments, we use np chart to visualize the fold change and pValue.

Our work is based on liver cancer [3] data. We choose total data and balanced data to generate sample sets. For fold

change equal to 1.5 or 2.0, and pValue equal to 0.05 and 0.01, we generate different data sets, as shown in Table 1. For each data set, we use np chart and moving range chart to visualize fold change and pValue, as shown in Figure 2 and Figure 3.

	Fold change	pValue
Total data	1.5	0.01
	1.5	0.05
	2.0	0.01
	2.0	0.05
Balanced data	1.5	0.01
	1.5	0.05
	2.0	0.01
	2.0	0.05

Table 1. Data Sets with Difference Values of Fold Change and pValue

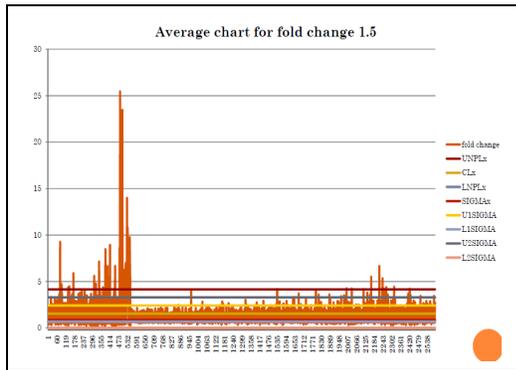


Figure 2. NP Chart for Total Data with Fold Change Set to 1.5

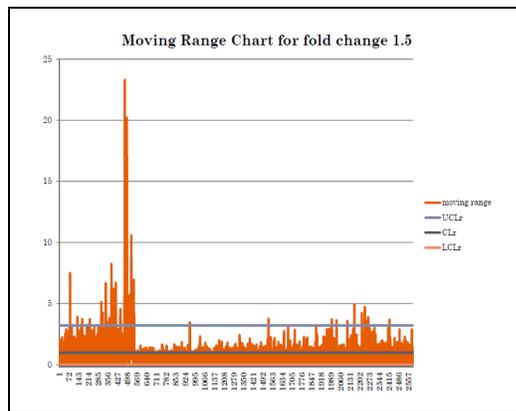


Figure 3. Range Chart for Random Data

In Total Quality Management (TQM), an organization would follow certain guidelines to scope an IQ project, identify critical issues, and develop procedures and metrics for continuous analysis and improvement [19]. Control charts provide us an easy way to compare the observed subgroup averages and subgroup ranges against the predicted limits. As shown in Figures 2, 3, we use upper and low control limits and 1 sigma and 2 sigma control limits to analyze the root cause. A sigma unit is a measure of scale for the data. Roughly 60% to 75% of the data will be located within a distance of one sigma unit on either side of the average. Usually 90% to 98% of the data will be located within a distance of two sigma units on either side

of the average. Approximately 99% to 100% of the data will be located within a distance of three sigma units on either side of the average. Figures 4(a)-4(d) show us the relationship between empirical rules and control charts. Any point outside of the upper and lower control limits is a clear example of a special cause variation. The other forms of special cause variation are called runs. Trends are special forms of a run. According to the average chart, we can see some data are out of upper and lower 2 sigma limits, and even out of upper and lower control limits, they are definitely out of control. That means those observations are not consistent with predictions and can make the process unstable. Since the average chart and range chart are for root cause analysis, out of range data are special causes of variation, we need to take actions to identify and remove them.

Rule 1: Any point beyond Zone A

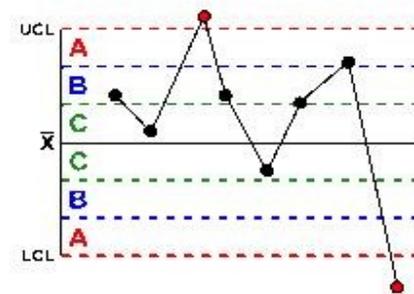


Figure 4(a). Interpretation of Western Electronic Rule 1 [20]

Rule 2: two out of three consecutive points fall Zone A or beyond

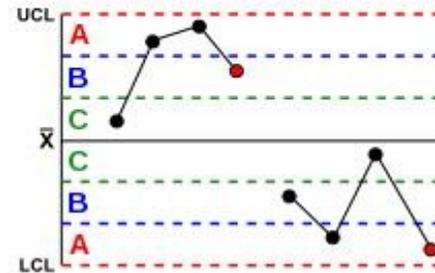


Figure 4(b). Interpretation of Western Electronic Rule 2 [20]

Rule 3: Four out of five consecutive points fall Zone B or beyond

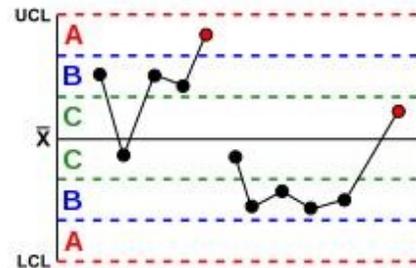


Figure 4(c). Interpretation of Western Electronic Rule 3 [20]

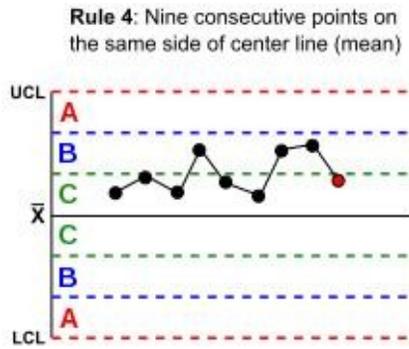


Figure 4(d). Interpretation of Western Electronic Rule 4 [20]

Table 2 lists the number of genes and the percentage of deletion by using fold change and t-test feature selection. In next section, we use prediction model to evaluate the output after removing special causes.

dataset	sample	genes	removing pe	total orde	sub orde
1	total	179	19536		
2	total_pvalue005	179	6660	0.6590909	9 24 8
3	total_pvalue001	179	4383	0.7756449	6 3 18 6
4	total_fc20	179	2717	0.8609234	3 12 4
5	total_fc20_pvalue005	179	1270	0.9349918	1 6 2
6	total_fc20_pvalue001	179	772	0.9604832	1 3 1
7	total_fc15	179	5793	0.7033681	4 1 19 7
8	total_fc15_pvalue005	179	2768	0.8583128	3 8 14 5
9	total_fc15_pvalue001	179	2181	0.8883599	3 1 9 3
10	balance3	50	19536		
11	balance3_pvalue005	50	5864	0.6998362	2 1 7
12	balance3_pvalue001	50	3939	0.7983722	3 6 16 6
13	balance3_fc20	50	2785	0.8574426	7 15 5
14	balance3_fc20_pvalue005	50	886	0.9546478	3 4 2
15	balance3_fc20_pvalue001	50	743	0.9619676	9 2 1
16	balance3_fc15	50	5900	0.6979934	4 8 22 8
17	balance3_fc15_pvalue005	50	2403	0.8769963	4 10 4
18	balance3_fc15_pvalue001	50	2135	0.8907145	7 3
19	balance2	50	19536		
20	balance2_pvalue005	50	5946	0.6955364	4 6 23 8
21	balance2_pvalue001	50	4006	0.7949426	7 17 6
22	balance2_fc20	50	2739	0.8597972	7 13 5
23	balance2_fc20_pvalue005	50	898	0.9540335	9 5 2
24	balance2_fc20_pvalue001	50	739	0.9621724	1 1 1
25	balance2_fc15	50	5851	0.7005016	3 8 20 7
26	balance2_fc15_pvalue005	50	2422	0.8760237	1 11 4
27	balance2_fc15_pvalue001	50	2148	0.8900491	4 8 3

Table 2. The Number of Genes Present after Sample Selection and Feature Selection

#### 4. PHASE III: IMPROVE THE INFORMATION PRODUCT

After generating different data sets, as the output of the process, we use data mining analysis to evaluate them. In TQM, knowledge has been created for Information Quality (IQ) practice [5, 6]. Precision model building includes two steps: model building and model validation. Model building involves in training data selection. Model validation involves in testing the built model with testing samples and measuring the precision and recall of the output of the generated model.

We use K-Nearest Neighbor (KNN) [4], Random Forest (RF) [7], Multipass-LVQ (MPL) [8], and Self-Organizing Map (SOM) [9] algorithms to calculate the precision and

recall on different data sets. KNN is based on the direct comparison of the distance between two neighbors. This algorithm is good for high dimensional vectors. Random Forest is based on decision tree theory. Since the best features are selected to build decision trees, the significance of different features are considered in this algorithm. Multipass-LVQ and SOM belong to neural network algorithm. Since samples can be randomly selected as input for many times, these algorithms are good for high-dimensional small size data sets, such as microarray expression data. Precision and recall of these algorithms on different data sets are shown in Figure 5(a) to 5(g) below.

In the below Figures 5(a)-5(e) and 5(g), KNN is used for #1-4, RF for #5-8, SOM for #9-12, and MPL for #13-16. In Figures 5(f) and 5(h), SOM is used for #9-16, and MPL for #17-20.

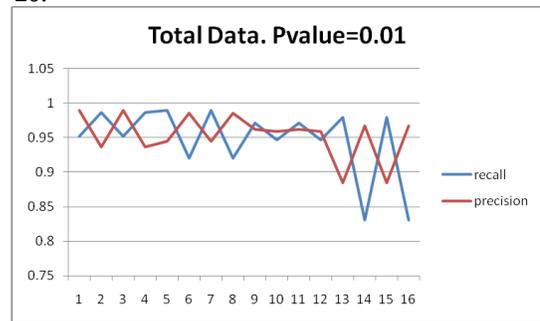


Figure 5(a)

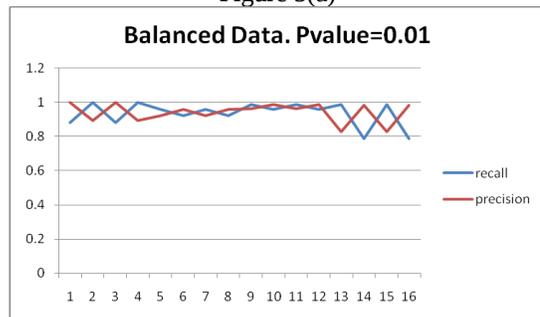


Figure 5(b)

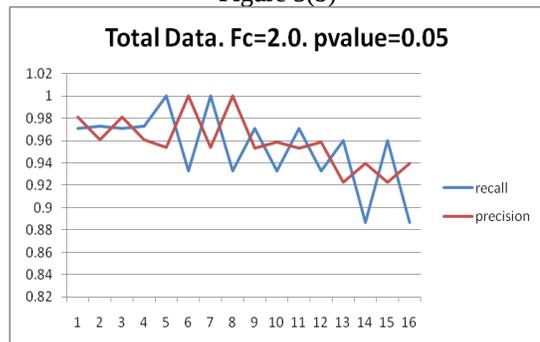


Figure 5(c)

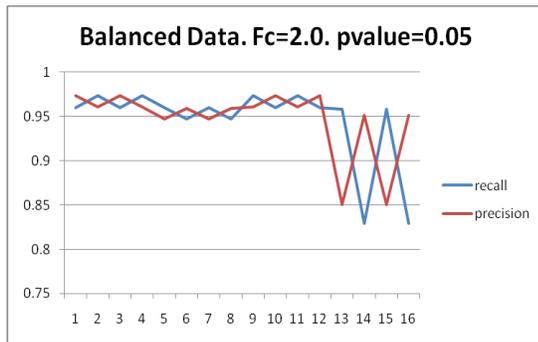


Figure 5(d)

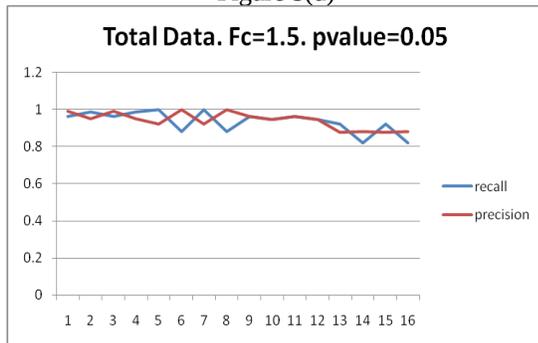


Figure 5(e)

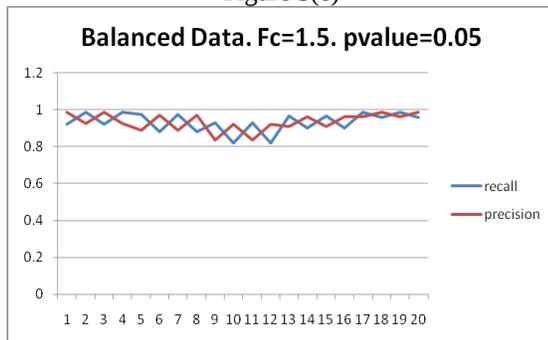


Figure 5(f)

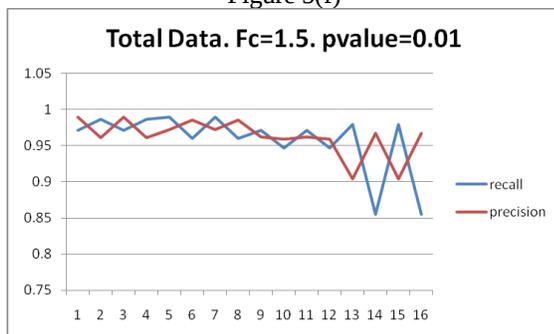


Figure 5(g)

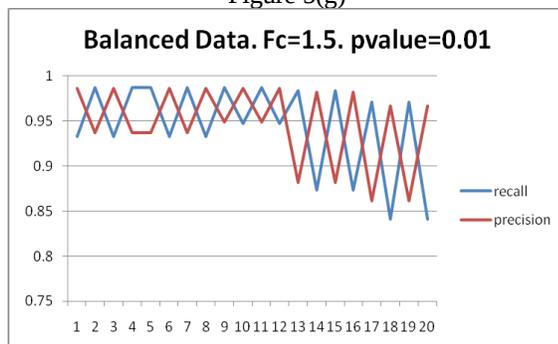


Figure 5(h)

## 5. CONCLUSIONS

This paper introduced the significance of data quality control in microarray experiments. According to different microarray comparisons, we collected data in different ways. A formal method was given to measure the possibility and problems about whether data are sufficient and can be used to generate, evaluate, and improve prediction model. We used T-test and fold change to select samples and genes, and used control charts to visualize the quality of the output. Four data mining algorithms, such as KNN, SOM, Random Forest, Multipass-LVQ, were used to build prediction models and to evaluate the quality of the data. The performance of the output showed us control charts are useful for the visualization of the root cause variation of the data. Selection of appropriate charts to visualize the output is very important for data quality control. Empirical root cause rules and analysis can be used to explain control charts and ensure control charts will yield very few false conclusions.

## 6. REFERENCES

- [1] M. M. Babu, "An Introduction to Microarray Data Analysis", Chapter 11 in **Computational Genomics** by R. Grant, Editor, Horizon Press, UK, 2004, pp. 225-249, <http://www.mrclmb.cam.ac.uk/genomes/madanm/microarray/chapter-final.pdf>
- [2] L. Breiman, "Random Forests", **Machine Learning** Volume 45, Number 1, 2001, pp. 5-32.
- [3] X. Chen, S. T. Cheung, S. So, S. T. Fan, C. Barry, J. Higgins, K. M. Lai, J. Ji, S. Dudoit, I.O. Ng, et al., "Gene Expression Patterns in Human Liver Cancers", **Molecular Biology of the Cell**. Vol. 13, Issue 6, pp. 1929-1939, June 2002.
- [4] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification." **IEEE Transactions of Information Theory**, Volume IT-13, Number 1, 1967, pp. 21-27.
- [5] P. Cykana, A. Paul, and M. Stern, "DoD Guidelines on Data Quality Management", **Proceedings of the 1996 Conference on Information Quality**, Cambridge, Mass., 1996, pp. 154-171.
- [6] C.P. Firth and R.Y. Wang, **Data Quality Systems: Evaluation and Implementation**, Cambridge Market Intelligence Ltd., London, UK 1996.
- [7] G. J. Gordon et al., "Translation of Microarray Data into Clinically Relevant Cancer Diagnostic Tests Using Gene Expression Ratios in Lung Cancer and Mesothelioma", **Cancer Research**, Vol. 62, No. 17, pp. 4963-4967, September 2002.
- [8] T. K. Kohonen, J. Kangas, J. Laaksonen, and K. Tokkola, **LVQ-PAK: The Learning Vector Quantization Program Package**, Helsinki University of Technology,

Finland, 1992.

[9] T. Kohonen, "The Self-Organizing Map", **Proceedings of the IEEE**, Volume 78, Issue. 9, 1990, pp. 1464-1480.

[10] J. Lapointe, C. Li, M. van de Rijn, J. P. Huggins, E. Bair, et al., "Gene Expression Profiling Identifies Clinically Relevant Subtypes of Prostate Cancer", **Proceeding of the National Academy of Sciences**. Vol. 101, No.3, pp. 3811-816, January 2002.

[11] S. Lu and R. S. Segall and S. Atiff, "Using Entity Resolution to Discover Scientific Numbers," Poster Abstract in **Proceedings of 15<sup>th</sup> International Conference on Information Quality (ICIQ-2010)**, University of Arkansas at Little Rock (UALR), Little Rock, AR, November 10-12, 2010.

[12] S. Lu and R. S. Segall, "Linkage in Medical Records and Bioinformatics Data", submitted for publication to **Proceedings of 2011 MidSouth Computational Biology and Bioinformatics Society (MCBIOS), Special Issue of MBC Bioinformatics**, College Station, TX, April 1-2, 2011.

[13] A. Rosenwald, G. Wright, W. C. Chan, et al., "The Use of Molecular Profiling to Predict Survival after Chemotherapy for Diffuse Large-B-cell Lymphoma", **The New England Journal of Medicine**. Vol. 346, pp. 1937-1947, June 2002.

[14] R. S. Segall, "Data Mining of Microarray Databases for Biotechnology," **Encyclopedia of Data Warehousing and Mining**, Edited by John Wang, Montclair State Uni-

versity, USA; Idea Group Inc., 2006, ISBN 1-59140-557-2, pp.734-739.

[15] R. S. Segall, and Q. Zhang, "Data Mining of Microarray Databases for Human Lung Cancer", **Proceedings of the Thirty-eighth Annual Conference of the Southwest Decision Sciences Institute**, vol. 38, no. 1, March 15-17, 2007, San Diego, CA.

[16] R. S. Segall and R. M. Pierce, "Data Mining of Leukemia Cells using Self-Organized Maps," **Proceedings of 2009 Conference on Applied Research in Information Technology**, sponsored by Acxiom Laboratory of Applied Research (ALAR), University of Central Arkansas (UCA), Conway, AR, February 13, 2009, pp. 92-98.

[17] R. S. Segall and R. M. Pierce, "Advanced Data Mining of Leukemia Cell Micro-Arrays", **Proceedings of 13<sup>th</sup> World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2009**, Orlando, FL, July 10-13, 2009.

[18] R. S. Segall and R. M. Pierce, "Advanced Data Mining of Leukemia Cell Micro-Arrays", **Journal of Systemics, Cybernetics and Informatics (JSCI)**, Volume 7, Number 6, 2009, pp.60-66.

[19] R. Wang, "A Product Perspective on Total Data Quality Control", **Communications of the ACM**, Vol. 41, No. 2, 1998, pp. 58-66.

[20] Wikipedia, Western Electric Rules, 2011, [http://en.wikipedia.org/wiki/Western\\_Electric\\_rules](http://en.wikipedia.org/wiki/Western_Electric_rules)

## Multiple measurement of physical quantity from the risk evaluation point of view

Ing. Jan Kamenicky, Ph.D., Ing. Hana Cermakova, CSc. – Technical University of Liberec, Liberec; Doc. Ing. Radovan Soušek, Ph.D. – University of Pardubice, Pardubice; Ing. Vladimír Němec, Ph.D. - Czech Technical University in Prague, Prague, Czech Republic

[Jan.Kamenicky@tul.cz](mailto:Jan.Kamenicky@tul.cz), [hana.cermakova@tul.cz](mailto:hana.cermakova@tul.cz), [radovan.sousek@upce.cz](mailto:radovan.sousek@upce.cz), [nemec@fd.cvut.cz](mailto:nemec@fd.cvut.cz)

### Abstract

Humankind operates e.g. petrochemical refineries, nuclear power plants or other relatively dangerous equipment. The main reason why people accept their processing is the benefit which these machineries make. Their dangerousness is wrongly understood by general public only as a consequence of dangerous event. But it is well known that risk is computed as combination of more factors, at least of consequences and expected number of occurrence of unwanted event. This mentioned unwanted event could be standardly understood as a failure but also e. g. as false action of safety system when its impact is not needed (so-called safe failure). The consequence of false impact is mainly “only” economical loss, resulting from process stoppage. More important impacts result from no-action state of safety system in case of its action demand, typically in case of failure or unwanted house event occurrence. Then consequences are not only economical, but there could be also human health and life exposure or environment damage danger.

Dilemma of systems related with safety, especially systems based on modern electronic components and digital signal computing, is the area which is really focused today. The quantity of international standards connected with reliability, risk and safety confirms this statement. Producers and suppliers of safety systems solve many problems related with request of dependability standards. Problems are solved in scientific way, but some of them are solved intuitively, based on “technical feeling” and experiences of designers. It means that solutions of similar problem, made by different designers, are often diverse and there is no reason why their access was chosen.

Submitted text focuses on partial problem with choosing of relevant data about physical variable related with safety, where exist more principles of solution. To be more concrete this work deals with problem of choosing relevant value of physical quantity with respect to reliability parameters and related long-term economical profitability of chosen solution.

### Key words

Safe failure, risk management, k-out-of-n system

### Acknowledgement

This research was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project No. 2B08011 BIOTRA.

### 1 Introduction

Humankind cannot solve all problems related with modelling of complex industrial processes. Risk assessment is one of such problems. People are used to take care of dangerous events, but there can be more losses caused by so-called safety failures.

Designers of sophistic industrial systems decrease the risk of disaster (among others) by multiplying of measurements of physical values. It is well known, that it is recommended to use measurements of physical parameters based on different physical principles. This can avoid occurrence of common cause failures. However, using this recommendation, there can occur many failures in each measuring channel. We have to split these failures into more than just two states - available and fault. This text works with five-state systems, see below. Moreover control systems are able to compute only with one value of physical magnitude, so it is necessary to join all information from all measuring channels into one result. First option, studied in this paper, how to do this is simply average all entering information, which are valid. Second option of making final signal is very often used “k out of n” system. The last reflected option of combining more signals to one is to use “k out of n” system with ability to validate measurement channel and reduce number of inputs due to its validity. This paper will tell us, how safety failures affect final risk of three channel measurement system.

### 2 Measurement channels configuration

This chapter describes studied methods of joining information into one final controlling signal in detail. For better orientation in problem, all options will have the same initial conditions:

- scoring algorithm has three inputs
- all measurements are independent
- scoring algorithm is hundred percent reliable
- verification of each signal is hundred percent reliable

Joining signal options are three:

- analog signal joining - average of all valid inputs
- out of 3 system
- adaptable 2 out of 3 system reducing invalid inputs

Following chapters describe important characters of joining options mentioned above.

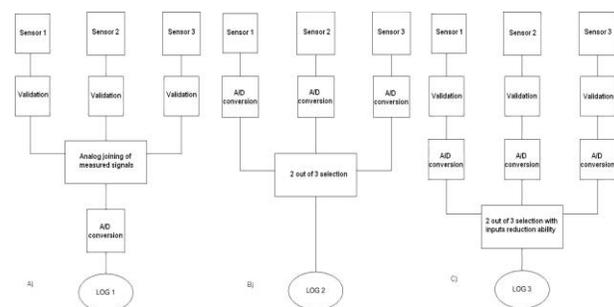


Figure 1: Three possible principles of redundant inputs joining

*Analog joining of signal*

An option, how to make one final signal from more redundant measurements of one physical value, is to combine all valid information into one result. This operation is mostly made as an average value of all inputs, but there are also different algorithms, e.g. geometrical average, median, weighted average, weighted sum of squared values etc. Resulting information is well usable as displayed value for operator or as input into the regulator. Schematic picture of figured measurement progress is shown at Figure 1 A) and we will mark this principle as LOG1 in further text.

*2-out-of-3 selection system*

Next option, how to determine one resulting value from multiple measurements, is selective principle, often used in praxis. Schematic view of logical sequence, leading to the result, is shown at Figure 1 B) and we will mark it as LOG2. In this case it is demonstrated by 2-out-of-3 system. In contrast to analog signal joining in case of selective system it is not taken into account data validity and all values are processed in selective unit. Impossibility of wrong individual measurement value recognizing is a disadvantage of this mechanism. On the other hand simple failure do not cause failure of whole measuring circuit, because this simple failure of individual measurement will be “outvoted” by the remaining useful values. Failure of whole measurement circuit will occur when at least two of individual measurements will fail simultaneously. This option seems to be improbable but one must notify that failure of one individual measurement channel will not be indicated, so the failure seems to be latent and its mean time to repair will be long.

*Adaptable 2-out-of-3 selection system reducing invalid inputs*

This option combines advantages of both previously mentioned algorithms. There is used a gate which evaluates validity of each individual input and moreover the selection system “votes” from all its inputs, so that one simple failure will not endanger resulting signal. Principle of this sketched connection is shown at Figure 1 C) and for our purpose will be indicated as LOG3. In case of one input failure state detection, smart selection system will change decision rules and control process will be modified (e.g. reduction of selection system to 1-out-of-2 or 2-out-of-2).

*Estimation of mentioned algorithms options*

There were three options of triplicate input connecting introduced above. In praxis there is in common usage a selection system 2 out of 3, which is ‘compromise’ between analog signal joining (poor reliability, single individual input failure leads to the failure of whole control system and depreciates resulting information) and technically difficult model of selection system with wrong inputs erasing option. This work justifies such access to ‘high backed up systems’ problem and compares reliability of all three logical options to make confirmation/negation of deep-seated opinion that 2 out of 3 selection system is the best of all possible ones. We will also see economical losses when using each algorithm in different industrial process.

**3 Comparison of logical signal joining possibilities**

Current praxis supposes only two possible states of individual measuring channel – able to operate system state (up state) and unable to operate system state (down state). The failure state is

commonly understood as something unwanted, potentially dangerous. This comprehension is only partially right. We are not allowed to think about all failures in the same way. Looking at the failure modes and their effects in detail we can split failure states at least on latent/obvious and dangerous/safe failures.[1] Safe failure is often called as false impact. This labeling came from safety systems, where safe failure did not lead to disaster but “only” to the false trip of production.

This chapter analyses three options of logical signal joining, described above. To start the analysis it is necessary to define complete list of possible states for each individual sensor.

- State 1 belongs to dangerous and for validation algorithm obvious failure. Its failure rate will be  $\lambda_1$ .
- State 2 belongs to dangerous and for validation algorithm latent failure. Its failure rate will be  $\lambda_2$ .
- State 3 belongs to safe and for validation algorithm obvious failure. Its failure rate will be  $\lambda_3$ .
- State 4 belongs to safe and for validation algorithm latent failure. Its failure rate will be  $\lambda_4$ .
- State 5 in this analysis is reserved for reliable (failure-free) state of sensor.

Now we have completely scheduled and indicated list of possible inputs into the decision-making algorithm. We have to mark possible outputs of resulting algorithm. Outputs will be similar marked for every logical operation (LOG1, LOG2 and LOG3). Outputs list is following:

- D - dangerous failure state (failure leads to non-operating system state)
- F - safe failure state (failure leads to false system impact )
- R - reliable system state (failure free)

Logic operations with signals are described in chapters before, but for better orientation let’s repeat that:

- LOG1 is a signal analog joining, in our case averaging
- LOG2 is a selection system, in our case 2 out of 3
- LOG3 is a selection system (2/3) with ability to reduce the number of inputs

To obtain complete list of states for each logical evaluating algorithm it is needful to set table of all possible states of each individual channel, recording to division of failures on latent/obvious and dangerous/safe. For five possible states and three individual channels we obtain table of 125 rows, each independent to another, which is impossible to be shown at this place. Following table shows only first five states.

Table 1: Example of output-states table

ID No.	S1	S2	S3	LOG1	LOG2	LOG3
1	1	1	1	D	D	D
2	1	1	2	D	D	D
3	1	1	3	D	D	D
4	1	1	4	F	D	F
5	1	1	5	R	D	R

If we need to decide about profitability of algorithm, we have to

know at least approximate values of failure rates and repair rates (in all cases - latent/obvious and dangerous/safety failure state of an item). If we want to think of latent and obvious failures (meaning for verification algorithm), we have to vary the rate between frequencies of their appearance, which are also known as failure rates. It is not necessary to know exact values of failure and repair rates for our purpose; it is enough to know their occurrence ratio.

As shown above, results of unavailability calculations for the comparing model do not depend on total failure rate. So we are allowed to select "some" failure rate, let's say 2.10-5h-1. Now we have to select also parameters of restoration. Let's say that repair rate of latent failure will be 4000h and repair rate of obvious failure will be 4h. This difference is because of the discovery time of the failure. We can presuppose scheduled every-year test of equipment, so mean time to discovery of latent failure will be one half of year (approximately 4000h).

On the other hand mean time to restoration of function in case of obvious failure will be one half of work shift, it is 4h. These are the initial conditions of our computations. Now we have to vary ratio between latent and obvious failure and calculate unavailability and mean time between (safety) failures for each case. Tab. 2 - 4 shows beginning of such a table. There are three similar columns, which are important for the comparing model. These columns are marked LOG1, LOG2 and LOG3, depending on the algorithm, which they belong to. Each column contains three information:

- Logical output state (see Tab. 1)
- Analyzed state unavailability
- Failure rate of analyzed state false impacts (safety failures)

Table 2: Unavailability and mean time between false impacts calculated for algorithm LOG1

ID No.	State	U	U.m [h-1]
1	D	8,00E-18	
2	D	8,00E-34	
3	D	8,00E-18	
4	F		4,00E-34
5	R		

Table 3: Unavailability and mean time between false impacts calculated for algorithm LOG2

ID No.	State	U	U.m [h-1]
1	D	8,00E-18	
2	D	8,00E-34	
3	D	8,00E-18	
4	D	8,00E-34	
5	D	4,00E-12	

Table 4: Unavailability and mean time between false impacts calculated for algorithm LOG3

ID No.	State	U	U.m [h-1]
1	D	8,00E-18	
2	D	8,00E-34	
3	D	8,00E-18	
4	F		4,00E-34
5	R		

Calculations in Tab. 2-4 derive benefit from Schneeweiss

formula. Now we have calculated partial unavailabilities and mean times between false impacts for every combinations of individual measurement channel state. When summarizing all values in each column we obtain total values for wanted reliability parameters. Next step in our comparison process is to repeat calculation, described above, over and over again for different ratio between failure rates of latent and obvious failures. As result we will get Tab. 5-7.

Table 5: Results of unavailability and mean time between false impacts calculations for the LOG1 algorithm

LOG1		
U [1]	T [h]	$\lambda_{latent}/\lambda_{obvious}$
1,20E-06	2,50E+09	1,00E-13
1,20E-05	2,50E+08	1,00E-04
1,20E-04	2,50E+07	1,00E-03
1,20E-03	2,50E+06	1,00E-02
1,20E-02	2,60E+05	1,10E-01
5,90E-02	5,60E+04	1,00E+00
1,10E-01	3,40E+04	9,00E+00
1,20E-01	3,10E+04	9,90E+01
1,20E-01	3,10E+04	1,00E+03

Table 6: Results of unavailability and mean time between false impacts calculations for the LOG2 algorithm

LOG2		
U [1]	T [h]	$\lambda_{latent}/\lambda_{obvious}$
4,90E-09	4,10E+08	1,00E-13
5,80E-09	3,80E+08	1,00E-04
1,90E-08	2,10E+08	1,00E-03
5,80E-07	3,80E+07	1,00E-02
4,90E-05	4,10E+06	1,10E-01
1,20E-03	8,20E+05	1,00E+00
3,80E-03	4,50E+05	9,00E+00
4,60E-03	4,00E+05	9,90E+01
4,70E-03	4,00E+05	1,00E+03

Table 7: Results of unavailability and mean time between false impacts calculations for the LOG3 algorithm

LOG3		
U [1]	T [h]	$\lambda_{latent}/\lambda_{obvious}$
1,90E-10	2,40E+14	1,00E-13
2,00E-09	2,40E+14	1,00E-04
2,40E-08	1,60E+13	1,00E-03
6,70E-07	3,40E+11	1,00E-02
5,00E-05	3,80E+09	1,10E-01
1,20E-03	3,90E+07	1,00E+00
3,80E-03	1,60E+06	9,00E+00
4,60E-03	4,00E+05	9,90E+01
4,70E-03	4,00E+05	1,00E+03

Because of space, the mean time between safety failures is in previous table marked just as "TLOGX". Tab. 5-7 contain results in dependence on ratio of latent and obvious failures.

The dependencies of unavailability and mean time between failures on the rate between obvious and latent failure rates are shown in graphical representation better than in table shape, see Figure 2 and 3.

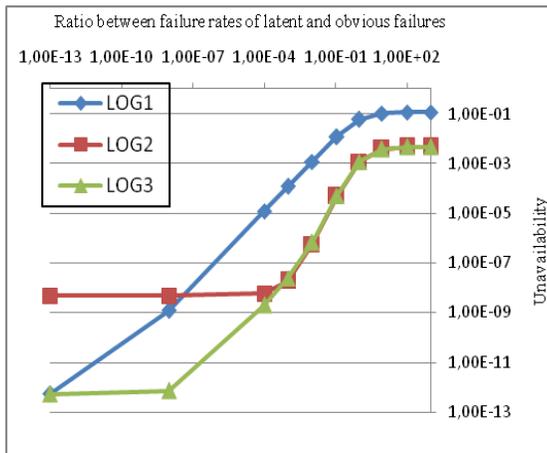


Figure 2: Graph of unavailability dependence on ratio between failure rates of latent and obvious failures

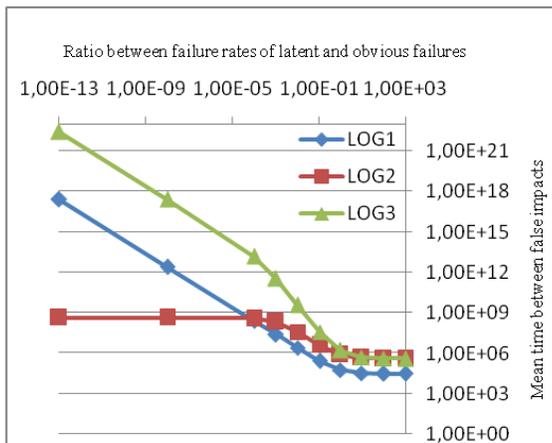


Figure 3: Graph of MTBF dependence on ratio between failure rates of latent and obvious failures

#### 4 Conclusion

The same trend in progress of MTBF and unavailability is visible from the graphs. Unavailability grows and MTBF sinks with growing majority of latent failures. This is not why we have done the work. Based on these values we can estimate total costs of such a measurement system operation in relation to the whole industrial company. Now we are able to compute total costs per hour of each measurement connection option and compare suggested algorithms on some level of ratio between latent and obvious failures. Let's say, that:

- Cost of false impact - 106 Euro
- Cost of dangerous failure - 109 Euro
- Mean time between dangerous events - 100 years

Making calculations on all levels of ratios mentioned above, we will get graph, shown at the Figure 4.

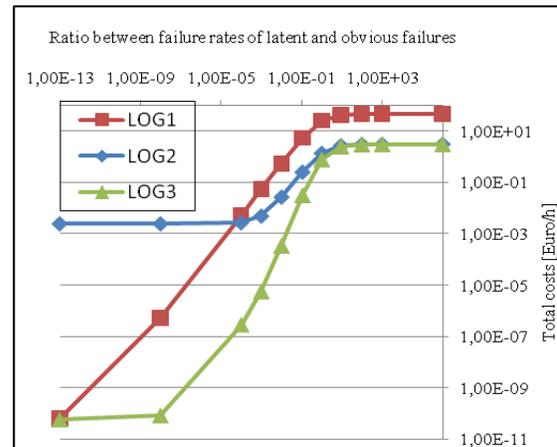


Figure 4: Comparing of total process costs

#### References

- [1] ČSN IEC 50(191) Mezinárodní elektrotechnický slovník - kapitola 191: Spolehlivost a jakost služeb
- [2] Marko M.: Validace individuálních kanálů měření, Praha, 1998
- [3] Fuchs P., Vališ, D.: Metody analýzy a řízení rizika. Liberec: Technická univerzita v Liberci 2004.
- [4] Schneeweiss W.: Teória spoľahlivosti, Bratislava, 1981
- [5] Jurčák V., Nečas P., Labuzík M., Ptáček D., Pazdziurko M.: Organizácie medzinárodného krízového manažmentu, Akadémia ozbrojených síl generála M. R. Štefánika, 2009, 234 s. ISBN 978-80-8040-387-4.
- [6] Duchoň, B., Říha, Z., Faifrová, V.: Technology, Environment, Economics, Management: 4 Factors of Engineering Education in Global World In: International Conference on Science, Technology and Education Policy. Hangzhou: Zhejiang University, 2008, p. 75-77.
- [7] Duchoň, B., Říha, Z. (ed.): Národohospodářské aspekty dopravního systému, Praha: ČVUT, Fakulta dopravní, Katedra ekonomiky a managementu v dopravě a telekomunikacích, 2007. 90 s. ISBN 978-80-01-03706-5.

# Monitoring and Debugging Distributed Autonomous Systems using Petri Nets

Joaquín López, Diego Pérez, Miguel Gayoso and Enrique Paz

Dep. Ingeniería de Sistemas y Automática,  
University of Vigo, Campus Universitario.  
C/Maxwell s/n 36200 Vigo, Spain  
{joaquin, dplosada, migayoso, epaz}@uvigo.es

## ABSTRACT

This paper describes the use of Petri nets to design, analyze, monitor, log and debug the run-time coordination of distributed control systems. The interaction between the distributed components is modeled using Petri nets. Before running the application, different Petri net analysis tools can be used to analyze the system. While running, interactions between the different components of the distributed system can be easily monitored watching the evolution of the different Petri nets. Besides monitoring, the system can be instrumented to log state changes at different levels of detail. Finally, graphical tools can be used for visualizing the evolution of the system at the same running pace, jumping to a specific point of the execution and visualizing the state of the global system.

**Keywords:** run-time verification, coordinating distributed systems, Petri nets, debugging distributed systems.

## I. INTRODUCTION

A distributed application is a set of processes executed on different machines across a network. The improvement in communication infrastructure has contributed to the development of more and more distributed systems. Debugging and verifying distributed control programs is notoriously difficult because of their distributed nature. Obtaining a snapshot of a large scale distributed system on an asynchronous communication infrastructure for complex applications might be very costly. Still, distributed control programs are becoming more and more common for complex applications that include process control [1], robot applications [2] and production plant control [3] among others.

There are various approaches to debugging distributed systems. They include: model checking, collecting and analyzing logs, replay debugging, and using virtual machines. Using model checking, a model of the program is created and then explored to verify the design specification of the distributed system and the implementation. Examples of model checking are MaceMC [4] and CrystalBall [5]. Collecting logs from different processes and analyzing them off-line to compare them with the expected behavior is used in Pip [6] and [7]. Replay debugging tools log the execution of the application nodes and replay them deterministically. In order to be able to replay the execution some tools such as liblog [8] need to log every message and application state. Debugging with virtual machines uses a debugger placed in a virtualization layer above the hardware but beneath the operating system.

One conventional approach to distributed programming [9] involves sequential processes that send and receive messages. In distributed control, these messages are mainly related to commands and events that coordinate their execution in order to carry out a defined task. In some applications such as that used in many production plant controls, one of the modules is in charge of the global coordination. For these cases, even though each process works autonomously, they will receive events and commands from the coordination module. Each individual process may be analyzed independently. Debugging existing tools can be used to debug local errors. Our approach provides Petri net based tools for designing, analyzing, logging and debugging the coordination between the different processes. This in turn determines which process is not working as expected. We propose the use of existing tools together with a coordination replay simulator to debug each individual process.

The rest of the paper is organized as follows. The next section overviews the general structure of the distributed control system. Section 3 describes the coordination

between the different control units. Section 4 presents the model checking as a first way of debugging. Section 5 explains the second way of debugging that consists on monitoring the system in execution. Section 6 describes the off-line debugging analyzing the logged data. And finally, section 7 ends the paper with the results and conclusions.

## II. OVERVIEW

In a distributed control system, the controller elements are not central in location. Rather, they are distributed throughout the system with each component sub-system controlled by one or more control units. The entire system of controllers is connected by networks for communication and monitoring.

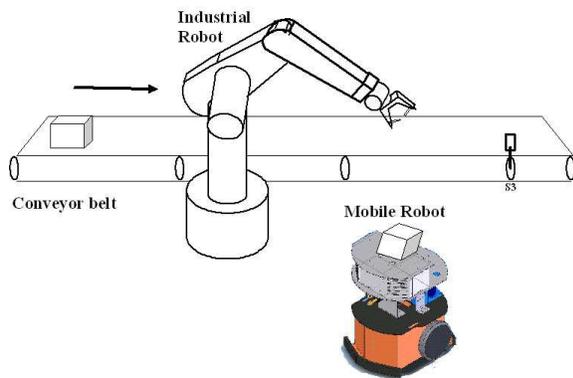


Fig. 1. Pick up cell. An industrial arm gets boxes from a conveyor belt and drops it on top of a mobile robot.

In the example on figure 1 an industrial robot picks up boxes from a conveyor belt and leaves them on top of a mobile robot. Then the mobile robot delivers the boxes throughout the building. There is a computer that controls the conveyor belt and sends signals to the industrial robot (pick up control unit). The mobile robot has its own onboard control unit. There is a third controller computer for the building automation system (BAS) that is in charge of opening and closing doors and controlling the elevators. The mobile robot needs to communicate with the BAS to request that doors be opened and closed and to manage the elevators. The pick up control unit needs to coordinate with the robot in order to wait for the mobile robot to be ready to drop the boxes.

Figure 2 shows the distributed control units and the coordination and debugging system presented in this paper. There are four different control units that autonomously control different elements of the system: the robot, the building devices, the user interface and the pick up cell (conveyor belt and industrial robot). The central server is in charge of the coordination of the distributed control units for the execution of different

tasks. There are three modules running on the central server:

- Petri net editor. This module allows the definition of tasks as Petri nets. These tasks include commands that might be executed by the four control units. We use hierarchical Petri nets where a command in one Petri net can be the execution of another Petri net. Events can be produced by any control unit and any other task (Petri net). Finally, Petri nets can be analyzed for properties such as safeness, boundedness and liveness.
- Petri net dispatcher. The tasks are executed by the dispatch that loads the Petri nets from the xml definition file. Then interprets the Petri net executing the commands according to the current marking and subscribes to events that might evolve the loaded Petri nets. The debugger can also log different information such as the evolution of the Petri nets, the commands sent and the events subscribed to at different levels of detail.
- Petri net debugger. While executing tasks, the evolution of the associated Petri nets can be monitored using this debugger. Besides, the debugger is an almost necessary tool for analyzing the logs and visualizing the system evolution at the execution running pace or jumping to a specific point.

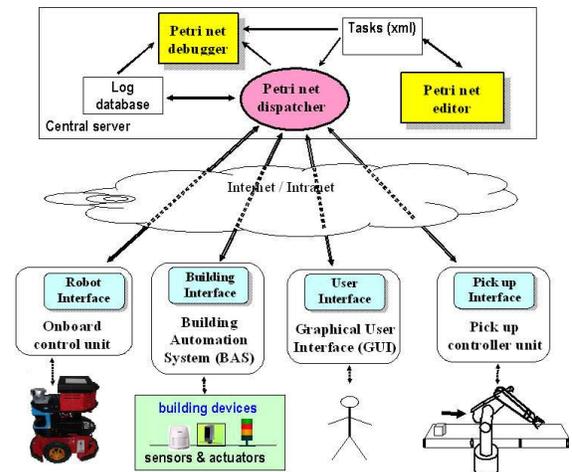


Fig. 2. Distributed system architecture. Four distributed control units are coordinated by the central server.

## III. COORDINATING THE CONTROL UNITS OF THE DISTRIBUTED SYSTEM

Petri nets have been widely used to model, design, execute and evaluate tasks in manufacturing dynamic systems. In this work we use hierarchical binary interpreted Petri nets.

As a simple example, figure 3 shows the Petri net that can be used for the “GET NEXT BOX” task. There is only one initial mark in the place labeled “INIT”, while the “END” place has been selected as a final place. The task ends when only the final places are marked. It can also end when there are no marks on the Petri net if no final places have been defined

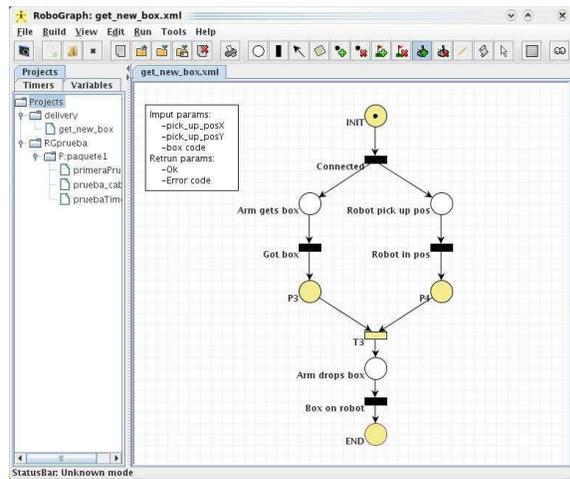


Fig. 3. RoboGraph GUI editing a Petri net that coordinates the mobile robot control unit and the pick up control unit to execute the GET NEW BOX task.

First transition (“Connected”) is fired when both control units are connected. In this case two places become marked: the “Arm gets box” commands the pick up control unit to get the next box and the “Robot pick up position” commands the mobile robot to move towards the pick up position. Both actions will execute at the same time. When the arm gets the box and is ready to release it, the corresponding event will be published and the “Got box” transition will be fired. On the other sequence, when the robot reaches the pick up position, it will publish the position reached event and the “Robot in position” transition will be fired. Transition T4 has no event associated. Therefore, when both incoming places are marked, meaning that the arm has the box and the mobile robot is in position, T4 will be fired and the arm will proceed to drop the box on top of the robot according to the command associated to “Arm drops box” place. When the box is on top of the mobile robot the transition “Box on robot” is fired. Firing this transition will remove the mark from “Arm drops box” place and sets a mark on “END”. The “END” place is the only final place; therefore, the task will finish publishing the corresponding “end task” message. Most of the commands associated to most of the places in figure 3 are messages that need to be sent to control units. However,

the “robot pick up position” place executes another Petri net that needs to coordinate the robot and the building control unit to navigate the robot from the current position to the pick up position.

Even though they are not included in figure 3, different Petri net mechanisms such as timers can be added to deal with some common problems.

In editor mode, the user can create new tasks using a simple and intuitive Petri net graphical editor. Fig. 3 shows the GUI while editing a Task. The Petri net structure is created by selecting and dragging the different elements: places, transitions, arcs and marks. Then the actions (associated to places and transitions) and conditions (associated to transitions) must be defined

Actions can be commands implemented in any control unit in the distributed control architecture of Fig. 2. These commands can be selected from a menu list automatically generated by the GUI. Each command is a message and the user must define the command parameters that will automatically appear in a new window when that command is selected in the editor.

When dispatch executes the Petri net, the messages assigned to places and transitions will be published as the net progresses. Also available are some special commands, such as start and stop another task (Petri net) or start a timer.

Conditions can be events produced by any module in Fig. 2. These events are selected from the menu list generated automatically by the GUI. An event can be the simple arrival of a message, a condition on some message parameter or any logical expression on several parameters over the same or different messages. RoboGraph GUI allows any logical expression to be defined over the message fields. However, complex conditions over message fields are sometimes more naturally expressed using other programming languages. For these cases, a Java-like editor is also integrated in the Petri net editor to program conditions and actions associated to places and transitions.

Timers are a tool widely used in automation that come in very handy here. In addition, in our applications, we have also used them as an error detection mechanism in order to time some actions of different modules. Actions can start a timer while conditions can test the value of a timer.

Global variables are used to get starting data and store information to share conditions and events in different places and/or transitions.

#### IV. ANALYZING THE COORDINATION MODEL

Petri nets have an origin dating back to 1962, when Carl Adam Petri wrote his PhD on the subject. Since that time, Petri nets have been accepted as a powerful formal

specification tool for a variety of systems including concurrent, distributed, asynchronous, parallel, deterministic and non-deterministic systems.

The ability to analyze Petri nets is generally considered to be the most important activity. Through the analysis of a Petri net, the designer can gain insight into the behavior and properties of the modeled system. There are two major types of analysis that may be performed on Petri nets. The first involves the creation of a reachability tree, and second involves matrix equations. There are tradeoffs involved with selecting a particular analysis technique, but the reachability tree method was first chosen for this project because it can effectively solve the safeness and boundedness properties.

The Petri net is declared safe if all of the places in the net are safe. A place is said to be safe if, for all possible markings, the number of tokens in that place never exceeds one.

A Petri net is k-bounded if, for all possible markings, the number of tokens in any individual place in the net never exceeds k. Since we use binary Petri nets, we want them to be 1-bounded which is equivalent to safe.

One interesting property regarding control systems is liveness which is a system that will be able to run continuously, i.e. a system which does not deadlock.

There are a significant number of Petri net tools that do provide analysis of the reachability tree, some of which address the properties of liveness and boundedness. Roméo [10] and CPN-AMI [11] are a couple of Petri net design and analysis applications of the many that can be found in tool databases such as [12].

We are more interested in behavioral than structural analysis because, even though it depends on the initial marking, the structural analysis is too restrictive for this application. The analysis modules implemented include:

- State space analysis: The Petri net is analyzed by building a tree of all the reachable markings. The reachability tree can then be analyzed to determine properties of the Petri net such as reachability, boundedness, deadlock, and safeness.
- Invariant Analysis: The Petri net is analyzed to determine both, the P-Invariant and T-Invariant vectors.

V. MONITORING

Even if a program is based on a fault-free algorithm, its implementation can contain errors arising, among other possible sources, from coding errors. Similar situations arise in distributed system coordination. This is, we might use safe and live Petri nets but still debugging tools are needed.

Dispatch schedules the different control unit actions and executive actions (other Petri nets), as well as the

synchronization with the events produced. The interaction with other control units in the architecture is performed by publishing and subscribing to messages. This way, local problems in a control unit, such as a deadlock problem, do not block dispatch. Also, simple mechanisms to detect and recover from a failure or exception situation can be established.

When starting, dispatch subscribes to the task execution or cancellation requests. Execution requests that cannot be executed at the reception time, are stored in different queues according to their priority. A task execution request can come from different modules (figure 2), such as user interface modules (user requests) and dispatch itself (command associated to other Petri net).

The execution of a new task starts loading the interpreted Petri net from a PNML file (Petri net markup language). Then dispatch subscribes to all the messages referenced in the events. Finally the initial marking is set and the actions associated to the marked places are executed. The Petri net can only progress with the arrival of messages or the end of a timer.

Each time a change in the status of a Petri net (start, stop, evolve) or in the waiting queues (new requests added or removed) is produced, a new message reporting that change is issued for monitoring (debugger module in figure 2) and stored in the log file for off-line debugging.

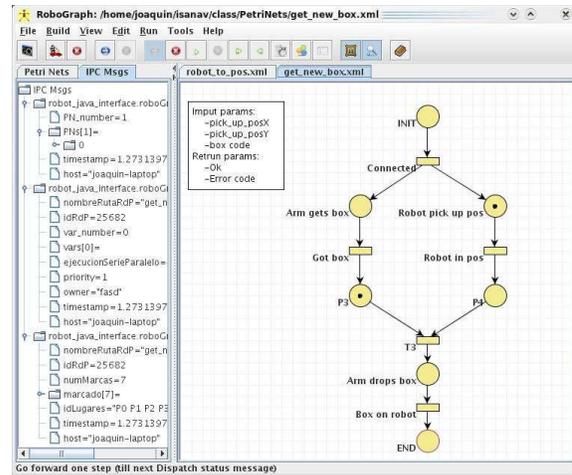


Fig. 4. Petri net debugger main window.

Debugger module in monitor mode subscribes to different dispatch messages that shows the status of the different running or waiting Petri nets. Every running Petri net is shown in a different tab with the current marking as in Fig. 4. When dispatch evolves a Petri net marking, a new message is issued and the debugger will update the monitor tabs. Therefore, using the monitor we can see the status of the system in a snapshot, since the

marking of the running Petri nets represents its status. This is a very helpful tool when debugging an application. An information window with the queued tasks (Petri nets) and messages can also be displayed on the left tabbed pane of Fig. 4.

## VI. LOGGING AND DEBUGGING

A general approach to debugging a distributed system is to create logs while the system is running and analyze them off-line to reduce interference with the execution. In the coordination system proposed here, the evolution of the coordination Petri nets, together with the messages received and published to the different control units, can be logged at different levels of detail:

- Level 1. Logging the status (marking) of a Petri net each time a transition is fired.
- Level 2. Logging data on level 1 and messages exchanged with the control units without the parameters
- Level 3. Logging data on level 2 and messages including all the parameters

When the amount of logging data is significant, then the debugging takes place in an offline manner. The search for a bug must employ effective strategies for choosing what is relevant and what is not.

The system administrator can then run the debugger in the play-logger mode, open the log file and play it at the same pace as in the real execution. Different tabs with the running Petri nets will be shown as in monitor mode. Besides the regular play option, the user can monitor the log file step by step. It can also jump to a defined place in “execution” as do many commercial programming development environments (C, Java, C++, etc.). Finally, the user can see different details about the IPC messages.

Figure 4 shows the main debugger while replaying an application. The panel on the left-hand side of figure 4 includes a couple of tabs: the Petri nets info in one tab and the messages info in another tab. The panel on the right-hand side of figure 4 shows two tabs corresponding to the two Petri nets that are active at that execution point shown in the figure.

Debugging the system directly from the logging data can be quite difficult. However, using the Petri net debugger is a very easy way to find the execution point where the evolution of the system does not correspond with the one for which it was designed. Besides, a great number of bugs end up with the Petri nets stuck with a fixed marking. It is quite easy to find this point and from it debug the application.

For example, when executing the Petri net of figure 3, the Petri net debugger can freeze with the marking of figure 4. With a simple look at the Petri net marking we can see that the problem is that the `get_new_box` Petri net

is waiting for the “mobile robot in pick up position” event. And this should be issued by the `robot_to_pos` Petri net. The Next step is to check why the `robot_to_pos` Petri net is not reporting the event and we can find one of the following: the building control unit is not working properly, the mobile robot control unit is not working properly or there is some design problem in the `robot_to_pos` Petri net.

## VII. RESULTS AND CONCLUSIONS

The main contribution of this paper is to show the debugging advantages of using Petri nets in the design and implementation of the distributed system coordination.

First, they can be used for model checking to gain insight into the behavior and properties of the modeled system. Research on Petri net analysis has produced a number of algorithms that can be used.

Second, monitoring the system using Petri nets is very helpful because we can see the status of the system in a snapshot because the marking of the running Petri nets represents its status.

The third possibility is to analyze the data logged during execution. A visual interface is used to present the evolution of the Petri nets. It is also possible to jump to a specific point in the execution and see the data of each message if necessary.

The fourth possibility is to use the log to replay the coordination system while we debug one of the units without the need to run all the others. This can be done only in the case that all the messages exchanged between the control units go through the coordination unit



Fig. 5. Surveillance application robot (WatchBot).

The system has been tested in a security and surveillance application recently finished and presented (Fernández & Souto 2009), even though a former version was already presented in (Fernández 2008). The building surveillance application is a Multirobot system monitored

through Internet. Each robot (figure 5) can handle some daily surveillance routine tasks. Sensor information such as real-time images captured by a camera on the robot with pan/tilt/zoom functions can be transmitted back to the central management office via a local area network. In this distributed application three different type of control units have been used:

- The mobile robot control unit manages mobile robot navigation and surveillance autonomous tasks. It is connected to the intranet via Wi-Fi.
- The building control unit manages the building devices including security sensors and alarms. We developed a scale model for the building sensors and actuators network. For the model, a few Modbus modules (Schneider OTB 1E0DM9LP) with some leds and buttons, connected to each Modbus interface as in figure 6, were installed.
- The user interfaces manage the interactions with users allowing them to monitor the mobile robots activities and command new surveillance tasks. The main window of a user GUI is shown in Fig. 7

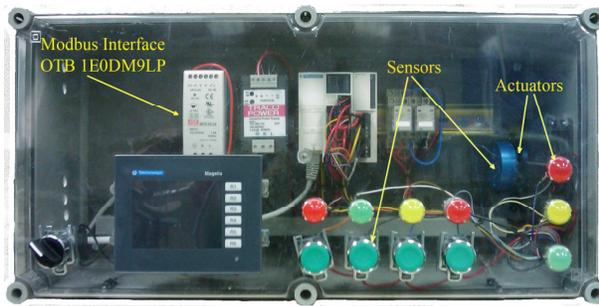


Fig. 6. One Modbus module used to test the transactions between building devices and the central control system.

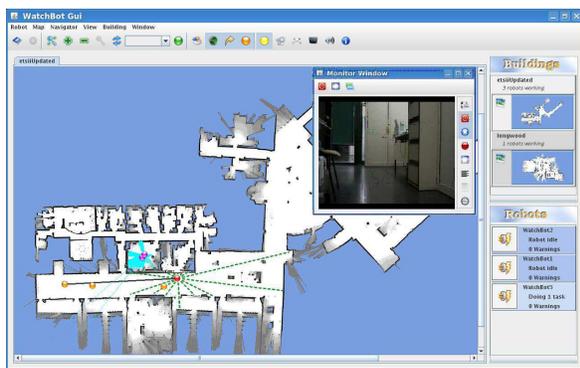


Fig. 7. Surveillance users GUI

Videos showing the applications together with examples using RIDE tools can be seen at [13].

#### ACKNOWLEDGMENT

This work has been partially supported by the Spanish Comisión Interministerial de Ciencia y Tecnología, (CICYT) under Project DPI2008-06738-C02.

#### REFERENCES

- [1] Bonasso, R.P., Kortenkamp, D., and Thronesbery, C.: Intelligent control of a water recovery system. *AI Magazine*, Vol. 24, No. 1, Spring 2003.
- [2] Simmons, R., Singh, S., Hershberger, D., Ramos, J. and Smith, T.: First results in the coordination of heterogeneous robots for large-scale assembly. *Proc. of International Symposium on Experimental Robotics*, Honolulu Hawaii, Dec. 2000.
- [3] Musliner, D.J. and Krebsbach, K.D.: Applying a procedural and reactive approach to abnormal situations in refinery control. in *Proc. of Conference on Foundations of Computer-Aided Process Operations (FOCAPO)*, Snowbird, Utah, July 1998.
- [4] Killian, C., Anderson, J. W., Jhala, R. and Vahdat, A.: Life, Death, and the Critical Transition: Finding Liveness Bugs in Systems Code. In *4th USENIX Symposium on Networked Systems Design and Implementation*, pages 243-256, April 2007.
- [5] Yabandeh, M., Knezevic, N., Kostic, D., and Kuncak V.: CrystalBall: Predicting and Preventing Inconsistencies in Deployed Distributed Systems. Technical report, EPFL, 2008. LARA-REPORT-2008-06.
- [6] Reynolds, P., Killian, C., Wiener, J. L., Mogul, J. C., Shah, M. A., and Vahdat A.: Pip: Detecting the Unexpected in Distributed Systems. In *3rd USENIX Symposium on Networked Systems Design and Implementation*, pages 115-128, 2006.
- [7] Kortenkamp, D., Simmons, R., Milam T., and Fernández J.: A Suite of Tools for Debugging Distributed Autonomous Systems. *Formal Methods in System Design*, Volume 24 , Issue 2 (March 2004), pp:157-188.
- [8] Geels, D., Altekar, G., Shenker, S. and Stoica I.: Replay Debugging for Distributed Applications. In *USENIX Annual Technical Conference*, pages 289-300, May 2006.
- [9] Armstrong, J.: Making Reliable Distributed Systems in the Presence of Software Errors. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, November 27 2003.
- [10] Gardey, G., Lime, D., Magnin, M. and Roux, O.: Roméo: A tool for analyzing time Petri nets. In *17th International Conference on Computer Aided Verification (CAV'05)*, Lecture Notes in Computer Science, Edinburgh, Scotland, UK, July 2005.
- [11] Kordon, F. and Paviot-Adet, E.: Using CPN-AMI to validate a safe channel protocol. in the proceedings of the International Conference on Theory and Applications of Petri Nets - Tool presentation part, Williamsburg, USA, June 21-25, 1999.
- [12] Petri net world web page. Petri Nets Tools Database. (<http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/quick.html>).
- [13] Robotics Integrated Development Environment (RIDE). <http://webs.uvigo.es/vigobot>.

# Robust following maneuver Design in a Simulated Autonomous Platooning Control System

Qing Wu

School of Logistics Engineering, Wuhan University of Technology, Wuhan, 430063, China

and

Yulin Ma<sup>2,3</sup>

Engineering Research Center of Transportation Safety (Ministry of Education), Wuhan University of Technology, Wuhan, 430063, China

Intelligent Transport System Research Center, Wuhan University of Technology, Wuhan, 430063, China

## ABSTRACT

Intelligent vehicle highway system (IVHS) have drawn much concern since they are assumed to be safe and effective methods to meet the increasing traffic demand without building more roads and resultant problems such as traffic accidents, traffic congestion and air pollution. A platoon of closely spaced vehicles is one of the important issues in IVHS. The advantages of the platooning system are increased safety and highway capacity mainly resulting from designing relevant following maneuver according to the platoon leader to form reasonably large platoons (five or more vehicles). In this paper, in order to simulate a platoon driving process, the vehicle dynamics as well as the safety distance policy are analyzed, and the following spacing controller is designed by applying fuzzy sliding mode control method into the spacing maintained between vehicles within the platoon. The proposed controller guarantees that vehicles in the platoon follow one another quickly and stably. Moreover, the achieved platoon stability is proven to be robust with respect to vehicle parameter uncertainties and unknown time varying disturbance. A seven vehicle platoon simulations illustrate the good performance of the proposed controller.

**Keywords:** Intelligent vehicle highway system, autonomous platoon, following maneuver, spacing policy, and fuzzy sliding mode control

## 1. INTRODUCTION

With ever increasingly improvement on design and implementation of intelligent vehicle and highway systems (IVHS), several transportation agencies in different nations are turning towards higher levels of intelligence and automation technologies for groups of vehicles to address urgent problems about traffic safety and congestion. A platoon of closely spaced vehicles becomes one of the important issues in IVHS. It is evident that organizing traffic in closely spaced platoons can increase capacity and safety.

In the one hand, the spacing between these platoons is large, platooning changes the average inter-vehicle distance to achieve capacity in terms of the traffic capacity. And in the other hand, the maintained distance between vehicles within a platoon is small, in case of a collision the relative impact velocity between colliding vehicles is also small. Because of platoon's size, speed and complexity, the design of the platooning control system need the development on new advance in communication, computer, sensors, and actuators and the analysis and configuration of hierarchical, hybrid control system [1,2].

Generally speaking, there are mainly five types of maneuvers performed in the platooning control system, including following maneuver, leader tracking maneuver, joining and splitting maneuver, and lane changing maneuver [2-4]. Due to some continuous maneuvers that regard "platoon" as the primary unit, that is, the platooning control system applies group collaboration rather than individual independence to perform maneuvers, designing following maneuver can line up several vehicles and have these vehicles the same properties, as a result form platoons. The key to following maneuver design is follower spacing control. When a vehicle is a follower, it must be controlled by a feedback law that maintains the required close spacing with the vehicle in front of it in its platoon. There are some encouraging recent developments in spacing control for a platoon of vehicles, many control methods, such as PID control, fuzzy control and sliding mode control, and multi agents system and hybrid system, are used for it [5-8,10,12]. In previous works, authors are apt to configure complex system structure to design controller, which has great difficulties in simulations and tests. The main contribution of this paper, compared with previous publications which also used sliding mode control method such as [9][10][11], is the following maneuver design for platoon by using fuzzy sliding mode control (FSMC) is simple and available in simulations and tests, even if in the presence of vehicle

parameter uncertainties and unknown time varying disturbance.

The rest of this paper has been organized as follows. Section II presents the platoon model and problem on the following maneuver design; Section III describes the developed follower spacing controller for platoon, Section IV presents following maneuver simulation for a seven vehicle platoon in the presence of different vehicle parameters and time varying disturbances. Section V presents the conclusions of our work.

## 2. PLATOON MODEL AND PROBLEM STATEMENT

A platoon of vehicles is traveling along the road sand table. By Newton's second law, the relationship between the acceleration of the vehicle, propulsion force, mechanical drag, and aerodynamic drag can be described as

$$m_i \ddot{x}_i = F_i - k_{d_i} \dot{x}_i^2 - k_{m_i} x_i + d_1(t) \quad (1)$$

where,  $m_i$  denotes the mass of the  $i^{\text{th}}$  vehicle,  $F_i$  denotes the propulsion force of the  $i^{\text{th}}$  vehicle,  $k_{d_i} \dot{x}_i^2$  denotes the aerodynamic drag of the  $i^{\text{th}}$  vehicle,  $k_{m_i} x_i$  denotes the mechanical drag of the  $i^{\text{th}}$  vehicle, and  $d_1(t)$  denotes the external disturbance bearing on the  $i^{\text{th}}$  vehicle, such as wind gust, tire-road friction, etc.

The propulsion system which represents the engine dynamics of a vehicle can be modeled as a first-order nonlinear differential equation

$$\dot{F}_i = \frac{1}{\tau_i(\dot{x}_i)} (-F_i + u_i) + d_2(t) \quad (2)$$

where,  $\tau_i(\dot{x}_i)$  denotes the  $i^{\text{th}}$  vehicle's engine time constant when the  $i^{\text{th}}$  vehicle is traveling with a speed equal to  $\dot{x}_i$ ,  $d_2(t)$  denotes the engine transmission variations, and  $u_i$  represents the output of the  $i^{\text{th}}$  vehicle's controller.

When a platoon is traveling, the following vehicle is required to keep a safe distance from its preceding vehicle. In terms of the traffic capacity, the desired safe distance should be as small as possible. However, this desired safe distance is not always constant, and it has a relationship with the velocity of the following vehicle. In other words, as the following vehicle's speed increases, the desired safe distance should be increased to improve traffic safety; while as the speed decreases, the distance should be reduced to improve the road capacity. Therefore, in this paper, we adopt the constant time headway policy for the following vehicle, where the desired safety distance defined by the velocity of the following vehicle.

Let  $x_i(x_{i-1})$ ,  $v_i(v_{i-1})$  and  $a_i(a_{i-1})$  be the position, velocity and acceleration of the  $i^{\text{th}}$  vehicle( the  $(i-1)^{\text{th}}$ , respectively). As shown in Fig. 1, the spacing deviation for the  $i^{\text{th}}$  vehicle from the desired safety distance is

$$\begin{aligned} e_i &= x_{i-1} - x_i - L_i - S_{d_i} \\ &= x_{i-1} - x_i - L_i - (\lambda_1 \cdot v_i + \lambda_2) \end{aligned} \quad (3)$$

where,  $L_i$  is the length of the  $i^{\text{th}}$  vehicle,  $S_{d_i}$  is the desired safety distance.

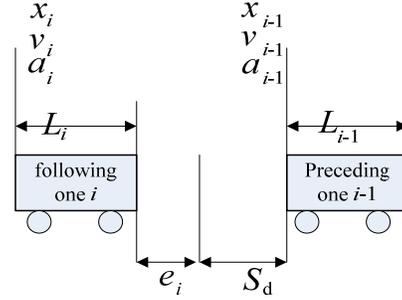


Fig. 1 Configuration of the following maneuver for a platoon

In practice, the vehicle parameters and the disturbances are partially or completely unknown. In this paper, we assume that the parameters of the  $i^{\text{th}}$  vehicle  $m_i, k_{d_i}, k_{m_i}, \tau_i(\dot{x}_i)$  are unknown, but satisfied with the following inequalities

$$\left\{ \begin{array}{l} m_i \leq M \\ k_{d_i} \leq K_D \\ k_{m_i} \leq K_M \\ \tau_i(\dot{x}_i) \leq \Gamma \\ 0 < b_1 \leq \frac{m_i \tau_i(\dot{x}_i)}{\lambda_1} \leq b_2 \end{array} \right. \quad (4)$$

where, the parameters  $M, K_D, K_M, \Gamma$  are positive constants. In addition, in view of the acceleration calculated, we assume that the upper bound of the vehicle's jerk is known

$$|\ddot{\dot{x}}_i| \leq K \quad (5)$$

where,  $K$  is also a positive constant.

To sum up, the objective of our following maneuver design is to design the follower spacing controller. So, our controller is to choose  $u_i(t)$ :

- 1) the spacing deviation is regulated to zero, i.e.,  $\lim_{t \rightarrow \infty} |e_i| = 0$ ,
- 2) the derivative of the spacing deviation is regulated to zero, i.e.,  $\lim_{t \rightarrow \infty} |\dot{e}_i| = 0$ ,
- 3) the relative velocity between the preceding vehicle and the following vehicle when the preceding vehicle drives at constant speed is regulated to zero, i.e.,  $\lim_{t \rightarrow \infty} |\dot{x}_{i-1} - \dot{x}_i| = 0$ , for constant  $\dot{x}_{i-1}$ .
- 4) the acceleration of the following vehicle when the preceding one drives at constant speed is regulated to zero, i.e.,  $\lim_{t \rightarrow \infty} |\ddot{x}_i| = 0$ , for constant  $\dot{x}_{i-1}$ .

## 3. FOLLOWER SPACING CONTROLLER DESIGN

In this section, we design a follower spacing controller by using FSMC method. Firstly, we present the spacing deviation model regarded as a second-order plant model that can be obtained by differentiating  $e_i$  in Eq. (3) twice

$$\ddot{e}_i = \ddot{x}_{i-1} - \ddot{x}_i - \lambda_1 \cdot \dot{v}_i \quad (6)$$

Differentiating  $\ddot{x}_i$  in Eq. (1) once, and putting Eq. (2) into it, yields

$$\ddot{x}_i = f(\dot{x}_i, \ddot{x}_i) + f(\dot{x}_i)u_i + \frac{d(t)}{m_i} \quad (7)$$

where,

$$\left\{ \begin{aligned} f(\dot{x}_i, \ddot{x}_i) &= -\frac{1}{\tau_i(\dot{x}_i)}(\ddot{x}_i + \frac{k_{d_i}}{m_i}\dot{x}_i^2 + \frac{k_{m_i}}{m_i}) - \frac{2k_{d_i}}{m}\dot{x}_i\ddot{x}_i \\ f(\dot{x}_i) &= \frac{1}{m_i\tau_i(\dot{x}_i)} \\ d(t) &= \dot{d}_1(t) + d_2(t) \end{aligned} \right.$$

Secondly, we define a sliding surface as

$$\begin{cases} S = ce_i + \dot{e}_i \\ \dot{S} = c\dot{e}_i + \ddot{e}_i \end{cases} \quad (8)$$

Additionally, In order to simplify the model solution, that is, to reduce the amount of calculation in iterative process and reduce impact the computing errors have on the task, we design a fuzzy sliding mode controller, the structure of which is as follows:

- {NB: negative big,
- NM: negative medium,
- NS: negative small,
- ZO: zero,
- PS: positive small,
- PM: positive medium,
- PB: positive big}

Input:  $S$  and  $\dot{S}$ ,

Output:  $u$

Fuzzy rule:

if  $S$  is A and  $\dot{S}$  is B, then  $u$  is C

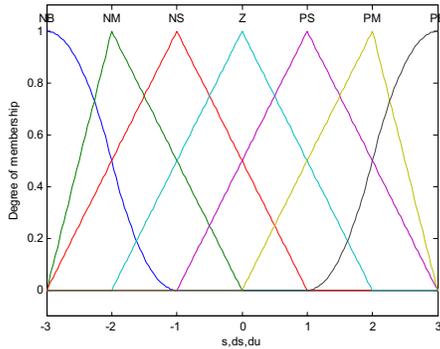


Fig. 2 The shape of membership functions

The rules, as shown in Table 1, state that if both  $S$  and  $\dot{S}$  are the positive big, which means  $S\dot{S}$  is the positive big, then  $u$

need a big positive variable to decrease  $S\dot{S}$  quickly; if  $S\dot{S}$  is less than zero, that is the desired state and  $u$  is equal to zero; if both  $S$  and  $\dot{S}$  are the negative big, which means  $S\dot{S}$  is the positive big, then  $u$  need a big negative variable to decrease  $S\dot{S}$  quickly. The rules ensure the stability of the fuzzy sliding mode control system.

Table 1 Fuzzy sliding mode control rules

$u$		$\dot{S}$						
		NB	NM	NS	ZO	PS	PM	PB
$S$	PB	ZO	PS	PM	PB	PB	PB	PB
	PM	NS	ZO	PS	PM	PB	PB	PB
	PS	NM	NS	ZO	PS	PM	PB	PB
	ZO	NB	NM	NS	ZO	PS	PM	PB
	NS	NB	NB	NM	NS	ZO	PS	PM
	NM	NB	NB	NB	NM	NS	ZO	PS
	NB	NB	NB	NB	NB	NM	NS	ZO

By the defuzzification method, the control law of the fuzzy sliding mode controller is

$$u_i = F_i + \left(\frac{1}{2} \frac{kM\Gamma}{\lambda_1} + c|\dot{\delta}_i|b_2\right) \frac{\sum_{i=1}^k \theta_i(\mu(S)\mu(\dot{S}))}{\sum_{i=1}^k \mu(S)\mu(\dot{S})} \quad (9)$$

where,  $\mu(S)$  and  $\mu(\dot{S})$  are the membership functions of normal distribution describing the fuzzy sets of  $S$  and  $\dot{S}$ , respectively;  $\theta_i$  is the principal values of corresponding fuzzy set of  $u$ ,  $k$  is a positive constant.

Then, we choose the following Lyapunov-like function candidate to prove the stability of the controller.

$$V_1 = \frac{1}{2} \frac{m_i\tau_i(\dot{x}_i)}{\lambda_1} S^2 \quad (10)$$

whose derivative according to Eq. (1), (2) and (8) is

$$\begin{aligned} \dot{V}_1 &= \frac{m_i\tau_i(\dot{x}_i)}{\lambda_1} S \left[ c\dot{\delta}_1 + a_{i-1} - a_i + \frac{\lambda_1 F_i}{m_i\tau_i(\dot{x}_i)} \right. \\ &\quad \left. + \frac{2\lambda_1 k_{d_i} v_i a_i}{m_i} - \frac{\lambda_1 \dot{d}_1(t)}{m_i} - \frac{\lambda_1 d_2(t)}{m_i} - \frac{\lambda_1 u_i}{m_i\tau_i(\dot{x}_i)} \right] \\ &= c\dot{\delta}_1 \frac{m_i\tau_i(\dot{x}_i)}{\lambda_1} S + a_{i-1} \frac{m_i\tau_i(\dot{x}_i)}{\lambda_1} S - a_i \frac{m_i\tau_i(\dot{x}_i)}{\lambda_1} S \\ &\quad + SF_i + 2\tau_i(\dot{x}_i)\lambda_1 k_{d_i} S v_i a_i - \tau_i(\dot{x}_i)\dot{d}_1(t)S \\ &\quad - \tau_i(\dot{x}_i)d_2(t)S - Su_i \end{aligned}$$

Taking the inequalities (4) and (5), we have

$$\begin{aligned} \dot{V}_1 &\leq |S| \left( c|\dot{\delta}_1|b_2 + |a_{i-1}|b_2 - |a_i|b_2 + 2K_D\Gamma|v_i||a_i| \right. \\ &\quad \left. - \Gamma\dot{d}_1(t) - \Gamma d_2(t) \right) + S(F_i - u_i) \end{aligned}$$

Substituting Eq. (9) into above inequality, we have

$$\begin{aligned} \dot{V}_1 &\leq |S|c|\dot{\delta}_1|b_2 - \frac{1}{2} \frac{kM\Gamma}{\lambda_1} S^2 - Sc|\dot{\delta}_1|b_2 \frac{\sum_{i=1}^k \theta_i(\mu(S)\mu(\dot{S}))}{\sum_{i=1}^k \mu(S)\mu(\dot{S})} \\ &\leq -\frac{1}{2} \frac{kM\Gamma}{\lambda_1} S^2 \leq -k \frac{1}{2} \frac{m_i \tau_i(\dot{x}_i)}{\lambda_1} S^2 = -kV_1(t) \end{aligned}$$

Therefore, we get the stability of the controller.

Finally, assuming that all vehicles have the same parameters, the longitudinal displacement, velocity and acceleration of the  $i^{\text{th}}$  vehicle are obtained by the following equations

$$\begin{cases} \ddot{x}_i(k) = f[\dot{x}_i(k-1), \ddot{x}_i(k-1)] + f[\dot{x}_i(k-1)]u_i(k-1) \\ \ddot{x}_i(k) = \ddot{x}_i(k-1) + \ddot{x}_i(k) \\ \dot{x}_i(k) = \dot{x}_i(k-1) + \ddot{x}_i(k) \cdot \Delta t \\ x_i(k) = x_i(k-1) + \dot{x}_i(k) \cdot k + \frac{1}{2} \ddot{x}_i(k) \cdot \Delta t^2 \end{cases} \quad (11)$$

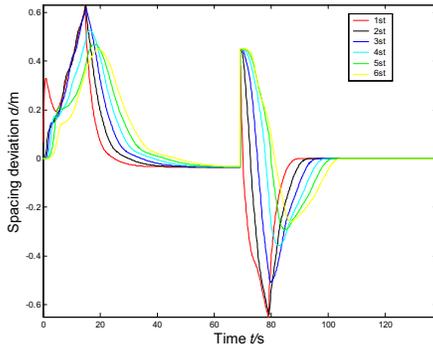
where,  $k$  is the number of the sampling time;  $\Delta t$  is the sampling time.

#### 4. SIMULATION TESTS

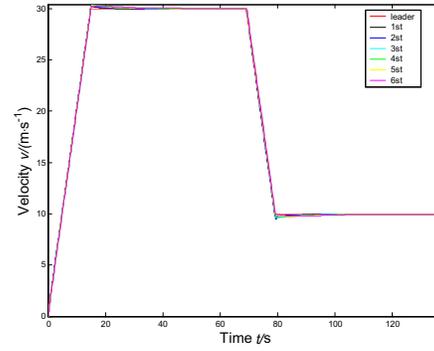
In order to test the proposed controller, we perform a series of simulation tests. The control parameters are chosen as follows:

$$\begin{aligned} m &= 1500, \quad k_d = 0.3, \quad k_m = 140, \quad \tau = 0.2, \quad \lambda_1 = 0.02, \\ c &= 2, \quad b_1 = 260, \quad b_2 = 340 \end{aligned}$$

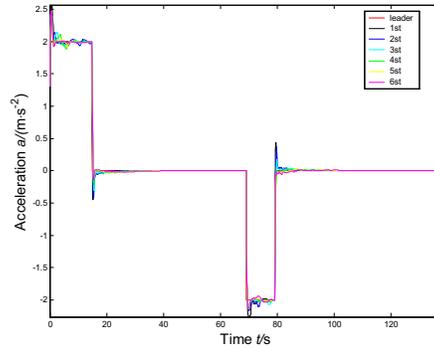
In the first simulation, we set about following maneuver simulation for a seven vehicle platoon based on FSMC algorithm. We assume that all vehicles in the platoon are identical with the control parameters above. In addition, we assume that no disturbance exists for all vehicles. The leading vehicle is assumed to accelerate from 0 m/s to 30 m/s at 2 m/s<sup>2</sup>. After the vehicle reaches 30 m/s, it begins to decelerate to 10 m/s at -2 m/s<sup>2</sup>. The simulation results are shown in Fig. 3.



(a) Spacing deviation profiles



(b) Velocity profiles



(c) Acceleration profiles

Fig. 3 Following maneuver simulation for seven vehicle platoon based on FSMC

In simulation 2, in order to investigate the platoon stability in the presence of parameters uncertainties and external disturbance, different vehicle has different mass  $m_i$ , aerodynamics drag coefficient  $k_{di}$  and engine time constant  $\tau_i$ , and the external disturbance is used in the simulations as shown in Fig. 4.

$$d(t) = \begin{cases} 0, & t < (27 + (k-1) \cdot T) \\ 1 \cdot [1 - e^{-0.2((t-k \cdot T)-27)}], & 27 \leq (t + (k-1) \cdot T) \leq 54 \\ 0, & t > (54 + (k-1) \cdot T) \end{cases}$$

where,  $T$  denotes the sampling period in the simulation, and  $T=0.15$ ,  $k$  represents the two scenario in which the platoon performs following maneuver, and  $k=1, 2$ .

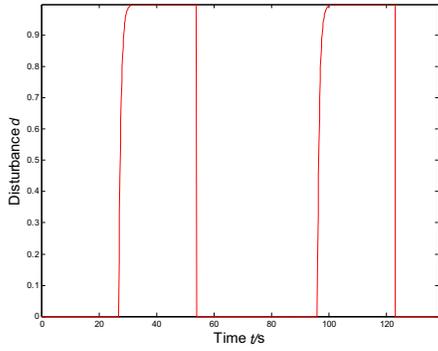
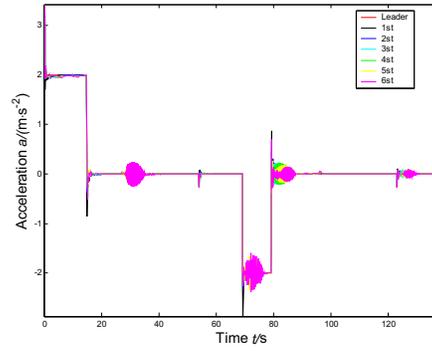


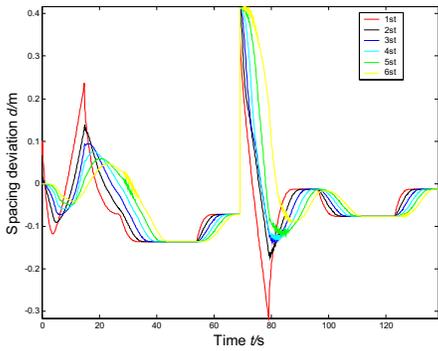
Fig. 4 Schematic diagram of the external disturbance



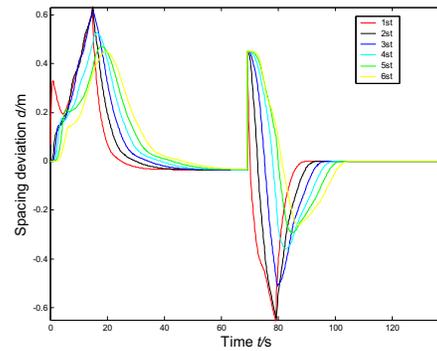
(c) Acceleration profiles

The simulation results are shown in Fig. 5-6. Compared with the ordinary SMC algorithm, it is obvious that, even with different parameters and external disturbance, the following maneuver using FSMC algorithm can still ensure the platoon stability.

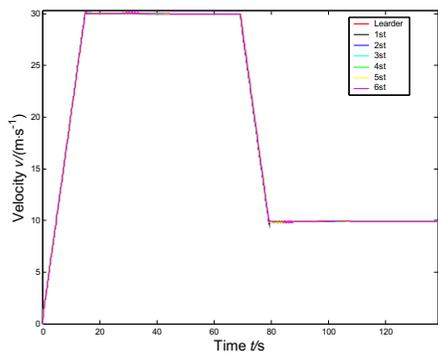
Fig. 5 Following maneuver simulation for seven vehicle platoon based on ordinary SMC under different parameters and external disturbance



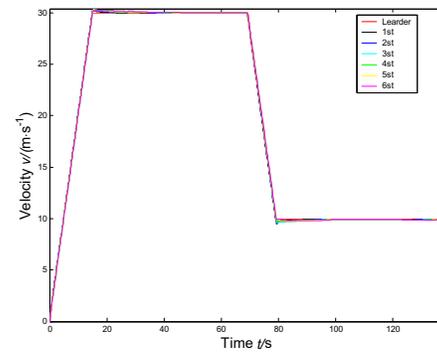
(a) Spacing deviation profiles



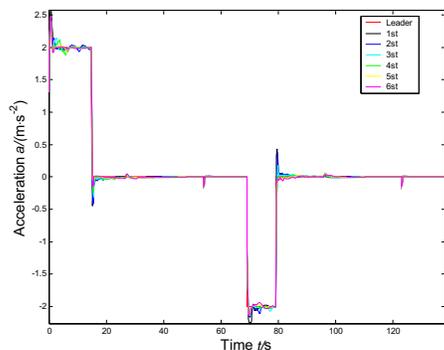
(a) Spacing profiles



(b) Velocity profiles



(b) Spacing profiles



(c) Spacing profiles

Fig. 6 Following maneuver simulation for seven vehicle platoon based on FSMC under different parameters and external disturbance

Seen from the simulation results, the following controller designed, through making the fuzzy rules, reduces the output's gain of FSMC, effectively reduces the chattering, even in the presence of different vehicle parameters and time varying disturbances, thus, achieves a good following performance along with the platoon scaling up.

## 5. CONCLUSIONS

This paper focuses on the problem presented in the following maneuver design for a seven vehicle platoon, proposes a fuzzy sliding mode method for the follower spacing control, which simplifies the platoon model solution, and the chattering in the output of the platooning system is reduced by making the fuzzy rules. The following maneuver performed in seven vehicle platoon is achieved, even if in the presence of different vehicle parameters and time varying disturbances. The next work is to explore the high-fault-tolerant following maneuver for platoon on road sand table testing.

## 6. ACKNOWLEDGMENT

The work is supported by the nation natural science foundation of China (No.60874081).

## 7. REFERENCES

- [1] Alvarez, L., Horowitz, R., "Hybrid controller design for safe maneuvering in the PATH AHS architecture," **Proceedings of the American Control Conference**, Albuquerque, New Mexico, USA, Jun.4-6, 1997, pp. 2454-2459.
- [2] Horowitz, R., Varaiya, P., "Control design of an automated highway system," **Proceedings of the IEEE**, Vol. 88, No. 7, 2000, pp. 913-925.
- [3] Nobe, S.A., Wang, F.Y., "An overview of recent developments in automated lateral and longitudinal vehicle controls," **IEEE International Conference on Systems, Man and Cybernetics**, Oct. 7-10, 2001, pp. 3447-3452.
- [4] Wu, Q., He, Z.W., Chu, X.M., et al, "Key Technologies Research Development of Vehicle Platoon Drive Control in Intelligent Vehicle-infrastructure System," **Computer and Communications**, Vol. 26, No. 4, 2008, pp. 154-157.
- [5] Rajamani, R., Shladover, S.E., "An experimental comparative study of autonomous and co-operative vehicle-follower control systems", **Transportation Research Part C: Emerging Technologies**, Vol. 9, No. 1, 2001, pp. 15-31.
- [6] Lee, G.D., Kim, S.W., "A longitudinal control system for a platoon of vehicles using a fuzzy-sliding mode algorithm," **Mechatronics**, Vol. 12, No. 1, 2002, pp. 97-118.
- [7] Seiler, P., Pant, A., Hedrick, K., Disturbance propagation in vehicle strings[J]. **IEEE Transactions on Automatic Control**, Vol. 49, No. 10, 2004, pp. 1835-1841.
- [8] Li, Y.N., Zheng, L., Qiao, Y.J., "Fuzzy-PID Control Method on Vehicle Longitudinal Dynamics System," **China Mechanical Engineering**, Vol. 17, pp. 99-103, Jan. 2006.
- [9] Ren, D.B., Zhang, J.Y., "Adaptive Algorithm to Vehicle Following Control in Intelligent Transportation System", **2008 International Conference on Information Processing**, pp: 46-50, May, 2008.
- [10] Wu, Q., He, Z.W., Chu, X.M., et al, "An application of the adaptive fuzzy control in the longitudinal control of the platoon," **Proceedings of Pacific-Asia Workshop on Computational Intelligence and Industrial Application**, vol. 2, pp. 344-348, Dec. 2008.
- [11] Xiao, L.Y., Gao, F., Wang, J.F., "On scalability of platoon of automated vehicles for leader-predecessor information framework", **Proceedings of the IEEE Intelligent Vehicles Symposium**, pp: 1103-1108, Jun. 2009.
- [12] Ma, Y.L., Wu, Q., Zhang, R., et al, "Research on Longitudinal Tracking Control for a Simulated Intelligent Vehicle Platoon", **The 3<sup>rd</sup> International Multi-Conference on Engineering and Technological Innovation**, Vol. 2, pp: 202-208, Jun. 2010.

# Control and Experimental Study of Flexible Needle Insertion Mechanism

Yong-de Zhang, Hao Chen and Yan-jiang Zhao

*Intelligent Machine Institute, Harbin University of Science and Technology  
Harbin, Heilongjiang, China*

and

Jin-you Wang

*Beijing Research Institute of Automation for Machinery Industry  
Beijing, China*

**Abstract** –In order to meet the needs of motion control of flexible needle insertion mechanism, a control system is designed based on Sunplus SPCE061A 16-bit processor which is the core of the system for insertion mechanism. Upper and lower level computer are used to control two stepper motors to realize various operation modes. Experiment result shows the control system operation is stable and the function is perfect, feeding control error is within 1.28mm. What's more, the maximum error and the root mean square between the experiment paths and the theoretic model are little enough to fit the ideal path well. All of these meet the requirements of flexible needle motion control.

**Keywords :** flexible needle insertion, motion control, mechanism, SPCE061A.

## 1. INTRODUCTION

Needle insertion is important means of many medical diagnosis, treatment and scientific research including interventional therapy and living tissue pathology. But as in the process of the needle insertion, sometimes it needs to drive round some important organs, nerves or bones. And the traditional sense of rigid straight needle is difficult to realize this kind of action. So undoubtedly using the flexible needle is the best choice if you want to achieve the ideal treatment purposes. In this article, needle with the bevel tip is chosen. During the needle insertion, the lateral force between the bevel tip and the tissue makes the needle axis bend. Such not only can control needle track more accurately, but also can bypass vital organs to achieve targets accurately. Using the flexible needles can make the needling insertion much more accurate and stable, and improve the diagnosis and treatment effect. Also, the technical has very wide scope of application, and it can be applied to some process of needling insertion into organs such as prostate, liver, lungs and brain, etc.<sup>[1-2]</sup>

In this paper the flexible needle insertion mechanism consists of feed motor, rotation motor, ballscrew mechanism, the needle and guiding device. (3D model shown in fig.1). Among them, feed motor is to realize the needle feeding and rotation motor can realize the needle rotation, ballscrew mechanism will change stepper motor's rotation movement into a straight-line

movement, meanwhile it has characteristics of high accuracy, high efficiency, and reversibility, guiding device can ensure the direction of flexible needles.

Nowadays, the market mainstream of stepper motor motion controller is still based on computer standard bus and the core of which is control card. The motion control card of the traditional stepper motor driver has following characteristics:

1. The board structure controller is easily oxidated and unfavorably longtime used because of the link of golden finger.
2. PC resource waste. User actually used only pieces of hardware resources because of bunding selling with upper computer. So it not only causes PC resource idle and waste, but also gives troubles to the computer maintenance.
3. Difficult to guarantee the reliability of the whole. Due to the PC could be selected from industrial computer or commercial computer, so the reliability of the system integration is very different to be guaranteed.
4. High price, low portability, there is no special-purpose control system to provide the information of flexible needle for doctor. But motion controller based on single chip microcomputer can be able to overcome the above shortcomings.

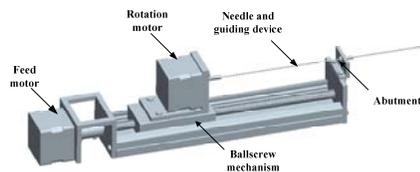


Fig.1 the flexible needle insertion mechanism model

In this article, a kind of control system for flexible needle insertion mechanism is designed which is feasible, reliable and easy to carry. It can realize the motion control of the flexible needle insertion mechanism. The whole of the system includes three parts: flexible needle insertion mechanism, upper computer control system and lower computer control system. (as shown in fig.2) The working flow of the System is: Use imaging system to obtain coordinates of obstacles and target, use MATLAB for path planning to get a set of parameters, then input these data to the PC system, send out the command to

lower computer system to drive the stepper motors. The parameters can be also directly input into the lower computer system to drive stepper motors to control needle insertion and rotating. Control needle feeding and adjust bevel tip angle to achieve a different path. This control system can through controlling the stepper motor running to realize the position control and speed control during the flexible needle feeding.

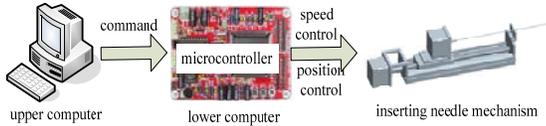


Fig.2 flexible needle insertion mechanism control system

**2. HARDWARE DESIGN OF THE CONTROL SYSTEM**

Control system uses modular design. It mainly composed of the following modules: core microcontroller module, driver module, man-machine interaction module(The framework of the hardware is shown in fig.3).Through keyboard preset the speed and swerve, man-machine interaction module shows the speed value. Finally, SPCE061A single chip sends corresponding time-series signals to control motor speed and swerve.

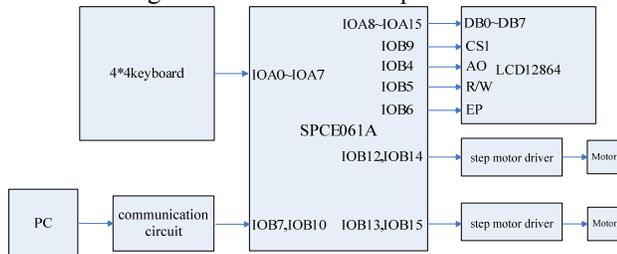


Fig.3 control system circuit connection diagram

*A. core microcontroller module*

Traditional 8-bit microcontroller has the limited hardware resources, computing and processing speed problems. The 32-bit microcontroller is a waste of resources and has higher costs. The Sunplus SPCE061A that is the structure of a 16-bit microcontroller is used. It has two 16-bit programmable timers/counters, thirty-two general-purpose programmable input/output ports A and B, fourteen interrupt sources, one universal asynchronous serial module(UART), built-in watchdog and other abundant hardware resources. Various function components are integrated into a single chip. It is bus structure which reduces the functional connection between the components and improves the reliability and anti-interference ability. The single chip supports more than ten interrupt-vectors and interrupt-sources. It is suitable for real-time application. Command system provides high speed of operation of 16-bitX16-bit multiplication instructions and inner product instructions. It also has DSP feature<sup>[3]</sup>.

*B. stepper motor driver module*

Flexible needle feeding accuracy requires 0.5 mm. Converted to motor shaft, each pulse stepper motor

rotates 7°. So motor stepper angle should be less than or equal to 7°. Considering the structural characteristics of flexible needle insertion mechanism is a simple structure, small size, reliable operation, so the model 42BYGH101 stepper motor is chosen. Stepper angle is 1.8°. In order to improve the accuracy of the flexible needle feeding, high fine fraction driver is chosen to solve this problem. Stepper motor drive type is selected SH-2H042MA. Fine fraction is chosen eight. It uses high-frequency chopping constant-current drive mode. It has a strong anti-interference, high performance, high start frequency, current options, simple structure, stable running and low noise. Control signals and internal signals plus the photoelectric coupling circuit to protect the digital circuit not to be damaged by high voltage drive circuit.

*C. man-machine interaction module*

The man-machine interaction module can be divided into keyboard module and LCD module. Keyboard module uses 4X4 matrix keyboard that uses low-level scan button input to set the parameters of the flexible needle feeding. LCD module uses 128x64 dot-matrix. The LCD panel displays flexible needle feeding speed, rotation speed, rotation direction, running time parameters. It makes users easy to real-time understand the situation of the flexible needle insertion.

**3. SOFTWARE DESIGN OF THE CONTROL SYSTEM**

Flexible needle movement consists of feeding movement and rotation movement. In order to achieve linear motion of the flexible needle, it requires the control system to control the feeding rate and the rotation speed respectively. In order to make flexible needle move forward and backward, it requires the feed motor to move clockwise and anticlockwise. The flexible needle insertion control system includes upper computer control system and lower computer control system. Microcontroller programming uses a mix of C language and assembly language programming. The characteristic of Sunplus microcontroller is that in the C language program, the assembly language program can be called<sup>[3]</sup>. The development tool uses Sunplus' unSP IDE which provides online by SPCE061A to make debugger program download and debug.

*A. lower computer software design*

Lower software consists of 4X4 matrix keyboard modular, 12864 LCD modular, stepper motor control modular, communication with upper computer modular. Fig.4 is the main program flowchart.

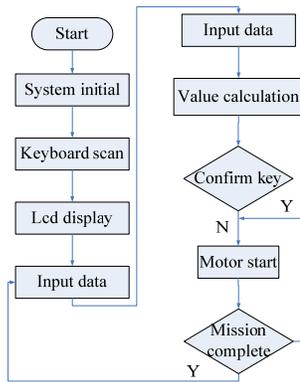


Fig.4 main program flowchart

Pulse signal is produced through timer interrupting to make I/O ports reverse. Variables can be set in the interrupt subroutine. Count the pulses to control the stepper motor position precisely. The stepper motor's speed can be controlled through setting timer's initial value to change time of the timer, then it produces different pulse signals and make different rotating speed. Using the relationship between input pulse  $f$  (HZ) and rotating speed  $n$  (revolution per minute.), any rotating speed which is allowed within the permission scope of stepper motor can be set. The mathematical relationship as follows:

$$n \times \frac{360}{a} = 60 \times f, \text{ (a is step angle of stepper motors)}$$

Stepper motor rotation direction can be controlled by setting the I/O port electrical levels, motor rotates clockwise under high level, and rotates anticlockwise under low level.

**B. upper computer software design**

The control software of flexible needle insertion mechanism is the upper computer software, which can obtain the function of communicate with the lower computer and the function of visual setting and debugging of the flexible needle inserting mechanism. The upper computer communicates with the lower computer through the serial port. The UART communication circuit module integrated by SPCE061A provides a full-duplex standard port, which can communicate with the upper computer and other microcontroller controllers and peripheral equipments. In order to connect with the computer interface or the terminal TTL device, the level and logical relation must be transferred with each other. The transition can be realized by disserting circuit element or using integrated circuit chip. During this paper, the electrical level is transferred by using the MAX232CPE chip. The program was developed based on VC++6.0 in the Windows XP environment, and the communication part is programming by using the MSCOMM ActiveX. The user's interface of control program of flexible needle inserting mechanism is shown in Fig.5. The whole user's interface is divided into feeding control zone, autorotation control zone, serial port setting zone and operation setting zone. In order to simplify

communication process and improve communication quality, the control interface just send data to the microcontroller and the microcontroller don't send the data back received from the upper computer. It needn't set receiving back display in MSCOMM control function. It is pretty good to just design the size of port sending buffering area.

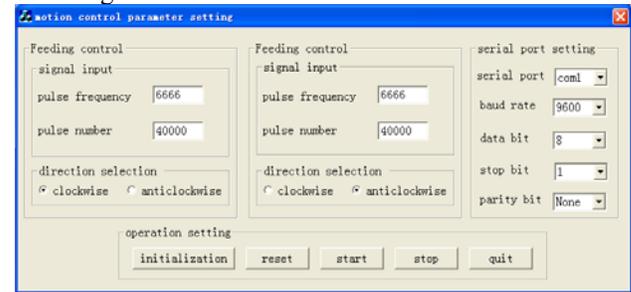


Fig.5 motion control program parameter setting interface

**4. EXPERIMENT RESEARCH**

**A. flexible needle insertion principle**

In flexible needle movement, through theoretical analysis of mechanics, needle axis bend through tissue giving lateral force to the bevel tip and rotation changes bevel tip direction then the needle's direction is changed and form the 3D trajectory. When the flexible needle inserts into the human body tissue, there is only feeding movement and no needle axis rotation movement. Needle bevel is effected by the force  $F$ . Force  $F$  makes needle flexible needle move along the arc of fixed radius  $r$  that is irrelevant with the needle feeding speed<sup>[4]</sup>. After feeding a distance, rotating some angles and feeding again, the needle will make another arc. When feeding and rotating at the same time, the needle will move along path of helix. When the rotating speed is much higher than the feeding speed, the needle's moving path will be a straight line. Movement rules conform to the resilience bicycle model<sup>[5-6]</sup>. Fig. 6 shows flexible needle insertion principle.

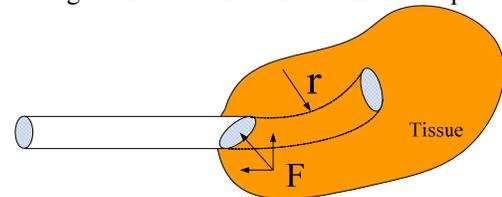


Fig.6 flexible needle insertion principle

In order to facilitate understanding and measurement, we adopt the needling track within the same plane. Three tracks is realized: straight line, in feeding movement performing a high-speed rotation movement simultaneously; Single arc, only feed movement, no needle axis of rotation movement; Double arcs, feed a distance and needle axis rotate 180°, then go on a long distance.

Fig.7 shows a complete process of the flexible needle insertion. First, flexible needle moves along a straight line, adjusts the bevel tip direction to make feeding movement and realize a period of circular arc. After this needle axis turn 180° and make another feeding

to realize another arc. The whole path is like “S” type and the needle avoids obstacles to achieve objective target<sup>[7]</sup>.

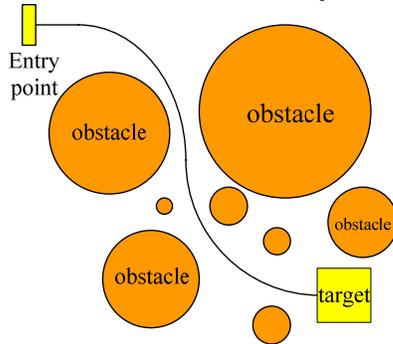


Fig.7 Flexible needles linear feeding, tip rotating 180°, flexible needles avoid obstacles to achieve objectives target

**B. experiment equipment**

Experimental platform consists of several parts as follows: flexible needle insertion mechanism, control system, simulation tissue, image acquisition device. It shows in fig.8. Feed motor drives ballscrew produces needle feeding movement. Rotation motor drives needle axis rotating movement. Agar is used to simulate biopsy tissue and high-resolution cameras to acquire images. Diameter of 0.4mm cylindrical nickel-titanium alloy silk simulate flexible needles. Tip angle of the needle is 30°<sup>[8]</sup>.

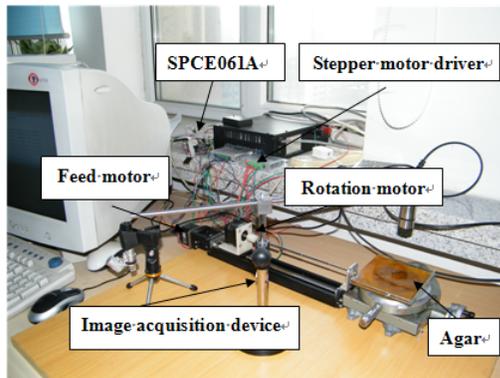
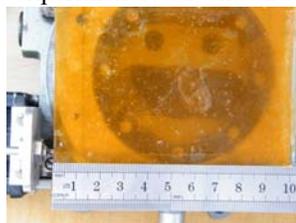


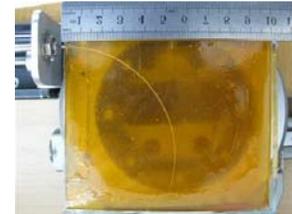
Fig.8 Flexible Needle Insertion system experiment equipment

**C. experimental results and analysis**

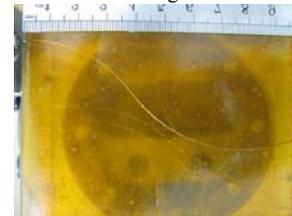
Experiment is divided into three groups. One group is linear feeding.  $v=0.02\text{m/s}$ ,  $\omega=3.5\text{rad/s}$ , insert tissue 80 mm. One group is arc.  $v=0.02\text{m/s}$ ,  $\omega=0$ , insert tissue 100mm. One group is double arc. Firstly,  $v=0.02\text{m/s}$ ,  $\omega=0$ , insert tissue 50mm. Then,  $v=0\omega=3.5\text{rad/s}$ , rotate 180°. Finally,  $v=0.02\text{ m/s}$ ,  $\omega=0$ , insert tissue 60 mm. Each group repeats 5 times, experiment result shown as fig.9.



(a) linear



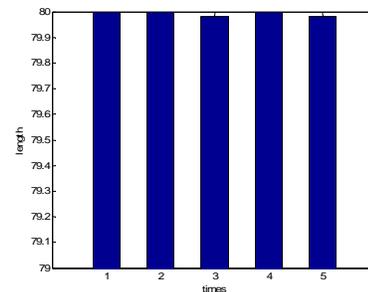
(b) single arc



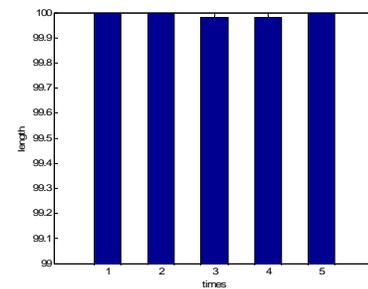
(c) two arcs

Fig.9 experiment result

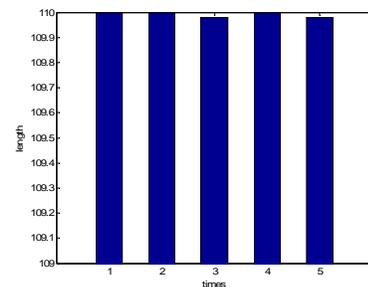
To illustrate this, the results of the experiment are analyzed from two aspects. On the one hand, we compare the actual displacement with the displacement we want to control theoretically when the control system controls three different paths (as shown in fig.10); on the other hand, we compare the track pierced by insertion system controlled by control system with the error of theoretical one (as shown in fig.11).



(a) : Linear



(b): Single arc



(C): Double arc

Fig.10 Actual displacements of three different paths

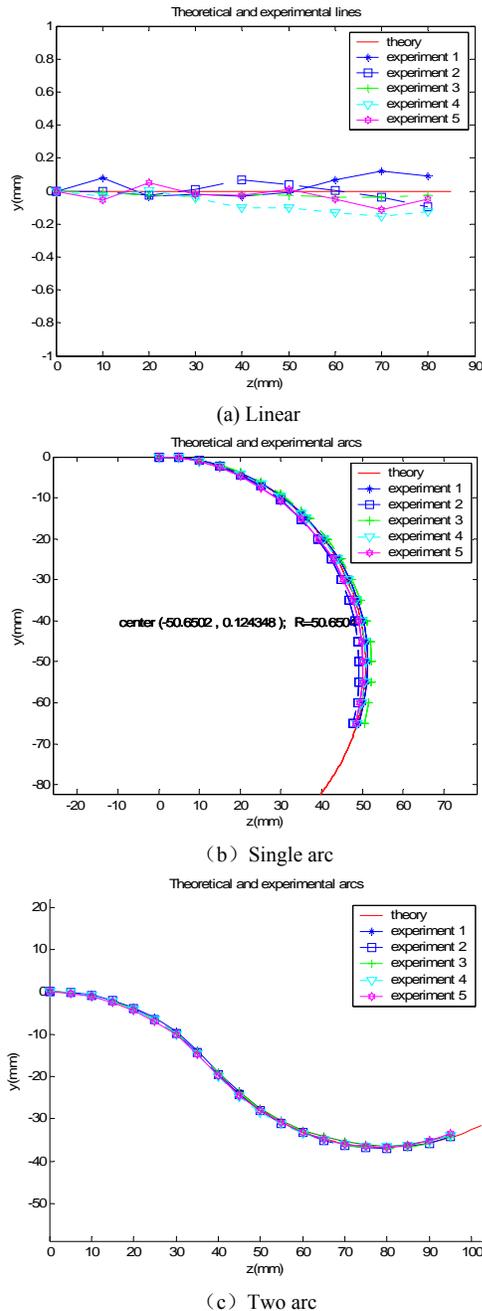


Fig.11 Actual and theoretically track error result of three different paths

From fig.10, we can see that no matter linear, single arc or double arc, the error of displacement control is very close. The max error is 0.02 mm. Due to the measuring tool's precision, actual error will be smaller.

From fig.11, we can see that experimental data and theoretical value are quite close and even superposition. The maximum error of linear is 0.54mm. The RMS from model fitting experimental data is 0.20mm. The maximum error of single arc and double arc is 1.28mm. The RMS from model fitting experimental data is 0.39mm. These meet biopsy needling accuracy and there is only four bigger error which is greater than 1 in single arc and double arc 10 group total 390 data. For the experimental

and theoretical value of line are more fit, almost superposition.

### 5. CONCLUSION

Through the flexible needle insertion principle research, motion control is designed based on SPCE061A for flexible needle insertion mechanism. Use upper and lower level computer to control two stepper motor to realize various operation modes. The controller is with friendly interface, stable operation, perfect function, high accuracy and portable feature. Based on it, the experiments of flexible needle are carried out. The experimental results are analyzed and compared. It turned out that displacement control error is less than or equal to 0.02mm. Maximum error and RMS between actual tracks and theoretical one are small enough. The actual paths fit the theoretical path well. It meets the needs of flexible needle insertion. But there are still a few data whose error is slightly higher than the measurement error in all experimental data. It is because the adjustment of orientation of bevel tip before insertion, and the characteristics of simulated tissue, which is different under different temperatures. This is direction that is needed to research later. Motion control and experiment study of flexible needle insertion mechanism set up a good foundation for next 2D and 3D of the flexible needle insertion.

### ACKNOWLEDGMENT

This work was supported by the Heilongjiang Province Education Bureau (No.20080820) and National Natural Science Foundation of China(No.51075105).

### REFERENCES

- [1] Webster III R J, Kim J S, Cowan N J, et al. Nonholonomic modeling of needle steering[J]. International Journal of Robotics Research, 2006, 25(5/6): 509-525.
- [2] Hamzavi N, Chui C K, Chui C C, et al. Flexible liver-needle navigation using fish-like robotic elements[C]//IEEE International Conference on Systems, Man and Cybernetics. Piscataway, NJ, USA: IEEE, 2008: 3491-3496.
- [3] Zhang Peiren, Zhang Zhijian, Gao Xiufeng. 16-bit single-chip microprocessor principle and application[M]. Beijing: Tsinghua university press, 2005: 10-11.
- [4] Joseph M. Romano, Robert J. Webster III, and Allison M. Okamura. Teleoperation of steerable needles[C]//2007 IEEE International Conference on Robotics and Automation Roma, Italy, 10-14 April 2007.
- [5] Zhao Yanjiang, Zhang Yongde, Shao Junpeng. Kinematic modeling and experiment study of flexible needle[J]. Robot. 2010, 05: 666-673.
- [6] Zhang Yong-de and Zhao Yan-jiang. Kinematic Modeling of Bevel Tip Flexible Needle[C]. The 2010 International Conference on Intelligent Robotics and Applications (ICIRA 2010), Shanghai, 2010, Part II, LNAI 6425: 405-416.
- [7] Engh J A, Podnar G, Khoo S Y, et al. Flexible needle steering system for percutaneous access to deep zones of the brain[C]//IEEE 32nd Annual Northeast Bioengineering Conference. Piscataway, NJ, USA: IEEE, 2006: 103-104.
- [8] Kyle B. Reed, Vinutha Kalleem, Ron Alterovitz, et al. Integrated Planning and Image-Guided Control for Planar Needle Steering[C]//IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechanics Scottsdale, AZ, USA, October 19-22, 2008 AZ, USA, October 19-22, 2008.

## Should Internet be Governed or Not: Case of Google in China

Satish Mahadevan Srinivasan  
College of IS&T, University of Nebraska at Omaha  
Omaha, Nebraska – 68106, USA

Abhishek Tripathi  
College of IS&T, University of Nebraska at Omaha  
Omaha, Nebraska – 68106, USA

Sambasivam Srinivasan  
VGSOM, Indian Institute of Technology Kharagpur  
Kharagpur, West Bengal – 721302, India

### ABSTRACT

“Google’s China Problem and China’s Google problem” as termed in [18] is one of its kind, an interesting case which has challenged the ethical issues based on which the e-commerce company operate. This paper presents a case study on the conflicts or dilemma presented by Google’s e-commerce activity and makes suggestions on what corrective steps Google can take to sustain on their e-commerce business objectives. Finally this paper takes an opportunity to take a stand on the debatable issue “Should Internet be Governed or Not?”

**Keywords:** Google’s China problem, ITU, ICANN, Amnesty International, Human Rights.

### INTRODUCTION

Google Inc. was launched in 1998 by two graduate students, namely, Sergey Brin and Larry Page when they came up with a design of a search engine that they believed would revolutionize the use of internet [17]. Initially the company was called BackRub based on their notion of search engine. The concept behind this search engine is that useful web sites can be visited merely using the backward links [17]. The company started with an investment of \$100,000. In 1999 Google Inc. received near about 25 times the first investment from other venture capitalist and finally grew up to be a company with an employee base of 60. This is where the Google started building up its culture [8, 17]. In the same year the company relocated themselves in to Googleplex, Mountain View, California. In 2001, Brin and Page appointed Eric Schmidt as the CEO of Google and both of them decided to be the president of Google. In 2004, Google Inc. went for IPO (Initial Public Offerings) which took place successfully and the company valued \$100 billion. Since then Google become the fifth most popular web site worldwide and accounts for more then 380 million visitors per month, with 50% international users. Today Google, Inc. is ranked No. 1 employer out of the fortune 500 US companies with \$10,605 million revenue in 2006 [7, 9].

Google is a C2C company with the objective of providing its customers with an experience of obtaining the best possible information rather than providing partial or no information about any keyword [17]. Other than the search engine Google

provides its customer with AdWords, AdSense (Google advertisers target users based on the key words used in the searches), Google Toolbar, Google Image Search, Froogle (Internet based shopping tool), and programs that can allow users to track DNA history for themselves online [17].

Of all the Google products, the search engine is the popular product used by people world-wide. The search engine was designed to make use of the structure present in the hypertext. The Google search engine crawls and indexes the web efficiently and produces more satisfying search results then the existing systems [3, 6]. This design ensured that the search engine provides its users with organized information which is universally accessible and useable. The company’s motto was coined with the term “Don’t Be Evil” [17]. According to Kai-Fu Lee, head of operations for Google in China, Google’s mission is to bring information to the masses and the value proposition of Google can be represented using an exuberant equation “Youth + Freedom + Equality + bottom-up innovation + user focus + don’t be evil = The Miracle of Google” [18].

### GOOGLE IN CHINA

Google considers China as their target market in the future. Statistically, China is a population of 1.3 billion and more with 1.11 million regular internet users. Again China is ranked the second largest internet market in the whole world. The only competitor for Google in China is the Baidu.com which owns the 48% of the search engine market [17, 6]. The Chinese market has also ranked Google higher then Baidu, Microsoft, and Yahoo. The Google experience has been better with the Chinese users and this experience will in future translate into increased market share [17].

Thou Google is popular in China but the prospects of their popularity seems to diminish. The problem that Google is currently facing is with respect to the censorship of information related to Tibet, Taiwan, Falun Gong, Dalai Lama, Tiananmen Square, massacre of 1989 etc [17]. These censorship restrictions were posed by the Chinese government as they felt that these information could disrupt the peace of the Chinese as well as give an opportunity for Chinese to form an organization and revolt against their own government [2]. The Chinese government requires all search results obtained from the Google

search be bypassed through a firewall set up by them. The objective is to filter the keywords and censor unclassified information. The firewall would also block undesired contents from the web site by examining the messages and search requests. This firewall also termed as the great firewall of China makes Google slower and at times unavailable to their customers in China [2]. The Chinese government itself has employed 30,000 people who monitor e-mail, web site, blogs and chat rooms around the clock. Due to this the information available to the Chinese internet user in China were much less than what the Chinese immigrants in USA have access to, thus going against the motto of the Google "Don't be Evil" [18].

The dilemma for Google is; should they go for the China market by agreeing to the conditions imposed by the Chinese government or leave the China market and strict with their motto. To address this Google came up with a strategic alliance scenario where they wanted to launch Google.cn for the China market and allow Google users in China to have access to Google.cn, all together having the servers in China. This would allow Google user's in China to use the same search engine with a similar speed as the counterpart USA users have access to at home [17]. Google in addition planned to comply with the censorship requirement posed by the Chinese government and used their own filters in their server to filter unwanted stuff, from the search result, that were deemed as objectionable by the Chinese Government. Google.cn did not provide other services like e-mail, blog sites, and chat rooms fearing that the Chinese government could demand user's personal information [4, 5].

Google's strategic alliance looked as a prospect for them but they were not far from other troubles. The US government exerted pressure on Google stating that Internet is a tool for an individual to raise their voice and make them think like democratic and Google being a leading internet search service provider should not make internet a tool for repression [2]. The Global Human Rights and International Operations exerted pressure on Google asking Google to have a license for exporting the technologies used for censorship [2]. Representative Tom Lantos, and Congressional Human Right, Caucus, demanded Google to develop new technologies to sicken the censorship requirements posed by the Chinese government [2]. The Amnesty International accused Google stating that Google is not doing ethically correct and is abusing the human rights of the individual [2]. Julian Pain, internet spokesman for campaign group, Reporters Without Borders (RWB), a Paris- based public interest group, condemned Google on their relationship with the Chinese government and remarked that this collaboration is a real "shame" for Google [4, 5]. According to RWB US based search engine companies are not doing justice by agreeing to the censorship policy posted by the Chinese government. They believe that these companies are compromising on their mission just to gain access to the international market. Julian Pain added to the criticism stating that when companies like Google collaborates with the Chinese government, then it becomes much easier for the Chinese government to control what is being said in the internet. The Human Rights Watch (HRW) argued that any government can not exert their control over the information medium without the willing collaboration of the private sectors and leaders of the industry [17]. According to them, it was Google's fault to entertain the Chinese government by agreeing to their censorship policy and limiting the amount of information to the Chinese user. HRW also added that if all the top search engine companies in USA including Google would have refused to

agree to Chinese government censorship policy, then there would be a strong position for all the companies to push for free access [17]. The Congressional Human Rights Caucus accused Google saying that Google should be "ashamed" of complying to the Chinese government censorship policy as they compromised their motto for the sake of profits [17]. The Representative Lantos after having a talk with the search engine representatives commented that the US based search engine companies were unprepared to admit their mistakes, nor where, willing to take the shame or responsibilities for what their behavior has brought. The Global Online Act of 2006 forbids any US based search engine companies to locate their content servers in China or any other country that abuses human rights and from cooperating with the officials of countries that believe on effecting political censorship of online contents [17].

Leaving the organizations aside Google had conflicts with their investors and shareholders. Pamela Geller Oshry, former publisher of the new York Observer newspaper made it clear that she would never do any search from Google and posed a message through blogs and e-mails urging investors to sell their Google shares [8, 19]. Randy Thomas, a conservative blogger dropped Google ads from his blog explaining that Google did not comply with its moral duties by bowing to China's wishes [19]. The Beijing news reported that Google was under investigation by the Chinese authorities as they suspect Google to be operating in China without proper license [17].

Google took a stand on this issue by making a policy that whenever it would sensor any information from the search result it would replace the censored location with a caption stating the reason for censoring. Google makes it clear to their users about the censorship policy it agreed with the Chinese government [17]. But the Chinese government demanded Google to remove the censorship caption out of their search result and produce only the remaining contents that resulted after filtering the censored contents. In reply Google disagreed to remove the censorship policy caption and also made a point that it has already availed a license to operate in China from the Chinese ministry of Information Industries. Now Google finds itself in a nowhere land, where it has just to get blamed and also blame itself for what it has done [17].

Their investors and shareholders are doubttable about their stand in the market and have their fingers crossed. The fall in the prices of the Google's share [8, 17] and boycotts organized by the American users [19] clearly expresses the conflicts or the dilemma faced by Google with their stakeholders.

## **GOOGLE'S COMMENTS ON THEIR ISSUES**

In reply Google concluded that it was important for them to take this decision as they wanted their services to be available in China no matter how limited their scope was. They added the fact that their search engine is so powerful that with a limited scope it could reap the seeds of "democratization in China" [17]. Google also added the point that there was no scope of negotiating with the Chinese government and lot of companies in China perform filtering and censorship [17]. Google thanked Caucus for the invitation and discussed the Google strategy with the Congressional Human Right Caucus. In that discussion Google made it clear that the Google notifies its user in China whenever it filters the search result and made it clear that it will not offer services like GMail, chat room and blogging sites, as they would be demanded to disclose personal information to the

Chinese government. Google also promised that it would make huge investments in China to encourage research and development within China and its willingness to provide a new Chinese version of the unfiltered Google.com [17]. Peter Norvig, director of research at Google, felt that the US government and the corporate America should join hands together and have stronger principles and laws [2].

Google is facing all these conflicts and dilemmas because of their mismatch between their value proposition and their actions [1]. So Google has to be careful and decide beforehand when ever they take any action. If there is any conflict between the values and the actions, the stakeholders will not go along with the company ahead. This is in accordance to the New York Times test rule of the basic ethical principle [14, 15]. Google has created bad reputations among its users, shareholders, investors and its other concerned stakeholders by initializing programs agreeing to the censorship policy imposed by China. Its now time for Google to take corrective steps to reduce the effect of the bad reputation and keep itself prospering in its e-commerce activities.

In the next section we have provided several remedies that Google can follow up to regain its reputations among its users, shareholders, and stakeholders.

### REMEDIES

There are certain remedies provided by the Amnesty International as stated below:

1. Google should put a public statement expressing the Chinese constitution rights (freedom of expression) and also lobby for the release of all cyber-dissidents and journalists imprisoned [1, 16]. This is in accordance to the golden rule of the basic ethical principle [14].
2. Google should make their filtering and censorship process transparent to their users explaining what words or phrases were filtered and how these words or phrases are selected [1]. This is in accordance to the risk aversion rule of the basic ethical principle [14].
3. Google should make all their agreement policies available online (digitally) to their users. In particular the censorship agreement of filtering information and suppression of dissent has to be available publicly [1]. This is in accordance to the risk aversion rule of the basic ethical principle [14].
4. Google should exhaust all judicial remedies and appeals in China and in different countries before complying with the state directives. Google has to make clear to the government about their human right implications. Google should also make government aware about their principled opposition regarding implementing their requests or directives when it feels that the government rules are breaching the norms of the international human rights [1, 16].
5. Google should develop an explicit human right policy that is in accordance with the Universal Declaration of Human Rights, UN Norms for Business and the UN Global Compact's principle. By doing it Google would be free from any complications caused due to violation of human rights [1]. This is in accordance to the Universal rule of the basic ethical principle [14].

6. Google should decide beforehand till what extent will there be a compromise or considerations taken against the human right principle when they need to assist government to censor access to Internet. Google should also make this decision open to public [1].
7. Google should promote human rights in China through lobbying the government to align their legislative and social reform with that of the international human right standards. Google can do this by seeking clarification of the existing legal framework and through adopting business practices that encourage China to comply with its human rights obligations [1, 16].
8. Google should participate in and support the outcomes of a multi-stakeholder process so that a set of guidelines is framed relating to the Internet and human rights issues [1]. This is in accordance to the Social Contract rule of the basic ethical principle [14].

### SHOULD INTERNET BE GOVERNED OR NOT?

It is really a debatable issue; whether to have internet governed or not? In our view it should be governed and our assertion is based on the values provided by ITU, ICANN and Amnesty International. Before getting in to this debate we would like to discuss about Internet Governance? The term Internet Governance signifies the collective actions taken by government (s) and/or private sector operators of the network connected by the Internet to establish agreements about the standards, policies, rules, and enforcement and dispute resolution procedures to apply to global internetworking activities [10, 13]. This definition of the Internet Governance makes it clear that government has to collaborate with the private sectors to establish their three main functionality namely the Technical Standardization, Resource Allocation and Assignment, and Policy Formation, Enforcement and Dispute Resolution [12, 13].

It is trivial, to understand the need for Technical Standardization and its benefits. ITU has taken up the responsibility of technical standardization along with other heterogeneous set of functionality (responsibility) like policy making, resource assignment and allocation, sector research and statistics gathering, education, and the promotion of telecom development and trade shows in developing country [13]. The job of resource allocation and assignment has been taken over by the ICANN. ICANN also engages itself responsible for resource assignment and policy making. ICANN and ITU have together agreed to pursue the responsibility of resource assignment but they had few occasions when they disagreed with each others strategy. Both of them agreed to have a multi-stakeholder governance regime rather than to have an intergovernmental one and for having a central coordination body to manage DNS [13]. But they disagreed on issues like inability of ICANN to incorporate most of the country code top-level domain managers (ccTLDs) in to its regime, disagreement of the root server operators, and the role of the US government as contracting authority for ICANN and supervision of the root zone [11, 13]. ICANN had also disagreement with the IETF regarding regulating the supplies of the domain name services and making of relevant standard documents for compliances [13]. These disputes are trivial and needs time to resolve. These issues do not affect the value provided by the Internet Governance. The Policy Formation, Enforcement and Dispute Resolution is related with balancing the Intellectual property

protection with fair use and free expression, trade and e-commerce, taxation, law enforcement and crime prevention, content regulations and freedom of expression, spam, data protection, privacy and surveillance, security, rights to domain names, competition policy in the domain name industry, and domain name user privacy. These responsibilities are shared between ITU (information and communication network security), WIPO (Intellectual properties), UN-OHCR (Human Rights), Amnesty International (Human Rights), UNESCO (For promotion and use of Multilingualism and Universal access to Cyber Space), WTO (International Trade Law), UNCITRAL (International Trade law), ASEAN (regulation and issues on WWW), ICRA (content rating), and ASTA [11, 13].

## CONCLUSION

These discussions suggest the fact that there is a need for the Internet to be governed and the collaboration between the government and the companies can make Internet Governance a success. A private sector can not alone look in to every legal issues and also can not come along with their partners/competitors as all or most of them are profit minded and can falsify illegal issues to be legal, if profit is a concern. Google case with China emphasis this fact. Google did not understand the need for its collaboration with the U.S government and the non-profit organization discussed above. At the end of the day it made Google to be shameful in front of others. Sometimes Government policies can be intriguing and at times be confusing. For example, Internet Governance has policies against Internet surveillance expressing the fact that the internet providers should withhold the privacy of their user. Google complied with the policy and did not provide information to the Department of Justice (US government) and were finally sued in the court. At the same time the law also states that if the government requires user privacy information to validate Child Online Protection Act then the concerned entity having this data should comply with what is required and provide the data to the government. Google has to work out such intriguing issues considering their value proposition and by understanding the legal boundaries, keeping aside their profit intentions. Again, Google agreeing to the Chinese government policy of censorship can be termed as the "height of hypocrisy" as Google did this because its competitor agreed to Chinese government censorship policies and Google never wanted to loose their market share in China [17].

## REFERENCES

1. Amnesty International. (n.d.) *Undermining freedom of Expression in China*. Retrieved March 5, 2008 from <http://irrepressible.info/static/pdf/FOE-in-china-2006-iores.pdf>.
2. Baron, D. P., "Google in China" [Electronic version]. *Stanford Graduate School of Business*, Case: P-54, March 2006, pp. 1-8.
3. Brin, S., & Page, L., "The Anatomy of a Large-Scale Hypertextual Web Search Engine" [Electronic version]. *Computer Networks and ISDN Systems*, Vol. 30, No.1, 1998, PP. 107-117.
4. BBC NEWS. (n.d.) *Google censors itself for China*. Retrieved March 3, 2008, from <http://news.bbc.co.uk/1/hi/technology/4645596.stm>.
5. BBC NEWS. (n.d.) *Google move 'black day' for China*. Retrieved March 4, 2008, from <http://news.bbc.co.uk/1/hi/technology/4647398.stm>.
6. Corporate Website marketing. (n.d.) *Strategies for Google Now Available*. Retrieved March 3, 2008, from [http://www.corporatewebsitemarketing.com/bhartzler/Jul05feed/Strategies for Google Now Available.rss.html](http://www.corporatewebsitemarketing.com/bhartzler/Jul05feed/Strategies%20for%20Google%20Now%20Available.rss.html).
7. CNN Money.com. (n.d.) *100 Best Companies To Work For 2008 Top 50 employers*. Retrieved March 2, 2008, from [http://money.cnn.com/galleries/2008/fortune/0801/gallery.bestcos\\_top50.fortune/index.html](http://money.cnn.com/galleries/2008/fortune/0801/gallery.bestcos_top50.fortune/index.html).
8. Ezine Articles. (n.d.) *Stakeholder Analysis and Stakeholder Management*. Retrieved March 5, 2008 from <http://ezinearticles.com/?Stakeholder-Analysis-and-Stakeholder-Management&id=6783>.
9. Google. (n.d.) *Corporate Information: Google Management*. Retrieved March 5, 2008 from <http://www.google.com/corporate/execs.html>.
10. Hellstrom, J. (2005, september). *WSIS, Internet Governance and Human Rights – Background material* [Electronic version]. *Collegium for Development Studies*, 1-9, Retrieved March 5, 2008, from [http://www.kus.uu.se/pdf/publications/Background\\_material\\_IG.pdf](http://www.kus.uu.se/pdf/publications/Background_material_IG.pdf).
11. International telecommunication Union. (n.d.) *International Telecommunication Union*. Retrieved March 5, 2008 from <http://www.genevabriefingbook.com/chapters/itu.pdf>.
12. INFORMATION WAR in China. (n.d.) *Lies, damned lies, and Brin spin*. Retrieved March 1, 2008 from <http://www.google-watch.org/china.html>.
13. Internet Governance.org. (n.d.) *Internet Governance: The State of Play*. Retrieved March 5, 2008 from <http://www.internetgovernance.org/pdf/ig-sop-final.pdf>.
14. Laudon, K. C., & Traver, C. G., *E-commerce: Business, Technology, and society*, Pearson Prentice Hall, 2007.
15. Laudon, K. C., & Traver, C. G., "Print the Library [Online]: is Google Playing fair, or Just Out to make a Buck? E-Commerce Business. Technology. Society III Edition, pp. 558-560. Pearson Prentice Hall, 2007.
16. Nobleprize.org. (n.d.) *Amnesty International: History of Organization*. Retrieved March 5, 2008 from [http://nobelprize.org/nobel\\_prizes/peace/laureates/1977/amnesty-history.html](http://nobelprize.org/nobel_prizes/peace/laureates/1977/amnesty-history.html).
17. O'Rourke IV, J.S., Harris, B., Ogilvy, A., "Google in China: government censorship and corporate reputation" [Electronic version], *The Journal of Business Strategy*, vol. 28, No. 3, 2007, pp. 12-21.
18. The New York Times. (n.d.) *Google's China Problem (and China's Google Problem)*. Retrieved March 3, 2008, from <http://www.nytimes.com/2006/04/23/magazine/23google.html>.
19. World News. (n.d.) *Google China censorship fuels calls for US boycott*. Retrieved February 26, 2008 from [http://www.boston.com/news/world/articles/2006/01/28/google\\_china](http://www.boston.com/news/world/articles/2006/01/28/google_china)

**Acknowledgement:** We would like to extend our thanks to Dr. Sajda Qureshi, Associate Professor, ISQA, University of Nebraska at Omaha for laying the initial interest and

foundations in this topic. As the instructor for the Electronic Commerce course she provided us with the *Google in China* (Baron, D. P. (2006)) case study material and gave us valuable suggestions to develop the case study presented in this paper. This paper is a short version of the assignment submitted for her course.

# Assistance in Similarity Classes Interpretation

Ismail BISKRI, Hassane HILLALI, Steve DESCOTEAUX, Louis ROMPRÉ, Boucif AMAR BENSABER

LABoratoire de Mathématiques et Informatique Appliquées  
Département de Mathématiques et Informatique  
Université du Québec à Trois-Rivières  
3351, boul. des Forges, C.P. 500, Trois-Rivières (QC) G9A 5H7 Canada  
{[ismail.biskri](mailto:ismail.biskri) ; [hassane.hillali](mailto:hassane.hillali) ; [steve.descoteaux](mailto:steve.descoteaux) ; [louis.rompre](mailto:louis.rompre) ; [boucif.amar.bensaber](mailto:boucif.amar.bensaber) }@uqtr.ca

## ABSTRACT

Text classification has been the focus of interest of many researchers for a long time. However, the results take the form of classes of similarities that we often do not know what to do with. The combination of text classification and maximal association rules will allow the extraction of hidden knowledge, often relevant from the text and allow the detection of dependencies and correlations between the relevant units of information (words) of different classes.

**Keywords:** Text classification; Maximal association rules.

## 1. INTRODUCTION

In recent years, a phenomenal growth in the amount of text information is noticed. Because of the rise of the Web or because of electronic documents in various institutions, text-mining has become of major importance. In addition, the operations of reading and analysis of texts, access to information, knowledge extraction, etc. must be supported by efficient computational processes to assist users in their tasks.

Text classification is the focus of interest of several researches in the field of text-mining. Among these algorithms, the best known include: Knn, Kmeans, ART, SOM, SVM. Classifiers are based on the common principle of a vector representation of documents with using a matrix of frequencies (possibly the presence / absence) of units of information in each document. The main goal of text classification is to group into "homogeneous" classes textual objects that share similar properties. The result is a set of classes of similarity usually displayed as lists of words (in fact called bag of words) that co-occur together. These classes can sometimes seem less significant or completely insignificant. They are often very large and despite many improvements, they are very noisy. Some classes share, also a part of their lexicon. This constitutes a major obstacle to an objective interpretation of the extracted knowledge made by a human. We believe that it is necessary to develop tools to facilitate the interpretation of classes, and thus, to enhance the interest of the classification.

The process of the maximal association rules extraction downstream of a classification operation is an interesting

avenue to enable the discovery of relevant lexical associations for an informed decision. The M-support and the M-confidence are two discriminating measures that expert user may consider to « clear » classes of similarity and accelerate the interpretation by assisting the expert. We believe, moreover, that the identification of maximal associations can play a major role in practical applications as information retrieval, construction and maintenance of ontologies, computer-assisted reading and analysis of texts, etc.

## 2. MAXIMAL ASSOCIATION RULES

A brief survey of the literature on data mining [3] teaches us that association rules allow for a representation of regularities in the co-occurrence of data (in the general sense of the term) in transactions, regardless of their nature. Thus, data that regularly appear together are structured in so-called association rules. An association rule is expressed as  $X \Rightarrow Y$ . This is read as follows: each time that X is encountered in a transaction, so is Y. There are also ways to measure the quality of these association rules: the measure of Support and the measure of Confidence.

The concept of association rule emerges mainly from the late 60 [8] with the introduction of the concept of the support and the confidence. Interest in this concept was revived in the 90s through the work of Agrawal [2] [1] on the extraction of association rules in a database containing business transactions. Currently, work is being done on how best to judge the relevance of association rules, as well as the quality of their interpretation [14] [9] [4], and their integration into information retrieval systems [5] and into classification processes for text mining [4].

To illustrate association rules, consider the definition of the principal elements in the following example:

- Three transactions to regroup the data that co-occurs: T1:{A, 1, K}; T2:{M, L, 2}; T3:{A, 1, 2}
- Two sets to categorize the data: E1:{A, M, K, L}; E2:{1, 2}
- X and Y: two separate sets of information units: X:{A}; Y:{1}.  $X \subseteq E1$  and  $Y \subseteq E2$ .

For a transaction  $T_i$  and a set of information units  $X$ , it is said that  $T_i$  supports  $X$  if  $X \subseteq T_i$ . The Support of  $X$ , noted as  $S(X)$ , represents the number of transactions  $T_i$  such that  $X \subseteq T_i$ . In the case of transactions  $T_1, T_2$  and  $T_3, S(X) = S(A) = 2$ .

The Support of the association rule  $X \Rightarrow Y$  is the number of transactions that contain  $X$  and  $Y$ . In the case of our example  $S(X \Rightarrow Y) = S(A \Rightarrow 1) = 2$ .

The Confidence of the association rule  $X \Rightarrow Y$ , noted as  $C(X \Rightarrow Y)$ , corresponds to the support of this association rule divided by the Support of  $X$  otherwise stated as  $C(X \Rightarrow Y) = S(X \Rightarrow Y)/S(X)$ . In the case of our example,  $C(X \Rightarrow Y) = C(A \Rightarrow 1) = 1$ .

Despite their potential, association rules cannot be established in the case of less frequent associations. Thus, certain associations are ignored since they are not frequent. For example, if the word *printer* often appears with the word *paper* and less frequently with the word *ink*, it is very probable that the association between *printer* and *paper* will be retained to the detriment of the association between *printer*, *paper* and *ink*. In fact, the confidence criterion associated to the relationship between *printer*, *paper* and *ink* would be too low.

The maximal association rules, noted as  $X \xrightarrow{\max} Y$ , compensate for this limitation. They are dedicated to the following general principle: each time that  $X$  appears alone,  $Y$  also appears. Note that  $X$  is reputed to appear alone if and only if for a transaction  $T_i$  and a category set  $E_j (X \subseteq E_j), T_i \cap E_j = X$ . In this case,  $X$  is maximal in  $T_i$  with regards to  $E_j$  and  $T_i$  M-Supports  $X$ . Note the M-Support of  $X$  by  $S_{\max}(X)$ , which thus represents the number of transactions  $T_i$  that M-Support  $X$ .

In the transaction  $T_1, X$  is not alone with regards to  $E_1$  since  $T_1 \cap E_1 = \{A, K\}$ . On the other hand, in the transaction  $T_3, X$  is alone since  $T_3 \cap E_1 = \{A\}$ .

The M-support of the maximal association  $X \xrightarrow{\max} Y$  noted as  $S_{\max}(X \xrightarrow{\max} Y)$  represents the number of transactions that M-support  $X$  and support  $Y$ . In the case of our example, only the transaction  $T_3$  M-supports  $X$  while  $T_1$  and  $T_3$  support  $Y$ . Consequently  $S_{\max}(A \xrightarrow{\max} 1) = 1$ .

The M-confidence noted as  $C_{\max}(X \xrightarrow{\max} Y)$  represents the number of transactions that M-support  $X \xrightarrow{\max} Y$  relative to the set of transaction that M-support  $X \xrightarrow{\max} Y$ . The M-confidence of the rule  $X \xrightarrow{\max} Y$  is thus calculated by the formula  $C_{\max}(X \xrightarrow{\max} Y) = S_{\max}(X \xrightarrow{\max} Y) / S_{\max}(X \xrightarrow{\max} E_2)$ . In the association  $A \xrightarrow{\max} 1$ , the M-Confidence is found to be equal to 0.5.

Finally, it should be noted that we must define the minimum thresholds for the M-support of a maximal association, as well as for its M-Confidence.

### 3. IDENTIFICATION OF MAXIMAL ASSOCIATION RULES IN SIMILARITY CLASSES

GRAMEXCO (n-GRAMs in the EXtraction of knowledge (COnaissance)) is our prototype that has been developed for the numerical classification of multimedia documents [11], particularly text documents. The numerical classification takes place by way of a numerical classifier. The unit of information considered in GRAMEXCO is the n-gram of characters, the value of  $n$  being configurable. The main objective is to provide the same processing chain, regardless of the corpus language, but with easily legible layouts in the presentation of the results. Recall that the use of n-grams of characters is not recent. It was first used in work by Damashek [6] on text analysis and work by Greffenstein [7] on language identification. The interest in n-grams today has been extended to the domains of images, and musicology, particularly in locating refrains [10]. A character n-gram is defined here as a sequence of  $n$  characters: bigrams for  $n=2$ , trigrams for  $n=3$ , quadrigrams for  $n=4$ , etc. For example, in the word *computer* the trigrams are: com, omp, mpu, put, ute, ter. We justify our choice of n-gram of characters as the unit of information by: (i) The cutting into sequences of  $n$  consecutive characters is possible in most languages. It is necessary that any approach can be adapted to several languages because of the "multilingual" nature of the web ; (ii) The necessary tolerance for a certain ratio of deformation or flexion of lexical units. The functioning of GRAMEXCO is not entirely automatic. The choice of certain parameters is made by the user according to their own objectives. GRAMEXCO takes a raw (non indexed) text as input in UTF format. There are then three first main steps where the user can customize certain processes.

The **first step** consists of building a list of information units and information domains (parts of texts to be compared for similarity). From the two operations carried out simultaneously, we retrieve an output matrix with a list of the frequency of appearance of each information unit in each information domain. The information units may be in the form of bigrams, trigrams, quadrigrams, etc. Obtaining information domains passes through the process of text segmentation which may be done in words, phrases, paragraphs, documents, web sites or simply in sections of text delimited by a character or a string of characters. The choice of the size of the n-gram and the type of textual segment is determined by the user according to the goals of their analysis.

The **second step** consists of reducing the size of the matrix. This operation is indispensable given the important cost in resources that an overly large matrix

would represent. During this step, a list of n-grams undergoes some trimming that corresponds to:

- (i) the elimination of n-grams whose frequency is lower than a certain threshold or above another threshold,
- (ii) the elimination of specific n-grams selected from a list (for example, n-grams containing spaces or n-grams containing non-alphabetic characters),
- (iii) the elimination of certain n-grams considered as functional, such as suffixes.

In the **third step**, the classification process takes place. The classifier used here is the neural network ART. The choice of classifier is not dictated by particular performance reasons since this is not our objective. We could have just as easily chosen another classifier that would have admittedly yielded different results. Such variations continue to be the focus of research such as was presented in Turenne [13].

At the end of this step, segments considered as similar by the classifier are regrouped into similarity classes. Furthermore, the lexicon of these segments forms the vocabulary of the classes to which they belong.

The classes obtained at the end of the classification operation will be the transactions of the process that will allow the extraction of maximal association rules. Finally, in order for the process to be carried out, it must be supervised by the user who will have to first determine the word for which the most probable associations will be found.

To illustrate this step, let us posit the following scenario that will allow us to discover maximal association rules  $X \xrightarrow{\max} Y$  based on the results of a classification.

The input of the classification is a text in which the vocabulary represents a category set E1: {x, a, b, c, d, e, f}. The classification outputs classes with their respective lexicon: C1 : {x, a, b, c}, C2 : {a, c, d}, C3 : {x, e, f, d}.

If the classes represent the transactions, the vocabulary of the input text represents a set E1 for categorizing the textual data (the vocabulary) in which set X is chosen.

This being established, the extraction process of maximal association rules is carried out in three steps:

1<sup>st</sup> step: choice of set X: it is the user who chooses the lexicon from a list of elements of E1 that will represent X. Let us assume for explanatory purposes that  $X = \{x\}$ .

2<sup>nd</sup> step: identification of set Y and set E2: the identification of the category set E2 in which Y would be a subset largely depends on the set X selected and on the classes of which X is a subset.

In the case of our illustration, X is included in C1 and in C3. Y may therefore be a subset either of {a, b, c} or of {e, f, d}. In other words, Y may represent one of the following subsets: {a}, {b}, {c}, {a, b}, {a, c}, {b, c}, {a, b, c}, {e}, {f}, {d}, {e, f}, {e, d}, {f, d}, {e, f, d}.

The measures of M-Support and of M-Confidence will be calculated with regards to these different possible values of Y. An iterative process would allow for testing the set

of these possibilities. We may, however, limit the number of iterations in order to avoid an overly prohibitive computational cost, for example, by fixing (via parameter) the cardinality of subset Y.

Let us suppose that  $Y = \{a, c\}$ ; in order to construct E2, the respective categories of elements a and c must first be established. These are obtained by uniting classes that contain a (or c, respectively). Consequently,  $E2 = \text{category}(Y) = \text{category}\{a, c\}$  would be obtained by intersecting  $\text{category}(a)$  with  $\text{category}(c)$ . Thus:

$$\text{category}(a) = \{a, b, c\} \cup \{a, c, d\} = \{a, b, c, d\}$$

and

$$\text{category}(c) = \{a, b, c\} \cup \{a, c, d\} = \{a, b, c, d\}$$

therefore :

$$E2 = \text{category}(Y) = \text{category}(a, c) = \text{category}(a) \cap \text{category}(c) = \{a, b, c, d\}$$

3<sup>rd</sup> step: once the sets E1, E2, X and Y as well as the transactions have been clearly identified, the calculation of the measures may be made.

Consider the association  $x \xrightarrow{\max} a, c$ . Using the classes C1: {x, a, b, c}, C2: {a, c, d}, C3: {x, e, f, d} as transactions, and  $E2 = \{a, b, c, d\}$ , it follows that M-support equals 1, since only Class 1 contains  $X = \{x\}$  and  $Y = \{a, c\}$ , and an M-confidence of 0.5 since two classes contain X while only one contains X and Y.

## 4. EXPERIMENTS

The whole of the theory presented here was implemented in C#. The results of the analyses are stored in XML databases.

The following experiments were applied to four corpora (three of them are extracted from web sites). Two corpora are in French and two are in Arabic. The first corpus is a collection of interviews with directors of small and medium Quebecois businesses in order to learn about their perspectives on the notion of *risk*. The second corpus addresses the history of the reign of *King Hassan II* of Morocco. The third corpus (in Arabic) addresses the *Organisation of the Petroleum Exporting Countries* (OPEC). Finally, the fourth and final corpus (in Arabic) summarizes the biography of the American President, *Barack Obama*. The domains are sufficiently different to draw conclusions on the efficacy of the methodology. Note : we limit ourselves to show just maximal associations and scores of each association (M-support and M-Confidence). We assume that the reader is sufficiently familiar with the methods of classification and we do not need to show classes of similarities.

1<sup>st</sup> experiment: the corpus, as mentioned above, addresses the perspective of directors of small and medium Quebecois businesses with regards to the notion of *risk*. One of the constraints during the interviews was the obligation put on the directors to use the word *risk* when they deemed it necessary. In our experiments, this aspect

is crucial since we need to know which words are associated to *risk* in the discourse of the directors. Thus, despite the presence of noisy data such as, for example, *Pause* and *X*, which were intentionally inserted into the text for ethical reasons (*X* represents the name of people who were questioned) and to represent silences (*Pause*), interesting results were still obtained. For example:

- $Risk \xrightarrow{\max} Project$  is an association that is found in 10 classes (M-support = 10) with a confidence of 100%.
- $Risk \xrightarrow{\max} Management, Project$  is an association that we find in 7 classes (M-support = 7) with a confidence of 70%. In other words, 30% of the time, it is possible to find the word *Risk* in classes where *Management* and *Project* did not occur together.
- $Risk \xrightarrow{\max} Management$  is an association that we find in 7 classes (M-support = 7) with a confidence of 70%.
- $Risk \xrightarrow{\max} Product$  is an association that we find in 5 classes (M-support = 5) with a confidence of 50%.

The following table summarizes the results obtained:

Y	M-Support	M-Confidence
Client	1	10%
Shareholders, Cost	1	10%
Client, Project	1	10%
Decision, Product	2	20%
Year	2	20%
Markets, Price	2	20%
Science	3	30%
Interview, Studies	3	30%
Function	4	40%
Manner, Level	5	50%
Product	5	50%
Question	6	60%
Interview, Risk	6	60%
Level, X	7	70%
Management	7	70%
Management, Project	7	70%
Project, Risks	8	80%
X	10	100%
Pause	10	100%
Project, X	10	100%
Pause, X	10	100%
Project	10	100%

Table 1: Results of the 1<sup>st</sup> Experiment

2<sup>nd</sup> experiment: For the second experiment, we chose a short 4-page text about the reign of *King Hassan II*. For this experiment, we intentionally chose to consider the cardinality of set Y equal to 1. For  $X = \{Hassan\}$ , we obtained the results summarized in table 2.

Note that, for example, the association  $Hassan \xrightarrow{\max} II$  is very strong. Its confidence is 100%. Likewise for the associations  $Hassan \xrightarrow{\max} Morocco$  and  $Hassan \xrightarrow{\max} King$ . Although their confidence is only 61.54%, this is sufficiently high to consider the two associations as maximal.

Y	M-Support	M-Confidence
Doctor	1	7.69 %
Professor	1	7.69 %
Spain	1	7.69 %
Tunisia	1	7.69 %
Spanish	2	15.38 %
Journalist	3	23.08 %
History	3	23.08 %
Prepare	3	23,08 %
Title	4	30,77 %
France	5	38.46 %
Politics	6	46.15 %
Year	7	53,85 %
King	8	61.54 %
Morocco	8	61.54 %
II	13	100 %

Table 2: Results of the 2<sup>nd</sup> Experiment

3<sup>rd</sup> experiment: For the third experiment, we chose an Arabic text regarding the *Organisation of the Petroleum Exporting Countries* (OPEC), the goal being to evaluate the validity of the method with regards to the Arabic language. For the purposes of the experiment, we chose  $X = \{OPEC\}$ . The following table provides a summary of the results (a translation of the Arabic words is provided):

Y	M-Support	M-Confidence
Mechanisms	1	9,09 %
Paris, Countries	1	9,09 %
Creation, prices	2	18,18 %
Petroleum	3	27,27 %
Countries, members	3	27,27 %
Prices	3	27,27 %
Organisation, prices	3	27,27 %
Creation	3	27,27 %
Members	4	36,36 %
Summit	4	36,36 %
World	4	36,36 %
Organisation, country	4	36,36 %
Organisation	6	54,55 %
Countries	7	63,64 %
In	9	81,82 %

Table 3: Results of the 3<sup>rd</sup> Experiment

The results obtained indeed show the tight relationship between the acronym *OPEC* and the two words *Organisation* and *Countries*. However, there is an association with a relatively high M-support and M-

confidence that relates *OPEC* to the function word *in*. We consider this association as being noise that may be eliminated if a post-process is added to suppress associations with function words.

4<sup>th</sup> experiment: The corpus studied here is a short biography of President *Barack Obama*. The text is written in Arabic. Upon reading the following table, it can be noted that in the text, *Obama* is strongly associated (M-confidence = 100%) to *Barack* even if the M-support is only 3. It is also noted that in terms of important values for M-confidence, *Obama* is strongly associated to the word pairs *origins, African* and *states, united*. However, there is a weak association of *Obama* with the function words *like* and *of* with an M-confidence of 66.67%. Once more, this type of noise can be eliminated with the addition of a post-process that would suppress the undesired associations.

Y	M-Support	M-Confiance
candidate, last	1	33,33 %
arms	1	33,33 %
president life	1	33,33 %
Washington, American	1	33,33 %
like	2	66,67 %
of	2	66,67 %
states, united	2	66,67 %
origins, African	2	66,67 %
Barack	3	100,00 %

Table 4: Results of the 4<sup>th</sup> Experiment

## 5. CONCLUSION

In general, the results of our experiments seem interesting. The configuration of the classification results seems, in fact, to discourage users who found themselves helpless in the face of “voluminous word lists”. The downstream use of the numerical classification of an extraction process of maximal association rules may help to better read the results of a classification.

In each experiment, the main topic of each document is represented in the extracted association rules, since the first keyword used. We can conclude, in general, that the maximum extraction rules capture all the main topics of the documents.

Maximal association rules are clues that can help the user to understand the content of classes. The M-support and the M-confidence indicate lexical proximity in the documents, but also in the language used and in the areas covered by the textual content of the documents. Of course, associations, M-support and M-confidence are only clues. The most important is that the user no longer has to go through all possible classes.

## 6. REFERENCES

[1] Agrawal, R., Srikant, R., « Fast algorithms for mining association rules in large databases », *In*

*Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of the 20th International Conference on Very Large Data Bases*. Santiago, Chile, 1994.

[2] Agrawal, R., Imielinski, T., Swami, A., « Mining association rules between sets of items in large databases », *In Proceedings of the ACM SIGMOD International Conference on Management of Data*. Washington, 1993.

[3] Amir, A., Aumann, Y., *Maximal association rules: a tool for mining association in text*. Kluwer Academic Publishers, 2005.

[4] Cherfi, H., Napoli, A., « Deux méthodologies de classification de règles d'association pour la fouille de textes », *Revue des nouvelles technologies de l'information*, 2005.

[5] Diop, C. T., Lo, M., « Intégration de règles d'association pour améliorer la recherche d'informations XML », *Actes de la Quatrième conférence francophone en Recherche d'Information et Applications*. Saint-Étienne, 2007.

[6] Damashek, M., « Gauging Similarity with n-Grams : Language-Independent Categorization Of Text », *Science*, 267, 1995, p. 843-848.

[7] Greffenstette, G., « Comparing Two Language Identification Schemes », *Actes des 3<sup>èmes</sup> Journées internationales d'Analyse statistique des Données Textuelles*, Rome, 1995.

[8] Hajek, P., Havel, I., Chytil, M., *The GUHA method of automatic hypotheses determination*. Computing, 1966.

[9] Lallich, S., Teytaud O., « Évaluation et validation de l'intérêt des règles d'association », *Revue des nouvelles Technologies de l'information*, 2003.

[10] Patel, N., Mundur, P., « An N-gram based approach to finding the repeating patterns in musical », *In Proceedings of Euro/IMSA*, Grindelwald, 2005.

[11] Rompré, L. Biskri, I., Meunier, F., « Text Classification: A Preferred Tool for Audio File Classification », *In Proceedings of the 6th ACS/IEEE International Conference on Computer Systems and Applications*, Doha, 2008.

[12] Srivastava, A. N., Sahami, M., *Text Mining, Classification Clustering and Applications*, Chapman & Hall/ CRC Press, 2009.

[13] Turenne, N., Apprentissage statistique pour l'extraction de concepts à partir de textes (Application au filtrage d'informations textuelles), Thèse de doctorat en informatique, Université Louis-Pasteur, Strasbourg, 2000.

[14] Vaillant, B., Meyer P., « Mesurer l'intérêt des règles d'association », *Revue des Nouvelles Technologies de l'Information (Extraction et gestion des connaissances: État et perspectives)*, 2006.

## Visual Assessment Design System

Phillip R. CHASE

Michigan Department of Education

Lansing, MI 48909 USA

**Abstract:** While computerized and Web-based systems for the delivery of assessments of various types are not new, the careful design, planning and development of assessments have not enjoyed a similar mode of operation. An interactive, user-friendly, and visual system of assessment design is now a critical need for the harmonization of the human-computer interface in large-scale assessment design.

**Keywords:** Visual System, Human-Computer Interface, Assessment Systems, Iconization

**Overview:** The advent of automated assessment design systems has taken hold and now is spreading to the state governments themselves, who are beginning to realize a great advantage in developing electronic storage and selection systems themselves. Besides simple storage of items in a database, most item banking systems track important item elements, attributes, and statistics, allowing a user to define multiple criteria along which to select items. The most recent of these systems also include automated algorithms for the selection of items given a bank of appropriate size and along user-chosen criteria.

This would appear to satisfy the needs of a state government to design appropriate assessments for statewide administration, since this would seem to remove the tasks associated with developing an assessment item-by-item in favor of automating selection given a defined blueprint or set of requirements in advance. However, no matter how sophisticated the algorithm, no computerized system can truly make the kinds of judgments that educators—human operators—must make

when considering the unique, complex, and changing needs of students, the “end users” of any reliable and valid assessment instrument.

Thus, I submit that a piece is missing from the puzzle—a critical one that has traditionally been misconceived as negligible and one that can be completely automated. I propose that assessments themselves have *content*—they are not purely mechanical instruments for collecting data. Quality assessments have a “glow and flow,” a way to engage the test-taker with a full appreciation for the whole person and with a respect for the way that information and tasks are accessed by that student. If sophisticated item banking systems include a myriad of tools for the storage, management, organization, selection, and export of passages and items, and if these same systems can automate the random selection of items along certain criteria selected by the user, then a danger exists where the human is no longer involved to address problems related to content and flow; problems arise where the delegation of item choice to a computerized system has damaged the assessment in terms of accessibility, relevance, and meaning.

What are these kinds of problems? Conversely, what opportunities or benefits arise when human operators of a system preserve their power to intervene at critical times? I believe these ideas can relate to assessment development on both a test level and a program level. Since the purpose of a visual assessment design system is to address these issues at a test level, I will not cover program-level issues in detail here—except to say that these are still critical, as they

involve compliance with federal and state laws, integration of stakeholder feedback in design and administration, systems that are adjunct to assessment development such as distribution/delivery and reporting, integrity of test administration security, and other such processes and tasks that are informed by test design but do not govern it.

Rather, there are several design, development, and planning decisions at the test level which can, if addressed by personnel trained in a system *outside* of a computerized item bank, avoid major problems when items are selected. Furthermore, considering these issues add a level of quality to an assessment, reducing measurement error, improving face validity, removing test anxiety, and dampening the effects of test fatigue. I propose that these issues can be categorized into eleven *critical design criteria* for items and passages.

1. Type, or "how the item is set up"
2. Content, or "what the passage or item is about"
3. Value, or "how much the item is worth"
4. Standard, or "what the item assesses"
5. Complexity, or "how rigorous the item is"
6. Access, or "whether students of varying demographics and disabilities can interact with the passage or item equivalently"
7. Context, or "what's around the item or passage"
8. Input, or "what passage or item requires"
9. Key, or "how to answer the item correctly"
10. Maturity, or "if the item is an anchor, operational, or field test item"
11. Difficulty, expressed as a Rasch b-value

While these are not new concepts to test developers, they are decision-making points around which the development of assessments turn. While a few of these can be tracked using item banking software (such as value, standard assessed, and correct

answer), to see any of these as existing in a vacuum at the time an assessment is planned and developed is misguided. Furthermore, to discount how these critical design criteria interact, working in concert to ameliorate student accessibility of items and passages on an assessment—especially given unique needs and abilities of vast and varied populations of students—would lead the test developer down a path fraught with measurement error, disapproving feedback from educators, and ultimately problems of reliability and validity. If computerized item banking systems are left to select items by themselves, many of these criteria will fail to be considered, as no selection algorithm exists to take into account all of these criteria at the same time.

However, this is a main tenet of the visual design system that follows. A human operator performs the oft-misunderstood role of simultaneous, real-time consideration—and often complex negotiation—of all of these critical design points when deciding passages and items to feature on an assessment. A computerized system should be seen as an adjunct to that process, as an assistive, but not independently authoritative, tool for automating certain processes. These processes can shorten the human operator's time in researching certain aspects of items and bring to light important data whose search would otherwise add unnecessary labor, but it is ultimately the role of the trained professional to ensure that all aspects of good assessment design have been considered—including content and flow.

If computer-assisted automation of assessment design is a goal, and since a human operator must be included, a system that ties visual icons or symbols to the eleven critical design criteria is superior. The rationale for this is simple: since these ten criteria must be

considered simultaneously and in real time, and since they interact in ways that make some items more desirable than others when considering the student as a whole, a way to display them to the user without the interruption of referencing multiple documents or sources of data must be held paramount. Even various reports generated by computerized system still require the user to correlate item-level data against other reports showing different, but essential, design criteria. This correlation, while seemingly easy, disrupts the operator's ability to make design decisions when done in rapid succession, such as during times that require the review of groups of items for placement on an assessment.

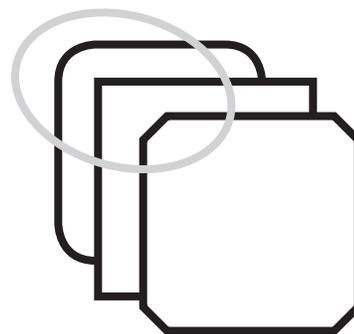
Besides what they can inform on an item-by-item level in terms of content, representing these eleven critical design criteria within a symbolic system allows the human operator to review how entire testlets, text contexts, or individual forms flow in relation to each other or within the assessment as a whole. Besides providing a coded and iconized way for the human operator to visualize aspects of individual passages and items, one of the main benefits of this system is found in the real-time review of inter-item clueing and control of inherent bias. Whereas random choices of items done by computerized system would tend to be insensitive to flow-related issues such as the possibility of one item providing a clue to another, the use of several items in a row featuring topics that favor one gender or the other, or the appearance of passage graphics that provide unintended help to passage-specific items, a visual systems such as this one can allow for easy identification of such issues. As icons for items and passages are laid out in a column, undesirable but correctable trends such as those mentioned here come easily into focus.

It should be noted that this visual system does not compete with the desired functionality of allowing a computerized system to randomize choices of items along pre-determined criteria. This visual system can be used either after or before items have been chosen in this way. First, set parameters of critical criteria can be determined on this system before randomization, translating a planned assessment design into values or ranges of values describing item criteria for randomization. Second, a visual system can be used to articulate with the output of a computerized system as a validation check done by content experts, who, based on the visual icons generated by this system, can make changes or modifications *ex post facto*. To be clear, this visual system is meant to provide the missing *piece* to a puzzle—not supplant the puzzle itself.

**Translating Critical Design Criteria Visually:**

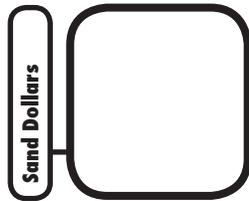
A basic tenet is that each item is represented as a box shape, a single, discrete unit that can be stacked in a column and which can contain other data and features.

The first critical design criterion, Type, is expressed by the corners of the box. A multiple-choice item has rounded corners; a constructed-response item has straight corners; and a performance-based task has truncated corners.



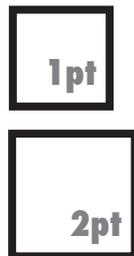
**Critical Design Criterion 1.**

The second design criterion, Content, is expressed by a keyword that appears next to the box. For items, it is suggested that the keyword is taken from the correct answer, or from a general "topic" to which the item relates.



**Critical Design Criterion 2.**

The third critical design criterion, Value, assumes that the size of the box is proportional to the number of possible operational points it scores. With a base size set as a value of one point, the area of the box can be multiplied to reflect a point value greater than one. This area difference can be represented by squares along a set grid (leaving items with odd numbers of point values as L-shapes) or along proportional height and width amounts (giving the impression of boxes of various sizes).

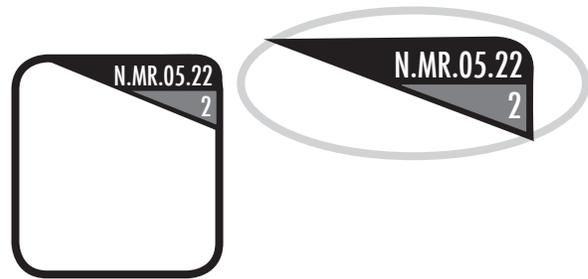


**Critical Design Criterion 3.**

The fourth critical design criterion, Standard, can be represented as a flag inside the box, bearing a code for the instructional domain, strand, and benchmark.

The fifth critical criterion, Complexity, can be expressed as a flag that bears a numerical code for a taxonomy of cognitive processing, such as the Webb Depth of Knowledge matrix

or a rating system of linguistic difficulty.



**Critical Design Criteria 4 and 5.**

The sixth critical design criterion, Access, can be at least partially understood as "bias," although strong bias that would affect a student's ability to answer an item is expectedly eliminated by educator committees. Rather, this is a mild type of bias or slant that is inherent in the wording or graphic that may cause some items and passages to bear a certain character that can be aligned with one gender or another, one ethnicity over another, or a certain age range. For example, the depiction of a light-haired girl in an illustration might slightly "slant" toward Caucasians and/or toward females.

Alternately, the concept of Access also covers the possibility that some items and passages can erroneously assume that students can understand them without consideration of a student's disabilities (for example, a Reading passage about colors provided on a Braille assessment).

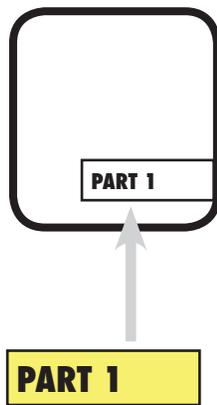
- Sample Codes  
 N: No Gender  
 N: No Ethnicity  
 M: Male  
 F: Female  
 C: Caucasian  
 Af: African-Am.  
 Ar: Arabic  
 L: Latino  
 As: Asian  
 etc.

**Critical Design Criterion 6.**

The seventh critical criterion, Context, provides a schema for associating items within a testlet. It might be used to mark parts of an assessment where students need

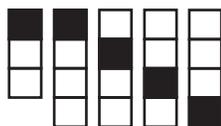
to be exposed to new directions, sample items, stimuli, prompts, or passages, and whose items fall within a predetermined set. This criterion can be understood as "Format" for certain assessments, such as English language assessments, where students encounter items that ask for different language processes within the same domain.

The eighth critical criterion, Input, is used to track the student's need for passage or item graphics, legends, notes, or formulae, and the use of materials such as a calculator in order to correctly access to passage or item. The Context text area can be color-coded or shaded to show the need for illustrations or photos, for example.



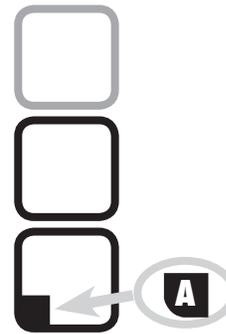
**Critical Design Criteria 7 and 8.**

The ninth critical criterion, Key, refers to the correct answer for a multiple-choice response, or the existence of a rubric for scoring constructed-response or performance-based items.



**Critical Design Criterion 9.**

A tenth critical criterion, Maturity, designates whether or not the item is a field-test item (an item whose point score does not contribute to a student's raw score but whose response data will help determine item information), an operational item (an item whose point score contributes to scoring) or an anchor item (an operational item shared between two cycles of an assessment and whose difficulty is used as part of a comparative model between the two cycles). An item that is an anchor item or an operational one will appear as a strong, bold icon, while a field-tested one will appear as a faded-out version. A small "A" mark distinguishes between anchor and operational items.



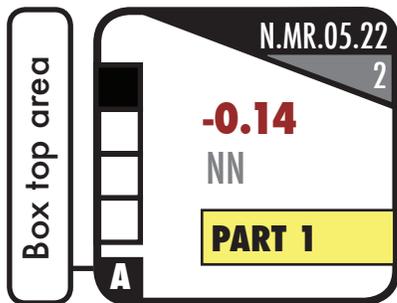
**Critical Design Criterion 10.**

Finally, the eleventh critical criterion, by no means the least important, is a Rasch b-value assigned to the item if it is operational or anchor. Displaying this along with all of the other item criteria in a visual format will help the user predict whether or not the items chosen are suitable for use in an equating model. A value displayed in red is handy for quick reference.

**-0.57**

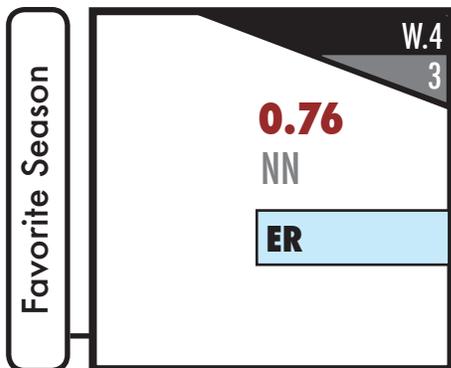
**Critical Design Criterion 11.**

**Sample MEAP Item:** A sample item from the Michigan Educational Assessment Program (MEAP) Mathematics assessment is shown here. From one glance, a developer can see that it is a multiple-choice item that scores 1 point; it assesses a certain Math GLCE at a depth of knowledge (DOK) of level 2; its correct answer, of four possible choices, is A; it has a b-value of -0.14 and is intended as an anchor item; it should have no “slant” towards one gender or one ethnicity; it is included in Part 1 of the assessment at this grade level and requires an illustration; and it the key word may help in catching inter-item clueing or give an idea of what the item asks the student to do.



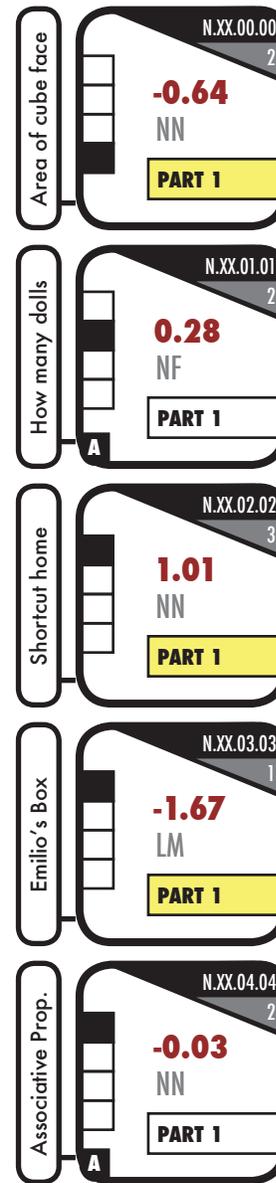
**Sample MEAP Mathematics Item as Visual Icon.**

**Sample Constructed-Response Item:** This might be an item a student encounters on the ELPA Writing assessment; it uses ELPA-specific codes (such as W.4 as the ELP Writing standard) and formats (such as ER, for Extended Response). It scores a maximum of four points, as obvious by its size compared to other icons. A photograph is required for this item.



**Sample ELPA Constructed-Response Item.**

**Sample Testlet:** By displaying items in a vertical line, entire testlets can be viewed at once. How their assigned items interact and function as a group can be thus better predicted by the developer. In addition, the developer can instantly see how many graphics are needed, whether or not correct answer keys are evenly distributed and are free of patterning or repetition, whether or not the testlet will appeal too much to one gender or ethnicity, and how “difficult” the testlet is overall.



**Sample Testlet in Vertical Format.**

## An information hiding technique based on visual cryptography

Te-Yu Chen<sup>1</sup>, Hsia-Hung Ou<sup>2</sup>, Yong-Jay Yang<sup>3</sup>, and Jing-Hang Tseng<sup>3</sup>

Department of Information Networking Technology, Hsiuping Institute of Technology<sup>1</sup>  
NO.11, Gongye Rd., Dali Dist., Taichung City 41280, Taiwan (R.O.C.)

Department of Computer Science and Information Engineering, Hwa-Hsia Institute of Technology<sup>2</sup>  
No.111, Gongzhuan Rd., Zhonghe Dist., New Taipei City 23568, Taiwan (R.O.C.)

Department of Information Management, Hsiuping Institute of Technology<sup>3</sup>  
NO.11, Gongye Rd., Dali Dist., Taichung City 41280, Taiwan (R.O.C.)

### ABSTRACT

Visual Cryptography was first proposed by Moni Naor and Adi Shamir in Eurocrypt'94. They demonstrated a  $(t, n)$  visual secret sharing scheme, in which visual information is encoded into  $n$  shares so that only someone with  $t$  or more than  $t$  shares could decrypt this visual information, while nothing would be revealed by any  $t-1$  shares. Particularly, the decryption is performed directly by stacking shares and the concealed information is gained by human visual system without any cryptographic computations. Additionally, a keyed  $(t, n)$  visual secret sharing scheme is also considered, in which some specific shares, called keyed shares, must be involved to recover the visual information. Without the help of the keyed shares, no information will be retrieved. Information hiding is the art of imperceptibly embedding some information into the digital work and getting it more or less modified as a result. In this paper, we design a scheme in which the redundancy in visual secret sharing scheme is employed for embedding additional information. According to the similar concept, information hiding technique based on the  $(2, 2)$ ,  $(2, 3)$ ,  $(3, 3)$  visual secret sharing schemes and  $(2, 3)$  keyed visual secret sharing scheme are proposed. From the experimental results, the proposed schemes not only contribute the functionality of visual cryptography, but also provide with the ability of information hiding.

**Keywords:** Visual cryptography, Information hiding, steganography, Information security, Watermarking, Visual secret sharing.

### 1. INTRODUCTION

Owing to the rapid development of information technologies, we are now living in a highly digitalized environment. However, along with all the convenience has brought, there have also been quite some security threats going on that need to be taken care of. For example, how to transmit confidential information over insecure channel is a primitive and crucial problem. Many sophisticated solutions have been proposed to deal with the security issues related to cryptography [1, 2, 3]. However, most of them are based on the strength of abstruse mathematics and tend to be extremely time-consuming. Therefore, some other solutions such as visual cryptography and information hiding techniques have attracted many interests.

Information hiding is the art of imperceptibly embedding some information into the digital work and getting it more or less

modified as a result. Because there is a limit to the sensitivity of human beings' sensory organs such as the eyes and ears, it is difficult for human beings to distinguish the subtle differences made to multimedia works, e.g., image, sound, video, 3D object, etc. [4]. Based on this property, some secret information can be embedded into a multimedia work in such a manner as to conceal the existence of the secret information from the irrelative parties. Therefore, information hiding technique can be considered as another admissible solution for the transmission of the confidential information [4, 5, 6].

A  $(t, n)$  Visual secret sharing scheme, a security mechanism which combines visual cryptography [7] and secret sharing [8], decomposes a secret image into  $n$  shadow images called shares such that in order to recover the secret image, one has to stack  $t$  or more than  $t$  shares. On the other hand, the recovery of the secret image does not need any arithmetic computation; it only needs human visual system to recover the image. And no one can get any information about the secret image while there are fewer than  $t$  shares obtained. Visual Cryptography was first proposed by Moni Naor and Adi Shamir in Eurocrypt'94 [7]. Up to now, the visual cryptography techniques have been well studied. A large number of variant visual cryptography techniques for various environments have been proposed [9, 10, 11, 12, 13, 14, 15]. Additionally, in this paper, we consider a certain application of  $(t, n)$  visual secret sharing scheme in which the weight of these shares may be not equal. Some specific shares must be involved to recover the secret image; these specific shares are called keyed shares. Without the help of the keyed shares, no information will be retrieved no matter how many non-keyed shares are collected. This kind of schemes is named as keyed  $(t, n)$  visual secret sharing scheme.

In this paper, we first consider  $(2, 2)$ ,  $(2, 3)$ ,  $(3, 3)$  visual secret sharing schemes and  $(2, 3)$  keyed visual secret sharing scheme. Combining the concepts of visual cryptography and information hiding, an information hiding technique based on visual cryptography is proposed in this paper. The proposed scheme is detailed in Section 2. Some experimental results and analysis are presented in Section 3. Finally, some conclusions are given in Section 4.

### 2. THE PROPOSED SCHEMES

In this section, the methods of how to hide additional information into the  $(2, 2)$ ,  $(2, 3)$ ,  $(3, 3)$  visual secret sharing schemes and  $(2, 3)$  keyed visual secret sharing scheme are discussed. There are two portions of information, the visual

information and the additional information, which can be concealed in these visual secret sharing schemes. The visual information is embedded by the general visual cryptography. On the other hand, the additional information is embedded by exploiting the redundancy while embedding the visual information.

**(2, 2) visual secret sharing scheme**

In this subsection, we design a technique in which the redundancy in a 2 out of 2 visual secret sharing scheme is employed for embedding additional information. There are two portions of information, the visual information and the additional information, which can be concealed in a (2, 2) visual secret sharing scheme. To embed the visual information, each pixel of this visual information is extended to a 2x2 block in both 2 shares, Share A and Share B. There are six states for a 2x2 block, that is, , , , , , and . In addition, each of the six states is represented for a digit of a senary (base-6) numeral system. A white pixel of the visual information is expanded to two identical blocks chosen from these states. One of the selected blocks is assigned to Share A, and the other is assigned to Share B in the corresponding positions. On the other hand, a black pixel of the visual information is expanded to two complementary blocks chosen from these states in the two shares. Similarly, one is assigned to Share A, and the other is assigned to Share B. Different from a common visual cryptographic system, the choice of expansion from the six states is referred to the additional information. To embed the additional information, the additional information is first converted into its senary numeral format. The order of these senary digits is then rearranged by a chaotic mapping using a secret key in terms of security. Thereafter, the choice of expansion from the six states is made according to the pixel of the visual information and the senary digit of the additional information. Based on the selection, Share A and Share B are generated accordingly. It is clear that one can obtain the visual information by stacking Share A and Share B. However, the additional information only can be retrieved by one who has the knowledge of the permutation secret key and the mapping between the states and the senary digits. Therefore, with the help of simple computation, the additional information can be retrieved. Note that, the additional information is performed by converting into senary and rearranging the order of its senary digits before being embedded. This preprocess makes a general data in a very irregular manner. Therefore, the selection of the states tends almost towards randomness, and in consequence, non information about the visual information or the additional information would expose to the shares.

**(2, 3) visual secret sharing scheme**

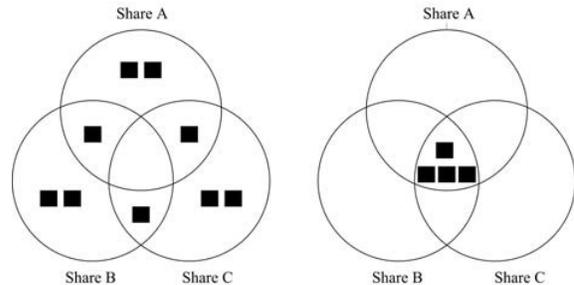
In this subsection, we first introduce our design of the (2, 3) visual secret sharing scheme for embedding the visual information. Then, the method of how to embed the additional information is proposed. The (2, 3) visual secret sharing scheme, in which visual information is encoded into 3 shares so that only someone collecting any 2 or more than 2 shares from them could decrypt this visual information, while nothing would be revealed by any single share. In order to improve the perception of the visual information, each pixel of this visual information is extended to a block of 3x3 pixels in all shares, Share A, Share B, and Share C. As shown in Table 1, the compositions of blocks in shares and their stacking result are defined as follows. (i) There are four black pixels and five white pixels in each block

among all shares. (ii) There are seven black pixels and two white pixels in the block of the stacking result if the corresponding pixel of the visual information is black. (iii) There are four black pixels and five white pixels in the block of the stacking result if the corresponding pixel of the visual information is white.

Table 1. The compositions of blocks in (2, 3) VSS

	Black	White
Block in shares	4B5W	4B5W
Block in stacking 2 shares	7B2W	4B5W
Block in stacking 3 shares	9B0W	4B5W

Venn diagrams [16, 17] usually used in set theory are adopted for designing the block of the visual secret shares in this paper. The union of sets which represent blocks in shares stands for the stacking result of blocks. As illustrated in the left-hand Venn diagram of Figure 1, each circle represents the block in shares in which the stacking result is intended for black. Four pixels in each block should be black. Only one pixel in the intersection of any two blocks should be black, i.e., there is one black pixel located at the same position in corresponding blocks for any two shares. None pixel in the intersection of all blocks should be black, i.e., none black pixel is located at the same position in corresponding blocks for all three shares. Therefore, after stacking any two blocks, there are seven black pixels and two white pixels in the stacking result. On the other hand, there are four black pixels in the intersection of these three blocks, i.e., four black pixels are located at the same position in corresponding blocks for all three shares if the stacking result is intended for white. Hence, there are four black pixels and five white pixels in the stacking results if any two blocks are stacked.



(i) Visual information pixel is black (ii) Visual information pixel is white

Figure 1. The block design of (2, 3) VSS

The method of how to embed the additional information is similar to that in (2, 2) visual secret sharing scheme. If the visual secret pixel is black, there are  $\frac{9!}{2! \cdot 2! \cdot 2!}$  arrangements possible for designating Share A, Share B, and Share C. Accordingly,  $\log_2 \frac{9!}{2! \cdot 2! \cdot 2!}$  bits of additional information can be embedded into each black block. On the other hand, if the visual secret pixel is white, there are  $C(9,4)$  arrangements possible for designating Share A, Share B, and Share C. Accordingly,  $\log_2 C(9,4)$  bits of additional information can be embedded

into each white block.

**(3, 3) visual secret sharing scheme**

The (3, 3) visual secret sharing scheme, in which visual information is encoded into 3 shares so that only someone collecting all of the three shares could decrypt this visual information, while nothing would be revealed if any one share is absent. Similar to the (2, 3) visual secret sharing scheme, each pixel of this visual information is extended to a 3x3 block in all shares, Share A, Share B, and Share C. Table 2 defines the compositions of blocks in shares and their stacking results; and Figure 2 shows the block design of this scheme. We can observe that after any two shares are stacked, the numbers of black pixels in every block are the same no matter whether the visual information pixel is black or white. For example, if Share A and Share B are stacked, there are seven black pixels in a block no matter whether the visual information pixel is black or white. Therefore, none of the visual information will be revealed from stacking any two shares. However, when all of the three shares are stacked, all pixels (9 pixels) are black in a block if the visual information pixel is black; while there are seven black pixels in a block if the visual information pixel is white. Hence, the contrast between the two types of blocks reveals the visual information.

Table 2. The compositions of blocks in (3, 3) VSS

	Black	White
<b>Block in shares</b>	4B5W/ 5B4W	4B5W/ 5B4W
<b>Block in stacking 2 shares</b>	6B3W/ 7B2W	6B3W/ 7B2W
<b>Block in stacking 3 shares</b>	9B0W	7B2W

The method of how to embed the additional information is similar to that in (2, 3) visual secret sharing scheme. If the visual secret pixel is black, there are  $\frac{9!}{2! \cdot 2! \cdot 2! \cdot 3!}$  arrangements possible for designating Share A, Share B, and Share C. Accordingly,  $\log_2 \frac{9!}{2! \cdot 2! \cdot 2! \cdot 3!}$  bits of additional information can be embedded into each black block. On the other hand, if the visual secret pixel is white, there are  $\frac{9!}{2! \cdot 2! \cdot 2!}$  arrangements possible for designating Share A, Share B, and Share C. Accordingly,  $\log_2 \frac{9!}{2! \cdot 2! \cdot 2!}$  bits of additional information can be embedded into each white block.

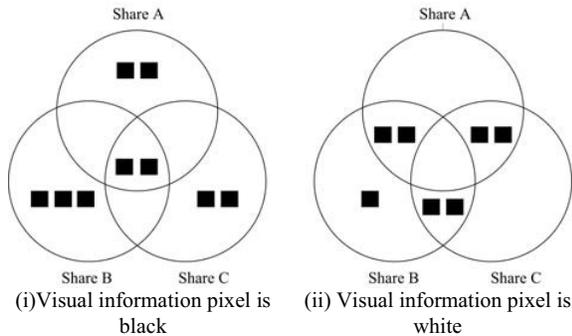


Figure 2. The block design of (3, 3) VSS

**(2, 3) keyed visual secret sharing scheme**

The (2, 3) keyed visual secret sharing scheme, in which visual information is encoded into 3 shares, where Share A and Share B are common shares and Share C is the keyed share, so that only someone collecting Share C and one of the common shares could decrypt the embedded visual information, while nothing would be revealed if the keyed share or both the common shares are absent. Similar to the above visual secret sharing schemes, each pixel of the visual information is extended to a 3x3 block in all shares. Table 3 defines the compositions of blocks in shares and their stacking results; and Figure 3 shows the block design of this scheme. We can observe that if Share A and Share B are stacked there are seven black pixels in a block no matter whether the visual information pixel is black or white. Therefore, none of the visual information will be revealed from stacking Share A and Share B. However, when the keyed share, Share C, joins in the decryption of visual secret no matter whether the other share is Share A or Share B, there are eight pixels are black in a block if the visual information pixel is black; while there are seven pixels are black in a block if the visual information pixel is white. Hence, with the help of Share C, the visual information could be revealed.

Table 3. The compositions of blocks in (2, 3) keyed VSS

	Black	White
<b>Block in Share A/B</b>	4B5W	4B5W
<b>Block in Share C</b>	6B3W	6B3W
<b>Block in stacking A and B</b>	7B2W	7B2W
<b>Block in stacking A and C/ B and C</b>	8B1W	7B2W

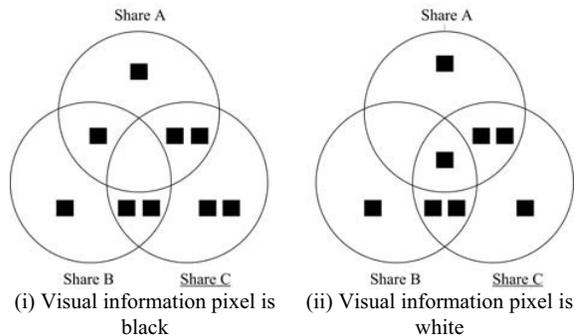


Figure 3. The block design of (2, 3) keyed VSS

The method of how to embed the additional information is similar to that in the previous visual secret sharing schemes. If the visual secret pixel is black, there are  $\frac{9!}{2! \cdot 2! \cdot 2!}$  arrangements possible for designating Share A, Share B, and Share C. Accordingly,  $\log_2 \frac{9!}{2! \cdot 2! \cdot 2!}$  bits of additional information can be embedded into each black block. On the other hand, if the visual secret pixel is white, there are  $\frac{9!}{2! \cdot 2! \cdot 2!}$  arrangements possible for designating Share A, Share B, and Share C. Accordingly,  $\log_2 \frac{9!}{2! \cdot 2! \cdot 2!}$  bits of additional information can be embedded into each white block.

### 3. ANALYSIS AND EXPERIMENTAL RESULTS

Some experiments are performed to validate the feasibility of the proposed schemes. Figure 4 and 5 show the experimental results of the hiding technique based on (2, 2) visual secret sharing scheme. As shown in Figure 4, the visual information and the additional information which has been transformed to the ternary numeral system are loaded in the top-left and the bottom of Figure 4 respectively; and the two shares are generated in the bottom of Figure 4.

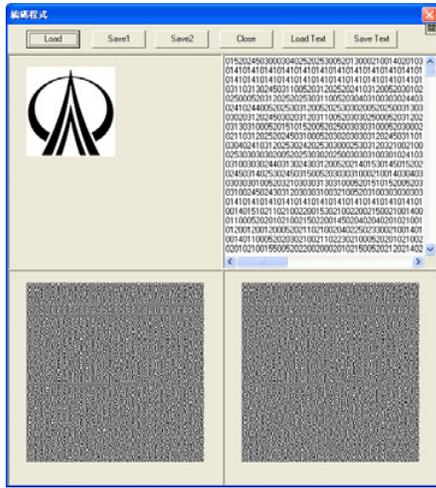


Figure 4. The encoding procedure of (2, 2) VSS

The two shares are loaded in the left of Figure 5. The stacked result of the two shares is illustrated in the bottom-right of Figure 5. The additional information is extracted and shown in the top-right of Figure 5.

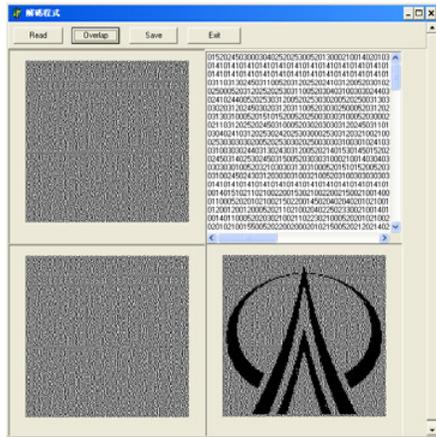


Figure 5. The decoding procedure of (2, 2) VSS

The binary logo image as shown in the top-left of Figure 4 is adopted as the visual information for all experiments in this paper. Owing to each pixel in the visual information is extended to 3x3 pixels, all the shares and stacking results are 3x3 times the size of the binary logo in the (2, 3), (3, 3) visual secret sharing schemes and (2, 3) keyed visual secret sharing schemes.

Figure 6 illustrates the experimental results of the proposed (2, 3) visual secret sharing scheme. All of them confirm the feasibility of our design.

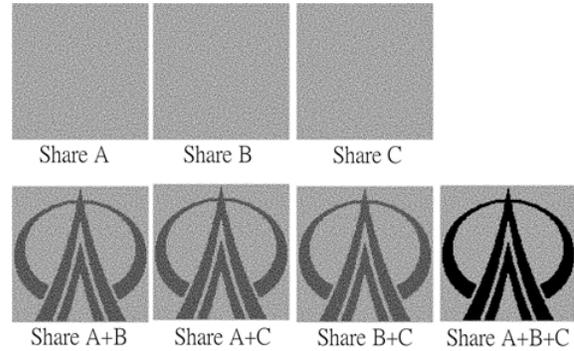


Figure 6. The experimental results of (2, 3) VSS

The experimental results of the proposed (2, 3) visual secret sharing scheme are demonstrated in Figure 7. Share B is darker than the other two shares which meets the block design of (3, 3) visual secret sharing scheme as shown in Figure 2. The stacking results of Share A and B as well as Share B and C, which are darker, also arise from this design. As the expectation, the stacking results of any two shares reveal nothing. However, the visual information is revealed only by stacking all the three shares.

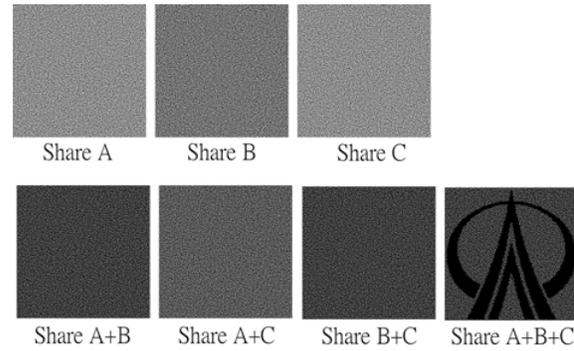


Figure 7. The experimental results of (3, 3) VSS

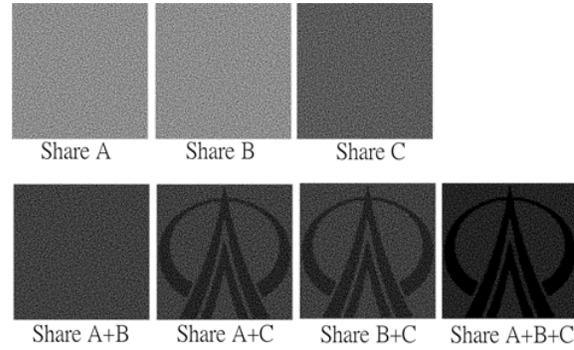


Figure 8. The experimental results of (2, 3) keyed VSS

Figure 8 shows the experimental results of the proposed (2, 3) keyed visual secret sharing scheme, in which Share C is the

keyed share. As shown in Figure 8, stacking Share A and Share B does not reveal any information. Nevertheless, the visual information could be retrieved only when the keyed share, Share C, is present. Owing to (3, 3) keyed visual secret sharing scheme does not makes sense, we discard it in this paper.

For the security concern, the additional information is embedded after performing cryptographic encryption. Because of the properties of cryptographic functions [1, 2, 3], the encrypted data is close to randomness. Hence, the selection of blocks in the share is nearly random. The redundancies of the decisions of blocks in the shares are employed for hiding the additional information. Therefore, the hiding capacity is fully depend on the number of options which the block could be. Table 4 demonstrates the hiding capacities of the additional information with respect to the proposed visual secret schemes.

Table 4. Hiding capacity (bits/block)

Scheme	Black block	White block	Average
(2, 2) VSS	$\log_2 6$	$\log_2 6$	2.58
(2, 3) VSS	$\log_2 \frac{9!}{2!2!2!}$	$\log_2 C(9, 4)$	14.68
(3, 3) VSS	$\log_2 \frac{9!}{2!2!2!3!}$	$\log_2 \frac{9!}{2!2!2!}$	14.18
(2, 3) keyed VSS	$\log_2 \frac{9!}{2!2!2!}$	$\log_2 \frac{9!}{2!2!}$	15.97

#### 4. CONCLUSIONS

In this paper, information hiding technique based on the (2, 2), (2, 3), (3, 3) visual secret sharing schemes and (2, 3) keyed visual secret sharing scheme are proposed. The analysis and experimental results prove that the proposed schemes not only contribute the functionality of visual cryptography, but also provide with the ability of information hiding. In the design of (3, 3) visual secret sharing scheme, Share B is darker than the other two shares. How to equally arrange the black pixels in all blocks of shares deserves to be taken into consideration carefully. Furthermore, only the information hiding technique based on some specific (t, n) visual secret sharing scheme and keyed visual secret sharing scheme are discussed in this paper. The generalization of (t, n) visual secret sharing scheme or keyed visual secret sharing scheme is worthy of further study.

#### 5. ACKNOWLEDGEMENT

This work was partially supported by National Science Council, Taiwan, R.O.C., under Grant No. NSC 100-2218-E-164-002.

#### 6. REFERENCES

[1] Stallings, W, Cryptography and Network Security, Prentice Hall, 4<sup>th</sup> edition, 2006.  
 [2] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone, Handbook of Applied Cryptography, CRC, 1996.  
 [3] B. Schneier, Applied Cryptography, John Wiley & Sons, 2<sup>nd</sup> edition, 1996.  
 [4] T. Y. Chen, M. S. Hwang, and J. K. Jan, Adaptive authentication schemes for 3D mesh models, International Journal of Innovative Computing, Information and Control, Vol. 5, No. 12(A), pp. 4561-4572, December 2009.  
 [5] P. C. Wang and C. M. Wang, Reversible data hiding for

point-sampled geometry, Journal of Information Science and Engineering, Vol. 23, No. 6, pp.1889-1900, 2007.  
 [6] N. I. Wu and M. S. Hwang, Data hiding: Current status and key issues, International Journal of Network Security, Vol. 4, No. 1, pp.1-9, 2007.  
 [7] M. Naor, A. Shamir, Visual Cryptography, Advances in Cryptography, Eurocrypt '94, pp. 1-12, 1995.  
 [8] A. Shamir, How to Share A secret, Communications of the ACM, Vol. 22, No. 11, pp. 612- 613, November 1979.  
 [9] T. H. Chen and D. S. Tsai, Owner–customer right protection mechanism using a watermarking scheme and a watermarking protocol, Pattern Recognition, Vol. 39, No. 8, pp. 1530–1541, 2006.  
 [10] W. P. Fang and J. C. Lin, Visual cryptography with extra ability of hiding confidential data, Journal of Electronic Imaging, Vol. 15, No. 2, pp. 023020-1- 023020-7, Apr. 2006.  
 [11] W. P. Fang, Friendly progressive visual secret sharing, Pattern Recognition, Vol. 41, No. 4, pp. 1410–1414, April 2008.  
 [12] S. J. Shyu, S. Y. Huanga, Y. K. Lee, R. Z. Wang and K. Chen, Sharing multiple secrets in visual cryptography, Pattern Recognition, Vol. 40, pp. 3633–3651, 2007.  
 [13] T. H. Chen and K. H. Tsao, Visual secret sharing by random grids revisited, Pattern Recognition, Vol. 42, No. 9, pp. 2203–2217, 2009.  
 [14] C. C. Wu, M. S. Hwang, and S. J. Kao. A new approach to the secret image sharing with steganography and authentication, The Imaging Science Journal, Vol. 57, No. 3, pp. 140-151, 2009.  
 [15] T. H. Chen and K. H. Tsao, Threshold visual secret sharing by random grids, Journal of Systems and Software, Vol. 84, No. 7, pp. 1197-1208, July 2011.  
 [16] R. P. Grimaldi, Discrete and Combinatorial Mathematics: An Applied Introduction, Addison Wesley Longman, Fifth Edition, 2003.  
 [17] F. Ruskey and M. Weston, A Survey of Venn Diagrams, Electronic Journal of Combinatorics, Dynamical Surveys DS5, June 18, 2005, <http://www.combinatorics.org/Surveys/#DS5>.

# Implementation of 3D Stereo Image Capture System Based on Multi-Segmented Method

Woonchul Ham<sup>1</sup>, Hyeokjae Kwon<sup>2</sup> and T. Enkhbaatar<sup>2</sup>

<sup>1</sup>Professor, <sup>2</sup> Ph.D student of Division of Electronics and Information Engineering

Chonbuk National University, Korea

Email: wcham@chonbuk.ac.kr

## ABSTRACT

In this paper, we suggest a new modeling of CMOS camera by using combinations of several pinhole camera models and check its validity by using synthesized stereo images based on OpenGL software. We introduce 3D image capturing hardware system built in my laboratory which consists of 5 motor controller and two CMOS camera module based on S3C6410 processor. We choose 9 segmentations and propose the method to find optimal alignment and focusing based on the measure of alignment and sharpness and propose the synthesizing fusion with the optimized 9 segmentation images for the best 3D depth feeling. From the experimental results concerning disparity values in each 9 segments, we can assure that the multi-segment method proposed in this paper may be one of good methods which improve the better feelings of 3D depth in stereo image.

**Keywords:** Pinhole, Segmented Image, Multi-segment, Disparity, CMOS camera module

## 1. INTRODUCTION

3D reconstruction from 2D images is an active research topic in the computer vision community. Some recent works are focused on interactive 3D reconstruction algorithms. The CMU 3D room propose an algorithm for view-dependent nonuniform sampling for image-based rendering. This algorithm may be needed for making a good 2D image for many viewers when each of them has his own different geometric view plane[1]. A taxonomy and evaluation system proposed by D. scharstein [2] is widely used in evaluating the disparity accuracy in stereo matching algorithms with a set of benchmark image data sets. There has been many researches for estimation of optical flow especially for estimating of an object by using 2D images under the assumption that the light source is invariant with respect to movement of the object which is not true in real life. Fleet and Weiss provide a tutorial introduction to gradient based optical flow [4] [5] under a image constraint equation which is not always true in 2D image. The perceived depth seen in a stereoscopic images varies not only with many geometric parameters but also with the many parameters with human eye such as separation between two eyes and focusing of each eye. It is well known fact that there is a limit to the range of perceived depth. Human can not detect the depth difference between two objects if two objects are located beyond 350m from the viewer. So it is useless to consider the scene depth of object if its distance is very far from the viewer. This is the reason why the research should be done concerning relationship and optimal mapping between scene depth and perceived depth for the best implementation of stereoscopic image. N.Hollman address this problem and propose multi-region algorithm and this algorithm allows different regions of the scene to be mapped to the different ranges of perceived depth on a target 3D display [6]-[8].

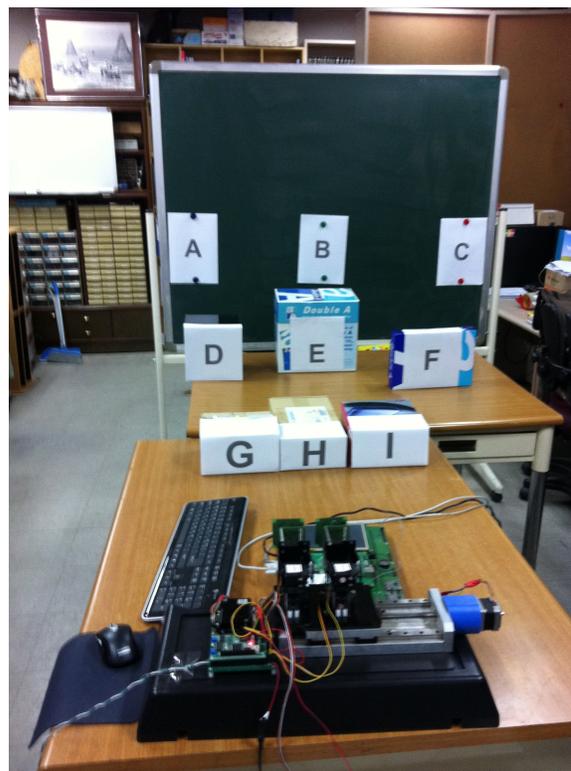


Fig. 1. Image capturing system

We also proposed new mapping algorithm between scene depth and perceive depth [9]. In this paper, we propose a theoretical and technical ideas for developing a 3D stereo video camera system based on previous research works. We comments on the important geometric parameters for synthesizing fusion with the multi-captured images. We simulate our research with OpenGL software to obtain 9 segmentation of the scene with the assumption that human do not rotate his head and just rolling his eye to watch the scene. We calculate and decide geometric parameters based on this assumption for obtaining virtual 9 segmentations of the scene. We introduce the 5 degree of freedom(DOF) stereo camera system which is the first version of our laboratory, as shown in figure 1. We use 4 DC servo motor and one stepper motor for mechanical control. We also develop a control algorithm to align the two left and right camera capture system by measuring the offset distances. We begin with a brief review and mathematical preliminary for the technology and theory for 3D image processing and propose an decision algorithm whether an object is within the trapezoidal region or not and also comment on the projection theory. In section 3, we discuss the technical issue for calibration and suggest

a simple control law for adjusting the focus and alignment for capturing several segmentations of image. We also propose a new performance index measure for focus and alignment problem in capturing the left and right images. In section 4, we suggest simple fusion algorithm to make a whole 3D stereo image from the several segments of image which are captured by stereo camera system operating with the proposed control algorithm. In section 5, we simulate our proposed algorithm with the virtual objects synthesized on circular polarized LCD screen by using OpenGL programming and we also introduce the hardware system and comment on the disparity. We conclude with a summary of this paper and a discussion of planned future work.

## 2. MATHEMATICAL PRELIMINARY

In this section, we introduce the mapping theory that is necessary to understand and design the parameters for the 2D projection view of 3D object. In stereo vision signal processing, we should synthesize the 3D object into the 2D circular polarized LCD screen. The key idea is that planar mapping function is dependent on the geometric parameters between LCD and the eyes of human who is sitting in front of LCD. This is the reason why we comments on the geometric parameters between object and camera. We insist that this geometric parameters play very essential role for synthesizing a good quality of 3D stereo image. Now we introduce the mathematics for projection process from the 3D image space to 2D image plane for synthesizing 3D stereo image into 2D polarized LCD screen. We also propose the algorithm whether or not a point is inside of trapezoidal region to decide whether a point should be considered or not in synthesizing a 3D stereo image.

### A. Geometric Parameters

The geometric coordinate relation between the eye of the viewer and virtual LCD screen can be expressed as follows.

$$T_{se} \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} = \begin{pmatrix} x_e \\ y_e \\ z_e \end{pmatrix} \quad (1)$$

where

$$T_{se} = R_{(x, -\theta_x)} R_{(y, \theta_y)} T_{(y, -d_y)} T_{(x, -d_x)} T_{(z, -d_z)} \quad (2)$$

In this paper, the virtual LCD denotes the 2D screen window through which a viewer watch the objects and we should synthesize this virtual scene into real LCD screen located at with the same orientation and position as those of virtual LCD. The scene synthesized with this suggested method is the the optimized scene under the assumption that a person is going to watch this scene with the same geometric relation with real LCD screen. In there,  $(x_s, y_s, z_s)$  denotes the coordinate of the virtual LCD screen coordinate system and  $(x_e, y_e, z_e)$  denotes the coordinate of the eye coordinate system of the viewer as shown in figure 2. The main problem to solve in this paper is to obtain the equation that makes the objects which is behind the virtual LCD to be projected onto the virtual LCD screen in the the direction of the eye view point of viewer. We define the trapezoidal region  $\mathfrak{R}$  which will be projected onto the virtual LCD by using the 4 corner points  $p_1, p_2, p_3, p_4$  and the range of scale  $k$  as follows in the eye coordinate system.

$$\mathfrak{R} = \text{trapezoidal}(p_1, p_2, p_3, p_4, k) \quad (3)$$

The trapezoidal region  $\mathfrak{R}$  is shown in figure 3. In the next

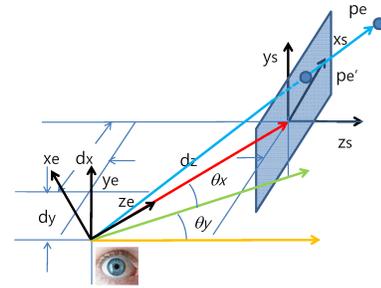


Fig. 2. Geometric Parameters

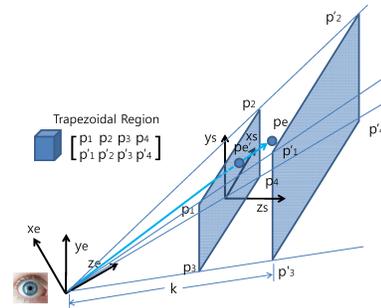


Fig. 3. Trapezoidal Region

subsection, we propose an algorithm whether a point is within the trapezoidal or not.

### B. Belonging Decision Algorithm

In this subsection, we propose an algorithm whether or not a point is inside of trapezoidal region to decide whether a point should be considered or not in synthesizing a 3D stereo image. This algorithm is needed to make a stereo 3D image from the synthesized scene and objects made by OpenGL software. Let a point in side of trapezoidal be  $P_0$  and its coordinate in eye coordinate system be  $(x_0, y_0, z_0)$ . We try to find 6 planar equations for the trapezoidal with its value at  $P_0$  be positive. The equation of each plane  $S_n$  is determined by any arbitrary three corner points of the corresponding surface of the trapezoidal. Let  $(a_n, b_n, c_n, d_n)$  be the parameter vector of each surface. Then the equation of each surface can be expressed as follow.

$$S_n(x, y, z) = a_n x + b_n y + c_n z + d_n = 0, \quad n = 1 \cdots 6 \quad (4)$$

Then we propose an algorithm whether a point is inside of trapezoidal or not as a theorem 1.

*Theorem 1:* If all the values of  $S_n(x, y, z)$  at a point  $P$  are positive, then the point  $P$  is located inside of trapezoidal.

### C. Projective Mapping Function

We propose a projective mapping as a theorem as follow under the assumption that a point is inside of trapezoidal region we are interested in. In here, we only consider a object which lie in side of the trapezoidal characterized by 4 vertex which denote the virtual LCD screen in the eye coordinate system and scale factor  $k$  expressed in equation (3).

*Theorem 2:* For a given 4 vertex  $p_1, p_2, p_3, p_4$  with the scale factor  $k$ , the projective point  $p'_e$  of the a point  $p_e$  which is located

TABLE I  
DYNAMIC RANGE OF CAMERA MODULE(UNIT: DEGREE)

$\theta_{LM}$	$\theta_{RM}$	$\theta_{MU}$	$\theta_{MD}$	$\theta_{LU}$	$\theta_{RD}$
-27.3	27.3	22.1	-22.1	33.3	-33.3

in the corresponding trapezoidal can be expressed as follow.

$$p'_e = P_{roj} p_e \quad (5)$$

Let  $p'_s$  is the corresponding coordinate of  $p'_e$  with respect to the coordinate system of virtual LCD screen. Then we obtain

$$p'_s = P_{roj}' p_s \quad (6)$$

where

$$P_{roj}' = T_{se}^{-1} P_{roj} T_{se} \quad (7)$$

and  $P_{roj}$  is the projective operation in the eye coordinate system with respect to trapezoidal region.

*Proof:* From the definition of  $T_{se}$  expressed in equation (2), we have

$$\begin{aligned} p'_s &= T_{se}^{-1} p'_e \\ &= T_{se}^{-1} P_{roj} p_e \\ &= T_{se}^{-1} P_{roj} T_{se} p_s \end{aligned} \quad (8)$$

Therefore, we prove equation(7).  $\blacksquare$

### 3. EXPERIMENTAL ENVIRONMENT

We build the stereo camera system using advanced embedded system based on the ARM1176 MPU core and we use MS WinCE as a operating system for developing camera device driver and application program based on USB OTG for sending image data file to the host computer system that is operating under the Windows XP and we use the OpenGL as a software tool and implement the 3D stereo image on the circular polarized 17 inch LCD Screen. We use the OV3640 CMOS camera module as key element for stereo image capture system. The main feature of the CMOS camera module is that the depth of field(DOF) is enough to have all the objects in the scene to be in focus when the objects are located 20cm far from the CMOS camera. We try to obtain the geometric parameters and the relations between camera capture system and 3D circular polarized LCD display system to make comfortable 3D real time video play. We assume that only one person is watching at center and 50cm in front of LCD which is 33.5cm width and 27.0cm height. We also assume the distance of the left eye and right eye of a person is 6.0 cm. In our camera capture system, we set the distance between the right and left camera as 6cm. The following table I shows the dynamic range of the view angle of OV3640 CMOS camera module which we find by analyzing the standard captured image set in our laboratory. Its dynamic range is enough to capture the left and right camera images for synthesizing its stereo 3D image on the 3D circular polarized LCD by using image processing and OpenGL programming.

#### A. Calibration of Pinhole Camera

In this paper, we briefly discuss about the calibration of pinhole under the assumption that the eye mechanism of human and the camera module we use in experiment are very similar to the pinhole camera[10]. The geometric relation among world

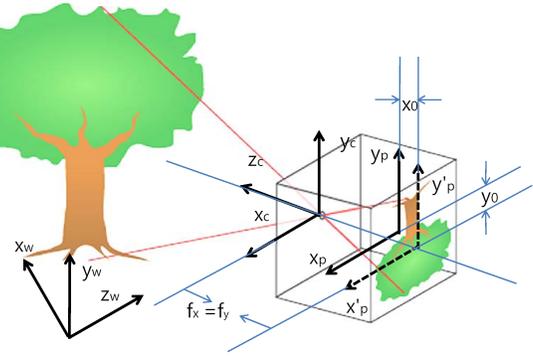


Fig. 4. Geometric Parameters of Pinhole Camera

coordinate system, camera coordinate system and pixel image coordinate system are shown in figure 4. Usually the focal length  $f_x$  and  $f_y$  of real camera module are not exactly the same and has a asymmetric optical property which is expressed by parameter  $\gamma$  between x and y axis in camera coordinate system. Also there exist a little planar offset  $(x_0, y_0)$  in pixel image coordinate system because of the misalignment between camera coordinate system and pixel image coordinate coordinate system. Let the 2D position in pixel image coordinate system and the 3D position in world coordinate system be  $(u \ v \ 1)^T$  and  $(x_w \ y_w \ z_w \ 1)^T$  by using notation of Homogeneous coordinate. Then we obtain following equation.

$$\begin{pmatrix} uz_c \\ vz_c \\ z_c \end{pmatrix} = A_{cp} A_{wc} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (9)$$

where

$$A_{cp} = \begin{pmatrix} f_x & \gamma & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_{wc} = (R : T) \quad (10)$$

and  $R$  and  $T$  are  $3 \times 3$  rotational matrix and  $3 \times 1$  translational vector between world coordinate system and camera coordinate system. The calibration problem in pinhole camera is to find the parameters for matrix  $A_{cp}$  and  $A_{wc}$ . The number of parameters of  $A_{cp}$  and  $A_{wc}$  are 5 and 6 respectively. Therefore the calibration problem in pinhole camera is to find or estimate these 11 parameters. In general case, CMOS camera module is designed and manufactured so that the values of  $f_x$  equals to  $f_y$  and  $\gamma$ ,  $x_0$  and  $y_0$  should be near to zero. In this paper, we assume that the values of  $f_x$  equals to  $f_y$  and  $x_0$  and  $y_0$  s are zero. In this case, the calibration problem is to find and calculate  $f_x$  and the 6 parameters of  $A_{wc}$ . Therefore, we try to control a camera system such that

$$\begin{aligned} A_{wc}^L &= \begin{pmatrix} 1 & 0 & 0 & +3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ A_{wc}^R &= \begin{pmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{aligned} \quad (11)$$

where  $A_{wc}^L$  and  $A_{wc}^R$  denote the  $A_{wc}$  for left and right camera module and 3 means 3cm in above matrices.

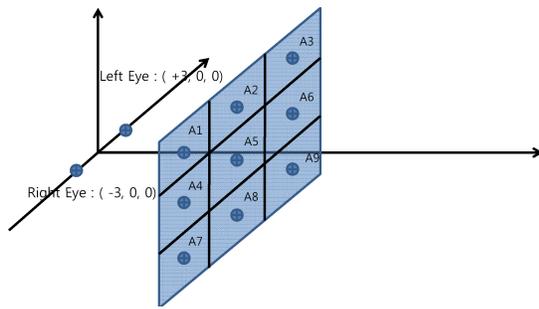


Fig. 5. Segmentation of Image

 TABLE II  
 ANGLE PARAMETERS OF HUMAN EYE. UNIT : DEGREE

Segmentation	$\theta_{L1}$	$\theta_{L2}$	$\theta_{R1}$	$\theta_{R2}$
LT	11.7	11.1	18.1	10.8
CT	-3.4	11.3	3.4	11.2
RT	-18.1	10.8	-11.7	11.1
LM	11.7	0	18.1	0
C	-3.4	0	3.4	0
RM	-18.1	0	-11.7	0
LB	11.7	-11.1	18.1	-10.8
CB	-3.4	-11.3	3.4	-11.3
RB	-18.1	-10.8	-11.7	-11.1

### B. Zooming and Alignment

We try to obtain and use 9 images from each left and right camera capture system for synthesizing fusion 3D image for the more comfortable 3d depth feeling. The geometric view of each segmentation of 9 image can be shown in figure 5. Each image is a captured projected image of left or right camera which is adjusted optimally with alignment and focusing of corresponding segmentation. In later, this 9 segmented images are synthesized for making one whole image. By considering the dynamic range of the rotation of human eye shown in table II, we choose the following pairs of dynamic range of each image shown in table III and we developed a control algorithm for finding the exact optimal capturing angles based on the sharpness of stereo images for each segmentation images and record the angle and distance offset parameters for rendering later. The angle parameters of human eye can be expressed as follows under the assumption that human right and left eyes are located at  $(-3, 0, 0)$ ,  $(3, 0, 0)$  respectively.

$$\begin{aligned}
 \theta_{R1} &= \text{atan2}(x+3, z) \\
 \theta_{R2} &= \text{atan2}(y, \sqrt{(x+3)^2 + z^2}) \\
 \theta_{L1} &= \text{atan2}(x-3, z) \\
 \theta_{L2} &= \text{atan2}(y, \sqrt{(x-3)^2 + z^2})
 \end{aligned} \quad (12)$$

Segmentation symbols of each images are left-top(LT), center-top(CT), right-top(RT), left-middle(LM),center (C), right-middle(RM), left-bottom(LB), center-bottom(CB),right-bottom(RB) respectively. In the above table,  $\theta_{L1}, \theta_{L2}, \theta_{R1}, \theta_{R2}$  denote the left-right, up-down rotating angle of the left and right camera capture system respectively. In this paper, we assume that a person who watches the synthesized LCD screen does not move his neck and just make rolling his eyes only. For more details, refer to [11].

 TABLE III  
 DYNAMIC RANGE OF SEGMENTATION. UNIT : DEGREE

Segmentation	$\theta_{L1}$	$\theta_{L2}$	$\theta_{R1}$	$\theta_{R2}$
LT	(10,13)	(10,13)	(17,20)	(9,12)
CT	(-5,-2)	(+10,+13)	(2,5)	(10,13)
RT	(-20, -17)	(9,12)	(-13,-10)	(10,13)
LM	(10,13)	(-2,+2)	(17,20)	(-2,+2)
C	(-5,-2)	(-2,+2)	(+2,+5)	(-2,+2)
RM	(-20,-17)	(-2,+2)	(-13,-10)	(-2,+2)
LB	(10,13)	(-13,-10)	(17,20)	(-12,-9)
CB	(-5,-2)	(-13, -10)	(2,5)	(-13,-10)
RB	(-20,-17)	(-12,-9)	(-13,-10)	(-13,-10)

### 4. FUSION BASED ON INTERPOLATION

In this subsection, we propose a simple fusion method that make a whole stereo 3D image with the 9 segmented/captured images from the left and right camera module. The simple method is to crop each segmented image with the same size, such as 1/9 of whole LCD image area, based on the angle data which are acquired when the corresponding segmented image is captured. Under the assumption that the angle data are the same as those of eyes of person who watch the real LCD screen, we just union the same pixel size of each segmentation to make a whole image for left and right stereo images as follows.

$$I_w = \bigcup_{i=1, \dots, 9} \text{Crop}(I_{seg,i}) \quad (13)$$

where  $I_{seg,i}$  denotes each captured segmented image and  $I_w$  denotes whole synthesized image. For example, let assume that the LCD Screen size is  $1280 \times 1024$ . Then we crop each segmented image into  $(1280/3) \times (1024/3)$  size image around its center and then combine these 9 cropped images into a whole  $1280 \times 1024$  image. We do this operation twice for obtaining left and right whole image. If the above assumption concerning the angle data is not satisfied, we should do adopt a filtering and smoothing signal processing with the captured images. In this case, we first calculate the geometric position vector  $p_v = (x_v, y_v, z_v)$  of center pixel of each captured segmented image in virtual LCD screen based on the capturing angle data and also calculate the corresponding position  $p_L = (x_L, y_L)$  in the real LCD screen. Then we crop the image as shown in figure 6 by considering this position  $p_L$  for each corresponding segmentation. In next future work, we try to find more advanced algorithm for interpolation of segmentation images for fusion.

### 5. SIMULATION AND EXPERIMENTAL RESULTS

We simulate our proposed algorithm based on the virtual objects made by programming with OpenGL software tool. We synthesize a 3D stereo image from the 9 segments of image captured by virtual left and right camera which is separated 6cm with each other. This separation length is the same length of distance between two left and right eyes of average person. We choose several cube as objects located around 1m behind of virtual LCD screen. In this simulation, we assume that a person who watch the synthesized image does not move and rotate his neck and moves only his eye balls. We use the angle parameters of human eye shown in table II for obtaining 9 segmentations of image. We also suggest a camera model for general CMOS camera. Camera model of general CMOS camera can be approximated as an image sum of several/many ideal pinhole cameras. Let  $I_{CMOS}$  and be the image of CMOS camera and let  $I_{pin}(x, y, z)$  be the image

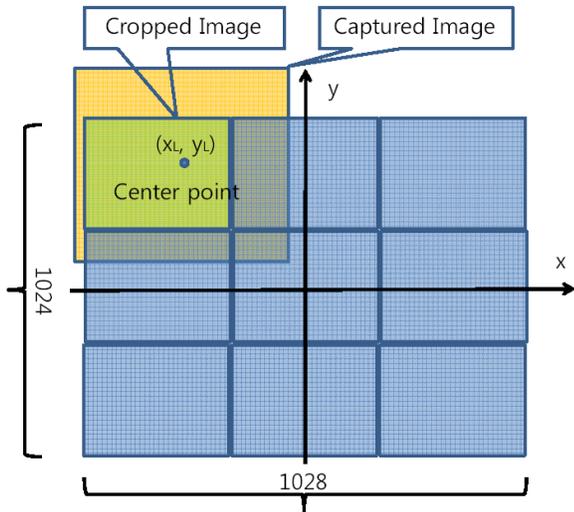


Fig. 6. Cropping and Union for Fusion

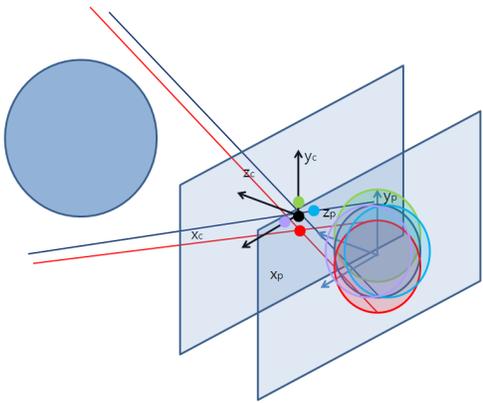


Fig. 7. CMOS Camera Model

of pinhole camera whose aperture is located at  $(x, y, z)$  from the center. We can approximate  $I_{CMOS}$  as follows by using several images obtained from the ideal pinhole cameras as shown in figure 7.

$$I_{CMOS} = \frac{1}{N} \sum_{i=1, N} I_{pin}(x_i, y_i, z_i) \quad (14)$$

For the simplicity of simulation, we assume that all the ideal pinhole cameras have the same focal length  $f_{x0} = f_{y0}$  and share the same image plane and have the same aperture plane, that is,  $z$  components of aperture are the same. We obtain camera parameters of each pin hole camera as follows.

$$A_{cp,i} = \begin{pmatrix} f_{x0} & \gamma & x_{i0} \\ 0 & f_{y0} & y_{i0} \\ 0 & 0 & 1 \end{pmatrix}, A_{wc,i} = \begin{pmatrix} 1 & 0 & 0 & -x_{i0} \\ 0 & 1 & 0 & -y_{i0} \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (15)$$

From the above equation, we can see the above transformation is not linear with respect to aperture offset  $x_{i0}$  and  $y_{i0}$ . This kind of nonlinear effect makes it difficult to find a good preprocessing method for the captured CMOS camera image. In this paper, we set  $N$  as 5 for the simplicity of simulation and  $(x_i, y_i), i = 1, \dots, 5$  be center point and 4 corner points of rhombus around the center of aperture. We set the distance of a vertex of rhombus  $d$

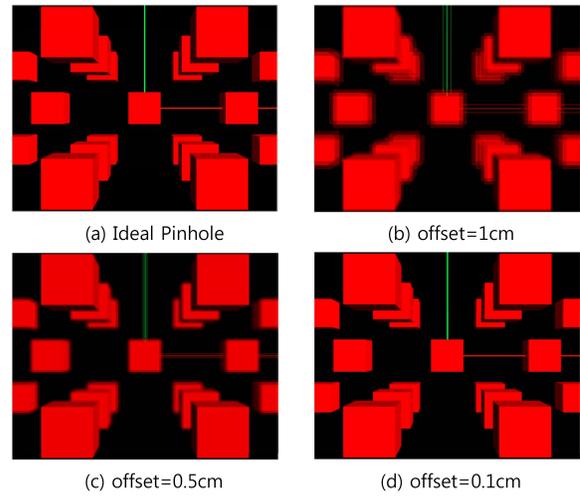


Fig. 8. Performance of pinhole models

from the center as 0.1[mm] in this paper. We set the focal length  $f_{x0}$  as 5[mm]. Therefore the percentage of the aperture offset compared with focal length is about 2 % in this simulation. We use the OpenGL real time graphics as a software tool. The test scene consists of same size of 19 cubes and 2 axes. The width, length and height of cube are 8cm each and surrounding 8 cubes are located 12cm far from the centered cube on each axis. We assume that distance between 3D LCD screen and human viewer is 50cm and the distance between the centered cube and human viewer is about 100cm. So we set the near and far clipping planes into the  $z$ -axis as follows.

$$z_n = -50cm \quad z_f = -200cm$$

Next we move the cubes 100cm in the direction of negative  $z$ -axis by using `glTranslatef()` API function of OpenGL. Figure 8 shows the image captured by using one ideal pin hole camera and the image synthesized by using 5 ideal pin hole camera model proposed above when the distance of a vertex of rhombus  $d$  from the center are 10.0, 5.0, 1.0[mm] respectively. We also calculate the difference of the image in region of interest  $R_{ROI}$  by using the following equation.

$$I_d = \sum_{(i,j) \in R_{ROI}} |I_{d,RGB}(i,j)| \quad (16)$$

where  $I_{d,RGB}(i,j)$  means the absolute sum of difference of RGB color components in pixel  $p(i,j)$  between two images. We select  $R_{ROI}$  as one segmentation ( about  $420 \times 330$  pixel size window patch) around the center of synthesized images. The simulation results in this region of interest are shown in table IV. From the data in table IV, we can see that the proposed CMOS camera module can be used for synthesizing general CMOS camera whose characteristics are very similar to the ideal pinhole camera. From the simulation results, we can simulate the general CMOS camera by choosing appropriate number of pinhole camera model and selecting appropriate offset value. We found that 1.0[mm] is appropriate for offset value in simulating CMOS camera model. Also another important characteristic of general CMOS camera is that the resolution of image with respect to the solid angle is nonlinear. Its image resolution at solid angle zero is better and

TABLE IV  
SIMULATION RESULTS OF CMOS CAMERA MODEL

offset [mm]	$I_d$	offset [mm]	$I_d$
10	2,502,362	2	645,503
5	1,365,189	1	300,149



Fig. 9. Centered segment

bigger than at any other angle. This is the same reason why we should control and focus our eye on an object to see more clearly. When our eye is focused at an object, we can get larger image by rotating eye with the same amount of linear angle. It means that the sensitivity of our eye with respect to rotating angle is relatively high when it is focused at an object. Therefore the disparity of the left and right captured images depends on the alignment of the left and right camera module with respect to an object which we are interested in. The disparity  $d_L$  of the left and right image for some region of interest can be decided such that the two left and right images match as closely as possible as follow.

$$I_L(i, j) \approx I_R(i - d_L, j) \text{ for } (i, j) \in R_{ROI} \quad (17)$$

where  $I_L(i, j)$  and  $I_R(i, j)$  is the left and right captured images respectively. Figure 1 shows the experimental environment of capturing system. We use 1 step motor and 4 DC servos for control the left and right CMOS camera by using ATCAN1280 micro-processor. We develop a main embedded board by using S3C6410 which was designed by Samsung electronic company based on ARM1176 main CPU core and we added some hardware circuit for multiplexing its CMOS capturing function. Figure 9 shows synthesized 480x800 sized images of combined left and right images for a sample environment which has 1 box object in each 9 segment of LCD screen where ROI is the centered segment. The values of  $d_L$  in each segment are shown in table V. From the data shown in the table, we can see that the segment alignment method between camera module and object is very important for the realization of 3 D depth synthesizing. These are the main key experimental results. From the analysis of experimental results, multi-focusing image capturing system may improve the depth quality in 3D image processing.

TABLE V  
DISPARITY TABLE OF CENTERED SEGMENT

$(x, y)$	$d_L$	$(x, y)$	$d_L$	$(x, y)$	$d_L$
(50,120)	-12	(50,400)	-11	(50,680)	-16
(220,120)	4	(220,400)	2	(220,680)	4
(390,120)	64	(390,400)	64	(390,680)	55

## 6. CONCLUSION

In this paper, we analyze the real experimental results obtained by using a fusion method to combine a whole stereo image from the 9 captured images. We also comment on the importance of the projective mapping for handling signal processing. We suggest real time 5 degree of freedom motor control system for capturing stereo images based on the measure of alignment and sharpness which is needed for calibrating camera module. In the next version, we will try to add 1 more DOF to each left and right camera module for zooming calibration by using voice coil motor. It is needed to have good knowledge concerning the geometric parameters not only in capturing a left and right image but also in synthesizing stereo image for designing of a good stereo vision system. We introduce mathematical preliminary which is needed to analyze and synthesize stereo image. We do simulate by using OpenGL software and get some results that verify the validness of our proposed algorithm. We have real experimental results based on ARM11 core embedded system. We capture nine segment stereo images and analyze the disparity parameters in each segment. From the analysis of 9 stereo segment images, we can see that this segmented method will be one of good methods for realization of the 3 D depth in stereo vision system. We will try to expand this method to more multiple segments than 9 segments for obtaining human-like vision capability.

## ACKNOWLEDGMENT

This research was supported partially by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the ministry of Education, Science and Technology(2010-0010531)and by the Technology Innovation Program of the Ministry of Knowledge Economy, Republic of Korea.

## REFERENCES

- [1] C. Zhang and T. Chen, "View-dependent Non-uniform Sampling for Image-based Rendering", ICIP 2004
- [2] D. Scharstein and R. Szeliski,"A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International Journal of Computer Vision, 47(1/2/3).pp.7-42, April-June 2002.
- [3] I.J.Cox, S.Roy, and S.L.Hingorani," Dynamic histogram warping of image pairs for instant image brightness, " IEEE International Conference on Image Processing, Vol.2, pp. 366-369, 1995.
- [4] Fleet, D.J. and Weiss, Y.,"Optical flow estimation," in "Mathematical models for Computer Vision: The Handbook," N. Paragios, Y. Chen, and O. Faugeras (editors.), chapter 15, Springer, 2005
- [5] J.L. Barron, D.J.Fleet, and S.S. Beauchemin," Performance of optical flow techniques," International Journal of Computer Vision, 12(1), PP. 43-77, 1994
- [6] N. Holliman,"Smoothing Region Boundaries in Variable Depth Mapping for Real Time Stereoscopic Images," *Stereoscopic Displays and Virtual reality System XII*, (2005), vol.5664A of *Proceedings of SPIE-IST Symposium on Electronic Imaging*.
- [7] N. Holliman,"3D Display Systems", Handbook of Opto-electronics, IOP Press, Spring, 2005.
- [8] N. Holliman,"Mapping Perceived Depth to Regions of Interest in Stereoscopic Images, " *Stereoscopic Displays and Virtual reality System XI, Proceedings of SPIE 5291*, 2004.
- [9] S.Kim and W.Ham," A New Algorithm for Depth Perception in 3D Screen and Its Implementation," *2007 International Symposium on Information Technology Convergence*, Nov, 2007.
- [10] T.Dang, C. Hoffmann, and C.Stiller, " Self-calibration for Active Automotive Stereo Vision," in IEEE Intelligent Vehicles Symposium, pp.364-369, 2006.
- [11] W.Ham, H. Kwon and T. Enkhbaatar, "Multi-focusing Image Capture System for 3D Stereo Image ", *Journal of KROS*, Vol.6, No.2, 2011

## Simulated Saccade for Saliency Sensitive Pattern Capturing in Naturally Complex Scenes

Kohji Kamejima

Osaka Institute of Technology

1-79-1 Kitayama, Hirakata 573-0196 JAPAN

e-mail: kamejima@is.oit.ac.jp

### ABSTRACT

A saliency-based scheme is presented for detecting landmark objects in naturally complex scenes. Based on the roadway segment down loaded from the satellite images, first, the generic ground-object structure is identified through scale space analysis of scene images. Next, the chromatic diversity of the scene images is articulated into a set of saliency patterns with respect to *as-is* primary. Finally, the saliency pattern is verified to yield a fractal attractor consistent with the ground-object structure.

**Keywords:** Saliency Pattern; Chromatic Complexity Generator; Simulated Saccade

### 1 INTRODUCTORY REMARKS

Despite infinite diversity of appearance, natural scenes exhibit environment specific landmarks to be identified within individual intention of viewers. To control the focus to such a landmark object, perception processes should gather randomly distributed image features and apply ‘feature integration’ schemes to generate ‘visual saliency’ associated with the complex scene [5]. Due to the redundancy of natural scene relative to decision makings by the viewers, computational integration processes easily fall into combinatorial explosion on symbolic description of visible features.

At the practical implementation of the basic capability of early vision, it is pertinent to transform the diversity of the image features into the space of implicit representation called ‘environmental saliency’ [10]: a distributed representation of universal rules nondeterministically governing physical-geometric structure. For instance, the scale-chromatic randomness has been matched with the universal rules to organize the environmental saliency as a transferable description of the ground-object structure. Such a randomness-based representation provides the theoretical basis of an anticipative road following system illustrated in Fig. 1; the environmental saliency extracted in a scene image is transferred to a cut of satellite image to extend a visible segment of the roadway pattern beyond the physical-geometric perspective.

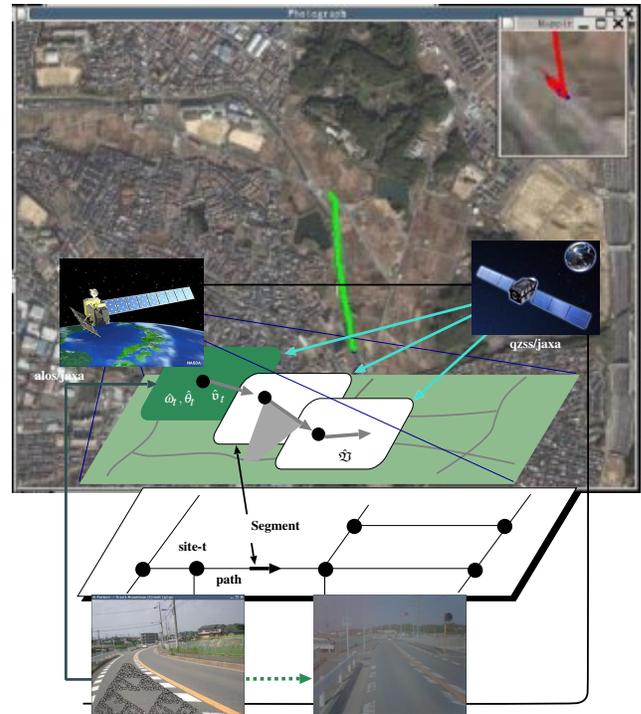
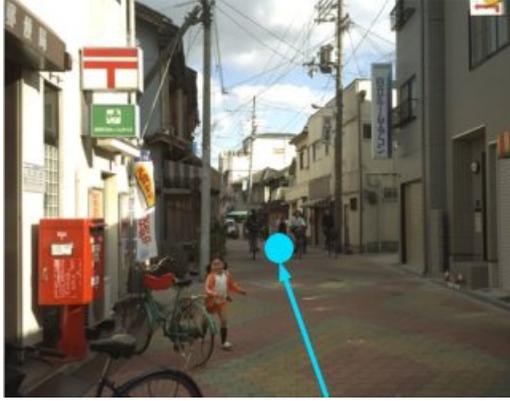
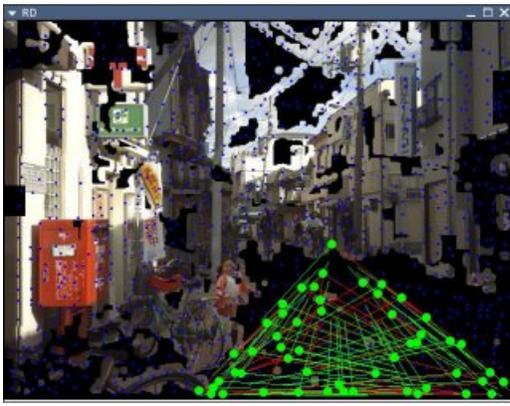


Figure 1: Anticipative Road Following System

Let  $(\hat{\omega}_t, \hat{\theta}_t)$  be the estimate of the current position and direction in the satellite image with local roadway segment  $\hat{v}_t$ . Suppose that the local segment is extended to generate a chain  $\hat{\mathcal{V}} = \{\hat{v}_t, t = 1, 2, \dots\}$  to a possible destination in the bird’s eye view. By this segmentation, geometric complexity of the local terrain is reduced to a graph on which we can constrain the diversity of future scenes *prior to* physical access. Adding to it, we can associate generic structure of ‘naturally designed scenes’ in the multi-viewpoint images spanning the temporal-spatio discrepancy. However, due to the essential uncertainty in the distribution of temporal and/or moving objects, it is not practical to utilize the bird’s eye view as the reference of the control of the vehicles. To expand the scope of the vehicle control systems to the graph spanning the *as is* local terrain, in this paper, we introduce a new method for analyzing vehicle specific scenes based on the transferable roadway segment  $\hat{\mathcal{V}}$ .


 (a) Scene to be Analyzed under *A Priori* Segment


(b) Fractal Coding of Connected Roadway Area

Figure 2: Ground-Object Structure

## 2 GROUND-OBJECT STRUCTURE INDUCED BY SCALE SPACE ANALYSIS

Let the *as is* route graph be down loaded as an *a priori* information of the generic structure of the scenes to be encountered. Following empirical knowledge of ecological optics [3] combined with recent advancements in machine perception [7] and emotional perception [4], we can expect that the generic structure of maneuverable scenes is described in terms of a set of fractal codes specifying the expansion of horizontal plane and the aggregation of boundary objects. Noting this, the randomness of the scale information is extracted as a version of latent images to identify the ground-object structure in the scene image to estimate the fractal code spanning a connected open space as shown in Fig. 2 [9]; down loaded segment  $\vec{v}_i$  is mapped into the scene image (a) to design a fractal code is confined by the distribution of boundary objects (b). To control vehicle mechanism, the boundary distribution should be articulated into a system of fractal codes to be associated with the landmark objects. As the cue to the multi-fractal coding, in what follows, another version of latent images,

random distribution of chromatic complexity, is extracted as the observables of not-yet-identified landmark objects.

## 3 CHROMATIC COMPLEXITY GENERATOR

Let R, G, B be three primaries, and suppose that the light is identified with a linear combination of the primaries by humans. This implies that the information conveyed by the spectral distribution be described as follows:

$$f_{\omega}^{\text{RGB}} = [f_{\omega}^{\text{R}} \quad f_{\omega}^{\text{G}} \quad f_{\omega}^{\text{B}}]^T,$$

where  $f_{\omega}^{(\cdot)}$  designates subjective weight of the primary  $(\cdot)$ . Define  $\phi_{\omega} = f_{\omega}^{\text{RGB}} / |f_{\omega}^{\text{RGB}}|$ . By identifying the totality of the chromatic information  $\phi_{\omega}$  with the positive part of a unit sphere, we can induce the following measure [8]:

$$g_{\alpha}(\phi_i | \phi_j) = \frac{1}{2\pi\alpha} \exp \left[ -\frac{|\phi_i - \phi_j|^2}{2\alpha} \right]. \quad (1)$$

Following experimental studies using roadway scenes and various natural objects,  $\alpha$  should be adjusted to  $1/10 \sim 1/100$ . For sufficiently small chromatic variation, the measure  $g_{\alpha}(\phi_i | \phi_j)$  approximates the Gaussian distribution on local tangential space at  $\phi_j$ .

Consider an inverse problem; detect a set of fixed points to regenerate the chromatic diversity as a fractal attractor in the color space. For such a global analysis, let the chromatic complexity index  $\phi_{\omega}$  be identified with the following planar representation of the color space:

$$\Gamma \ni \gamma = e^{\text{RGB}} \phi_{\omega}, \quad e^{\text{RGB}} = [e^{\text{R}} \quad e^{\text{G}} \quad e^{\text{B}}], \quad (2)$$

$$e^{(\cdot)} = [\cos \theta_{(\cdot)} \quad \sin \theta_{(\cdot)}]^T,$$

where the primary is assigned as follows  $\theta_{\text{R}} = \pi/2$ ,  $\theta_{\text{G(B)}} = \theta_{\text{R}} + (-)\pi/3$ . The diversity of the incoming light  $f_{\omega}^{\text{RGB}}$  is represented in the color space  $\Gamma$  through the linear transform (2). Thus, the chromatic information  $\phi_{\omega}$  is restored through the following procedure:

$$\hat{\phi}_{\omega} = \tilde{\phi}_{\gamma} + \bar{\phi}_{\omega} \mathbf{1}^{\text{RGB}}, \quad \mathbf{1}^{\text{RGB}} = [1 \quad 1 \quad 1], \quad (3)$$

where  $\tilde{\phi}_{\gamma} = \frac{2}{3} (e^{\text{RGB}})^T \gamma$  and  $\bar{\phi}_{\omega}$  designates a nominal brightness level given as the solution to the following

$$3\bar{\phi}_{\omega}^2 + 2\tilde{\phi}_{\gamma}^T \mathbf{1}^{\text{RGB}} \cdot \bar{\phi}_{\omega} + |\tilde{\phi}_{\gamma}|^2 = 1.$$

Let a set of samples  $\mathfrak{s} = \{\phi_i, i = 1, 2, \dots\}$  be collected in a scene image to generate the following field on  $\Gamma$  [6]:

$$\frac{\partial \varphi_{\rho}(\gamma | \mathfrak{s})}{\partial t} = \frac{1}{2} \Delta \varphi_{\rho}(\gamma | \mathfrak{s}) + \rho [\chi_{\mathfrak{s}} - \varphi_{\rho}(\gamma | \mathfrak{s})],$$

where  $\chi_{\mathfrak{s}}$  denotes the aggregation of Dirac's delta measure distributed on the set  $\{\gamma(\phi_{\omega}) \mid \phi_{\omega} \in \mathfrak{s}\}$  and  $\rho$  is adjusted to the complexity of the primary set. By identifying the

distribution  $\chi_s$  with the invariant measure associated with not-yet-identified fractal attractor, we can specify a set of control parameter  $\hat{\Pi} = \{\hat{\pi}_i\}$  for regenerating the diversity of the chromatic information in the color space  $\Gamma$ . To this end, first, a possible fixed point is located as an initial set of the control point,  $\tilde{\Gamma}_0^f$ , on the Laplacian-Gaussian boundary  $\partial^s \chi_s$  [6]. The initial set is expanded via the following successive scheme:

$$\tilde{\Gamma}_{t+1}^f = \tilde{\Gamma}_t^f \cup d\tilde{\Gamma}_t^f, \quad (4)$$

where the increment is selected with respect as follows

$$\begin{aligned} d\tilde{\Gamma}_t^f &= \left\{ \tilde{\gamma}^f \mid \forall \tilde{\gamma} : \overline{\eta}(\partial \tilde{\gamma}^f, \tilde{\Gamma}_t^f) \geq \overline{\eta}(\partial \tilde{\gamma}, \tilde{\Gamma}_t^f) \right\}, \\ &\quad \tilde{\gamma}^f, \partial \tilde{\gamma} \in \partial^s \chi_s - \tilde{\Gamma}_t^f, \\ \tilde{\Gamma}_0^f &= \left\{ \tilde{\gamma}_0^f \mid \forall \tilde{\gamma}_0 : |\partial \tilde{\gamma}_0^f| \geq |\partial \tilde{\gamma}_0| \right\}, \\ &\quad \partial \tilde{\gamma}_0^f, \partial \tilde{\gamma}_0 \in \partial^s \chi_s \end{aligned}$$

with respect to  $\overline{\eta}(\gamma, \Lambda) = \min_{\lambda \in \Lambda} |\gamma - \lambda|$ . Next, a sub set  $\{\hat{\gamma}_k\}$  satisfying the following conditions is selected as an estimate of the vertices

$$\begin{aligned} \forall m, k : \quad \theta_{mk} - \theta_{nk} &< \pi, \\ \hat{\gamma}_{(\cdot)} - \hat{\gamma}_k &= |\hat{\gamma}_{(\cdot)} - \hat{\gamma}_k| e^{j(\theta_{(\cdot)k} + \theta_k)}, \\ \hat{\gamma}_{(\cdot)} \in \tilde{\Gamma}^f, \hat{\gamma}_k &= |\hat{\gamma}_k| e^{j\theta_k}. \end{aligned} \quad (5)$$

Finally, the distribution of  $\tilde{\Gamma}$  is expanded along the following repulsive force:

$$d\hat{\gamma}_k = \sum_{\hat{\gamma}_j \in \tilde{\Gamma}} (\hat{\gamma}_k - \hat{\gamma}_j) g_\alpha(\phi_k | \phi_j), \quad (6)$$

within the possible coloring circle  $|\hat{\gamma}_k| \leq 1$ . Following the repulsive force, the vertices  $\{\hat{\gamma}_k\}$  are separated each other to yield a set of *as-is* primaries. The scheme (4) combined with (5) yields a set of fixed points to be associated with a set of contraction mapping for regenerating the distribution  $\chi_s$  in the color space  $\Gamma$ . By adding the dynamics (6), we have a meta-process: a generator of the fractal dynamics controlled by the *as-is* primaries  $\hat{\Pi} = \{\hat{\pi}_i\}$  given by  $\hat{\pi}_i = \tilde{\pi}_i + \bar{\pi}_i 1^{\text{RGB}}$  where

$$\begin{aligned} \tilde{\pi}_i &= \frac{2}{3} (e^{\text{RGB}})^T \hat{\gamma}_i, \\ 3\tilde{\pi}_i^2 + 2\tilde{\pi}_i^T 1^{\text{RGB}} \cdot \bar{\pi}_i + |\bar{\pi}_i|^2 &= 1. \end{aligned}$$

The implication of the chromatic complexity generator is demonstrated in Fig. 3; the distribution of the samples  $s$  in the scene image (Fig. 2) is indicated in (a); in (b) the associated field  $\phi_p(\gamma|s)$  is displayed with the distribution of possible fixed points  $\tilde{\Gamma}^f$ . In this figure, four vertices are selected and separated to yield the *as-is* primaries  $\hat{\Pi}$ .

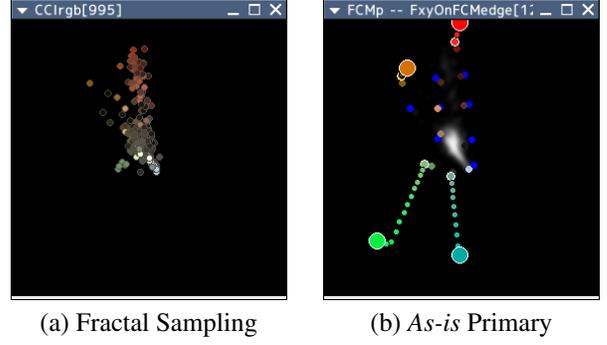
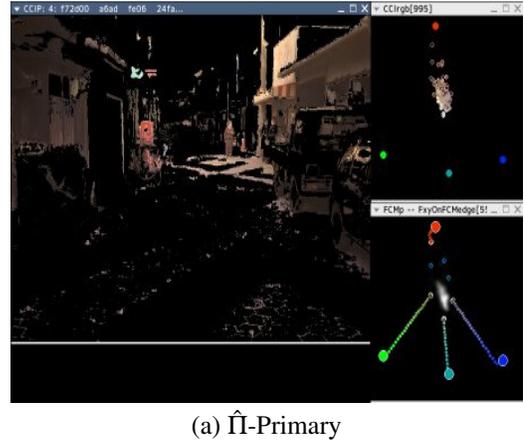


Figure 3: Chromatic Complexity Generator



(a)  $\hat{\Pi}$ -Primary

(b) RGB-Primary

Figure 4: Complexity Reduction via  $\psi_\omega$  Channel

#### 4 SALIENCY PATTERNS VIA $\psi_\omega$ -FILTERING

Let the probability for selecting a primary in the set  $\hat{\Pi}$  be evaluated by

$$p_i(\gamma_\omega | \hat{\Pi}) = \frac{g_\alpha(\gamma_\omega | \hat{\pi}_i)}{\sum_{\hat{\pi}_i \in \hat{\Pi}} g_\alpha(\gamma_\omega | \hat{\pi}_i)}.$$

Table 1: Complexity Reduction via  $\hat{\psi}_\omega$ -Filtering

scene	$dS_G$	$dS_H$	$\ \hat{\Pi}\ $
shopping street	1.137650	< 2.078783	4
post office	0.647912	< 2.172450	4
street view at night	1.340637	< 2.681494	3
industrial park	1.055658	< 1.259931	4

By identifying the distributed representation  $p_i(\gamma_\omega|\hat{\Pi})$  with the Kolmogorov's complexity in the primary selection, we can induce the following saliency index in the scene image:

$$\begin{aligned}\hat{\psi}_\omega^{\hat{\Pi}} &= \exp\left[-\hat{\mathcal{J}}_\omega^{\hat{\Pi}}\right], \\ \hat{\mathcal{J}}_\omega^{\hat{\Pi}} &= -\sum_{\hat{\pi}_i \in \hat{\Pi}} p_i(\gamma_\omega|\hat{\Pi}) \log p_i(\gamma_\omega|\hat{\Pi}).\end{aligned}\quad (7)$$

By definition, the saliency index yields a pixel-wise evaluation for discriminating 'easy to predict' patterns in the naturally complex scene. The effectiveness of the  $\hat{\psi}_\omega^{\hat{\Pi}}$ -filtering is illustrated in Fig. 4; in (a) and (b), saliency indices with respect to the estimate  $\hat{\Pi}$  and the RGB primary were applied to the scene image (Fig. 2), respectively. As shown in this figure, various landmark objects can be detected via  $\hat{\psi}_\omega^{\hat{\Pi}}$ -filtering. The comparison of (a) and (b) reveals that the *as-is* primary can extend the detectability of saliency patterns wrapped by complex colors; the *as-is* primaries (a) is effective to extend the focus of perception channel to considerably weak landmark pattern.

The significance of the  $\hat{\psi}_\omega^{\hat{\Pi}}$ -filtering was verified by using various types of scene images as summarized in Table 1 where the reduction of the unpredictability is evaluated in terms of relative entropy  $dS_{(\cdot)} = S_\theta - S_{(\cdot)}$ ;  $S_\theta$  and  $S_G$  designate the Shannon's entropy with respect to the uniform distribution and the gray level distribution  $f_\omega$ , respectively;  $S_H$  stands for the following entropy

$$S_H = -\int_{\Omega} \bar{\psi}_\omega \log(\bar{\psi}_\omega) d\omega,$$

with respect to the normalized version of the saliency index

$$\bar{\psi}_\omega = \frac{\hat{\psi}_\omega^{\hat{\Pi}}}{\int_{\Omega} \hat{\psi}_\omega^{\hat{\Pi}} d\omega}.$$

As shown in Table 1, the  $\hat{\psi}_\omega^{\hat{\Pi}}$ -filter concentrate the information distributed in the image plane into a set of saliency patterns of average size  $e^{dS_H} \sim 7.38$ ; this amount to 1/2 of detectable patterns through the analysis of  $f_\omega$ .

## 5 SIMULATED SACCAD E ON SALIENCY INDEX

We can exploit the saliency index  $\hat{\psi}_\omega^{\hat{\Pi}}$  to capturing a landmark object observed as a saliency pattern. Let a pixel in a saliency pattern be located at  $m_\omega$  in the image plane  $\Omega$  and consider 2D Gaussian distribution  $g_\sigma(\omega|m_\omega)$ ,  $\omega \in \Omega$  around  $m_\omega$ . Noticing the following generation mechanism

$$\begin{aligned}g_\sigma(\omega|m_\omega) - \delta_{m_\omega} &= \int_0^\sigma \frac{1}{2} \Delta g_s(\omega|m_\omega) ds \\ &\sim \frac{\sigma}{2} \Delta g_\sigma(\omega|m_\omega),\end{aligned}$$

with Dirac's delta measure  $\delta_{m_\omega}$ , we can evaluate the probability for capturing a Brownian motion process with origin  $m_\omega$  at some time  $t \leq \sqrt{\sigma}$  in terms of  $g_\sigma(\omega|m_\omega)$ . This evaluation can be extended for the capturing of a fractal pattern with invariant measure  $\hat{\psi}_\omega^{\hat{\Pi}}$ ; in this case, the capturing probability  $\varphi(\omega|\hat{\Pi})$  is obtained as the solution to the following partial differential equation [6]:

$$\frac{\sigma}{2} \Delta \varphi(\omega|\hat{\Pi}) + [\hat{\psi}_\omega^{\hat{\Pi}} - \varphi(\omega|\hat{\Pi})] = 0, \quad (8)$$

where  $\hat{\psi}_\omega^{\hat{\Pi}}$  is identified with the aggregation of the delta measures distributed on entire the saliency patterns. In this equation,  $\sigma$  is adjusted to the complexity factor, i.e.,  $\sigma \sim 2/\log\|\hat{\Pi}\|$ .

Suppose that an *a priori* estimate of the object is given as a pixel  $\omega_m$  in the saliency pattern. By evaluating the deviation from the nearest delta measure  $\hat{\delta}$  in terms of the deviation factor given by

$$n_\omega^{\hat{\Pi}} = \sqrt{-2\sigma \log \left[ \frac{\varphi(\omega|\hat{\Pi})}{\max_{\omega \in \Omega} \varphi(\omega|\hat{\Pi})} \right]},$$

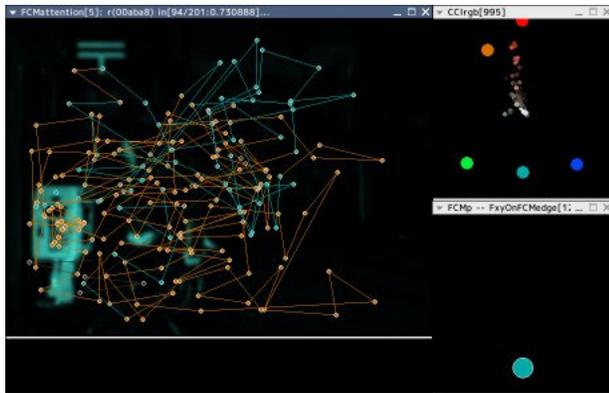
we have the following model update scheme:

$$\omega_m \leftarrow \omega_m + 2n_{\omega_m}^{\hat{\Pi}} \frac{\nabla \varphi(\omega_m|\hat{\Pi})}{|\nabla \varphi(\omega_m|\hat{\Pi})|}. \quad (9)$$

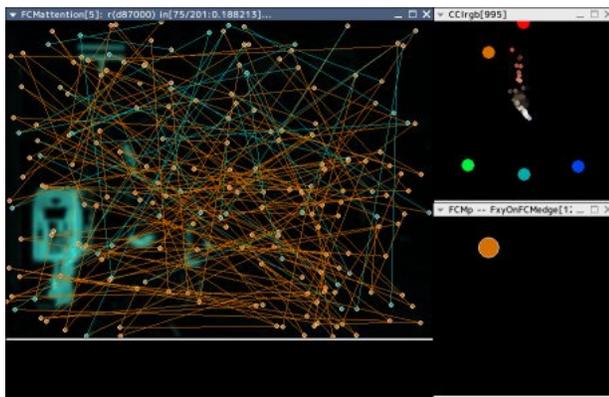
By iterating the update process (9), we have a saliency sensitive scanning mechanism in the scene image.

## 6 PERCEPTUAL EQUIVALENCE IN SALIENCY PATTERN CAPTURING

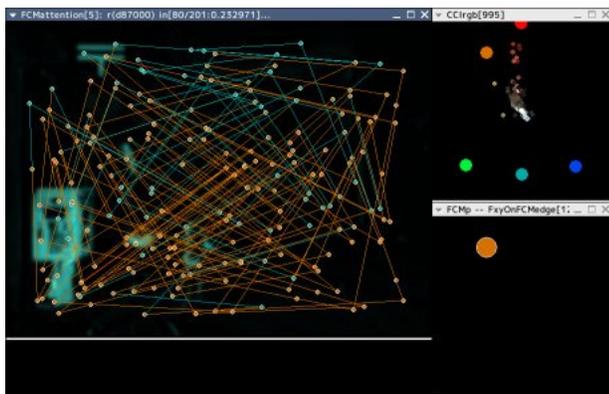
By adjusting the scale  $\sigma$  in (8) to the resolution for object detection, we have the capturing probability  $\varphi(\omega|\hat{\Pi})$  associated with the fractal attractor spanning a saliency pattern. This implies that saliency patterns can be efficiently scanned via the simulated saccade process as shown in Fig. 5; the dynamics (9) detect a set of saliency patterns successively as demonstrated in (a); based on the information  $\varphi(\omega|\hat{\Pi})$ , the scanning process is controlled to 'hit' the saliency patterns more efficiently than the scanning based



(a) Saliency Sensitive Process



(b) Random Sampling Process



(c) Fractal Sampling Process

Figure 5: Saliency Pattern Scanning

on 2D uniform random number and a fixed fractal mechanism indicated in (b) and (c) respectively. Thus, we can select a sequence of saliency patterns to articulate the scene image into a multi-fractal system [1]. By applying the fractal articulation scheme [11] to the scene image through  $\hat{\psi}_{\omega}^{\hat{\Pi}}$ -filtering, we have a set of fractal code  $\mathcal{C}^{\nu}$  confined by the saliency patterns.

The consistency of the multi-fractal model  $\mathcal{C}^{\nu}$  designed

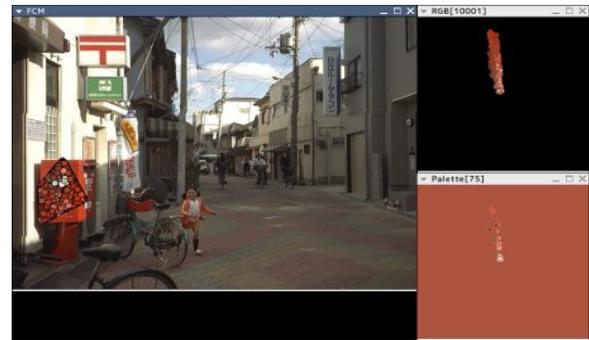


Figure 6: Multi-Fractal Coding

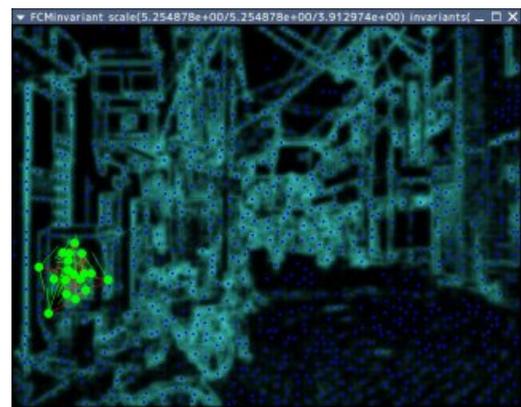
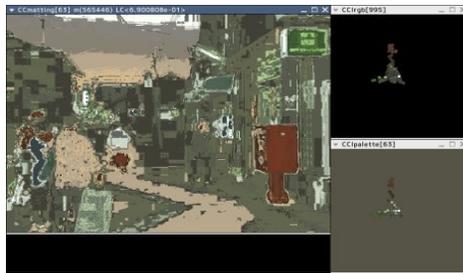


Figure 7: Finite Invariance Test on Scale Space Image

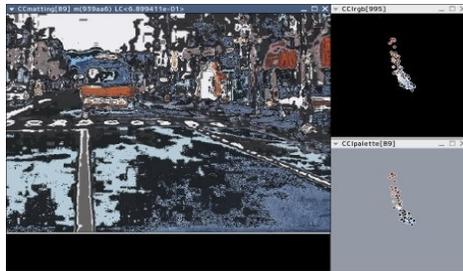


Figure 8: Contextual Visualization

on the 2D patterns of saliency colors with the ground-object structure identified via scale complexity analysis is discussed. As shown in Figs. 2 and 5, the scale and chromatic aspects of the environmental saliency are intrinsically consistent to yield a system of the fractal models: one is the ground model for separating a connected open space from boundary objects; the other is saliency pattern model for concentrating computational resources to the de-



(a) Street View in Night



(b) Industrial Park

Figure 9: Contextual Visualization

tection of landmark objects. The multi-fractal articulation process combined with the simulated saccade is effective to separate landmark objects in complex background successively as shown in Fig. 6. Furthermore, the fractal code is verified to be consistent with the ground-object structure identified through scale complexity analysis; the fractal model of the saliency pattern yields an invariant features as demonstrated in Fig. 7. This implies that the saliency pattern can be exploited as the basis for the contextual visualization of naturally complex scenes. For instance, Fig. 8 displays a version of the maneuvering context: the left side of boundary objects are articulated to confine an area of a post office; a post should be separated from a sign board to follow the open space.

The multi-fractal articulation scheme has been applied to various natural scenes to detect landmark objects to notice for pedestrians and vehicles going through the roadways [11]. A part of experimental results is shown in Fig. 9 where various types of landmark objects required for the mental development of an intelligent vehicle [2] are visualized through the chromatic saliency detection. These results imply that a kind of emotional expression, including the values and/or special taste, the expression of space and movement, and feeling of fear, can be computed on the environmental saliency.

## 7 CONCLUDING REMARKS

Various types of landmark objects can be detected via chromatic saliency analysis in naturally complex scenes.

The probability distribution evoked by boundary objects can be articulated within a color space coordinated by *as-is* primary. The saliency patterns yield a system of fractal attractors consistent with the ground-object structure.

## REFERENCES

- [1] M. F. Barnsley. *Superfractals*, Cambridge University Press, 2006.
- [2] V. Braitenberg. *Vehicles: Experiments in Synthetic Psychology*, The MIT Press, 1987.
- [3] J. J. Gibson. *The Ecological Approach to Visual Perception*, Houghton Mifflin Company, 1979.
- [4] C. Hagelhall, T. Purcell, and Taylor R. P. Fractal dimension of landscape silhouette as a predictor for landscape preference. *Journal of Environmental Psychology*, 24:247–255, 2004.
- [5] L. Itti, C. Kock, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. and Machine Intell.*, PAMI-20:1254–1259, 1998.
- [6] K. Kamejima. Generic representation of self-similarity via structure sensitive sampling of noisy imagery. *ENTCS*, 46:225–244, 2001.
- [7] K. Kamejima. Laplacian-gaussian sub-correlation analysis for scale space imaging. *Int. J. Innov. Comp. Info. Con.*, 1:381–399, 2005.
- [8] K. Kamejima. Chromatic information adaptation for complexity-based integration of multi-viewpoint imagery – a new approach to cooperative perception in naturally complex scene –. *Int. J. Innov. Comp. Info. Con.*, 4:109–126, 2008.
- [9] K. Kamejima. Anticipative coding and *In-Situ* adaptation of maneuvering affordance in a naturally complex scene. In *Advances in Human-Robot Interaction*. In-Teh, 2010.
- [10] K. Kamejima. Environmental saliency: Stochastic representation and dynamic perception. In *Proc. IIIS-WMSCI'10*, pages II:92–97, 2010.
- [11] K. Kamejima. Randomness-based saliency arising in naturally designed scene. In *Proc. D&E 2010*, page #21:12pages, 2010.

# A Fuzzy Multistep Filter for Video Noise Removal

Mike Nachtegaele

Ghent University, Dept. of Applied Mathematics and Computer Science  
Krijgslaan 281 (Building S9), 9000 Ghent, Belgium  
mike.nachtegaele@ugent.be

**Abstract**—Images and video are among the most important information carriers. They can contain a lot of information in the blink of an eye, and thanks to technological evolution they are easy to capture, transmit and store. The application area is very wide, from industrial and military applications to commercial and medical imagery. Due to (technical) errors during capturing or transmitting an image or video, noise can occur. Many techniques have been developed to deal with this noise, and the challenge still remains. In this paper we present a noise removal algorithm specifically developed for the removal of random impulse noise in video, using fuzzy logic as tool to obtain the desired results. The main ideas are relatively generic, and could be extended to specific application areas. Basically, to obtain a good trade-off between enough noise removal and detail preservation, the filter consists out of 4 steps. In every step we calculate for every pixel a noise degree and a noise-free degree. If the noise degree is greater than the noise-free degree, a motion-compensated filtering is applied. The different degrees are obtained using fuzzy rules that are the mathematical translation of expert knowledge, i.e., they represent the (approximate) way in which a human would detect noise in an image or video. The visual and numerical results are very good, and outperform several existing alternative algorithms.

## I. INTRODUCTION

Fuzzy set theory and fuzzy logic are mathematical models to deal and to reason with imprecise information. As noise detection is uncertain (to what degree is a pixel contaminated?) and noise removal is imprecise (to what extent do we correct a pixel value?), fuzzy set theory and fuzzy logic turn out to be very valuable tools to develop new algorithms for image and video denoising. In that regard, many papers show that fuzzy-inspired filters often outperform classical approaches, both numerically (e.g., in terms of Mean Square Error or Peak-Signal-to-Noise-Ratio) and visually (evaluation by a panel of experts). For example, in [4] we have discussed the variety of fuzzy filters for denoising, both for impulse noise (where a fraction of the pixel values is replaced by either fixed noise values or random noise values) and gaussian noise (additive noise), that were developed in cooperation with our research group. In each case, the proposed fuzzy filter was compared to existing classical and other fuzzy filters to show the better performance. Table I contains the filters and references for still images. Figure 1 shows some visual results of the FIDRM filter, indicating the power of this fuzzy filter for impulse noise removal.

Compared to images, it is quite clear that video (image sequences) makes things much more complicated: the intro-

duction of a temporal component implies that good filtering can only be achieved if the temporal aspect is taken well into account, i.e., if motion information is incorporated in the filter design. In this paper we present a 4-step fuzzy filter for the removal of random impulse noise in video. After an introduction to the basic required concepts from fuzzy set theory and fuzzy logic (Section 2), we outline the construction of the proposed filter (Section 3). Experimental results are discussed in Section 4.

	gray	color
fixed impulse	FIDRM [7]	FIDRMC [8] HFMRC [5] HFC [9]
random impulse	FRINR [10]	HFC [9]
gaussian	GOA [12] FuzzyShrink [6]	FCG [11]

TABLE I  
THE DIFFERENT FUZZY FILTERS (WITH REFERENCES) FOR STILL IMAGE NOISE REDUCTION, DEVELOPED IN OUR RESEARCH GROUP.

## II. BASIC NOTIONS FROM FUZZY SET

A crisp set in a universe  $X$  is characterized by an  $X \rightarrow \{0, 1\}$  mapping, where 1 indicates that an element belongs to the set and 0 indicates it doesn't. A fuzzy set  $A$  in a universe  $X$  is characterized by an  $X \rightarrow [0, 1]$  mapping  $\mu_A$ , called the membership function [15], where  $\mu_A(x)$  indicates the degree to which the element  $x$  in  $X$  belongs to the set  $A$  or satisfies the property expressed by the set  $A$ . In other words, fuzzy sets allow membership degrees between 0 and 1 and thus a more gradual transition between "belonging to" and "not belonging to". This makes fuzzy sets very useful for the processing of human knowledge, where linguistic values (e.g. large, small, ...) are used. For example, a difference in gray level is not necessarily small or not small, but can be small to some degree. A possible membership function of the fuzzy set `small` is given in Figure 2.

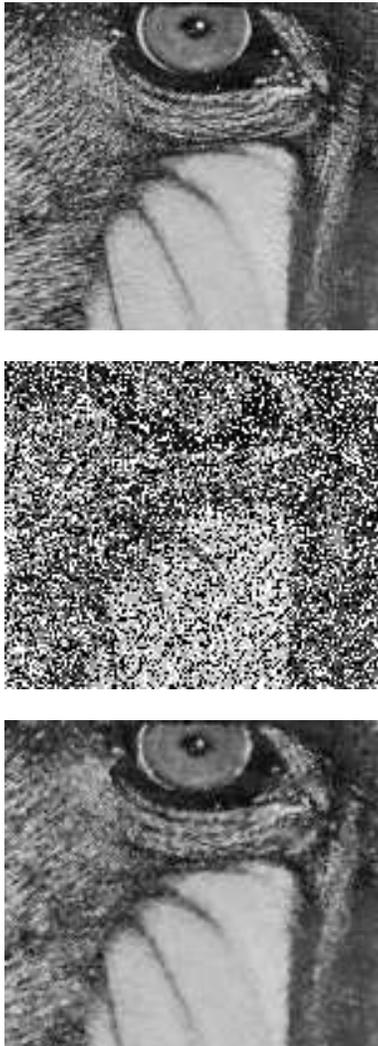


Fig. 1. Noise removal from the Mandrill image: top = part of the original image, middle = image contaminated with 50% impulse noise (salt & pepper noise), bottom = denoised result with the FIDRM filter.

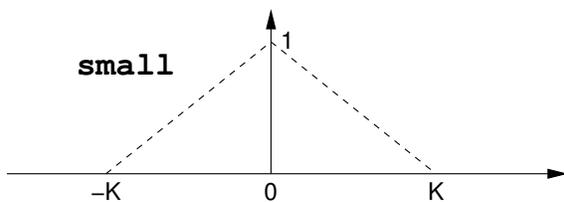


Fig. 2. A possible membership function of the fuzzy set *small* (the parameter *K* can be chosen by the user, depending on the application). The closer an element is to 0, the higher its membership value is.

The extension from crisp to fuzzy sets comes along with an extension of the underlying binary logical framework to fuzzy logic. In fuzzy logic, expressions can be true or false to a certain degree, and consequently we should be able to connect such expressions (with the logical NOT, AND, OR, ...) using fuzzy logical operators that extend their binary counterparts. This can be achieved by using fuzzy logical operators, such as negators (NOT), conjunctors (AND) and disjunctors (OR).

Formally [13], a negator  $N$  is a decreasing  $[0, 1] \rightarrow [0, 1]$  mapping that satisfies  $N(0) = 1$  and  $N(1) = 0$ , a conjunctor  $C$  is an increasing  $[0, 1] \times [0, 1] \rightarrow [0, 1]$  mapping that satisfies  $C(0, 0) = C(1, 0) = C(0, 1) = 0$  and  $C(1, 1) = 1$ , and a disjunctive  $D$  is an increasing  $[0, 1] \times [0, 1] \rightarrow [0, 1]$  mapping that satisfies  $D(1, 1) = D(1, 0) = D(0, 1) = 1$  and  $D(0, 0) = 0$ . The boundary conditions ensure that these fuzzy operators are real extensions of the binary NOT, AND and OR. Popular examples are  $N_s(a) = 1 - a$ ,  $C_M(a, b) = \min(a, b)$  and  $C_P(a, b) = a \cdot b$ ,  $D_M(a, b) = \max(a, b)$  and  $D_P(a, b) = a + b - a \cdot b$ , respectively, with  $a, b \in [0, 1]$ .

Having fuzzy sets to model linguistic values and fuzzy logic to reason with them, fuzzy rules can be used to model human reasoning and to derive new (imprecise) knowledge from given (imprecise) knowledge. An example of a fuzzy rule is an expression of the form

IF (( $p$  is  $P$  AND  $q$  is  $Q$ ) OR ( $r$  is NOT  $R$ )), THEN ( $s$  is  $S$ ),  
with  $P, Q, R, S$  fuzzy sets (modeling linguistic values) and  $p, q, r, s$  elements from the corresponding universes. The degree  $S(s)$  to which “ $s$  is  $S$ ” (e.g., to which a pixel is considered noisy) is given by the degree to which the antecedent of the rule (i.e., the IF-part) is true. This degree is given by

$$S(s) = D(C(P(p), Q(q)), N(R(r))),$$

using a disjunctive  $D$ , a conjunctor  $C$  and a negator  $N$ .

With the above tools we are able to create a mathematical model for human reasoning with imprecise knowledge (= approximate reasoning). In this paper we apply this to the construction of a noise removal filter. In general, the added value of fuzzy set theory and fuzzy logic is demonstrated in the scheme in Figure 3.

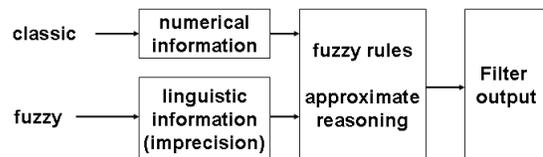


Fig. 3. Fuzzy filters not only use numerical information to filter out the noise in images, but can also work with linguistic information. Furthermore, fuzzy logic allows us to reason with this linguistic information and enables us to better approximate human reasoning.

### III. A FUZZY MULTISTEP FILTER FOR RANDOM IMPULSE NOISE REMOVAL FROM VIDEO

We consider the situation in which an original image sequence  $I_o$  is corrupted by impulse noise with random noise values coming from a uniform distribution, resulting in a noisy image sequence  $I_n$ :

$$I_n(x, y, t) = \begin{cases} I_o(x, y, t), & \text{with probability } 1 - pr, \\ \eta(x, y, t), & \text{with probability } pr. \end{cases}$$

In this expression,  $I_n(x, y, t)$  and  $I_o(x, y, t)$  stand for the grey value of the pixel with spatial coordinates  $(x, y)$  in the  $t$ -th frame of the noisy and original sequence and  $pr \in [0, 1]$  denotes the probability that a pixel is corrupted and replaced by a random noise value  $\eta(x, y, t)$  coming from a uniform distribution.

Numerous 2D techniques have been developed for denoising still images corrupted with impulse noise. These 2D filters can be applied on each frame of the sequence consecutively, but because temporal correlations between successive frames are neglected temporal inconsistencies will arise. Among the available 3D impulse noise filtering schemes, that also take into account pixels from neighbouring frames, are the 3D rational filter (RAT) [1], the adaptive 3D median filter (A3DM) [2], the weighted 3D median filter (W3DM) [2], the lower-upper-middle smoother (LUM) [3] and the peak-and-valley filter (PAV) [14]. The main drawback of these filters is that they either filter too many noise free pixels (detail loss) or too less noisy pixels (by trying to preserve the details).

In our proposed method the noise is removed step by step, and in every step fuzzy rules are used to model the human reasoning to distinguish noise-free from noisy pixels. As the experimental results will confirm, this leads to a good trade-off between noise removal and detail preservation.

#### A. Filter Layout

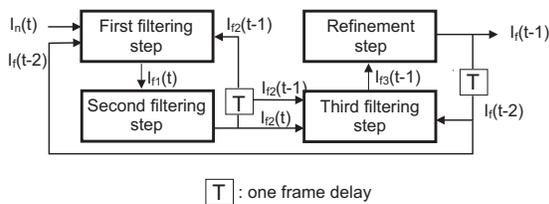


Fig. 4. Overview of the different steps in the proposed algorithm.

We have specifically chosen for a multistep layout. By removing the noise in successive filtering steps we can achieve a much better preservation of details, because it is easier to distinguish noise from small details if a considerable part of the noise has already been removed in a previous step. The filter that we discuss here has four filtering steps (Fig. 4).

In the first step, both a degree to which a pixel is considered noise-free and noise is calculated. If the noisy degree is larger

than the noise-free degree, the pixel is filtered in a motion compensated way.

The second step filters pixels for which the difference in grey value to all its neighbours is large positive or large negative. If there is one neighbour for which the difference is not large positive and not large negative, it is checked whether the considered pixel and that neighbour might belong to a line. If not, the pixel will be filtered also.

In the third step, the remaining noise is filtered based on temporal information.

The fourth and final step is intended to remove small impulses that may not have been detected up to then, but that are relatively large in a homogeneous (spatially) or non-moving (temporally) area.

In the next subsections we will discuss how the above scheme is translated into fuzzy rules (e.g., to determine whether a pixel might be corrupted by noise or not), effectively implementing approximate reasoning.

#### B. Filter Construction

1) *First Step:* We adopt the following reasoning. A pixel can be considered noise-free if it is similar to the pixel at the same spatial location in the previous and preprevious frame and has also some similar neighbours. If the noise-free situation of the pixel is not confirmed in time, possibly due to motion, we will require more spatial confirmation (similar neighbours).

We can translate this (approximate) reasoning into the following fuzzy rule to obtain the noise-free degree.

##### Fuzzy Rule 1:

IF ( the absolute differences  $|I_n(x, y, t) - I_{f_2}(x, y, t - 1)|$  and  $|I_n(x, y, t) - I_f(x, y, t - 2)|$  are NOT LARGE POSITIVE AND there are two neighbours  $(x + k, y + l, t)$  ( $-2 \leq k, l \leq 2$  and  $(k, l) \neq (0, 0)$ ) for which  $|I_n(x, y, t) - I_n(x + k, y + l, t)|$  is NOT LARGE POSITIVE

OR there are four neighbours  $(x + k, y + l, t)$  ( $-2 \leq k, l \leq 2$  and  $(k, l) \neq (0, 0)$ ) for which  $|I_n(x, y, t) - I_n(x + k, y + l, t)|$  is NOT LARGE POSITIVE

THEN the pixel at position  $(x, y, t)$  is considered NOISEFREE.

In a similar way, we adopt the following reasoning to detect noisy pixels. A pixel can be considered noisy if the difference in grey value compared to the pixel at the same spatial location in the filtered previous frame ( $I_{f_2}$ , see Fig. 4) is large positive or large negative and if this does not hold for its neighbours (then we can assume that the difference is not caused by motion) and if this is confirmed spatially by the fact that there is a direction in which the differences in grey level between the considered pixel and the two respective neighbours are both large positive or large negative and if the difference in grey value of those two neighbours is not large positive (then we can assume that the pixel is an impulse between two pixels that are expected to belong to the same object).

We can translate this (approximate) reasoning into the following fuzzy rule to obtain the noise degree.

##### Fuzzy Rule 2:

IF ( ( the difference  $I_n(x, y, t) - I_{f_2}(x, y, t - 1)$  is LARGE

POSITIVE AND NOT for five neighbours  $(x+k, y+l, t)$   $(-2 \leq k, l \leq 2$  and  $(k, l) \neq (0, 0))$  the difference  $I_n(x+k, y+l, t) - I_{f_2}(x+k, y+l, t-1)$  is LARGE POSITIVE) OR (the difference  $I_n(x, y, t) - I_{f_2}(x, y, t-1)$  is LARGE NEGATIVE AND NOT for five neighbours  $(x+k, y+l, t)$   $(-2 \leq k, l \leq 2$  and  $(k, l) \neq (0, 0))$  the difference  $I_n(x+k, y+l, t) - I_{f_2}(x+k, y+l, t-1)$  is LARGE NEGATIVE))

AND (( in one of the four directions the differences  $I_n(x, y, t) - I_n(x+k, y+l, t)$  and  $I_n(x, y, t) - I_n(x-k, y-l, t)$   $((k, l) \in \{(-1, -1), (-1, 0), (-1, 1), (0, 1)\})$  are both LARGE POSITIVE OR both LARGE NEGATIVE) AND the absolute difference  $|I_n(x+k, y+l, t) - I_n(x-k, y-l, t)|$  is NOT LARGE)

THEN the pixel at position  $(x, y, t)$  is considered noisy.

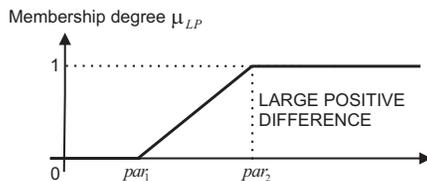


Fig. 5. Membership function of the fuzzy set *large positive*.

We use the membership function  $\mu_{LP}$  shown in Fig. 5 to represent the linguistic value *large positive*. Its reflection is the membership function of the linguistic value *large negative*. Optimal values for  $par_1$  and  $par_2$  ( $par_1 = 11$ ,  $par_2 = 43$ ) have been determined experimentally, based on the results of different training sequences each corrupted with different noise levels. In this paper,  $N_s$ ,  $C_P$  and  $D_P$  are used as logical operators because they yield the best results.

The degree to which the pixel at position  $(x, y, t)$  belongs to the fuzzy set *noisefree* corresponds to the degree to which the antecedent in fuzzy rule 1 is true, and is denoted as  $\mu_{nf}(x, y, t)$ . The degree to which the pixel at position  $(x, y, t)$  belongs to the fuzzy set *noisy* is denoted as  $\mu_n(x, y, t)$ .

All pixels for which  $\mu_n(x, y, t) > \mu_{nf}(x, y, t)$  are filtered, the other pixels remain unchanged. The filtering is performed in a motion compensated way. For the filtering of a pixel  $(x, y, t)$ , we determine the displacement vector  $(r, s)$  for the best matching  $(2 \cdot W_1 + 1) \times (2 \cdot W_1 + 1)$  block in a search region of size  $(2 \cdot W_2 + 1) \times (2 \cdot W_2 + 1)$  in the previous frame (a noise adaptive mean absolute difference is used to calculate this match). The pixel is filtered as the corresponding motion compensated pixel in the previous frame  $I'(x+r, y+s, t-1)$ , if it exists. Otherwise a spatial filtering is performed, taking the average of all noisefree pixels in the block. Optimal values for  $W_1$  and  $W_2$  ( $W_1 = 2$ ,  $W_2 = 8$ ) were obtained experimentally.

2) *Second Step*: The second filtering step follows a similar scheme as the first step. Only a different reasoning is used to detect the remaining noise: if for all 8 neighbours  $(x+k, y+l, t)$   $(-1 \leq k, l \leq 1$  and  $(k, l) \neq (0, 0))$  the difference  $I_{f_1}(x, y, t) - I_{f_1}(x+k, y+l, t)$  is large positive to

some degree (i.e.,  $\mu_{LP}(I_{f_1}(x, y, t) - I_{f_1}(x+k, y+l, t)) > 0$ ) or for all 8 neighbours this difference is large negative to some degree, then the pixel  $(x, y, t)$  is considered noisy and should be filtered.

We also try to avoid the situation in which we (wrongly) filter a pixel that belongs to a line. We adopt the following reasoning for this. If  $\mu_{LP}(I_{f_1}(x, y, t) - I_{f_1}(x+k, y+l, t)) > 0$  for 7 neighbours or  $\mu_{LN}(I_{f_1}(x, y, t) - I_{f_1}(x+k, y+l, t)) > 0$  for 7 neighbours and for the remaining neighbour both  $\mu_{LP}(I_{f_1}(x, y, t) - I_{f_1}(x+k, y+l, t)) = 0$  and  $\mu_{LN}(I_{f_1}(x, y, t) - I_{f_1}(x+k, y+l, t)) = 0$ , then this neighbour and the central pixel  $(x, y, t)$  might belong to a corrupted line. It is then checked whether for the opposite neighbour or one of the neighbouring neighbours or one of the pixels in their prolongation the difference in grey level compared to the central pixel or the neighbour for which the difference in grey value was not large positive and not large negative, is also not large positive and not large negative (i.e., has degree zero). If this is the case, then we have detected a possible line and the pixel will not be filtered. Otherwise no line is detected.

3) *Third Step*: The third filtering step is designed to remove the little clusters of similar noisy pixels that can still be present in the image frame. We try to detect these impulses based on temporal information, by comparing a pixel to the corresponding pixels in the previous and the next frame. The detection will only work well if the next frame is also more or less noisefree, and therefore a frame delay is applied. Again, a noise degree and noisefree degree is obtained by applying some fuzzy rules. If the noisy degree is larger than the noisefree degree, the pixel will be filtered.

We consider a pixel  $(x, y, t-1)$  to be noisefree if its grey value is not very different from that of both its two temporal neighbours  $(x, y, t-2)$  and  $(x, y, t)$  and if the pixel has a similar neighbour (not a large difference in grey value) for which the same holds.

We consider a pixel to be noisy if it differs in grey value from both its temporal neighbours and if there are not 7 neighbours in a  $7 \times 7$  neighbourhood that have a similar neighbour (and are thus more reliable) and that differ in grey value from both their temporal neighbours. If 7 such neighbours can be found, the pixel is expected to belong to a moving object. The number has been chosen small enough not to detect lines and large enough to detect the small noise clusters.

4) *Fourth Step*: In the final steps, the result from the previous step is further refined both temporally and spatially. Some very small impulses might not have been detected yet. However, such impulses might be relatively large in non-moving and homogeneous areas. Since the pixels in non-moving areas will correspond to the pixels in a previous or next frame, remaining isolated noisy pixels can be detected more by performing a temporal detection. Additionally, also a spatial refinement is performed.

### C. Experimental Results

The performance of the proposed method is compared to the above mentioned 3D filters RAT, A3DM, W3DM, LUM and PAV, as well as to the 2D fuzzy random impulse noise reduction method (FRINRM, [10]) to show that our filter takes real advantage from the temporal information. The FRINRM filter is a good representative, since in [10] it is shown that it outperforms all other compared state-of-the-art 2D methods. All methods have been processed on several image sequences (Salesman, Trevor, Tennis, Deadline, Miss America, Foreman), with random impulse noise levels between 5% and 30%.

Numerical evaluation is based on the peak-signal-to-noise ratio (PSNR). The higher the PSNR value, the more similar the images. Figure 6 shows a graphical representation of the numerical results of some of our experiments. From this and other graphs we see that the proposed method outperforms all other methods.

Visual evaluation is based on the inspection of several corrupted and restored sequences (available online upon request), which we illustrate here in Figures 7 and 8. Figure 7 shows the original frame, the noisy frame (25% noise) and the filtering result for all of the compared methods for the 20-th frame of the Deadline sequence. It can be seen that the RAT filter results in a moderate detail preservation and also leaves a lot of noise. The PAV filter succeeds in removing the noise better, but for higher noise levels little groups of impulse pixels remain present. Further, also some detail loss arises due to the filtering of too many noise free pixels. This is also the case for both the A3DM and W3DM filter. On the other hand, both filters have an acceptable denoising capacity. The LUM-smoother seems to preserve details quite well, but has problems with moving objects and fails to remove the noise adequately at higher noise levels. For the 2D FRINRM method, we see the expected drawbacks of only using spatial information. The filter has very good PSNR results frame per frame, but when watching the sequences, a lot of temporal inconsistencies arise. Further, the filter also performs less for sequences with a detailed background (e.g., Deadline, Salesman). Figure 8 shows the visual results for the 110-th frame of the Tennis sequence corrupted with 5% noise.

It can be concluded that the proposed fuzzy filter combines good detail preservation with good noise removal.

Regarding complexity of the filter, the largest computational cost is the motion compensated filtering, which increases with the noise level (more noise means that more pixels will need filtering). The average running time on our machine ranged between 0.48 and 3.43 seconds per frame, resp. for 5% and 35% noise. This could be improved by using fast motion estimation techniques.

### IV. CONCLUSION

Fuzzy set theory and fuzzy logic allow us to deal with approximate reasoning. The discussed filter for random impulse noise removal for video, which filters every frame in four steps, demonstrates the power of this approach by outperforming other (classical) methods in terms of PSNR and

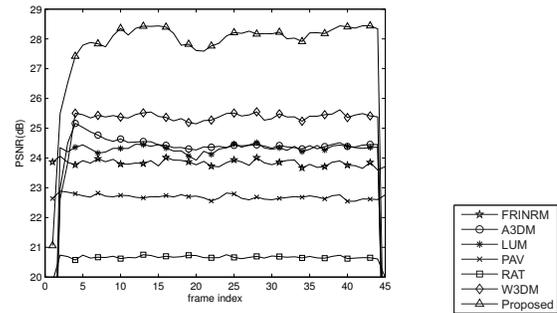


Fig. 6. PSNR for the different methods applied on the Deadline sequence (30% noise)

visually, at the same time obtaining a good balance between noise removal and detail preservation.

### REFERENCES

- [1] F. Cocchia, S. Carrato, G. Ramponi, Design and real-time implementation of a 3-D rational filter for edge preserving smoothing, in: IEEE Trans. on Consumer Electronics, Vol. 43(4), 1997, pp. 1291-1300.
- [2] J.-S. Kim, H.W. Park, Adaptive 3-D median filtering for restoration of an image sequence corrupted by impulse noise, in: Signal Proc.: Image Communication, Vol. 16, 2001, pp. 657-668.
- [3] R. Lukac, S. Marchevsky, LUM smoother with smooth control for noisy image sequences, in: EURASIP J. on Applied Signal Proc. 2001(2), pp. 110-120.
- [4] M. Nachtgeael, T. Mélange, E.E. Kerre, Advances in Image and Video Filtering using Fuzzy Logic, in: Proceedings of ICMV 2010 (3rd International Conference on Machine Vision), IEEE Press, 2010, pp. 215-219.
- [5] S. Schulte, V. De Witte, M. Nachtgeael, D. Van der Weken, E.E. Kerre, A Novel Histogram Based Fuzzy Restoration Method for Colour Images, in: Lecture Notes In Computer Science, Vol. 3708 (Proc. of ACIVS 2005), 2005, pp. 626-633.
- [6] S. Schulte, B. Huysmans, A. Pizurica, E.E. Kerre, W. Philips, A New Fuzzy-Based Wavelet Shrinkage Image Denoising Technique, in: Lecture Notes in Computer Science, Vol. 4179 (Proc. of ACIVS 2006), 2006, pp. 12-23.
- [7] S. Schulte, M. Nachtgeael, V. De Witte, D. Van der Weken, E.E. Kerre, A Fuzzy Impulse Noise Detection and Reduction Method, in: IEEE Trans. on Image Proc., Vol. 15(5), 2006, pp. 1153-1162.
- [8] S. Schulte, V. De Witte, D. Van der Weken, M. Nachtgeael, E.E. Kerre, Fuzzy Two-Step Filter for Impulse Noise Reduction from Color Images, in: IEEE Trans. on Image Proc., Vol. 15(11), 2006, pp. 3567-3578.
- [9] S. Schulte, V. De Witte, M. Nachtgeael, D. Van der Weken, E.E. Kerre, Histogram-Based Fuzzy Colour Filter for Image Restoration, in: Image and Vision Computing, Vol. 25(9), 2007, pp. 1377-1390.
- [10] S. Schulte, V. De Witte, M. Nachtgeael, D. Van der Weken, E.E. Kerre, Fuzzy Random Impulse Noise Reduction Method, in: Fuzzy Sets and Systems, Vol. 158, 2007, pp. 270-283.
- [11] S. Schulte, V. De Witte, M. Nachtgeael, T. Mélange, E.E. Kerre, A New Fuzzy Additive Noise Reduction Method, in: Lecture Notes in Computer Science, Vol. 4633 (Proc. of ICIAR 2007), 2007, pp. 12-23.
- [12] D. Van De Ville, M. Nachtgeael, D. Van der Weken, E.E. Kerre, W. Philips, I. Lemahieu, Noise Reduction by Fuzzy Image Filtering, in: IEEE Trans. on Fuzzy Systems, Vol. 11(4), 2003 pp. 429-436.
- [13] X. Wang, D. Ruan, and E.E. Kerre, Mathematics of Fuzziness - Basic Issues, in: Studies in Fuzz. and Soft Comp., Vol. 245, Springer, 2009.
- [14] P.S. Windyga, Fast impulsive noise removal, in: IEEE Trans. Image Proc., Vol. 10(1), 2001, pp. 173-179.
- [15] L.A. Zadeh, Fuzzy Sets, in: Information and Control, Vol. 8, 1965, pp. 338-353.



Fig. 7. 20-th frame of the Deadline sequence (top-left to bottom-right): original, noisy (25%), RAT, PAV, A3DM, W3DM, LUM, FRINRM and Proposed.

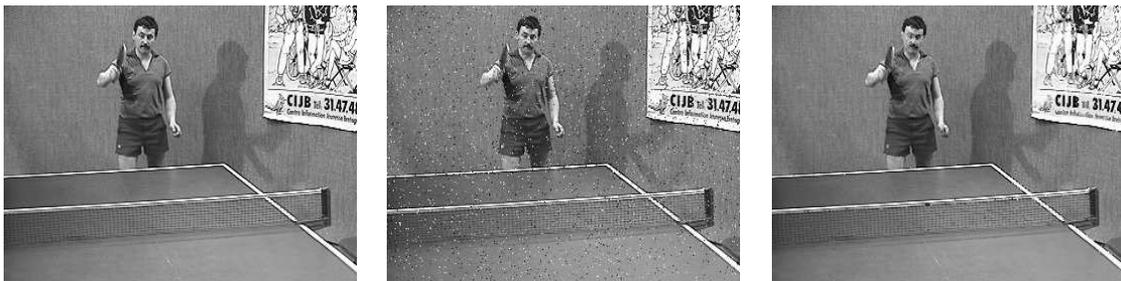


Fig. 8. 110-th frame of the Tennis sequence (left to right): original, noisy (5%) and Proposed.

# Super Resolution: A Database Driven Inference Approach

Monica TRIFAS, Jeremy STRAUB

Department of Mathematical, Computing and Information Sciences

Jacksonville State University

700 Pelham Road N., Jacksonville, AL 36265, USA

## ABSTRACT

A revised super-resolution technique is presented and evaluated in this paper. This technique incorporates prior knowledge gained via training subject domain specific or general purpose images prior to presenting an image for super-resolution image resolution enhancement. When presented with an image to enhance, the engine selects candidate patterns using a low-resolution search mechanism and then uses the higher source-image resolution to select a winning candidate for inclusion in the super-resolved image. This proposed technique is evaluated from an application-agnostic perspective using several evaluation metrics. The failure of common super-resolution evaluation metrics to adequately measure application goal success is discussed. This failure is also directly tied to engine performance as, in many cases, the success metric and the pattern selection metric are tightly aligned. Thus, by developing new super-resolution evaluation metrics and other enhancements, not only can evaluation of training-based super-resolution be improved but also the super-resolved image product as well.

**Keywords:** digital images, resolution, super-resolution, evaluation of super-resolution, image enhancement.

## 1. INTRODUCTION

Super resolution techniques, generally, produce one or a set of high-resolution images from one or a sequence of low-resolution frames. Super-resolution has applications in machine vision, remote sensing, intelligence, medical and a plethora of other fields. Each application has its own requirements; however, the techniques that can be used to enhance an image are applicable across multiple application fields.

A super resolution technique is, generally, evaluated based on its ability to produce a replica of the original high resolution image. Many studies involve down-scaling images, enhancing the produced lower-resolution versions and comparing the super-resolved output image to the original on a pixel-by-pixel basis. Alternate approaches, which take in to account the desired application outcome may, however, be more suitable metrics.

## 2. BACKGROUND

Image resolution refers to the number of pixels contained in an area of an image. The resolution of a digital image is determined by how it is collected (resolution of camera, etc.) and any transformations applied to it (e.g., resizing, etc.). The physical resolution of an image refers to the actual size of the image as stored on disk. Images can have a lower effective resolution than their physical resolution if they have been

blurred or resized (without super resolution techniques) to a larger size than the original. Super resolution is used to increase the effective resolution of the image (or set of images).

Super resolution approaches can be classified in to two main categories: those which infer data and those which simply piece together existing data. Inference-based super resolution may draw on patterns in the image format, prior knowledge or other sources to create an output image that is similar (to some extent) to what would be produced by capturing the original scene or subject at the output resolution. Piecing solutions [e.g. 3, 4], on the other hand, use several images of the scene or subject and take advantage of camera shifts or movement to allow a higher resolution image to be constructed from several lower resolution ones. Most piecing techniques are computationally complex and numerically ill-posed problems. However, early work [e.g., 7] in this area showed that the aliasing effects in the high-resolution fused image can be reduced or removed, if a relative sub-pixel motion exists between the under-sampled input images. Both approaches produce an approximation of the actual high resolution image. Problems can include false positive inclusion and false negative exclusion.

## 3. DISCUSSION OF PROPOSED METHOD

The proposed approach makes an inference from a single source image based on prior knowledge. It uses a database of known patterns that it is trained with prior to being presented with the source image. Each database record contains a high, medium and low resolution version of the applicable pattern. When an image is presented for super resolution, each area of the image is compared to the low and medium size patterns in the database and the best match is selected to be the large size pattern in the final image. The proposed approach is loosely based on [5].

### A. Pattern Sizing

A critical set of decisions that must be made with this super resolution approach is the size of the small, medium and large patterns. These patterns can, theoretically, be of any size; however, for the super resolution methodology to be effective (for most applications), pattern size should be quite small compared to the features in the image. The pattern should represent a part of an image feature (so that patterns will recur throughout various similar features) as opposed to covering a whole feature or (worse) a set of features. In a remote aerial/satellite sensing application, for example, the pattern could be a small part of a building or terrain feature.

Because the proposed technique involves inference, it may create false positives or false negatives when the closest match is not an accurate representation of the true image. If a pattern comprised the whole image feature or set of features, this miss-

match could result in the inclusion of a non-existent feature in to the image, the removal of an actually-existing feature, or the replacement of one feature with another. By keeping the pattern size small, relative to the size of image features, the amount of damage that a single miss-match can do to an image is minimized which decreases the likelihood of an undetected false negative or false positive.

Given this, it is necessary to choose the size of the patterns with the details of application and source images in mind. Multiple databases of trained patterns may be required to handle images with different effective resolutions and feature sizes.

## B. Training

During training, one or more high resolution images are presented to the system and used to populate a database of corresponding high, medium and low resolution patterns. Thus, each source image is split into sections of the large pattern size; medium and low sized patterns are produced by down-scaling the large size pattern. The software iterates through the image (from left to right and top to bottom) with a one pixel offset for each row and column. Thus, the image will produce significantly more than  $H/L \times W/L$  (where H is the image height, W is the image width and L is the large pattern size) patterns. Allowing for the fact that incomplete patterns will not be produced on the bottom and right of the image, the total number of patterns that will be collected from a source image is:

$$N = (H - (L - 1)) \times (W - (L - 1)) \quad \text{Eq. (1)}$$

It is necessary to collect all of the possible patterns (as opposed to simply cutting the image up so that each pixel is part of only a single pattern, a technique referred to as simple slicing) as there is no reason to presume that the image that will later be presented will match the artificial grid created by a simple slicing approach. Thus, simple slicing, would potentially fail to collect the patterns that may later serve to provide the best match as part of the image enhancement process.

In many cases, training will involve the presentation of multiple images. From a logical perspective, the training process involves simply adding additional patterns to the database from each image that is presented. The actual implementation of this, however, must be somewhat different. This is due to the fact that the training time for each subsequent pattern increases as the database must be searched to ensure that the pattern is non-duplicative (and, potentially, to increment an occurrence counter if a duplicate pattern is detected). This becomes particularly time consuming when the database size increases to the point where the index searches are requiring disk access (as opposed to simply searching data in RAM). To resolve this, all patterns could be loaded without checking for duplication and then duplicates could be removed (and occurrence counts set) later. Alternately, the approach that was taken was to train a single image at-one-time and then use the SQL MERGE command to combine the single image databases (and increment the occurrence counter, if desired).

The use of the merging approach has several benefits. First, it eliminates the need to create a routine to search for duplicates and to disable and later re-enable and populate the database index. Second, and most importantly, it allows the various training image datasets to be combined in various combinations for testing and evaluation without requiring re-training.

For a deployment of the proposed method, the database could be customized to contain patterns particular to the specific application, patterns from a wide variety of sources (for general use), or a hybrid of the two. The use of a customized database would, generally, be recommended for applications where images are constrained to one subject. For example, there would be little value in having up-close pictures of plant life in a remote sensing application as they would likely not match (or produce an erroneous match). Some previous research [e.g. 10], however, has demonstrated that patterns can recur at various levels of nature and human construction. This could, potentially, make patterns from one source useful in another. A general database, conversely, would be of greatest use when a particular image subject is not enhanced frequently enough to warrant the creation of a specialized database or a specialized database is not available. A hybrid approach combines a specialized database with a general one. This allows superior coverage of the specialized subject and provides for coverage of elements of the image that are not included in the specialized database.

## C. Image Enhancement

When an image is presented for super resolution, it is processed through in  $M \times M$  sized chunks (where M is the size of the medium pattern). In each instance, the medium pattern from the source image is resized to the small size. This produced small-size pattern is used to search the database for candidate match patterns. These patterns are selected by selecting patterns which fall within a range of  $V_s \pm T$  (where  $V_s$  is the source image value and T is the selected tolerance value) on a pixel by pixel basis (that is, each pixel in the pattern must be within the tolerance range of the source image pattern for it to be selected as a candidate). If a set of candidates is not found within tolerance, the tolerance value is incrementally increased until a match or matches are found (once an expanded tolerance value is set, all matches at that level of tolerance, not just the first identified, are considered).

Once candidate patterns are identified, the medium sized image for each candidate pattern is then compared to the original medium-sized pattern from the source image. The best match from the candidates is then used in the output super-resolved image. An optional enhancement (which was used in the results presented in this paper) is the consideration of adjacent areas when matching (as opposed to considering only the presented medium-sized pattern). This is accomplished by overlapping the pattern with previously enhanced areas. In this case, a weight is applied to the match of the pattern to the candidate and the candidate to the overlap area. The candidate with the best score for the combined selection metric (SM) is then incorporated in the output image.

$$SM = (a \times D_M) + (b \times D_O) \quad \text{Eq. (2)}$$

An additional metric can be added to this selection process for larger data sets. Additional weight can be used to emphasize patterns that appeared more frequently in training. This can be incorporated to create a more robust difference evaluation value (DEV) of

$$DEV = (a \times D_M) + (b \times D_O) + (c \times O) \quad \text{Eq. (3)}$$

where  $a$ ,  $b$ , and  $c$  are the weights for the difference between the compared medium patterns ( $D_M$ ), difference between the overlap areas ( $D_O$ ) and occurrence level ( $O$ ) respectively.

If an occurrence value is to be considered in pattern selection, a significant number of images must be presented to generate a robust set of occurrence values. Alternately, the training process could be modified to group similar patterns together under an averaged pattern and increment its occurrence value. This potentially would have the added benefit of reducing database size while providing additional weight to patterns that are common (while not necessarily recurring identically each time).

#### **D. Use of Commercial Database Product**

The approach is designed to incorporate a commercially available database product. Microsoft's SQL Server 2010 was used during testing; however, any commercially available database product could potentially be substituted. Therefore, the approach, takes advantage (without development cost) of features in commercial database products which include searching, caching and optimization. The availability of the MERGE command, discussed previously, is one such example of previously created functionality which was used to speed the development and testing of the software.

The use of a common database format may also allow greater interoperability with other potential providers or users of patterns. Database tools also aided the combination of various training image configurations that were used in testing, as discussed below.

#### **E. Configuration Learning**

Three key factors influence the successfulness of the super-resolution process, under this approach. First, the size of the small, medium and high resolution patterns can be varied as desired. It is important to consider each pattern size both relative to the image feature size and to the other pattern sizes. The pattern versus feature size, as previously discussed, can have a significant impact on the occurrence (and detection) of false positive and negative inclusions. The relative size of the medium and large patterns determines the amount by which the resolution of presented images will be able to be increased. The enhancement factor ( $EF$ ) is computed as  $EF = L / M$ .

Second, the pattern composition of the database that is used for enhancement can be altered. A database that is domain specific, as discussed previously, will generally perform better than a generic one. Finally, metrics related to matching (how many candidates to select) and closeness (what is considered a possible / acceptable match) can be changed. The adjustment of these factors should allow maximization of the two key super resolution performance evaluation metrics: speed and accuracy. A potential enhancement and topic for future research could involve using an automated neural network (ANN) with feedback (based on speed and accuracy metrics) and backpropagation to optimize these settings.

### **4. DISCUSSION OF SUPER RESOLUTION EVALUATION TECHNIQUES**

It was desired to evaluate the effectiveness of the proposed super resolution method's application without regard to application.

This however, is quite problematic as the success metrics vary from application to application. In a military remote sensing scenario, avoiding false positives and negatives would in many cases be of primary importance. In this context, a false negative could result in an enemy position being missed (possibly rendering friendly forces exposed to an unexpected attack). Alternately, a false positive could result in (in the best case) a waste of munitions on a non-existent target and at worse the unnecessary and undesired loss of human lives.

On the other hand, research has been previously conducted regarding using super resolution to reduce bandwidth needs for streaming video [e.g., 3]. In this context, the presence of false negatives and positives is less important and dwarfed by the need to avoid video discontinuities.

A third scenario to consider would be when super resolution is being used to enhance an image for navigation route-planning purposes. In this instance, discontinuities are less important and even false positive and negative readings are not problematic unless they impact the route generation process. Obviously, both a false positive and false negative inclusion have the possibility to do this. However, the impact of this may be able to be mitigated by the incorporation of a confidence value for each part of a super-resolved image.

Given the above, it is clearly not appropriate to judge all super resolution techniques by a single application-agnostic standard. The conventionally used pixel-difference metric values are exactly this type of standard. As such, this approach fails to capture the multitude of possible application impacts and instead renders judgments that may be of little use.

## **5. EXPERIMENTAL METHODOLOGY**

With the goal of evaluating the proposed super resolution technique in a way that is applicable to at least several (if not most) applications, several experimental scenarios were developed. An implementation of the super-resolution technique (as previously described) was created. Next, training was conducted. The training was common to all experimental scenarios. The scenarios, however, differ in the number and set membership of the trained images and in the evaluation technique used.

The first scenario involved using (and for the two and three image cases, combining – using the SQL MERGE command, as discussed previously) images from the same set as the image that would be presented. The images used in this study were collected during research related to facial recognition. As such, a set of images consists of images of a single individual's face taken at different angles. Thus, in this first scenario, the super-resolution engine was trained with one, two or three images from a given set and then another (not trained) image from the set was presented for super-resolution.

In the second scenario, one, two or four images from different sets were trained (one image per set was selected for the one and two image trainings and two from two sets were trained for the four image training). An image from an alternate (not-trained) set was presented for super-resolution. The subjects of these images, thus, had different skin completions and features as compared to the subject of the image that was presented for super-resolution.

A final control scenario involved training an image and then re-presenting the image for super-resolution. This scenario is significant as it allows evaluation of the performance of the engine (as opposed to the dataset, as the presentation of a trained image guarantees that all necessary patterns for super-resolution will be present in the pattern database).

The super-resolved result images from each of these scenarios (and each number of trained images) were evaluated in three ways. First, the conventional average pixel difference comparison approach was used. This involved calculating the difference in value for each pixel and averaging them. These numbers are presented in table 1 as a percent of the maximum possible difference. This allows the results to be compared across images of various color depth.

Second, in recognition of the fact that for many applications a close enhancement is a complete (not partial) success, a threshold-adjusted pixel difference comparison was computed. This computation treats any value within  $\pm V$  (acceptable variance) as a perfect 0-difference match. Any value outside of this threshold is averaged as per the standard average pixel difference computation.

Finally, a visual appraisal of the resulting images was conducted. This evaluation technique was included to evaluate the proposed technique's utility for applications (such as entertainment) where human perception of how good the resulting images look is the primary success metric.

**6. RESULT SET**

# Trained	Pixel Comparison Value	
	Same Set	Across Sets
	1	5.926173%
2	5.877611%	6.129308%
3	5.876358%	N/A
4	N/A	6.104011%

# Trained	Threshold Px Comparison Value	
	Same Set	Across Sets
	1	0.3939018%
2	0.377937%	0.4959146%
3	0.3775162%	N/A
4	N/A	0.473204%

Table 1 Evaluation of the super-resolved images

**7. DISCUSSION OF RESULTS**

**A. Visual Evaluation**

The images that are produced by the proposed approach are not as visually attractive as one would desire (see Figure 1). The proposed approach performs very well over patches of continuous pattern (for example, hair, skin texture and the background texture). It also performs well when there are features within this type of a texture (for example, wrinkles and fold lines on a face). It even performs well on boundaries between textures (for example, between the hair and face or between face and chin) when a shadow is not present. The technique requires improvement for the images having dramatic

edges (particularly those with shadows), in order to be used in an application where human viewing pleasure is a key metric. Specifically, the approach produces visible discontinuities in enhancing the eyes and the face-shadow-wall boundary.

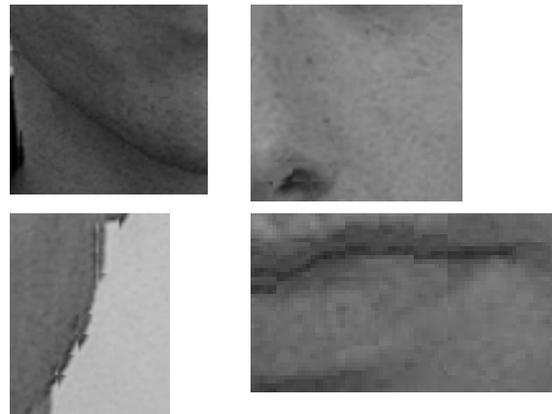


Figure 1a-d The effect of using super-resolution on four areas of an image.

**B. Pixel Comparison**

The pixel comparison evaluation technique directly mirrors the mechanism that is used to select the candidate patterns and choose (from amongst the candidates) the pattern that will be incorporated in the final super-resolved image. Given this, it is not surprising that the proposed approach performs very well with regards to this metric. Thus, by the conventional method of evaluating super-resolution performance, the proposed approach is performing acceptably.

**C. Pixel Comparison with Threshold**

Given that the pixel comparison with threshold approach is actually just a relaxed version of the pixel comparison approach, superior performance for the proposed approach under this metric would be expected (given the performance previously discussed for the pixel comparison approach). This is exactly what the test results have confirmed. For applications where achieving a proximal result represents a complete success, this metric is nearly perfect. As previously discussed, application-specific testing is necessary to ensure that the thresholds selected and results are appropriate to the particular use of the engine.

**D. Alternate Evaluation Techniques**

The pattern selection mechanism that is used has a significant bearing on the outcome of a super-resolution technique for any given metric. If the pattern selection mechanism and the evaluation technique are aligned (such as with the alignment that is demonstrated between the selection technique used in this proposed approach and the pixel comparison evaluation method), then it is obvious that the engine will perform well for the given metric. The identification of metrics, thus, is absolutely critical to any super-resolution application. If a computer-evaluable metric can be developed for a super-resolution goal, it is likely that the engine can be optimized and thus perform well for the application. Given this, it appears future studies of super-resolution would benefit from the

development of different metrics for success in addition to those presented in this paper and currently in use.

## 8. SOURCES OF ERROR

Virtually all super-resolution research uses experimental techniques which are necessarily somewhat contrived. Small and large source images are required, for many performance analysis techniques, requiring that an image be down-scaled, super-resolved and compared to the original high-resolution image. In this instance, images collected for previous facial recognition research were used to test the super-resolution image. These images incorporate a significant amount of foreground (skin) and background area (the painted wall) which has a similar texture. This may skew the results. Given that, for many applications, a significant amount of a quasi-similar pattern may exist, this is not an error per-se; however, it is one way that the results of this testing may not be applicable to other applications. Similarly, the choice of images used may make the results presented uncharacteristic of various applications.

## 9. CONCLUSION

The general suitability of the proposed approach and the engine developed to implement has been demonstrated in this paper.

As an inference-based methodology, the proposed approach is susceptible to producing both false positive and false negative inclusions. However, this can be mitigated through the use of appropriate pattern size values. Additionally, the training occurrence heuristic should (generally) push the engine towards avoiding these conditions as long as the training data is representative of (and inclusive of the feature types within) the image or images being presented for enhancement.

Unlike pure un-trained algorithm approaches, the proposed approach is also able to incorporate two benefits available only via training-based engines. First, it is able to provide a real-world-based occurrence heuristic with the benefits previously discussed. Second, the results can be customized and optimized for various image types via the customization of the training images presented. Output performance can also be improved via the adjustment of the pattern size settings and coefficients discussed previously.

## 10. TOPICS FOR FUTURE RESEARCH

Several areas of possible future research are suggested by the research that has been conducted. First, additional work is needed to develop metrics for evaluating super-resolution performance. This includes developing application-domain-specific metrics and other general-purpose metrics such as a surrogate metric for human image discontinuity perception.

Second, an artificial neural network (ANN) could potentially be used to optimize the engine's configuration parameters by incorporating metric feedback on engine performance across image enhancement runs. Research should be conducted into the feasibility of augmenting the proposed approach with a self-learning capability.

Third, a change to the training process may improve performance. Specifically, grouping together and storing highly similar patterns as a single pattern (as incrementing the

occurrence counter correspondingly) may result in a smaller database size (and thus faster super-resolution) as well as making the occurrence heuristic more useful. A closeness threshold parameter would need to be incorporated. This parameter's setting could also be optimized using an ANN, as previously discussed.

Finally, for super-resolved images to be useful as an input to other artificial intelligence (AI) systems (such as collision avoidance, route planning and such), it would be useful for them to include information regarding how confident the engine is about the enhancement of each particular area. Future research can focus on the development of a confidence value to convey this information.

## 11. ACKNOWLEDGMENT

Credit is hereby given to the Massachusetts Institute of Technology and to the Center for Biological and Computational Learning for providing the database of facial images.

## 12. REFERENCES

- [1] Bailey, Donald G. "Super-resolution of bar codes." *Journal of Electronic Imaging*, 2001: 213-220.
- [2] Bishop, Christopher M., and Andrew Blake. "Super-resolution Enhancement of Video." *Proc. Artificial Intelligence and Statistics*, 2003.
- [3] Borman, S., and R.L. Stevenson. "Super-resolution from image sequences-a review." *1998 Midwest Symposium on Circuits and Systems*, 1998: 374-378.
- [4] Elad, M., and A. Feuer. "Super-resolution reconstruction of image sequences." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999: 817-834.
- [5] Freeman, William T., Thouis R. Jones, and Egon C. Pasztor. "Example-Based Super-Resolution." *IEEE Computer Graphics and Applications*, 2002: 56-65.
- [6] Glasner, D., S. Bagon, and M. Irani. "Super-resolution from a single image." *2009 IEEE 12th International Conference on Computer Vision*, 2009: 349-356.
- [7] Huang, T.S. and Tsai, R.Y. "Multi-frame Image Restoration and Registration", *Advanced Computer Vision Image Processing 1*, p. 317-339.
- [8] Irani, M., and S. Peleg. "Super resolution from image sequences." *10th International Conference on Pattern Recognition*, 1990: 115-120 vol. 2.
- [9] Oktem, Rusen, and Levent Oktem. "A Superresolution Approach for Bar Code Reading." *Bilkent University Department of Electrical and Electronics Engineering*. <http://www.ee.bilkent.edu.tr/~signal/defevent/papers/cr1113.pdf>
- [10] Wang, Qiang, Xiaou Tang, and H. Shum. "Patch based blind image super resolution." *Tenth IEEE International Conference on Computer Vision*, 2005: 709-716.
- [11] Yang, Jianchao, John Wright, Thomas Huang, and Yi Ma. "Image super-resolution as sparse representation of raw image patches." *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008: 1-8.

## Sensor for the Tactile Information

Volf Jaromír, Oliva Jan, Vlcek Josef

Department of Control and Instrumental Engineering, Czech Technical University in Prague,

Faculty of Mechanical Engineering

Technická 4, 166 07 Prague 6, Czech Republic

email: Jaromir.volf@fs.cvut.cz, volf@tf.czu.cz, Jan.Oliva@fs.cvut.cz

### ABSTRACT

The qualitatively new Functional Model of Sensor is introduced – to be caught the mechanical part of the Tactile Information, what is the Slip by the 3D-Force Effect determination, what is the Slip by the 3D-Force Effect determination, to be more better analyzed – the force effect, the conductive elastomer is used; and for the slipping – the contact less optical way. The output values digitizing enables the intended using possibilities, mainly for robotics and biomedical purposes as e.g.: the object grasp or touch.

**Keywords:** Tactile Information, Slipping Sensor, Force Sensor, Gripping Force, Conductive Elastomer, Object Grasp, Object Touch.

### 1. INTRODUCTION

The idea of the using of the tactile information for the technical or biomedical purposes is found out the second part of the 20th century. What the Tactile Information is? For the first point of view – only rather complicated to be defined, what is it? - for a mixture of mechanical, physical, chemical quantities, which can be attacked by a number of surrounding's influences, so that the behavior the object of our interest can be partly or totally exchanged.

From technical branches – especially from Robotics - well known fact, what problem is e.g.: to learn and to control the robot touch / grasp; to be oriented this one - in a mixture of pressures, temperatures, surface elasticity, flexibility, moisture etc.; to be transported the object from one place to the other one, unless this one would be damaged - in worse case (usually, no allowed reversible / irreversible deformations, prolongation etc.), in “better case” – an inadequate gripping force is the trouble-cause for the object slipping. Then, as an exclusive performance by the well trained robot is - the test by the gentle an egg's gripping, or the fragile glasses transport unless this one is slipping by any various liquid levels, or the gentle flower's offer – as the test of very fine gripping force.

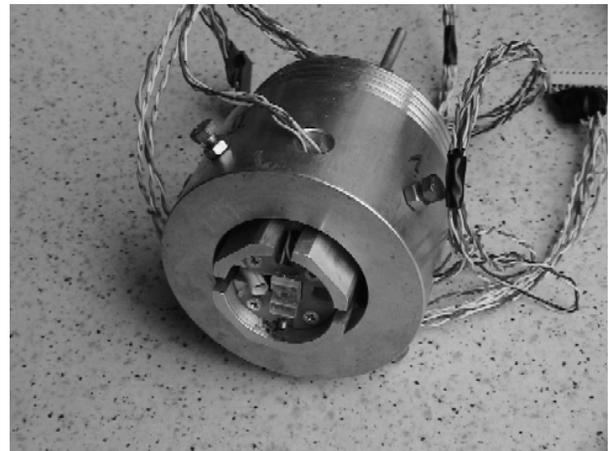


Fig. 1: Sensor for the Tactile Information

Not only technical branches are interested in the Tactile Information, but the biomedical branches, too. Especially in Medicine, the great trend is to help to handicapped people. At present, very popular - the “Artificial Touch”- to try – to return back for handicapped patients, who lost the touch feeling due to injuries, accidents, and for the total prosthesis, too – to realize the feeling of touch by means of electrical signals, which would be converted in the future – into the “understanding” signals for patient by means of the rest of nervous system (only partly, enough – what it could be considered as a success). Seems, to be more then probable – that similar as for robot, it would be necessary, to ensure the learning time for human patient, too. Unfortunately, as would be said above, the tactile information consists not only from the measurable values, but in case of human beings from the only heavy measurable qualities, too – as detection of pain, the detection of mechanical suggestions, and detection of the relative position of parts of the human body, etc.

Of course, for our Sensor - we simplify the conditions only for these mechanical ones, what enables the determination of the slipping and the strength. Mainly the Robotics, and Medicine development demand better results what concerns scanning, transmission and processing of tactile information - for what the above mentioned Sensor is intended.

## 2. TACTILE SENSOR DEVELOPMENT

Only very briefly, the complete sensor philosophy can be seen in block diagram - Fig.2.



Fig. 2: Scheme of the Sensor

Originally, we thought the nine sensors to be used for the detecting of the forces for the Cartesian coordinate system  $[0,x,y,z]$  – eight sensors in the plane  $xy$  (the angle  $8 \times 45^\circ$ ) and the ninth sensor for the detecting of the force of the direction of  $z$ -axis. Later, for simplification – we reduced

the number of the sensors in the “slipping” plane  $xy$  - we used only four sensors ( direction  $+x, +y, -x, -y$  ) for the determination direction of the elementary forces – what can be seen in principle diagram - Fig. 3.

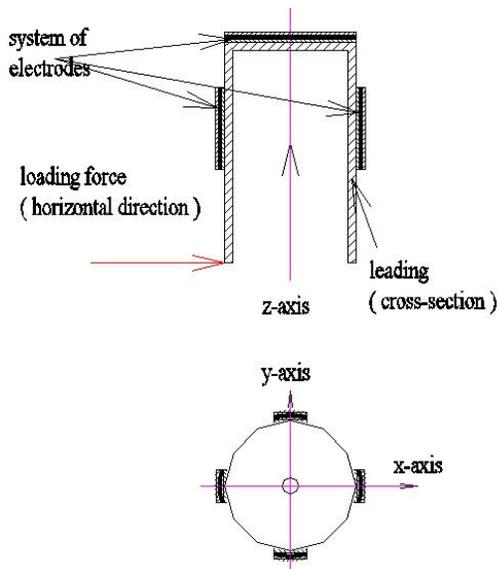


Fig. 3: Elementary Forces Determination

The system of the force sensors consists of two electrodes and a slice of conductive elastomer between them, as sensing material. The conductive elastomer changes its resistivity against the loading force. The output value of this type of sensor is analogue value of the voltage  $U_2$  for

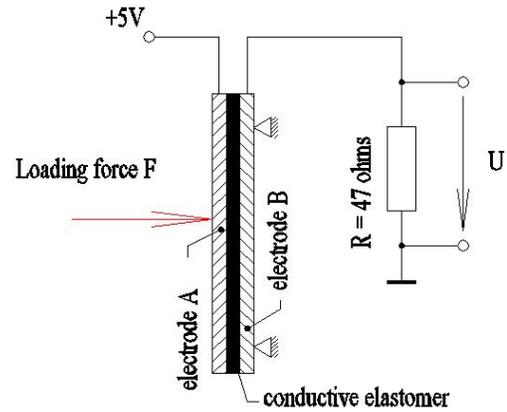


Fig. 4: Force Sensor Arrangement

each sensor. This analogue voltage is turned into the digital value by means of an analogue-digital converter, so we can obtain digital data. The force sensor arrangement can be seen in Fig. 4.

Slipping sensor – there is used an optical principle for the detection of the slipping. The advantage of this system is no contact between the sensor and the gripped object. The slipping can be detected unambiguously instead of mechanical detection. The LED diode ( red light - wavelength  $\lambda = 639-835 \text{ nm}$  ) is used as a light supply emitter. The sensor ADNS-2610 was used as a receiver of the red light. This sensor can catch the beam of the reflected light from the surface of the touched part. This reflected light can turn into the picture. The reflected light is captured in the periods of the time. The sensor can compare the  $(k)$  and  $(k+1)$  picture and calculate the slipping. The sensor sends data of the motion of the direction  $x$  and direction  $y$ . The output value of this type sensor is data of the motion for the direction  $x$  and  $y$ . The optical signal processing is introduced in block diagram - Fig. 5.

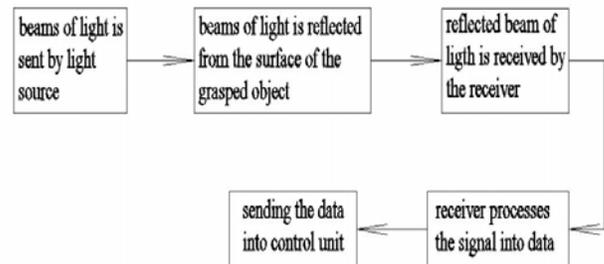


Fig. 5: Slipping Sensor - Optical Signal Processing

All sent data is captured by microprocessor AT89C2051. This microprocessor can send the data into the PC for further processing.

### 3. SENSOR ASSEMBLY TESTING

The experimental verification of sensor functionality was tested after assembling of the sensor and the printed circuit board finishing. The test of the sensor was divided into four parts – test communication between microprocessor and PC, test of the forces sensors, test of the conversion analogue voltage  $U_{2i}$  into digital value, and test of slipping.

#### Test of communication

The program for AT89C2051 was very simple. Each register of microprocessor (register R0-R7) consisted from 8-bit value and this value was sent into PC. The master of this communication was set PC, the slave was a microprocessor. The principle of the communication was set as a half-duplex communication.

#### Test of the force sensors

Each sensor was loaded by weight. The range of the weight was  $<0;900>$ g - and voltmeter reading was set in the range of analogue voltage  $<0;1>$  V.

#### Test of A/D converter

The analogue value of the voltage was set and the data of A/D converter was sent to PC and the value was read. The analogue value was calculated, too and the result was compared the PC value, with.

#### Test of the slipping

The Cartesian coordinate system  $[0,x,y]$  was drawn on the pad. The assembled sensor was moved in direction  $+x$ ,  $+y$ ,  $-x$  and  $-y$  and the data from the sensor of slipping was read. All measured data was in accordance with the datasheet [2].

Being successfully finished these partial tests, and on the base of experimental experiences, any small construction arrangements have been done. All adjustable parts were fixed and marked as “fixed setting” – to be ready for the calibration tests.

#### Loading Calibration Test

All sensors of force for each direction of the loaded force were individual loaded in accordance with the principle of superposition. The sensor was loaded in range of the weight  $<0;900>$ g and the analogue value  $U_{2i}$  was read. It is necessary to indicate the range of loaded force due to correct setting of the upper reference analogue value for the A/D converter, for what - it is better to measure the analogue value  $U_{2i}$ . So, this analogue value was set as output of the  $i$ -force sensor. Each force sensor was measured ten times for each value of the weight and the

uncertainties for each force sensor were calculated. The output value of the forces sensors shows Fig.6.

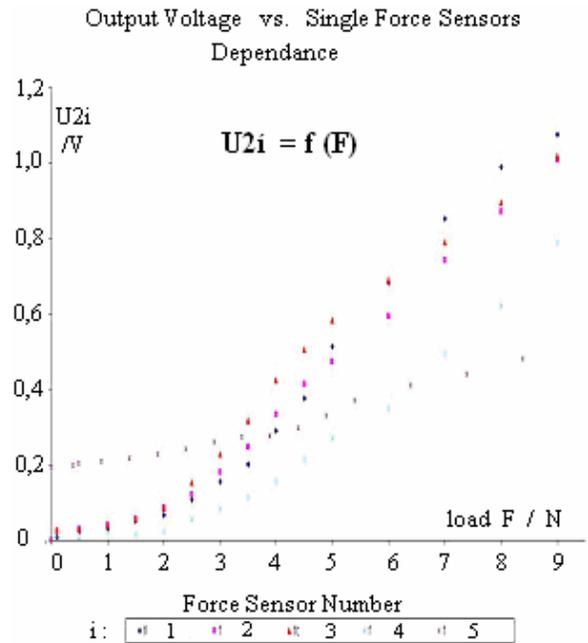


Fig 6: Force Sensors - Sample Ranges Values by Load

The upper reference value can be set suitable on the basis of the range of values of  $U_{2i}$ . If, the upper reference value is set unsuitable, the sensor will be not more sensible due to error of the discretization.

#### Slipping Calibration Test

The sensor of the slipping was calibrated too, but only the  $x$  and  $y$  direction. The measured values were the same like the values from datasheet [2]. For our purposes and for simplification, we don't need to calibrate the speed and the distance of the slip.

### 4. RESULTS EVALUATION

The Functional Model of the Sensor for the Tactile Information has been realized – can be seen in Fig.1. At first, the functionality verification was preferred, and then the principle possibility to be used as a system component (e.g. as sensor, feed-back element, or control element); later, to be received the experiences for the Sensor setting – to be caught the first results.

#### Using possibilities

This qualitatively new Sensor can be used, mainly where, the 3D-Force Effect versus Slip(tendency) is demanded to be followed – as for the Technical branches (as mentioned Robotics – by grasp, touch, gripping etc.); further – for the

Biomedical branches, too (as mentioned Medicine – for handicapped people – to be improved “touch feeling”, including the total prosthesis – however, this one is partly accompanied with the other problem – what the “electrical signal / neural signal converter” is.

#### **Adequate forces limits setting**

By determination the Tactile Information, very frequently – no important so much, what precise an elementary force in each axis is used, but the whole 3D-Force Effect, the result of which can be e.g.: object transport – in case of any inadequate forces - its damage, or slipping. Then, the criterion for successful transport / grasp / gripping is to be the right setting of the adequate forces limits – what means to be excluded, either the slip, and either the object destruction. The others criteria – (what concerns displacement and the number of the acting elements (as “fingers”) – round the object of our interest) - won't be given in this contribution.

#### **Main advantages and disadvantages**

To this time, mostly the main advantages have been introduced for tested Sensor – as: Delivering Information about the Object Slipping via contact less way by the Optical principle using - as result the 3D-Force Effect evaluation; a great number of using possibilities; analogue and digital output signals . The main disadvantages, seems to be: - rather large and rough scale, for the mentioned purposes. (Overall dimensions: diameter  $D = 85$  mm and the height  $H =$  cca 75 mm. The reason for it – to be done - the significant simplification for the principle Sensor's verification); a lot of Adjustable parts (for “well Sensor's tuning” – after any experiences collecting, these parts won't be used for more sophisticated construction, at all. Sensor's minimization is supposed in the future, for what no more details measurements results are given here). The detail bottom side view - can be seen in Fig.7.

#### **Outlook**

The other minimizing can be expected by nanotechnology using. The tested Sensor seems to be useful for technical branches – if, the mechanical part of Tactile Information is followed. In close future, can be expected – to be integrated the others sensing elements, to be improved the total superficial sensation for the Tactile Information determining; so that besides pressures analyzing, could be evaluated e.g.: temperature, moisture etc. – what would be useful, especially for Medicine – to be completed the better comfort of the “real touch” for handicapped people.

In the same time, this problem is the Challenge for the others interdisciplinary projects, too. As a top, for instance, seems - to be no Science fiction, in far future - to realize qualitatively new approach by substitution of “Artificial Touch”, by means of the Genetic engineering technology – to be “grown” the adequate substitution of an afflicted part.

## **5. CONCLUSION**

We assembled the qualitatively new Functional Model of the Sensor for the Tactile Information. This Sensor can detect Tactile Information – the Force Information and information of the Slipping. The output signals digitizing enables full comfort of PC - processing. Signal output can be determined for the human body in the future.

## **6. REFERENCES**

- [1] P. Králíček., “Úvod do speciální neurofyziologie”, Nakl. Karolinum. Praha (in Czech) “Introduction to Special Neurophysiology”, Prague, 2002.
- [2] AGILENT. “Agilent ADNS-2610 Optical mouse sensor.” Datasheet. [online]. <http://cp.literature.agilent.com/litweb/pdf/5988-9774EN.pdf>
- [3] ATMEL: “AT89C2051. 8-bit microcontroller with 2K bytes flash.” Datasheet. [online] [www.atmel.com/atmel/acrobat/doc0368.pdf](http://www.atmel.com/atmel/acrobat/doc0368.pdf)



Fig. 7: Sensor for the Tactile Information - detail bottom side view

## Bidirectional Evolution of Morphological Level Set for Fast Image Segmentation

Guopu Zhu  
 School of Information Science and Technology, Sun Yat-Sen University  
 Guangzhou, China, 510006

and

Shuqun Zhang  
 Department of Computer Science, College of Staten Island / CUNY  
 Staten Island, NY 10314

### ABSTRACT

We propose a novel level set method for fast image segmentation, which evolves level set functions using simple binary morphological operations. The proposed method is superior to a previously reported morphological level set method in capable of bidirectional evolution of level set functions, i.e., the interface of a level set function can either expand or shrink toward the object boundary. The experimental results on image segmentation show the high performance and fast computation of the proposed level set method, which also facilitates parallel hardware and optical implementation.

**Keywords:** Level set method, morphological operation, image segmentation

### 1. INTRODUCTION

The technique of level set methods is a numerical device for modeling dynamic interfaces to partition a domain into several sub-domains. Since proposed by Osher and Sethian [1], the applications of level set methods have ranged from fluid mechanics, epitaxial growth, to image processing and computer vision, and so on. This paper focuses on the application of level set methods in image segmentation, which was first introduced by Caselles et al. [2]. Due to their capability of automatically handling topology changes, level set methods have become very popular for image segmentation [2-5]. However, the expensive computational costs associated with re-distancing level set functions prevent the level set methods from being applied in real-time applications. To reduce the computational cost, some level set methods [6-9] were recently proposed for image segmentation by employing piecewise constant level set functions, especially the binary level set functions. The level set methods introduced in the papers [6, 7] apply a gradient decent procedure to evaluate the binary level set functions for efficient image segmentation. However, they lose the curve evolution property, i.e., the sign change of the binary level set function value is no longer restricted to the neighborhood of the interface of the binary level set function

when the binary level set function evolves. Shi and Karl [8] also presented a novel efficient scheme for evolving the piecewise constant level set functions without the need of solving partial differential equations, but their scheme is a sequential method. Zhang [9] proposed a parallel method by employing basic morphological operations to evolve the binary level set function for image segmentation. It attempts to utilize the real-time processing capability of existing specialized digital morphological processors [10-12], parallel array computers [13,14], or optical parallel morphological processors [15,16] for fast implementation. Whereas, in Zhang's level set method the interface of the level set function can only evolve outward or inward, which requires the initial interface to be set completely inside or outside the object boundary and thus is not convenient. In this letter, we propose a novel morphological level set method, which can bi-directionally evolve the level set function toward the object boundary. Furthermore, we implement a two-phase image segmentation algorithm on the proposed bidirectional evolution scheme. The experiments show the improved performance and the potential for real-time computation.

### 2. PROPOSED MORPHOLOGICAL LEVEL SET METHOD

In level set methods, the level set function  $\phi$  is usually set to be a signed distance function to its interface  $\Gamma$  to prevent the level set function from being too steep or flat near its interface. The signed distance function is defined as

$$\begin{cases} \phi(x) = d(\Gamma, x), & \text{if } x \text{ is inside } \Gamma \\ \phi(x) = 0, & \text{if } x \text{ is at } \Gamma \\ \phi(x) = -d(\Gamma, x), & \text{if } x \text{ is outside } \Gamma \end{cases} \quad (1)$$

where  $d(\Gamma, x)$  denotes the Euclidian distance from the position  $x$  to the interface  $\Gamma$ . A re-distance procedure is usually needed to keep the level set function  $\phi$  to be a signed distance function during its evolution. However, the re-distance procedure is an expensive operation, which prevents the level set methods from being used in real time

applications. Recently, in order to improve the computational efficiency of the level set methods, the signed distance function is replaced by the following binary level set function

$$\begin{cases} \phi(x) = 1, & \text{if } x \text{ is inside } \Gamma \\ \phi(x) = -1, & \text{if } x \text{ is outside } \Gamma \end{cases} \quad (2)$$

Therefore, the re-distance procedure is no longer needed by the binary level set function, and the expensive computational cost can be saved.

For two-phase image segmentation problem, the binary level set function formulated by Eq. (2) can use its interface  $\Gamma$  to partition the image domain  $\Omega$  into two sub-domains, which are inside and outside the interface, respectively. Given a gray-value image  $I : \Omega \rightarrow R^+$ , we assume that image  $I$  can be approximated by a binary function

$$u(\phi, c_1, c_2) = \frac{c_1}{2}(1 + \phi) + \frac{c_2}{2}(1 - \phi) \quad (3)$$

where  $c_1$  and  $c_2$  are two constants. As pointed out in the papers [6, 7], the problem of two-phase image segmentation can be modeled to minimize the energy functional

$$E(\phi, c_1, c_2) = \frac{1}{2} \int_{\Omega} |u(\phi, c_1, c_2) - I|^2 dx + \beta \int_{\Omega} |\nabla \phi| dx, \quad (4)$$

subject to

$$\phi^2 = 1 \quad (5)$$

where  $\beta$  is a nonnegative tuning parameter. The first term in Eq. (4) measures how well the function  $u$  approximates  $I$ , and the second term measures the length of the interface  $\Gamma$ .

It is somewhat difficult to directly solve the constrained minimization problem given by Eqs. (4) and (5). Therefore Tai *et al.*[7] proposed an efficient level set method to obtain the approximate solution of the optimization problem. The efficient method is an iterative procedure, in which each iteration contains three steps: First, evolve the function  $\phi$  by

$$\frac{d\phi}{dt} = \beta \nabla \cdot \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - [u(\phi, c_1, c_2) - I] \frac{\partial u}{\partial \phi} \quad (6)$$

which is obtained by minimizing the energy functional  $E$  in Eq. (4) with respect to  $\phi$  without considering the constraint of Eq. (5). Then, update the parameters  $c_1$  and  $c_2$ , respectively, by

$$c_1 = \frac{\int_{\Omega} I(1 + \phi) dx}{\int_{\Omega} (1 + \phi) dx}, \quad c_2 = \frac{\int_{\Omega} I(1 - \phi) dx}{\int_{\Omega} (1 - \phi) dx}, \quad (7)$$

Finally, compulsively set  $\phi$  as a binary image, i.e., let  $\phi = 1$  if  $\phi \geq 0$  and  $\phi = -1$  if  $\phi < 0$ . Unfortunately, their level set method loses the curve evolution property, i.e., the sign change of the binary level set function value is no longer restricted to the neighborhood of the interface when the binary level set function evolves.

The curve evolution property is preserved by Zhang's level set method [9], which evolves the binary level set function by the basic morphological operations. Since morphological operations are essentially parallel operations and there exist many parallel morphological processors [10-16], Zhang's level set method can take advantage of these computing powers for real-time implementation. However, in Zhang's level set method, the interface of the level set function can only evolve inward or outward, which requires the initial interface to be set completely inside or outside the object boundary. To overcome this problem, we propose a novel morphological level set method in which the interface can evolve bi-directionally. Further, we implement the proposed level set method on the work of Tai *et al.* [7].

Before describing the proposed morphological level set method, we first introduce the basic morphological dilation and erosion operations. Given two binary images  $A$  and  $B$ , we denote  $A \oplus B$  as the dilation of  $A$  by  $B$  and  $A \ominus B$  as the erosion of  $A$  by  $B$ , which are defined, respectively, by

$$A \oplus B = \{p : p = x + b, x \in A, b \in B\} \quad (8)$$

$$A \ominus B = \{p : p + b \in A, \text{ for every } b \in B\}$$

where  $B$  is named as structuring element. Essentially, the dilation and the erosion operations are to expand and shrink the boundary of  $A$ , which are similar to evolve a level set function outward and inward, respectively.

The main principle underlying our method is that an annular region along the interface of the binary level set function is built by the dilation and erosion operations, and then the sign of the level set function value changes only in the annular region to form a new interface for the level set function. In accordance with this principle, the interface iteratively evolves until it has converged. Based on the preceding description of the proposed morphological level set method, we detail the complete scheme as follows:

**Step 1:** Initialize a binary level set function  $\phi$ , and let  $A = (\phi + 1) / 2$ .

**Step 2:** Build an annular region along the interface by  $E = (A \oplus B - A \ominus B)$ .

- Step 3:** Compute the parameters  $c_1$  and  $c_2$  according to Eq. (7).
- Step 4:** Evolve the function  $\phi$  according to Eq. (6),
- Step 5:** Update the speed image by  $V=1$  if  $\phi \geq 0$ ,  $V=0$  if  $\phi < 0$ .
- Step 6:** Perform  $D = E \text{ AND } V$ , and then let  $A = (A \odot B) \cup D$  and  $\phi = 2A - 1$ . Go to Step 2.

There are several points that should be pointed out. First, the preceding procedure proceeds until the evolution of the level set function  $\phi$  has converged or the number of iterations has reached. Second, the morphological operation is implemented on the binary image, so the binary level set function  $\phi$  described by Eq. (2) needs to be converted into a binary image according to  $A = (\phi + 1)/2$ . Third, in Step 4 the value of the level set function  $\phi$  is not restrict to be -1 and 1, it is determined by the initial value and Eq. (6). Fourth, in Step 6 ‘AND’ denotes a logical ‘AND’ operation. Fifth, Tai *et al.*’s level set method [7] on two-phase image segmentation is a special case of our algorithm when the annular region in Step 2 extends to the whole image, i.e., all pixels in the binary image  $E$  are 1. Finally, when the speed image is fixed and the rectangular structuring element with size  $L \times L$  pixels is used, the iteration steps of the preceding procedure will be no more than  $\lceil \max\{M, N\} / \lfloor (L-1)/2 \rfloor \rceil$ , where  $M \times N$  is the size of the binary image  $A$ .

### 3. EXPERIMENTS

We conducted experiments on some real images to compare the proposed level set method with Zhang’s level set method as well as Tai *et al.*’s level set method. In all our experiments, the structuring element  $B$  was set as a rectangular structuring element with size  $3 \times 3$  pixels, and the tuning parameter  $\beta$  was chosen as  $0.01 \times 255^2$ .

Two of the experiments are presented here. Figure 1 shows the segmentation of the cup image. The size of the cup image is  $240 \times 240$  pixels. Fig. 1(a) is the original image, and Fig. 1(b) shows the initial interface of the level set function superimposed on the original image for all of the level set methods. Figures 1(c) -1(f) show the segmentation results obtained, respectively, by the level set method of Tai *et al.*, Zhang’s level set method with the interface shrinking, Zhang’s level set method with the interface expanding, and the proposed level set method. Due to losing the curve evolution property, the level set method of Tai *et al.* performs the image segmentation like an adaptive threshold segmentation method. This can be seen from Fig. 1(c) that the segmented regions by this level set method include some undesired regions with relative small gray-values, and exclude some desired regions with relative large gray-values. Since Zhang’s level set method can only move in one

direction, i.e., inward or outward, the initialization of the interface that is neither completely inside nor completely outside the cup boundary, leads to the incorrect segmentation as shown in Figs. 1(d) and 1(e). It is seen from Fig. 1(f) that the proposed level set method can accurately segment the cup from the background.

The segmentation of another image, the hand image, is shown in Fig. 2. It also demonstrates that the proposed level set method outperforms the other level set methods.

All the above experiments were run on a 1.5GHz PC with 1GB Memory, and the corresponding algorithms were implemented in C++ without any code optimization. The computation times of the morphological-based evolution methods for general fixed speed function are approximately 0.03 seconds, which is comparable to Shi and Karl’s sequential method<sup>8</sup> and faster than the Insight Toolkit (ITK) software implementation. For a changing speed function, the proposed morphological level-set method requires approximately 0.3 seconds. We expect to achieve at least several orders of speedup when the morphological level-set method is implemented by a digital or parallel morphological processor. Therefore, it has the potential for real-time processing.

### 4. CONCLUSIONS

We have described a novel morphological level set method for fast two-phase image segmentation. The proposed level set method has the advantage of either shrinking or expanding the interface of a level set function to conform to the object boundary. The experiment results show the potentials of the proposed level set method in terms of both performance and real-time processing. In the future work, we will apply the proposed level set method to the multi-phase image segmentation.

### REFERENCES

- [1] S. Osher and J.A. Sethian, “Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations,” *J. Comput. Phys.* **79**, pp. 12-49 (1988).
- [2] V. Caselles, F. Catte, T. Coll, and F. Dibos, “A geometric model for active contours in image processing,” *Numer. Math.* **66**, pp. 1-31 (1993)
- [3] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” *Int. J. Comput. Vis.* **22** (1), pp. 61-79 (1997).
- [4] T. Chan and L. Vese, “Active contours without edges,” *IEEE Trans. Image Process.* **10** (2), pp. 266-277 (2001).
- [5] G. Zhu, Q. Zeng, and C. Wang, “Dual geometric active contour for image segmentation,” *Opt. Eng.* **45** (8), Art. No. 080505 (2006).
- [6] J. Lie, M. Lysaker, and X.C. Tai, “A binary level set model and some applications to Mumford-Shah image

segmentation," *IEEE Trans. Image Process.* **15** (5), pp. 1171-1181 (2006).

[7] X.C. Tai, O. Christiansen, P. Lin, and I. Skjælaaen, "Image segmentation using some piecewise constant level set methods with MBO type of project," *Int. J. Comput. Vis.* Published online, July 10 (2006).

[8] Y. Shi and W.C. Karl, "A fast level set method without solving PDEs," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 97-100 (2005).

[9] S. Zhang, "A fast level set implementation method for image segmentation and object tracking," *Proc. SPIE*, vol. 6312, Art. No. 63121R (2006).

[10] P. P. Jonker, *Morphological image processing: architecture and VLSI design*, Kluwer Academic Publishers, Amsterdam, The Netherlands, 1992.

[11] L. Abbott, R. M. Haralick and X. Zhuang, "Pipeline architectures for morphologic image analysis," *Machine Vision and Applications* **1** (1), pp.23-40 (1988).

[12] C. Chen and D. L. Yang, "Realisation of morphological operations," *IEE Proceedings: Circuits Devices and Systems* **142** (6), pp. 364-368 (1995).

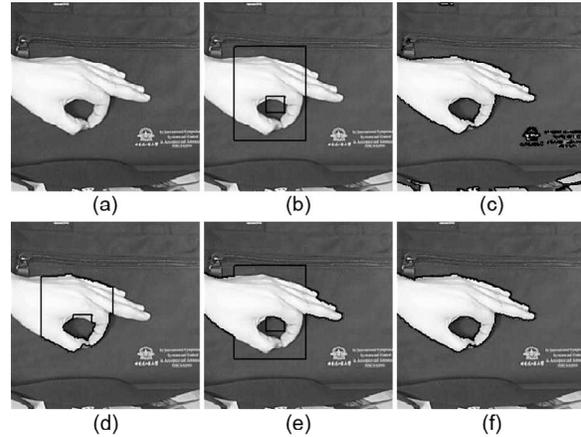
[13] I. Diamantaras and S. Y. Kung, "A linear systolic array for real-time morphological image processing," *Journal of VLSI Signal Processing* **17** (1), pp. 43-55 (1997).

[14] T. J. Fountain, K. N. Matthews and M. J. B. Duff, "The CLIP7A image processor," *IEEE Trans. Pattern Anal. and Machine Intell.* **10** (3), pp.310-319 (1988).

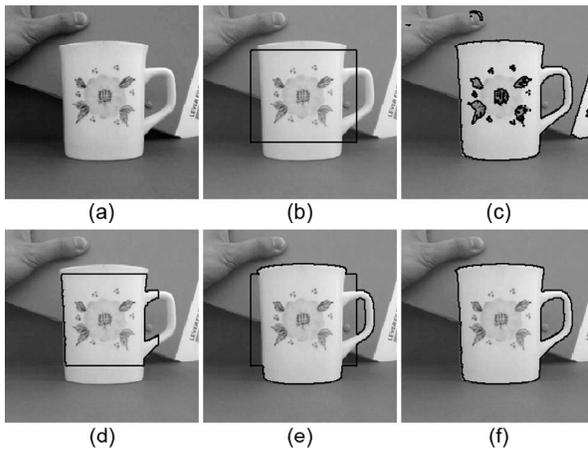
[15] D. Casasent, "General-purpose optical pattern recognition image processors," *Proc. of the IEEE* **82** (11), pp. 1724-1734 (1994).

[16] S. Zhang and M. A. Karim, "Image Distance Transforms that Use Optical Correlation," *Appl. Opt.* **39** (5), pp. 796-802 (2000).

**Fig. 1** Segmentation of the cup image. (a) The original cup image. (b) The initialization of the level set methods. (c)-(f) The segmental results obtained by, respectively, the level set method of Tai *et al.*, Zhang's level set method with the interface shrinking, Zhang's level set method with the interface expanding, and the proposed level set method.



**Fig. 2** Segmentation of the hand image. (a) The original hand image. (b) The initialization of the level set methods. (c)-(f) The segmental results obtained by, respectively, the level set method of Tai *et al.*, Zhang's level set method with the interface shrinking, Zhang's level set method with the interface expanding, and the proposed level set method.



## SonicFields: an immersed sonic experience (sonicfields.net)

Selcuk ARTUT  
Visual Arts and Visual Communication Design, Sabancı University  
Istanbul, TR

and

Ahmet Güzererler  
Visual Arts and Visual Communication Design, Sabancı University  
Istanbul, TR

### ABSTRACT

Sonicfields explores an immersive sonic experience in a virtual 3D environment via a 2D flat screen surface. Users of the World Wide Web are able to visit the system and navigate through a bewildering sonic architecture. Simultaneously, pre-registered occupants are allowed to introduce and manipulate their own sounds in the system in order to create personal soundscapes. With its user centric accumulation hubs, Sonicfields, as its name proposes, is a remediated 3D soundscape, a model of a sonic-world with myriad sounds emerging from various directions. This research exposes a distinct technique for referencing and browsing sonic material in a hyper mediated space without explicit addresses as an alternative method to currently available online media browsing structures.

### Keywords

Digital Communities, Soundscapes, Web 3D, Interaction on Web

### 1. INTRODUCTION

Sonicfields embodies an experience of navigated movement within a three-dimensional sound world by giving expression to visitors' awareness through ear-witnessing a vast collection of sound objects distributed in that world. The state of being continuously confronted by the assorted sounds from multiple sources evokes a feeling of exploration. According to R. Murray Schafer:

Soundscape is a sonic environment. Technically, any portion of the sonic environment regarded as a field for study. The term may refer to actual environments, or to abstract constructions such as musical compositions and tape montages, particularly when considered as an environment.<sup>1</sup>

As Schafer's definition suggests, Sonicfields exhibits a collectively intelligent environment that is attentively configured and constructed by its inhabitants. However, exploration in Sonicfields is a distinguished activity when considering the physical criterion of exploration's typical ontological definition. Navigating in Sonicfields is not a physical movement of a body within a metric universe, but rather the movement of a browsing activity. That is to say, a metric in Sonicfields is not based on

the notion of a physical distance determined by the parameters of an object's position in a three-dimensional world, as one would suppose. Alternatively, the term "Sonicfields" stands for a physical distribution of sounds in a field that is dynamic with its constantly shifting architecture driven by user interactions.

### 2. MAPPING OF 3D WORLD ONTO 2D SURFACES

Since the beginning of the use of flat screens as visual interfaces for computer graphics, there has been extensive research carried out in exploration for innovative methods of mapping 3D worlds onto 2D surfaces. A computer screen (better to referred to as a visual display unit), while nothing more than a circuitry employed to illuminate rendered objects, inevitably frames an intangible reality within our material physicality. While our experience with flat representation easily dates back to Lascaux, this fascination still fuels our curiosity today. As Manovich states:

The imprisonment of the body takes place on both the conceptual and literal levels; both kinds of imprisonment already appear with the first screen apparatus, Alberti's perspectival window. According to many interpreters of linear perspective, it presents the world as seen by a singular eye, static, unblinking and fixated.<sup>2</sup>

Throughout the history of visual representation, there have been many techniques developed that specifically cater to our visual curiosity—techniques such as stereoscopic imaging and anaglyph imaging, to name only two. Such varied techniques all share the same aspiration of mapping a three-dimensional space onto a two-dimensional plane.

Compared to the representation of still images, time-based media is a dynamic sequence of imagery that conveys the notion of temporality, and this temporality subsequently signifies another aspect of the animation of images: navigation. Navigation is a word used extensively in reference to the

---

<sup>1</sup> Schafer, R.M. (1994). *The soundscape: our sonic environment and the tuning of the world*, 274-275. Vermont: Destiny Books.

---

<sup>2</sup> Lev Manovich's "An Archeology of a Computer Screen," published in Kunstforum International, Germany, 1995; [http://www.manovich.net/TEXT/digital\\_nature.html](http://www.manovich.net/TEXT/digital_nature.html) (data of access: 19 April 2010)

guidance of ships and airplanes as they move through space, but more recently the word has been progressively associated within World Wide Web (WWW) terminology. In the initial phases of online exploration, a WWW navigator was a person who wandered across web space in a predominantly serial fashion. But with the increasing proliferation of three-dimensional web platforms, the term navigation is reasserting itself with the approximate meaning of the term referenced in aeronautical spaces: as the event of moving through space. As stated by Chiu et al. (2005) user interface that is familiar to way-finding conventions may help users to navigate more effectively among multimedia documents. Therefore navigation, in a three-dimensional sense, is a potential experience that now can take place within our acknowledged frame-screen. Yet it is here on the edge between frame-screen and experience where the quandary of bodily imprisonment is revealed.

Although navigation is undeniably a physical activity of movement in material reality, screen-based interaction redefines it as a physically stationary activity undertaken in front of a screen. Yet this semantic conundrum of physically static navigation, and its accompanying questioning of reality, is quite literally located in our own tunnel vision. For when we examine the proficiency of our senses to determine directionality from that which we sense, it becomes obvious that our auditory abilities are far superior to our visual abilities. This comparative richness in directional determination enables our sense of hearing to localize objects in space quite easily. And in reference to this auditory ability, the imprisonment issue we confronted visually between the frame-screen and experience while engaging two-dimensional representation is, to a certain extent, disentangled by the notion of encompassing sound environments. In this respect, sound is unconfined and vagabond. This research does not address issues of technical surround sound because current Internet manifestations are incapable of utilizing such technologies. In this research stereo panning and volume manipulations were implemented to partially localize sounds in space. The issues of the third dimension (z-depth) of a sound object in Sonicfields are also addressed with visual clues such as changing the depth of visual appearances of sound objects and by observing their presence and absence. As Lokki and Grön states "3D navigation in a virtual environment is possible with auditory cues alone. However, the fastest and most accurate navigation is obtained when both auditory and visual cues are available."<sup>3</sup>

### 3. PROJECT DESCRIPTION

Navigation in a virtual environment displayed on a two-dimensional flat surface can be a tedious task. There have been several attempts to find an effective and straightforward method for users to feel as if they are comfortably moving in a 3D environment. In Sonicfields, keyboard and mouse interaction, functioning in tandem, were taken as the main interface devices. Correspondingly, First-Person Shooter game controls that enable users to experience action through the eyes of a protagonist were adapted to the navigation technique implemented in this project.

In order to move in the space, visitors are able to use the conventional keyboard combination:

W - Move forward, A - Move right, S - Move downward, D - Move Left

Additionally, for the purpose of changing the protagonist's line of sight, users are required to first press and hold their left mouse button and then move their mouse accordingly.

In Sonicfields, all of the inhabitants in the system are surrounded with their own registered sounds and enclosed within a semi-transparent, wireframe sphere; this single entity is called a sonicsphere. Every sonicsphere contains a nucleus positioned at its center that indicates the name of its particular owner. Synchronously, surrounding sound particles are emitting their interactive contents as audio played through loudspeakers of a sound system (see Figure 1)



Figure 1. Example of a Sonicsphere

Each time a visitor connects to the Sonicfields web site (<http://www.sonicfields.net>), the system reconfigures and redistributes sonicspheres. That is to say, the structure appears different each time one enters the website. Basically, the system detects the most recently updated sonicsphere, and then this sonicsphere is automatically centered on the screen as the initial object of focus. As the visitor navigates in the Sonicfields, he/she is able to hear and track various sounds coming from several other sound particles within a sonicsphere neighborhood. Sonicspheres placed on the initial screen are selected from a set of most recently updated sonicspheres. However, as soon as a visitor leaves the scope of the initial sonicsphere neighborhood, he/she encounters a new sonicsphere that is correlated by similarity analysis to the nearest sonicsphere in the existing neighborhood. The method of similarity analysis will be discussed later in this paper. As one may infer, the whole world is never completely created; on the contrary, it is always partially created. The partial creation process is based on the idea of a dynamic browsing activity instead of a linear exploration. (see Figure 2)

<sup>3</sup> Lokki, Tapio, and Matti Grön. "Navigation with Auditory Cues in a Virtual Environment." *IEEE Multimedia* 12.2 (2005): 80-86. Print.

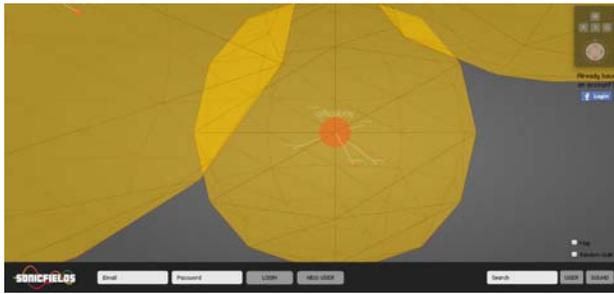


Figure 2. Sonicfields navigation environment

#### 4. INTERACTION IN SONICFIELDS

Every person who is connected to the Sonicfields website can either be a visitor (a guest) or a preregistered user who has already been logged into the system. Both guests and preregistered users are able to visually focus on to any sonicsphere just by clicking on its surface. Double clicking on a surface results in a focus upon the sounds from that particular particle. Further interactions include the ability to solo or mute any audio source. Also, logged-in users are allowed to add any registered user to their favorite lists and are able to include any sound particles into their own respective sonicspheres.

In Sonicfields, every logged-in user is able to edit his/her sonicsphere provided that user is within the bounds of his/her own sonicsphere. In edit mode, one can upload new sounds, edit volume levels and place/replace existing sounds in three-dimensional space through the use of intuitive mouse movements and keyboard combinations. Subsequently, these actions for manipulating the architectural layout of the distribution of sound particles enable a user to create his/her soundscape structure. The user can also tag his/her sounds with terms that he/she feels best describe those sounds.

Through a panel screen that appears when the mouse pointer hovers on a sound particle, visitors/users are able to rate the tags based upon their relevance to the accompanying sound. One does not need to be a registered user to participate in the rating process, and as a consequence the tag rating procedure is accessible to all users. It is this open rating system that is used as a weight to calculate the similarity of sonicspheres, and it is these determinations of similarity among sound particles that are in turn utilized to generate relevance values for tags and their referent sounds.

#### 5. COLLECTIVE INTELLIGENCE AND SIMILARITY FUNCTIONS

The similarity analysis used in building this collectively intelligent system is based on the evaluation of data derived from users' inputs on existing sound materials. Its purpose is to define specific similarities among sonicspheres, not users. The function is determined from the calculations based upon the tags the user has submitted and upon relevance values submitted to the system by its participating voters. The relevance values are used to verify the descriptive relation of a sound object with its corresponding tag. Such a collective autonomous control apparatus is utilized in this system to balance the subjective

efforts needed to describe a sound event semantically and to develop an effective filtering mechanism that operates as a transparent system moderator.

In Sonicfields, Pearson Correlation Score Analysis is used to determine the similarity between two sonicspheres. Pearson Correlation Score takes two sets of data and determines the slope of the best fitting line as a coefficient for each data set. The result of Pearson Correlation Score varies from -1 to 1 where 1 represents a perfect increasing linear relationship, -1 represents a perfect decreasing linear relationship and 0 represents total lack of correlation (see Figure 3). In this study, similarity is accepted as a positive value linear relationship. For any two arbitrary users, the data sets for Pearson Correlation Score calculations are derived from the frequency (occurrence) values of tags that are common to each user's respective sonicspheres. (see Appendix: Pseudo Code)

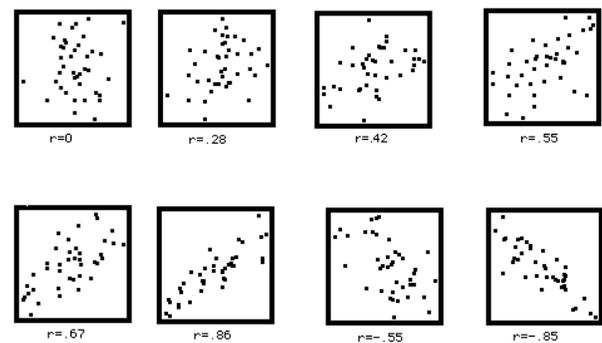


Figure 3. Pearson Correlation Score Analysis Results

#### 6. CONCLUSION

The current architecture of Sonicfields is purposely constructed with user centric sonicsphere structures. Alternatively, tag-based architecture could be offered as an additional feature to the existing user centric model. In such tag architecture, a semantic similarity approach such as Google Similarity Distance that offered by Cilibrasi, Rudi and Vitanyi, Paul at 2007 must be used for determining the equivalence of tags since the meaning of a word (tag) becomes a crucial element of the analysis.

Although Sonicfields users are able to create various individual soundscapes, these soundscapes are not visually distinguishable, searchable or taggable as a whole. In order to improve Sonicfields as a soundscape design platform, the feature of grouping sound particles as a soundscape could be introduced, and as a result, soundscapes could have rating scores in addition to inheriting sound particle properties.

Today, extensive use of diversified media and widespread access to data resources have resulted in a demand for finding fast and effective methods for cataloging, ordering and grouping information in interactive browsing systems. Sonicfields offers users the opportunity to participate in certifying relevance of information with the use of a dedicated collective voting system. Beyond this autonomous moderation capability (qualified by the browsing activity operating with words as inputs), navigation in the sound library is distinctive when compared with existing commercially available search engines such as Google, Yahoo, Bing, etc. In order to perform a search for sound material with an existing search engine, users are required to submit

corresponding text clues into relevant search fields. These search results are then received as lists with secondary reference links for accessing the intended sound materials. Alternatively and in addition to text-based search, Sonicfields offers its users a mode to navigate a space filled with suspended sound materials. That is to say, the user is able to immerse him/herself in the innate universe of the material itself and transcend the limitations of secondary linking to the targeted resource. Moreover, navigation in Sonicfields engenders a stimulating and organic search experience as every sonic neighborhood is in a permanent state of creation influenced by user-driven determinations.

## 7. APPENDIX

### Pseudo Code for Similarity Analysis

```

for each tagA
  where   tagA is a distinct element for user1 &
         tagA is a distinct element for user2

  F1(tagA) = relevance(tagA) * frequency of
            distinct(tagA) / # of tags in user1

  F2(tagA) = relevance(tagA) * frequency of
            distinct(tagA) / # of tags in user2

  insert F1(tagA) to dataset1;
  insert F2(tagA) to dataset2;
    
```

Similarity = Pearson Correlation Score (dataset1, dataset2)

## 8. REFERENCES

- [1] Schafer, R.M. 1994. *The soundscape: our sonic environment and the tuning of the world*. Vermont: Destiny Books.
- [2] Manovich, L. 1995 *An Archeology of a Computer Screen* published in *Kunstforum International*, Germany; [http://www.manovich.net/TEXT/digital\\_nature.html](http://www.manovich.net/TEXT/digital_nature.html) (data of access: 19 April 2010)
- [3] Lokki, Tapio, and Matti Grön. "Navigation with Auditory Cues in a Virtual Environment." *IEEE Multimedia* 12.2 (2005): 80-86. Print.
- [4] Patrick Chiu, Andreas Girgensohn, Surapong Lertsithichai, Wolf Polak, and Frank Shipman. 2005. *MediaMetro: browsing multimedia document collections with a 3D city metaphor*. In *Proceedings of the 13th annual ACM international conference on Multimedia (MULTIMEDIA '05)*. ACM, New York, NY, USA, 213-214. DOI=10.1145/1101149.1101182 <http://doi.acm.org/10.1145/1101149.1101182>
- [5] Rudi L. Cilibrasi, Paul M.B. Vitanyi, "The Google Similarity Distance," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 3, pp. 370-383, Mar. 2007, doi:10.1109/TKDE.2007.48

# Automatic Scoring in a Task of Retelling Stories for Language Learners

Meng Chen, Dean Luo, Lan Wang

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences/The Chinese University of  
HongKong

{chenmeng, da.luo, lan.wang}@siat.ac.cn

## ABSTRACT

In this paper, we propose a novel method for automatic scoring of retelling stories. By taking consideration of possible paraphrases, new scoring features on contents are proposed with the use of ASR in addition to traditional evaluation measures. Linear models are introduced to combine different features for automatic scoring. We evaluated machine scores by correlating them with human scores which were manually rated by an expert. Experimental results show the correlation between machine scores and human scores can be improved.

**Keywords:** Automatic Scoring, Retelling, Computer-Assisted Language Learning

## 1. INTRODUCTION

Recently, systems for computer-assisted language learning (CALL) have shown great advantages over traditional methods. It would be a much cheaper alternative, which is accessible at any time and at any place, and certainly tireless. In these systems, the main task is how to provide the type of feedback that a human teacher would provide. From a pedagogical point of view, a score for the overall assessment of the language learners is important. In order to provide feedback of score without the presence of a human teacher, methods for automatic scoring are required.

Many researchers have studied automatic methods based on Automatic Speech Recognition (ASR) for evaluating the speaking ability of language learners. There are mainly two kinds of tasks in previous work. One is for the restricted speaking tasks such as reading aloud. For example, Neumeyer and Franco presented a system for automatic evaluation of the pronunciation quality in task of reading English texts aloud [1] [2]. Cucchiarini et al. developed a system for Dutch pronunciation scoring along similar lines [3] [4]. The other is for the unstructured, unrestricted, and spontaneous speech. For example, Educational Testing Service (ETS) investigates

the automatic scoring of unrestricted, spontaneous speech of non-native speakers in the task of questions and answers [5].

In this study, we focus on the task of retelling stories, which has been proved an efficient way to improve oral proficiency of a language learner [6] [7]. In the test, students listen to a monologue of story (200~300 words) spoken by a native speaker, and then retell the story with their own words. The responses of students are spontaneous speech with lexical and syntactic errors. This means the vocabulary of ASR should include variations that are extended from the original story but semantically similar.

We approach automatic scoring by extracting scoring features from the output of ASR, and linear regression models to combine different scoring features. Considering the uniqueness of retelling, content related features are critical for the automatic scoring of the retelling task. Therefore, we compute the similarity to represent the content correctness of speech by comparing recognition hypothesis from ASR with all possible paraphrased expressions extended from original story. We also improve the feature of keyword coverage rate based on the vocabulary extended rules. Experiments showed a higher correlation between machine scores and human scores than the traditional methods.

In the remainder of this paper, the scoring features and linear regression model will be introduced in Section 2. The experiments and results will be presented in Section 3. Section 4 is the conclusions and discussions.

## 2. AUTOMATIC SCORING FEATURES AND METHODS

### 2.1 Scoring features

The traditional scoring features are mainly based on the intelligibility and fluency of non-native speech such as the global and local log-likelihood derived from the HMM log-likelihood [1], the rate of speech, and silence

length etc. However, content related scoring features are not taken enough consideration for automatic scoring. The possible reason may be it's not necessary for the task of reading text because the prompts are known to the learner. And for spontaneous speech in the task of questions and answers, the content of speech is difficult to predict. However, for the task of retelling, because the content of speech is related to the topic and the reference story, the extraction of content related scoring features is feasible and critical for automatic scoring.

For the task of retelling, the students are required to repeat the story as possible as what they hear, however, they would try to express with their own words when they can't remember clearly. That means the vocabularies and sentences in the speech of students are not the same as the original story. Therefore, the correct answers are not the only one. For the vocabulary, synonyms and near synonyms should be considered. And for the sentence, different sentence patterns but semantically similar should also be included.

According to this uniqueness, we made the extended rules to generate possible paraphrased expressions from the original story. Firstly, we extended the original story manually based on different sentence patterns. Then, we replaced the keywords with their synonyms and near synonyms automatically. As current stage, we didn't consider pruning during extending. After this process, the possible paraphrased expressions are generated, and used in the automatic scoring scheme.

Based on the extended rules above, we introduce new content-related scoring features, which are given as below,

- 1) Similarity, which shows the content correctness of speech,

$$Similarity = \max(S_1, S_2, \dots, S_i, \dots, S_n), \quad (1)$$

where  $S_i$  is the similarity between the speech of student with the  $i_{th}$  possible paraphrased expression.

- 2) Keyword Coverage Rate (KCR),

$$KCR = \frac{\sum_{i=0}^n cover(k_i)}{n}, \quad (2)$$

where  $k_i$  is the  $i_{th}$  keyword and  $n$  is the number of all keywords, and

$$cover(k_i) = \begin{cases} 1, & \text{if } k_i \text{ occurred in the speech} \\ 0, & \text{if } k_i \text{ didn't occur in the speech} \end{cases} \quad (3)$$

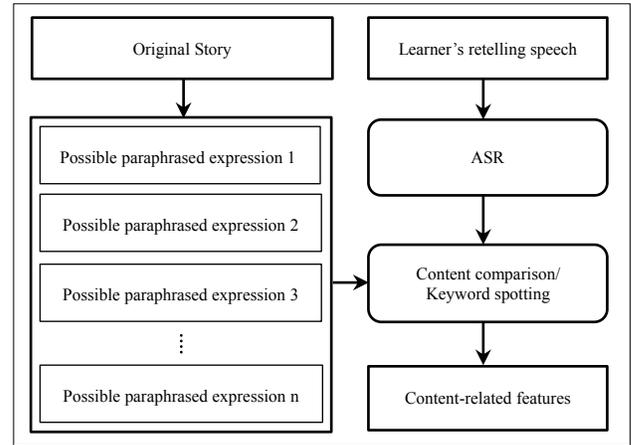


Figure 1: Block diagram of content-related features extraction.

Figure 1 shows the block diagram of content-related feature extraction. We proposed to compute the similarity between recognition hypothesis and all possible paraphrased expressions. Based on the vocabulary extended rules and the content of the presented story, we extended the possible paraphrased expressions first. Then a Dynamic Programming-based string matching was processed between the recognition hypothesis and each possible paraphrased expression (including the original story) for similarity computation. The maximum value was chosen to represent the similarity between the recognition hypothesis and paraphrased expressions, which showed the content correctness of the speech.

We also proposed a novel way of keyword spotting to calculate keyword coverage rate. We extended the keyword set based on the vocabulary extended rules to include more possible keywords which might occur in the speech of retelling task. We also used content-related features of word number and unique word number.

Our scoring feature set, extracted from the ASR results, can be categorized as follows: (1) Content-related, (2) Intelligibility, and (3) Fluency. Table 1 shows a complete list of the scoring features we computed, along with a brief explanation.

## 2.2 Linear regression model

In order to estimate learners' proficiency, linear regression models are used to combine different features. We established various independent variables  $\{x_i\}$  as parameters and the value  $Y$  as the human's score, and

**Table 1 List of features with definitions**

Feature name	Description	Category
Similarity	Similarity between recognition hypothesis and possible paraphrased expression	Content-related
KCR	Keyword coverage rate	Content-related
WN	Number of words in recognition hypothesis	Content-related
UWN	Number of unique words in recognition hypothesis	Content-related
GL	Global log-likelihood	Intelligibility
LL	Local log-likelihood	Intelligibility
ROS	Rate of speech in phoneme level	Fluency
SN	Number of silences in recognition hypothesis	Fluency

the linear regression model was defined as:

$$Y = \sum_i \alpha_i \times x_i + \varepsilon, \quad (4)$$

where  $\varepsilon$  is the intercept. The coefficients  $\{\alpha_i\}$  were estimated by the optimal least-squares of  $\varepsilon$ . Once the coefficients and intercept were determined, the machine score could be predicted with extracted features.

### 3. EXPERIMENT

#### 3.1 Speech recognition system

Our speech recognition system is based on speaker independent continuous Hidden Markov Models. We used the TIMIT/WSJ database to train the acoustic models [8] [9]. The speech is downsampled to 16kHz and preemphasized, and then a Hamming window with a width of 25ms is applied every 10ms. The acoustic features include 12 MFCCs (Mel Frequency Cepstrum Coefficient), energy, delta and acceleration features. The HMMs are composed of three emitting states, each of which has sixteen mixed Gaussian distributions with full covariance matrices. The inter-word triphone HMMs are trained.

In order to enhance the robustness of our ASR system, we conducted the acoustic model adaptation and language model adaptation based on the specific task of retelling.

Based on the extended rules, we could predict the vocabulary in the speech of students. One class is synonymy of the words in the original story, which are

correctly used but not occurred in the original story. The other class is wrong variants of the words in the original story such as the wrong tense of verbs, the wrong singular or plural forms, and the wrong adjectives or adverbs etc. Although these wrong variants are not used correctly, they are much likely to occur in the speech of non-native speaker. With the predicted vocabulary from the original story, we could adapt the general language model to our specific task of retelling.

#### 3.2 Database

We used RETELL data set, which contains 280 responses from 280 speakers and each response is two minute long, for our evaluation experiment. For each response, there are 100-200 words in total. All the speakers are native Chinese high school students. The data was collected in classrooms when the students were using our application for retelling task. An overall proficiency score is given for each learner by an expert in English education. A discrete score scale (from 1 to 6) indicates the overall proficiency (from least proficient to most proficient). These scores are used as reference scores.

#### 3.3 Results

We used leave-one-out cross validation method for experiments on our dataset. Each time we used the data of one learner as test data, and the others for training the scoring models. The Pearson coefficient of correlation between machine scores and human scores is used here as the measure of the agreement between raters (human or machines).

Table 2 shows the correlation of each feature with human scores. From this table, we can see the feature of Similarity, KCR (keyword coverage rate), and UWN (unique word number) have the highest correlation with human scores compared with the traditional features. Table 3 shows the correlations between machine scores and human scores with different combinations of features. In Table 3, we can see the performance was improved when the features of Similarity, KCR, UWN and WN (word number) were chosen. And we obtained higher correlation of 0.621 with human scores by combining all the features. Figure 2 illustrates the percentage of students for different absolute score difference between machine scores and human scores. According to Figure 2, there are 92.9% students whose absolute score difference of machine scores and human scores is less than 2. The average absolute score difference is 0.92. This confirms the effectiveness of our method.

**Table 2 List of features and correlation with human scores**

Feature name	Correlation
Similarity	0.462
KCR	0.496
WN	0.285
UWN	0.481
GL	0.231
LL	0.306
ROS	0.328
SN	-0.148

**Table 3 Correlation between machine scores and human scores for different feature set**

Feature set	Correlation
GL, LL, ROS, SN	0.484
GL, LL, ROS, SN, Similarity, KCR, WN, UWN	0.621

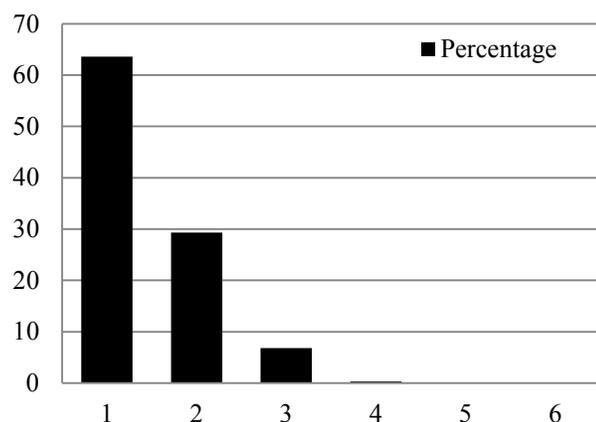


Figure 2 Percentage of students for different absolute score difference between machine scores and human scores

#### 4. CONCLUSIONS

We presented a novel approach to automatic scoring of non-native English speech by taking consideration of the uniqueness of retelling. With improvements on content-related scoring features, we obtained a high correlation of 0.621, which is rather higher than the traditional methods.

An important step for future work will focus on improving speech recognition. We plan to adapt the acoustic model with more data of non-native speakers. And more spoken styles will also be considered in the adaptation of language model. Furthermore, scoring

features that evaluate grammar correctness of speech in retelling task will be explored in order to obtain a broader coverage of communicative competence.

#### 5. ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (No.90920002), and The Knowledge Innovative Project of Chinese Academy of Sciences (No.KJ CX2-YW-617).

#### 6. REFERENCES

- [1] L. Neumeyer, H. Franco, M. Weintraub, and P. Price, "Automatic Text-Independent Pronunciation Scoring of Foreign Language Student Speech", in Proc. of ICSLP96: 1457-1460, 1996.
- [2] L. Neumeyer, H. Franco, Vassilios Digalakis and M. Weintraub, "Automatic scoring of pronunciation quality". Speech Communication 30(2-3): 83-93. 2000.
- [3] Catia Cucchiari, Helmer Strik, and Boves L., "Automatic evaluation of Dutch pronunciation by using speech recognition technology". In Proc. of Automatic Speech Recognition and Understanding, IEEE Workshop. 1997.
- [4] Catia Cucchiari, Helmer Strik, and Boves L., "Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms". Speech Communication 30(2-3): 109-119, 2000.
- [5] Derrick Higgins, Xi Xiaoming, Klaus Zechner and David Williamson, "A three-stage approach to the automated scoring of spontaneous spoken responses". Computer Speech & Language 25(2): 282-306. 2011.
- [6] Morrow L, "Retelling stories: A strategy for improving young children's comprehension, concept of story structure, and oral language complexity". Elementary School Journal, 647-661. 1985.
- [7] Stein, Nancy L., Glenn, Christine G, "An analysis of story comprehension in elementary school children". In R. Freedle (ED), New Directions in Discourse Processes, Vol. 2, pp.53-120. Norwood, NJ: Ablex. 1979.
- [8] J.S. Garofolo, L.F. Lamel, W.M. Fisher, J.G. Fiscus, D.S. Pallett, N.L. Dahlgren, "The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM". NTIS order number PB91-100354, 1993.
- [9] D. Paul, J. Baker, "The Design for the Wall Street Journal-based CSR Corpus". DARPA Speech & Nat. Lang. Workshop, Arden House, NY, Feb. 1992.

## Color Compression Optimization Techniques for Localized Aeronautical Charts

Michael Trenchard and Maura Lohrenz  
Geospatial Sciences and Technology Branch  
Naval Research Laboratory  
Stennis Space Center, MS 39529 USA

Tom Gray  
QinetiQ North America  
Technology Solutions Group  
Stennis Space Center, MS 39529 USA

### ABSTRACT

Aeronautical chart data used widely by the U.S. military and its Allies is typically compressed using vector quantization in a multi-stage approach. The first stage is a simple red-green-blue (RGB) 3:1 color compression, which reduces the chart data from 24-bit color to 8-bit color, using a color palette to map the resulting color indices back to 24-bit RGB. For the compressed U.S. military aeronautical charts discussed in this paper, the color palette is limited to 216 indices to accommodate additional colors for mission-specific overlays. Reducing color from 24-bit to 8-bit is a “lossy” compression technique, since there is potential for color data loss. Optimizing the selection of a color palette to best represent a given set of aeronautical charts is key to minimizing color loss during compression. This paper evaluates two color palette selection techniques and their associated color compression schemes, and compares each with two different color spaces, to determine which combination of techniques will result in the least amount of color loss. The first technique selects the most prevalent 216 color values in a given set of charts as the color palette, and then color-clusters the remaining colors to their closest color palette value by Euclidean distance in the chosen color space. The second technique progressively increases Euclidean distances to generate 216 color clusters and selects the weighted centroid value of each color cluster to produce the color palette. The color palette selection techniques are performed both in RGB (red-green-blue) color space and the Commission Internationale de l’Eclairage (CIE)  $L^*a^*b^*$  (CIELAB) color space. CIELAB color space is designed to be perceptually uniform, meaning that a change of the same amount in color value should produce a change of about the same visual importance. A high color variability aeronautical chart is used to evaluate each technique in terms of average distortion per pixel between the color compressed image and its source. Visual color loss is

expected to be reduced by performing color clustering and color matching calculations in CIELAB color space.

**Keywords:** color compression, digital mapping, aeronautical charts, CIELAB, color palette.

### INTRODUCTION

As part of a development effort for the Canadian Forces, the Naval Research Laboratory (NRL) was tasked with a software development project that would process source files consisting of aeronautical chart data and various mission planning overlays into Compressed Aeronautical Digital Raster Graphic (CADRG) map files [1]. The primary purpose was to fuse mission planning overlays (primary vector overlays) with the underlying, raster aeronautical chart data. In turn, this would enable Navy pilots to access both the underlying aeronautical chart data and their mission planning overlays directly in the embedded digital map systems installed on Navy aircraft.

Typically, the background aeronautical chart data is already in CADRG format [2]. Fusing new map content such as mission planning overlays to the background aeronautical map within the constructs of the mission planning system used by Navy pilots requires that both data types (chart and overlay) first be saved as geotiff images. Once the geotiffs are created, the aeronautical chart and overlays are effectively fused together in a raster image. However, geotiffs are not compatible with the embedded map systems used in most Navy aircraft. These geotiffs must then be converted back into CADRG format for proper display and management on the aircraft map systems.

Converting geotiffs back into CADRG requires a re-compression of the source data with the newly added map content (e.g. mission planning overlays). Vector quantization is a lossy compression technique in which color and spatial data loss may occur. Since the map data

is being re-compressed, special care must be taken to minimize the visual impact of any further data loss.

In this paper, we present and compare two compression techniques in an effort to minimize color loss during color compression. First, we attempt to optimize the selection of color values to construct a color palette. Second, color compression is performed by color clustering in one of two different color spaces (RGB and CIELAB) [3]. We compare our results by calculating average distortion per pixel. A histogram of all unique RGB color values contained within a localized set of aeronautical chart data and the frequency of each occurrence is generated. The first technique selects the most prevalent 216 color values and then color clusters the remaining colors to their closest color palette value by Euclidean distance in RGB color space. The second technique progressively increases Euclidean distances to generate 216 color clusters and selects the weighted centroid value of each color cluster to produce the color palette used for color compression in RGB color space. The third technique again selects the most prevalent 216 color values with the restriction that no two values are closer than the just noticeable difference (JND) value of 2.3 in CIELAB color space [4]. Once those most prevalent colors are determined, the remaining colors are clustered to their closest color palette value in Euclidean distance in CIELAB color space. The fourth technique is identical to the second technique except all calculations are made in CIELAB color space.

### COLOR PALETTE GENERATION

For the purpose of this study, a color palette is generated by the dynamic creation of 216 RGB color values selected from a sample source aeronautical chart that contained a total of 198,972 unique colors (see Figure 1). A histogram of all the unique RGB color values contained within the sample chart with the frequency of each occurrence is obtained. A color palette is constructed for each of the four techniques outlined above and detailed in the following subsections. While two of the techniques are calculated in CIELAB color space, the resultant color palettes are converted and stored as RGB color values.

For the CIELAB color space techniques, we use the 1976 color space definition specified by the CIE [5]. This color space was designed to be a device-independent model describing all the colors visible to the human eye. The three coordinates of CIELAB represent luminance ( $L^*$ ), the position between the hues red/magenta and green ( $a^*$ ), and the position between hues yellow and blue ( $b^*$ ). CIELAB is a nonlinear space modeled after the nonlinear response of the eye to various color stimuli. Raw image data are presented in RGB space. CIELAB is

derived from RGB via the intermediate color space CIE XYZ (ITU-Rec. 709):

$$\begin{aligned} X &= 0.412453R + 0.357380G + 0.180423B \\ Y &= 0.212671R + 0.715160G + 0.072169B \\ Z &= 0.019334R + 0.119193G + 0.950227B \end{aligned} \quad (1)$$

The individual components of CIELAB are then calculated as follows:

$$\begin{aligned} L^* &= 116f(Y/Y_n) - 16 \\ a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[f(Y/Y_n) - f(Z/Z_n)] \end{aligned} \quad (2)$$

where

$$\begin{aligned} f(q) &= q^{1/3} && \text{if } q > 0.008856 \\ f(q) &= 7.787q + 16/116 && \text{otherwise} \end{aligned} \quad (3)$$

$X_n, Y_n, Z_n$  identifies a reference white value in CIE, calculated by applying the RGB-to-XYZ transform to  $R=G=B=255$  (i.e. white), resulting in:

$$\begin{aligned} X_n &= 242 \\ Y_n &= 255 \\ Z_n &= 278 \end{aligned}$$



Figure 1. High color variability aeronautical chart

Each technique is evaluated by a calculation of average color distortion per pixel. Color distortion is calculated by the following measure:

$$\left[ \sum_{j=1}^k \left( \sqrt{(A'-A)^2 + (B'-B)^2 + (C'-C)^2} \right) / k \right] / D \quad (4)$$

$A', B',$  and  $C'$  are component color values from the original image in the selected color space (RGB or  $L^*a^*b^*$ );  $A, B,$  and  $C$  are the pixel values from the color compressed image; and  $k$  represents the total number of

pixels in the image. Distortion values are normalized to the maximum possible distortion (D) in each color space. For RGB, this is the Euclidean distance between white (255,255,255) and black (0,0,0), which is 442. For  $L^*a^*b^*$ , this is the distance between indigo blue (approximately  $L^* = 32, a^* = 79, b^* = -108$ ) and bright green (approximately  $L^* = 88, a^* = -86, b^* = 83$ ), which is 259.

### Technique 1: RGB Most Prevalent

The first technique generates a color palette based on the 216 most frequently occurring RGB color values in the sample image [6]. Since RGB color space is not perceptually uniform, the 216 most prevalent colors were chosen for this color palette regardless of how close (in Euclidean distance) any two colors might be. Once the color palette is generated, a color conversion table is created that acts as an index between the RGB color space and the color palette. Since RGB color space is linearly uniform, each pixel from the source image is color compressed to its nearest color palette match in RGB space by the following:

- 1) Determine the unique 24-bit RGB colors and the number of occurrences of each unique color in the source image.
- 2) Sort the unique colors in descending order by number of occurrences (i.e., the most prevalent color in the image is ranked first).
- 3) Write the top 216 color values to the palette file.
- 4) For every fourth red value between 0 and 255
  - a) For every fourth green value between 0 and 255
    - i) For every fourth blue value between 0 and 255
      - (1) For each RGB value in the palette
        - (a) Calculate the Euclidean distance between the current RGB value and the current palette RGB value.
        - (2) End Loop
        - (3) Determine the smallest Euclidean distance calculated above and write the RGB value and palette index value to the conversion file.
      - ii) End Loop
    - b) End Loop
- 5) End Loop
- 6) When performing color compression, the RGB value from the source file is mapped to the palette by dividing each of the red, green and blue values by four, which is an index into the conversion file. The value pointed by this index is the palette file index for the current RGB color.

Figure 2 shows an example comparison between the original image and the RGB Most Prevalent color

compression technique. A qualitative comparison between the original image and the color compressed image with this technique produced noticeably lighter shades in large homogeneous areas as highlighted in Figure 2. Further, distinct, stepladder transitions between map objects (especially with darker backgrounds) were clearly visible using the RGB Most Prevalent technique. Average normalized distortion per pixel in RGB space for this technique was 0.032, with a maximum distortion of 0.125 and standard deviation of 0.011.

### Technique 2: RGB Weighted Clusters

The second algorithm is also based in RGB color space. However, color clustering is used to determine the most prevalent 216 color clusters. Once the required set of color clusters is acquired, the centroid of each cluster (weighted by the frequency of each color in a cluster) is calculated. The color palette is constructed from the weighted centroids of each of the 216 color clusters.

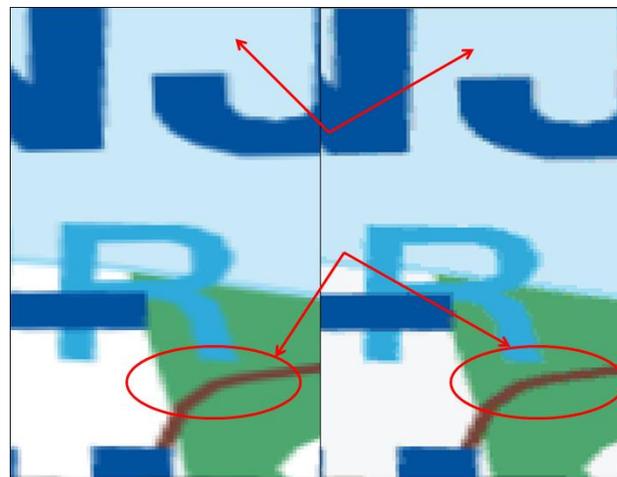
Since the full set of colors that comprise the sample data set are used in the clustering process to build the color palette, each color is indexed via a hash table to its clustered centroid value to color compress the image.

The following steps are used to generate the color palette and color compress the source image for this technique.

- 1) Determine the unique 24-bit RGB colors and the number of occurrences of each unique color from the source image.
- 2) If there are more than 216 unique colors then
  - a) Sort the colors in descending order by the RGB integer value.
  - b) Set an initial tolerance value that determines if two RGB colors are close enough to be clustered.
  - c) Loop through each color and cluster with other (non-clustered) colors whose Euclidean distance is within the tolerance value.
    - i) Calculate the Euclidean distance in RGB space for the current color and all remaining colors in the list.
      - (1) If the Euclidean distance is less than or equal to the tolerance value then
        - (a) Flag the color as clustered to the current color.
        - (2) End if
        - (3) If the number of non-flagged colors is equal to 216 then
          - (a) Exit both preceding inner and outer Loops.
        - (4) End if
      - ii) End Loop
      - iii) Increment the tolerance value.

- d) End Loop
- e) For each non-flagged color (ie. colors that are not marked as clustering to another color).
  - i) If the current color has colors that cluster to it, then
    - (1) Calculate a new RGB color value by taking the weighted average (based on number of occurrences of each color within the source images) of the clustered colors.
  - ii) End if
- f) End Loop
- 3) End If

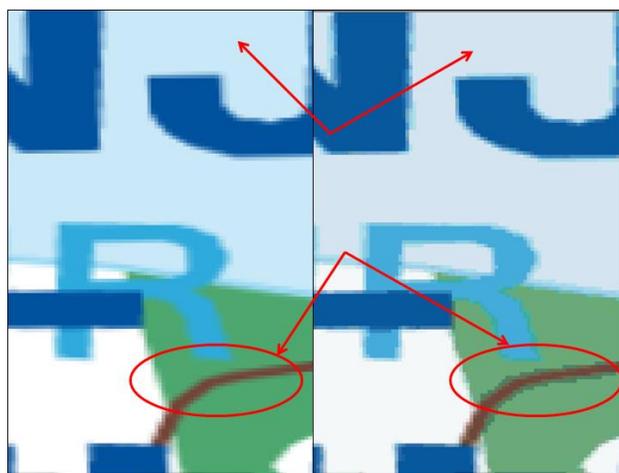
Figure 3 shows an example comparison between the original image and the RGB Weighted Clusters technique. A qualitative comparison between the original image and this color compressed image with this technique produced a much closer reproduction of the original colors. Further, while some color loss is noticeable within the transition between road object and the green background (circled in Figure 3), the transition between the two map objects is much smoother than the stepladder transition shown in Figure 2.



Original Image                      RGB Weighted Clusters

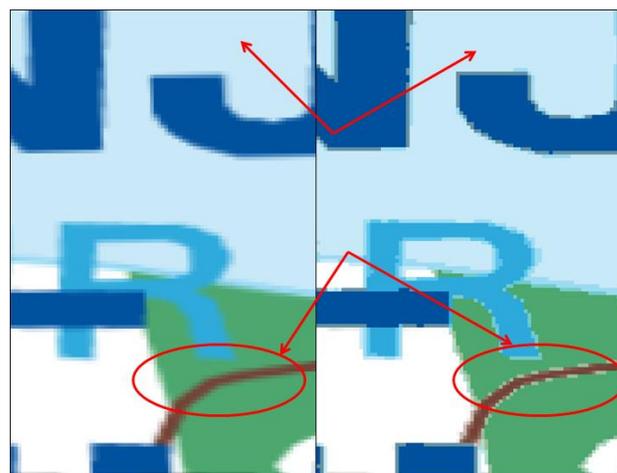
Figure 3. Original vs. RGB Weighted Clusters at 2X zoom. Notice how much closer the background colors are (top arrows) and how much smoother the road feature is (bottom arrows), compared with the RGB Most Prevalent method (Fig. 2).

Average distortion per pixel as measured in RGB space for this technique was 0.022 with a maximum distortion of 0.073 and standard deviation of 0.014.



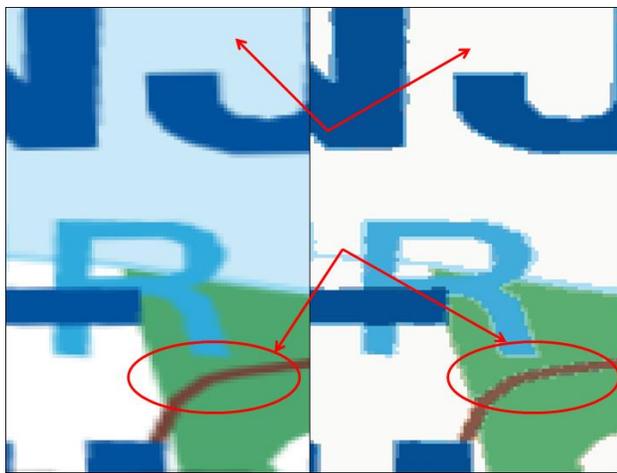
Original Image                      RGB Most Prevalent

Figure 2. Original vs. RGB Most Prevalent at 2X zoom. Notice the lighter background color (top arrows) and jagged edges for the brown road (bottom arrows) in the compressed image (right), compared with the original (left).



Original Image                      CIELAB Most Prevalent

Figure 4. Original vs. CIELAB Most Prevalent at 2X zoom.



Original Image CIELAB Weighted Clusters

Figure 5. Original vs. CIELAB Weighted Clusters at 2X zoom.

### Technique 3: CIELAB Most Prevalent

The third technique again uses the most prevalent set of colors resident within the source image. However, all color values once converted in the CIELAB color space are further filtered to ensure that the Euclidean distance between every possible pair of the most prevalent colors (in CIELAB space) is greater than the JND value of 2.3.

Color compression is calculated by clustering all remaining colors to their nearest neighbor by Euclidean distance in CIELAB color space. Once these clusters are generated, the color compression is performed by indexing clustered color values via a hash table to the most representative prevalent color.

The following steps describe how this technique is implemented.

- 1) Determine the unique 24-bit RGB colors and the number of occurrences of each unique color from the source image.
- 2) If there are more than 216 unique colors then
  - a) Order the unique colors in descending order by the number of occurrences value (ie., the highest number of occurrences will be ranked first).
  - b) Convert each RGB value to a CIELAB value.
  - c) Set an initial tolerance value that determines if two CIELAB colors are close enough to be clustered.
  - d) Loop through the first 216 colors in the list
    - i) Calculate the Euclidean distance in CIELAB space for the current color and all non-clustered remaining colors beginning with color 217 through the end of the list.
      - (1) If the Euclidean distance is less than or equal to the tolerance value then
        - (a) Flag the color as clustered to the current color.
      - (2) End if
    - ii) If the number of non-flagged colors for colors between 217 and the end of the list is equal to 0 then
      - (1) Exit both preceding inner and outer Loops.
    - iii) End if
  - e) End Loop
  - f) Increment tolerance value.
- 3) End Loop
- 4) End if

Figure 4 shows an example comparison between the original image and the CIELAB Most Prevalent technique. A qualitative comparison between the original image and this color compressed image with this technique again produced a much closer reproduction of the original colors than the RGB Most Prevalent technique. However, edge transitions between different map objects were much more distinct.

Average distortion per pixel as calculated in CIELAB color space for this technique was 0.012 with a maximum distortion of 0.298 and standard deviation of 0.018.

### Technique 4: CIELAB Weighted Clusters

The fourth algorithm performs the weighted clustering method used in technique 2 with all color values converted to CIELAB and Euclidean distances calculated within that color space.

Figure 5 shows an example comparison between the original image and the CIELAB Weighted Clusters technique. Surprisingly, this technique performed the worst by far. In fact, color clustering within CIELAB space forced the tolerance level of Euclidean distance measure to cluster the light blue background from the source image with the primary white background of the overall map. In addition, this technique still produced very distinct edge transitions similar to the CIELAB Most Prevalent method.

Average distortion per pixel as calculated in CIELAB color space for this technique is 0.005 with a maximum distortion of 0.095 and standard deviation of 0.015.

## RESULTS

Since distortion numbers are measured within the color space in which the technique was performed, one would not expect distortion numbers to correlate across both

color spaces. Table 1 summarizes the distortion numbers given for each technique. The distortion values did seem to correlate well with the visual representations for the RGB color space techniques [7]. Within CIELAB space, it was difficult to make a clear determination whether the distortion measures correlated to an improvement in the visual representation of the color compressed image [8]. This is true especially in light of the severe color loss of the light blue background in the CIELAB Weighted Cluster technique.

Table 1. Summary of distortion measure results.

<i>Technique</i>	<i>Norm. Avg</i>	<i>Norm. SD</i>	<i>Norm. Max</i>
<b>RGB Prevalent</b>	0.032	0.011	0.125
<b>RGB Clusters</b>	0.022	0.014	0.073
<b>LAB Prevalent</b>	0.012	0.018	0.298
<b>LAB Clusters</b>	0.005	0.015	0.095

From a qualitative standpoint, the RGB Weighted Cluster technique appears to provide the most visually accurate reproduction of the source image of the four techniques evaluated.

### DISCUSSION

The most visually accurate reproduction of the source aeronautical chart is key to the evaluation of techniques to reduce color loss. While transformation into CIELAB color space was expected to improve the visual representation of the source image, the RGB Weighted Clusters technique proved a better solution for color compression in this case.

The CIELAB techniques discussed in this paper tended to accentuate edges in the map features of the chart. From a purely color-compression perspective, this was not a desirable effect. However, color compression is only the first of two steps needed to complete the vector quantization compression to CADRG. The CIELAB techniques may prove beneficial in enhancing edge detail during the second phase: spatial compression.

From a quantitative standpoint, the calculation of average distortion per pixel provided a useful measure in predicting an improvement in color fidelity within RGB color space. Within either color space, the weighted average calculations used to cluster colors would be expected to degrade as the image becomes more complex. Degradation appeared to be more prevalent in CIELAB color space, even though the distortion values were actually lower, on average, for this color space. This

could possibly be due to the nonlinear nature of CIELAB space when attempting to cluster by means of Euclidean distances.

### ACKNOWLEDGEMENTS

This research was funded by Program Element 602435N by the NRL 6.2 Base Program. The authors wish to thank Dr. Fred Petry, Mr. Alfredo Fajardo, and Ms. Maivan Ly for their guidance and support of the investigation of this research and the transition of this research to mission planning applications.

### REFERENCES

- [1] Trenchard, M.E. and Gray, T.D. (2010). **Geotiff to CADRG Conversion Dynamic Link Library Specification – Version 1.1**. NRL/MR/7440—10-9306.
- [2] National Geo-Intelligence Agency. **Military Standard – Compressed ARC Digitized Raster Graphics**, MIL-C-89038. 6 October 1994.
- [3] Lohrenz, M.C. and Gendron, M.L. (2008). “A 3D clustering algorithm to model clutter in electronic geospatial displays”, **Journal of Management and Engineering Integration**, 1(2), 83-88.
- [4] Gaurav Sharma (2003). **Digital Color Imaging Handbook** (1.7.2 ed.). CRC Press. ISBN 084930900X.
- [5] Paschos, George (2001). “Perceptually Uniform Color Spaces for Color Texture Analysis: An Empirical Evaluation”, **IEEE Transactions on Image Processing**, Vol. 10, No.6, pp.932-937.
- [6] Trenchard, M.E., Lohrenz, Tyskiewicz, Rosche (1993). **Color Palette Considerations for Digital Map Displays**. NRL/FR/7441—92-9415.
- [7] Limb, J.O (1979). “Distortion Criteria of the Human Viewer”, **IEEE Transactions on Systems, Man, and Cybernetics**, pp. 778 -793.
- [8] X.Zhang, Setiawan and Wandell (1997). “Image Distortion Maps”, **Final Program and Proceedings of the Fifth IS&T/SID Color Imaging Conference. Color Science Systems and Applications**, pp.120-125, IS&T, SID.

## **An Oil and Gas Regulatory Data Mapping Portal**

Robert Balch  
Petroleum Recovery Research Center  
New Mexico Tech  
801 Leroy Place, Socorro, NM 87801, USA

and

Andrew Parker  
Petroleum Recovery Research Center  
New Mexico Tech  
801 Leroy Place, Socorro, NM 87801, USA

### **Abstract**

The oil and gas regulatory data mapping portal was introduced to New Mexico in 2009. The purpose of the mapping portal was to help oil and gas operators, natural resource environmental managers, and other related sector professionals evaluate natural resources and man-made features within a subject area. While initially designed for the New Mexico Pit Rule, the data provided by the portal has proven useful for other regulatory applications and beyond the borders of New Mexico and the mapping portal has since expanded to include data for surrounding states. The mapping portal includes spatial search capabilities that allow querying data to determine proximity of natural resources to a subject site. Before the inception of the mapping portal, obtaining the required data described in the siting criteria for a number of regulatory forms was not an easy task; the operator would have to invest many hours gathering the data from a variety of online and offline sources. The mapping portal solved this problem by compiling the majority of data required into a Geographic Information System, and locating the data at one internet location accessible via any standard web browser.

**Keywords:** Regulatory Data, GIS, Online Data, Mapping Portal, Drawing Tools.

### **Introduction**

The concept behind the generalized mapping portal is to function as an online geographical information system (GIS). A GIS is a computer mapping application that displays spatial information on a map of the earth rather than tables and spreadsheets; and allows queries and statistical analysis that displays the results on a map. A GIS can plot point and vector data (i.e., oil and gas wells; water wells with chemistry and ground water measurements; air quality monitoring stations; wind

direction) and display the data on aerial photographs, USGS topographic maps, and many other types of imagery such as a water table map, karst topography, geologic map, surface soils, etc. Currently, the generalized mapping portal shows the following national data: USGS topographic maps; USGS aerial photos for 1996-98 and 2008; aerial photography, terrain, and streets from Google Maps; street data from OpenStreetMaps; karst areas; major rivers, streams, and water bodies; elevation, and geology. While USGS ground water data; oil and gas wells; detailed river, stream, and water bodies; and domestic water supply wells are available for New Mexico. The portal is now expanding the detailed surface and groundwater data; oil and gas wells; local geology and soil data into Texas, Oklahoma, Colorado, Utah, Arizona, and Wyoming. After these states are mapped, the detailed data will further expand into the Gulf States, and ultimately the entire U.S.

The portal allows the user to enter a location and perform spatial queries. For example, a user can enter a location and perform a query to determine the distance to the nearest river or stream or to return a list of all water wells within a given distance. Data can be extracted as tables and/or highlighted on maps. A sample question to illustrate how the mapping portal can be used an example is presented: an operator wants to drill an oil well at a specific location. However, the oil and gas regulations do not allow a well to be drilled within a karst area nor within 1,000 feet of a watercourse. Using the mapping portal, the operator can quickly locate this well, zoom to the location, select the appropriate layers to display, and determine if the subject location will pass the regulatory requirements. The map can then be printed and submitted as an attachment to the appropriate forms and the map can be easily verified by a regulator using the same online system. A full example is presented later in the utilization section of this paper.

## Background

The oil and gas regulatory data mapping portal was introduced to New Mexico in 2009. The purpose of the mapping portal was to help oil and gas operators, natural resource environmental managers, and other related sector professionals evaluate natural resources and man-made features within a subject area. While initially designed for the New Mexico, the mapping portal has been expanded to include data for surrounding states. Mapped natural resources include, but are not limited to: depth to water and ground water elevation from USGS gauging stations; topography; aerial photos, digital elevation models; surface geology; karsts; and surface water. In addition, the mapping portal includes spatial search capabilities that allow querying data to determine proximity of natural resources to a subject site. The mapping portal is available at <http://saturn.nmt.edu>.

The beta version of the oil and gas regulatory data mapping portal was designed to help oil and gas operators comply with the New Mexico Pit Rule (Form C-144). Before the inception of the mapping portal, obtaining the required data described in the Siting Criteria (see NMOCD's Form C-144) in Sections 10 and 17 was not an easy task. To determine whether a proposed oil or gas site meets the Siting Criteria, the operator would have to invest many hours gathering the data from a variety of online and offline sources. The mapping portal solved this problem by compiling the majority of required data into one location that is accessible via the internet. Users can access this information using any standard web browser and view the data on a variety of base maps.

## A Generalized Mapping Portal

The purpose of the generalized mapping portal is to expand the beta mapping application to include data beyond the borders of New Mexico. The concept behind the generalized mapping portal is to function as an online geographical information system (GIS). In simple terms, a GIS is: a computer mapping application that displays spatial information on a map of the earth rather than in tables and spreadsheets; and allows queries and statistical analysis that displays the results on a map.

Currently, the generalized mapping portal shows the following national data: USGS topographic maps; USGS aerial photos for 1996-98 and 2008; aerial photography, terrain, and streets from Google Maps; street data from OpenStreetMaps; karst areas; major rivers, streams, and water bodies; elevation, and geology. While USGS ground water data; oil and gas wells; detailed river, stream, and water bodies; and domestic water supply wells are available for New Mexico. The generalized mapping portal will expand the detailed surface and

groundwater data; oil and gas wells; and local geology and soil data into Texas, Oklahoma, Colorado, Utah, Arizona, and Wyoming. After these states are mapped, the detailed data will further expand into the Gulf States.

## Utilization

In addition to viewing the above listed data, the mapping portal allows the user to enter a location and perform spatial queries. For example, a user can enter a location and perform a query to determine the distance to the nearest river or stream. The mapping portal can then return the result in tabular format and/or highlight the selected feature(s) on the map. Or the user can perform a query to return a list of all water wells within a given distance and gives the user an option to highlight these features on the map. Figure 1 shows a top-level map of the GIS showing political boundaries and topography.

To further conceptualize how the mapping portal can be used in the oil and gas industry, the following example is presented: an operator wants to drill an oil well at Latitude: 32.36909 Longitude: -103.88245 in southeastern New Mexico. However, the oil and gas regulations do not allow a well to be drilled within a karst area nor within 1,000 feet of a watercourse. Using the mapping portal, the operator can quickly locate this well (Figure 2), zoom to the location, select the appropriate layers to display (Figure 3), and determine if the subject location will pass the regulatory requirements. As shown in Figure 4, the subject location is near, but not within a karst area (gray shading at left in figure), and is not within 1,000 feet of a watercourse (dotted lines representing intermittent streams). Therefore, location passes the two regulatory requirements presented in the example. The map can then be printed and submitted as an attachment to the appropriate forms and the map can be easily verified by a regulator using the same online system.

The generalized oil and gas mapping portal is available at <http://saturn.nmt.edu>.

The beta version of the mapping portal for the New Mexico Pit Rule is available at <http://pitrule.source3.com>.

## Summary

The intersection of expanding digitized spatial data, widespread availability of data via the internet, and increasingly complex regulatory requirements provided an opportunity to demonstrate the viability of on-line interactive GIS software to speed regulatory processes. While demonstrated with an oil and gas industry example, similar systems could show utility in a variety of industries and regulatory environments.

### Acknowledgements

This work was funded by the Research Partnership to Secure Energy for America (RPSEA) and was performed

with regulatory oversight from the the New Mexico Oil Conservation Division and industry input from the Independent Petroleum Association of New Mexico.

### Figures

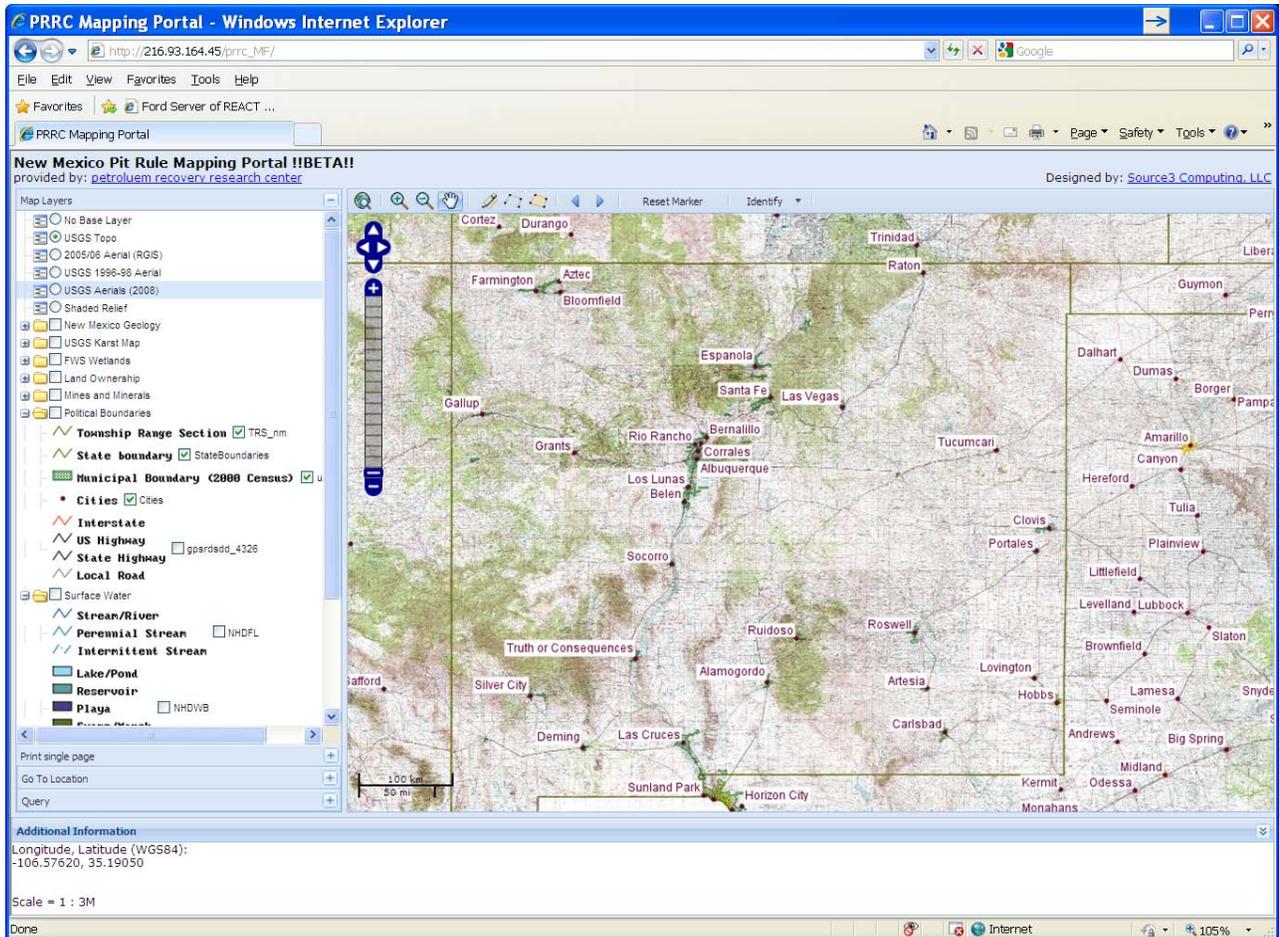


Figure 1. Top-level view of mapping portal web page.

Map Layers		+
Print single page		+
Go To Location		-
T-R-S	<b>Lat-Long</b>	Address
Latitude:	<input type="text" value="32.36909"/>	
Longitude:	<input type="text" value="-103.88245"/>	
Submit		

Figure 2. Zoom to location input box from the mapping portal

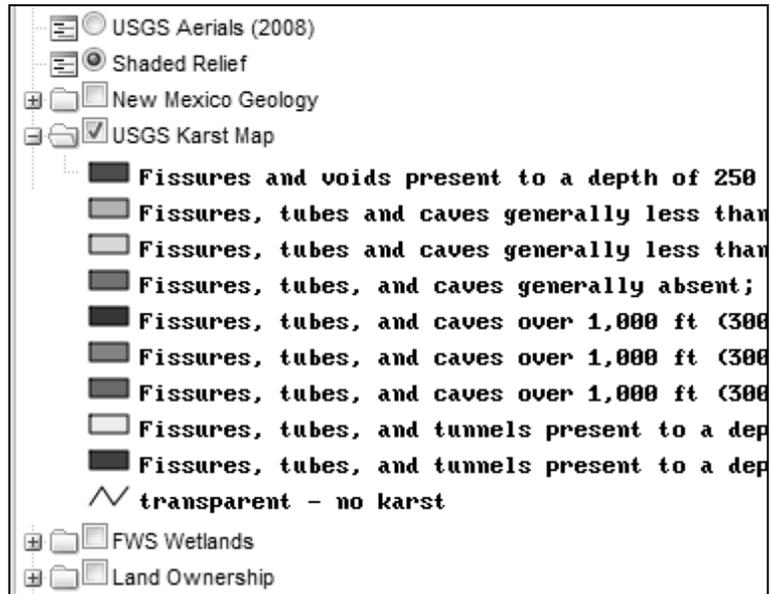


Figure 3. A small sample of available layers in the mapping portal.

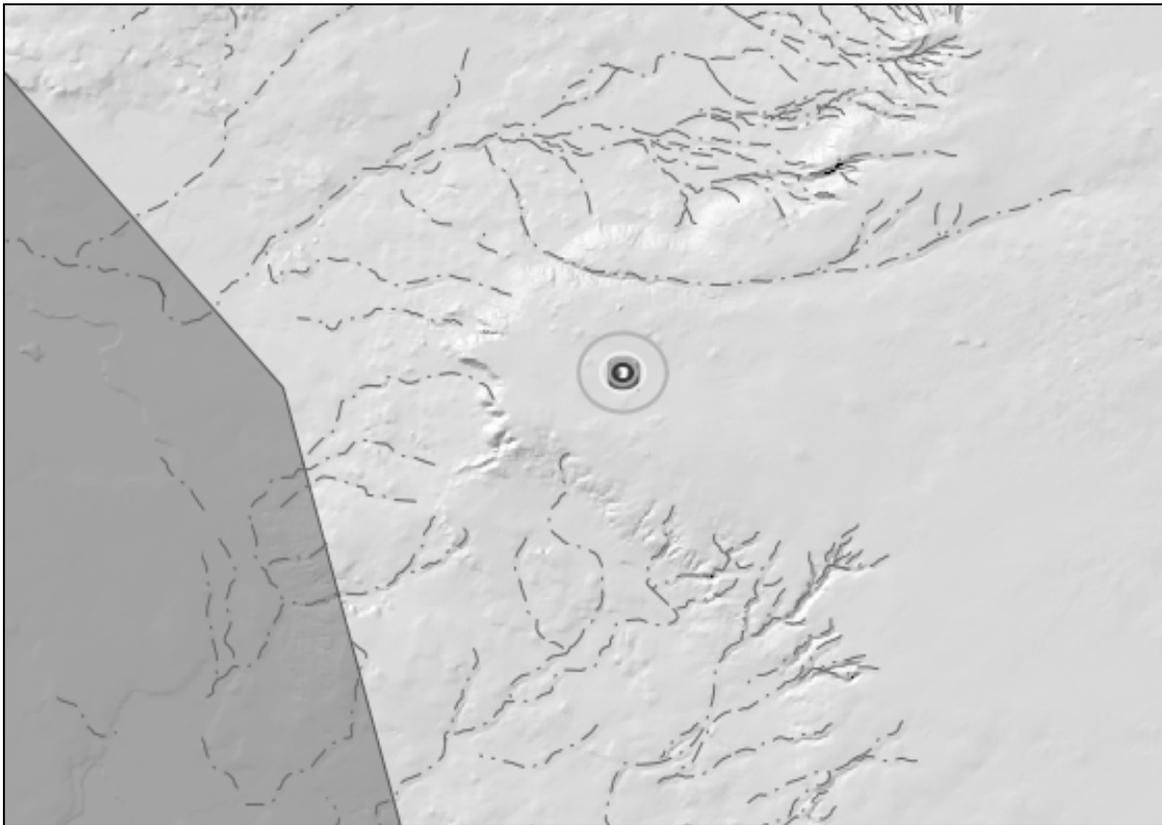


Figure 4 - Proposed oil well location plotted in the mapping portal. The map shows that the subject site is not within 1,000 feet of a watercourse (dashed gray lines) or within a karst area (dark gray shading on left of map) The dotted lines represent intermittent streams. The marker in the middle represents the subject location. The outer ring around the site marker represents a distance of 1000-feet from the subject site.

## **The virtual workplace: A new alternative merging technology and policy for more environmental, economic and social sustainability**

Naglaa FATHY EL DESSOUKY, Ph.D.  
Post-doctoral fellow, EPI, Grenfell Campus,  
Memorial University of Newfoundland  
Corner Brook, NL, Canada  
neldessouky@swgc.mun.ca

### **ABSTRACT**

At present, both developed and developing countries alike are facing challenges of a growing "virtualized" world. Organizations and policymakers have to understand the importance of the Virtual Workplace (VW) and constantly make efforts to "re-think" their strategies and policies to deal with this new reality. There is a growing body of literature focusing on the topic of the Virtual Workplace. However, to date there have been few theoretical and empirical studies related to the application of this topic especially in developing countries. This paper represents a first attempt to explore the challenges related to the implementation of the VW concept in the Middle East and North Africa (MENA) region, mainly Egypt. To investigate these challenges empirically, we examined a number of the well-known Egyptian commercial public sector banks that have already taken serious steps towards an e-banking system. The results of an open-question survey distributed to a random sample of employees at the top, middle and executive levels of management suggest that VW has several potential advantages and drawbacks correlated with the environmental, economic, technological and social context in Egypt. For the analysis of textual data and the administration of the large qualitative data sets, we used ATLAS.ti software. The research highlights the fact that female employees have different perspectives than males with respect to the VW concept. Egyptian women tend to be environmentally and socially oriented, while men give more attention to the economic and

technological aspects of the VW. The paper concludes with some theoretical and empirical proposals for corporate strategies, as well as for public policies to aid the adoption of an innovative VW programs, in particular to satisfy the requirements of environmental, economic and social sustainability in developing countries.

**Keywords:** Virtual Workplace (VW), MENA region (Egypt), Egyptian commercial public sector banks, Systemic framework.

### **1. INTRODUCTION**

The Virtual Workplace (VW) today represents a new reality facing organizations and policymakers in both developed and developing countries. According to Deloitte Touche Tohmatsu (DTT) TMT group (2006), "the number of employees using mobile e-mail access is estimated to increase from the current millions to at least tens of millions by 2010" [1]. Another study by Gartner, Inc. (2007), estimated that, "by 2011, 46.6 million corporate employees globally will spend at least one day a week teleworking, and 112 million will work from home at least one day a month" [2]. Moreover, surveys by various U.S. business groups confirm that "the number of employees working through virtual means will jump up to reach over 100 million workers in the upcoming years". These current trends are driving employees, organizations and governments alike to "re-invent" the major core believes, techniques, places and terms of work.

The VW can be generally described as an embryonic area of academic study, particularly in developing countries. A literature survey reveals the scarcity of the empirical studies dedicated to the practice of the VW especially in the Middle East and North Africa (MENA) region. This article provides a first step to explore the challenges related to the development of the VW concept mainly in Egypt. First, the paper starts by a presentation of the VW concept. It then proposes a systemic framework to study the virtual workplace and it investigates empirically the challenges facing

the implementation of the VW in Egypt by examining a number of the best-known Egyptian commercial public sector banks. The paper ends with some theoretical and empirical proposals for corporate strategies, as well as for public policies to help advance their understanding of the VW concept and practice.

## 2. THE VIRTUAL WORKPLACE

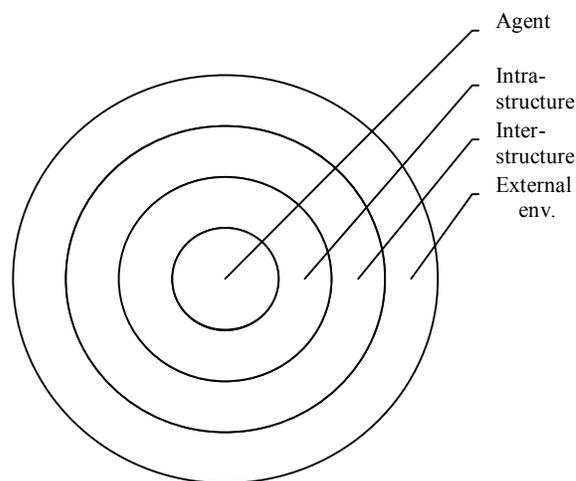
Although research in the Virtual Workplace (VW) is increasing, the concept and its practices remain unclear. A proliferation of definitions has existed in the literature and has caused great confusion. Some of these definitions, which often parallel the purpose of our research, are as follows. In a broad meaning, "the virtual workplace (VW) is a workplace where a group of individuals, from the same country or from all over the world, are working together using different communication technologies regardless of where they are or when precisely they start or finish their tasks or how they get done their everyday job, and that, regardless of their culture, nationality, race, gender, disability or age" (for more details see, among others, [3], [4], [5], [6], [7], [8], [9], [10], [11] and [12]). Other authors put forth the concept of "virtual organizations", a term that, for them, represents an "entities without borders or limits" [13]. This group of authors study "the mutual relationship between the development in information and communication technologies and the organizational structure, behavior, design, functions, practices, culture, values and continuous learning process" (for more details see, among others, [14], [15], [16], [17] and [18]).

## 3. A SYSTEMIC FRAMEWORK FOR VIRTUAL WORKPLACE STUDY

This paper represents a first attempt to explore the challenges related to the implementation of the VW concept in the MENA region, mainly Egypt. A review of literature dealing with the virtual workplace reveals that researchers are specific in their analysis and that, it is difficult

to determine all variables related to the study of this topic (for more details see, among others, [19] and [20]). Most writings examine the VW, mainly at the individuals "micro level" and the intra-organizational context "meso level". Few authors dedicate attention to the inter-organizational context "field level" and/or the external environment "macro level" as crucial variables for studying opportunities and threats related to VW. After a review of literature, we propose the following diagram to present the VW as a continuous process where agent, intra-structure, inter-structure and external environment affect each other in a mutually beneficial manner.

**Diagram (1)**  
Systemic framework to study the virtual workplace (VW)



1. **Agent** denotes the individuals working in a virtual environment. These individuals include employees and managers. This level of "micro" analysis would highlight "agents' core beliefs, interests, culture, skills and performance".
2. **Intra-structure** analyzes the framework that any organization would adopt to facilitate a VW. This level of "meso" analysis puts emphasis on procedural, social and technological processes.

Procedural processes include "laws, rules and norms of the organization". Social processes represent all "ties, networks and trust elaborated between the organization's agents". Technological processes embody "all technological facilities for improving the VW".

3. **Inter-structure** refers to the "field" level of analysis which studies "the relationship between organizations or entities collaborating with each other through virtual means".
4. **External environment** is the "macro" level of analysis that takes into consideration the "environmental, economic, technological and social context, at the country level", which may represent opportunities or threats for the application of the VW.

#### 4. RESEARCH METHODOLOGY

This work represents exploratory research that investigates the challenges related to the implementation of the VW concept in Egypt. To explore these challenges empirically, we examined a number of the well-known Egyptian banks that have already taken serious steps towards an e-banking system. We have limited our research to study only the commercial public sector banks in Egypt. These include the following banks: Banque Misr, National Bank of Egypt, Bank of Alexandria, and Banque Du Caire. An open-question survey has been distributed to a random sample of employees (500 anonymous candidates) at the top, middle and executive levels of management, mainly in the branches existing in Cairo. For the analysis of textual data and the administration of the large qualitative data sets, we used ATLAS.ti software.

#### 5. ANALYSIS AND INTERPRETATION: THE EGYPTIAN CASE

As general interpretation, it is interesting to mention that the research highlights the fact that the perspectives of female employees differ from males with respect to the concept of the VW. Egyptian women tend to be environmentally and socially oriented, while men give more attention to the economic and technological aspects of the VW.

##### 1- The VW environmental concerns:

Unit of Analysis	Advantages	Challenges
<b>Agent</b>	-Less energy consumed for daily transportation -Less CO <sub>2</sub> emissions per capita -Improved health conditions & fewer diseases -Less waste coming from daily routine transactions per person	-Changing core beliefs toward the environmental issues -Convincing people about the connection between the VW & the environment -Resistance of individuals to change
<b>Intra-Structure</b>	-Use of friendly environment technologies -More effective Corporate Social Responsibility (CSR)	-More electricity consumption resulting from working 24/7
<b>Inter-Structure</b>	-Less energy and fewer CO <sub>2</sub> emissions coming from transportation between organizations -Less waste coming from daily routine transactions	-Accommodation of organizations to the VW world

	between organizations	
<b>External Env.</b>	-Better air quality -Adherence to Kyoto Protocol -Improved Waste management systems -More sustainable cities	-Accommodation of people and organizations to the VW/environment relationship - Formulation and implementation of new environmental policies and programs, taking into consideration the VW at the organizational & governmental level

**2- The VW economic and technological concerns:**

Unit of Analysis	Avantages	Challenges
<b>Agent</b>	-Empowerment of women in the marketplace -Empowerment for people with special needs in the marketplace -Individuals can easily combine different jobs at the same time -Individuals are able to search for highest available wages	-No accurate calculation for productivity -No accurate reward system for real hours spent on the job - Special skills and knowledge are required to deal with advanced technology
<b>Intra-Structure</b>	-Increased productivity due to flexible work hours -Decrease of facilities and administration costs -More accessible employees	-Cost increase for capacity buildings and training - Establishing for a new system for performance reward -Increased costs for expanding an advanced internal technological system

<b>Inter-Structure</b>	-Easier and faster inter-organizational relationships, local & worldwide -Easier & more profitable to contract work -Geographical expansion	-Increased costs for elaborating an advanced external technological system to network organizations
<b>External Env.</b>	-More efficient markets & increased productivity -Faster & more flexible responses to consumer needs at the national & international levels	-Dependency of developing countries on developed countries, international organizations and foreign aids for hi-tech industries -Inability of developing countries budget to afford hi-tech -Formulation and implementation of new economic and technological policies and programs, taking into consideration the VW at the organizational & governmental level

**3- The VW social concerns:**

Unit of Analysis	Avantages	Challenges
<b>Agent</b>	-Women balance careers & domestic responsibilities -Time taken for daily transportation is invested in more family & social life -Ability to work anytime at anyplace	-Lack of human contact with colleagues -Employee difficulties regulating work time -Employee self-control, especially in the absence of a supervisor

<p><b>Intra-Struct. &amp; Inter-Struct.</b></p>	<p>-Openness to different cultures at the international level</p>	<p>-Lack of privacy -Elaboration of strong ties, networks and trust in virtual teamwork -Elaboration of specific laws, rules and norms for virtual teamwork -Elaboration of new social and benefits system while taking the VW into consideration</p>
<p><b>External Env.</b></p>	<p>-More social &amp; cohesive society -Less stressful societies -Increased civic contentment</p>	<p>-Formulation and implementation of new social policies and programs, taking into consideration the VW at the organizational &amp; governmental level</p>

## 6. CONCLUSIONS

The existing literature demonstrates the need of a systemic framework to study the VW concept, especially in developing countries, taking into considerations all possible variables related to this subject. To overcome this issue, our research has elaborated a framework for this purpose. It is interesting to mention that most theoretical and empirical research places great emphasis on the agents/individuals "micro level" and the intra-structure variables "meso level" as key units of analysis in their study of the VW. Few works have been done to date to study the inter-structure "field level" and/or external environment variables "macro level" of the VW. This paper represents a first attempt to explore the challenges related to the implementation of the VW concept in the MENA region, mainly Egypt. The research

highlights the fact that female employees have different perspectives than males regarding the VW concept. Egyptian women tend to be environmentally and socially oriented, while men give more attention to the economic and technological aspects of the VW. For the adoption of an innovative VW program in developing countries, organizations and policymakers have to understand the importance of the VW and constantly make efforts to "re-think" their strategies and policies in order to deal with this new reality.

## 7. REFERENCES

- [1] "Eye to the Future - How Technology, Media and Telecommunications Advances Could Change the Way We Live in 2010". **Deloitte Touche Tohmatsu (DTT) TMT group**, 2006.
- [2] "Teleworking, The Quiet Revolution", **Gartner Inc.**, May 14, 2007 (Update).
- [3] Cristina B. Gibson and Jennifer L. Gibbs, "Unpacking the Concept of Virtuality: The Effects of Geographic Dispersion, Electronic Dependence, Dynamic Structure, and National Diversity on Team Innovation", **Administrative Science Quarterly**, Vol. 51, No. 3, Sep. 2006, pp. 451-495.
- [4] Gabriele Piccoli and Blake Ives, "Trust and the Unintended Effects of Behavior Control in Virtual Teams", **MIS Quarterly**, Vol. 27, No. 3, Sep. 2003, pp. 365-395.
- [5] Bradley L. Kirkman, Benson Rosen, Cristina B. Gibson, Paul E. Tesluk and Simon O. McPherson, "Five Challenges to Virtual Team Success: Lessons from Sabre, Inc. ", **The Academy of Management Executive**, Vol. 16, No. 3, Aug. 2002, pp.67-79.
- [6] Kirsten J. Broadfoot, "When the cat's away, do the mice play? Control/Autonomy in the Virtual Workplace", **Management Communication Quarterly: McQ**, Vol. 15, No. 1, August 2001, pp. 110-114.
- [7] J. N. Pelton, **E-Sphere: the rise of the worldwide mind**, Westport, CT: Quorum Books, 2000.
- [8] Wayne F. Cascio, "Managing a virtual workplace", **Academy of Management Executive**, Vol. 14, No. 3, August 2000, pp. 81-90.

- [9] Sujata Gothoskar, "Teleworking and Gender", **Economic and Political Weekly**, Vol. 35, No. 26, Jun. 24-30, 2000, pp. 2293-2298.
- [10] Sirkka L. Jarvenpaa and Dorothy E. Leidner, "Communication and Trust in Global Virtual Teams", **Organization Science**, Vol. 10, No. 6, Special Issue: Communication Processes for Virtual Organizations, Nov.-Dec. 1999, pp. 791-815.
- [11] N. F. Crandall and M. J. Wallace, **Work and reward in the virtual workplace: a new deal for organizations and employees**. New York: American Management Association, 1998.
- [12] Anthony M. Townsend, Samuel M. DeMarie and Anthony R. Hendrickson, "Virtual Teams: Technology and the Workplace of the Future", **The Academy of Management Executive**, Vol. 2, No. 3, August 1998, pp.17-29.
- [13] J. R. Galbraith, **Designing Organizations**. San Francisco, CA: Jossey-Bass, 1995.
- [14] Gerardine DeSanctis and Peter Monge, "Introduction to the Special Issue: Communication Processes for Virtual Organizations", **Organization Science**, Vol. 10, No. 6, 1999, pp. 693-703.
- [15] Nancy B. Kurland and Terri D. Egan, "Telecommuting: Justice and Control in the Virtual Organization", **Organization Sciences**, Vol. 10, No. 4, July-August 1999, pp. 500-513.
- [16] B. Hedberg, G. Dahlgren, J. Hansson and N.-G. Olve, **Virtual Organizations and Beyond: Discover Imaginary Systems**, New York: Wiley, 1997.
- [17] S. E. Bleecker, "The virtual organization", **The Futurist**, Vol. 28, No. 2, March-April 1994, pp. 9-14.
- [18] Ann Majchrzak, Ronald E. Rice, Arvind Malhotra, Nelson King and Sulin Ba, "Technology Adaptation: The Case of a Computer-Supported Inter-Organizational Virtual Team", **MIS Quarterly**, Vol. 24, No. 4, Dec. 2000, pp. 569-600.
- [19] Alain Verbeke, Robert Schulz, Nathan Greidanus and Laura Hambley, **Growing the virtual workplace: the integrative value proposition for telework**. Edward Elgar Publishing, 2008.
- [20] Anne Powel, Gabriele Piccoli and Blake Ives, "Virtual teams: a review of current literature and directions for future research", **The DATA BASE for Advances in Information Systems**, Vol. 35, No. 1, 2004, pp. 6-36.

# **Technology platforms in distance learning: Functions, characteristics and selection criteria for use in higher education**

Rocío FERNÁNDEZ  
Universidad Católica de Valencia  
Valencia, Spain

Ignacio GIL  
Department of Management. Universidad Politécnica de Valencia  
Valencia, Spain

Daniel PALACIOS  
Department of Management. Universidad Politécnica de Valencia  
Castellón, Spain

and

Carlos DEVECE  
Department of Management. Universidad Politécnica de Valencia  
Valencia, Spain

## **ABSTRACT**

In the space of only a few years, e-learning platforms have managed to achieve universal presence in higher education centres. As well as changing strictly distance learning, e-learning has begun to play a major role in classroom learning, giving rise to new learning typologies such as the partial-classroom learning “Blended Learning”. Technological advances have, without a doubt, triggered this change and although research into methodologies and learning in this new environment still has a long way to go, e-learning platforms have reached a sufficient degree of maturity and standardization regarding their functions and tools. This article is dedicated to analysing e-learning platforms as a technological element necessary in any Virtual Training Environment (VTE) system. In it the following are studied: different platform types and their main tools, minimum requirements an institution needs to take into account when selecting a particular platform, as well as its advantages and drawbacks. Finally, the article carries out a brief review of some of the existing and better-known platforms within educational environments.

**Keywords:** e-learning, technology platforms, higher education, virtual training environment, integrated learning platforms

## **1. INTRODUCTION**

Virtual Training Environments are present in all Universities. Not only have they revolutionised distance learning, yet they already make up a substantial part of classroom learning. E-learning platforms are the technical element that has facilitated this development. These platforms are the main element of any

VTE system. They include tools to facilitate learning, communication and collaboration; course management tools; and tools for designing the user’s interface [1]. Even though creating materials for the Web is relatively easier, more flexible and less expensive, developing programmes for other environments, elaborating and managing online didactic courses or modules continues to be an activity that requires a great deal of effort. From the mid-nineties onwards, integrated platforms started to spread with the objective of facilitating the online courses creation and reducing costs [2]. They are also known as Virtual Platforms or Integrated Learning Environments [3]. The use of an Integrated Virtual Platform allows complete courses for the Web to be created and managed without the need of in-depth knowledge about programming or graphic design. They are also known as Distance-Learning Telematic Systems or Learning Management Platforms (Learning Management Systems-LMS). In each case they are integrated tools which are used for creating, managing and distributing formative activities via the Web ie: they are applications which facilitate the creation of teaching-learning environments, integrating didactic materials and communication, collaboration and educational management tools.

According to Bri et al. [4] e-learning platforms can be defined as “the hardware and software environment designed to automate and manage development academic formation activities. The software must administer, distribute and check activities for face-to-face formation or e-learning in an organization. The main functions are: manage and register users, resources and formation activities, access check, control and monitoring learning process, doing evaluation, informs and managing communication services.” It is a technological response that integrates different management tools, facilitating learning

development, distributed using information of a very diverse nature.

Different conceptualizations as to what e-learning platforms or learning management systems (LMS) are could be formulated emphasizing theoretical, technical or formal aspects to a greater or lesser extent. However, by rejecting this type of approach and accepting the common element in technical means, an e-learning platform, or a learning management system in the network, is a computing and telematic tool, organized in accordance with some formative objectives of integral form and of some psychopedagogical and organizational intervention principles. In short, eLearning platforms are the server software which is mainly in charge of user management, management of courses and communication services, providing the eLearning scene with its necessary support [5][6].

This definition of e-learning platform needs to be widely developed regarding the multiplex components and characteristics that can be brought together. Whilst still evolving, e-learning platforms have reached a certain standardization of functions and tools. The objective of this article is to answer questions as to what basic characteristics (both technical and functional) e-learning platforms should have and what the criteria for selecting a e-learning platform are, in particular in the case of university and higher education. The article's structure follows the order of these questions. In the first place, and in general, the basic characteristics of e-learning platforms are considered, in order to then revise their historical evolution. Afterwards, the tools which they should rely on, are listed. Finally, having analysed advantages and drawbacks, selection criteria are established according to higher education needs and final conclusions are discussed.

## 2. BASIC CHARACTERISTICS OF E-LEARNING PLATFORMS

There are four basic and essential characteristics that any platform should have [5][6]. The first is Interactivity: ensuring that the person using the platform is conscious of playing the leading role in his or her training. The second is Flexibility, defined as the set of functionalities that allow the eLearning system to be easily adapted into the organization wishing to implement it. This adaptation must address different levels, such as adaptability to the institution's structure. It must also allow for adaptation to the study plans of the institution wishing to implement the system and also to the content and pedagogical styles of the organization. The third characteristic is scalability, allowing the platform to function equally with a small or large number of users. The last is standardization or the capacity to use courses carried out by a third party. In this way courses are available for the organization that has created them and for others that comply with the standard. The durability of courses is also guaranteed, avoiding them becoming obsolete and last of all a follow-up of students' behaviour throughout the course can be carried out.

An essential part of e-learning platforms is that they must make remote access possible for both teachers and students in any given moment and from any place with connection to Internet or networks with TCP/IP protocol. For this, a browser is used, allowing users to access information via standard browsers and using http communication protocol.

## 3. EVOLUTION OF E-LEARNING PLATFORMS

The origin of Web-based eLearning platforms, is due to a specialization of content management systems (CMS), in systems oriented to content management for distance learning [7]. Content management systems (CMS), are software which is mainly used in order to facilitate web management, whether that be Internet or Intranet. However in the case of Distance Learning or eLearning, the application of these CMS systems is not focused on the Web, but on educational content (resources, documents, evaluation tests...). In these cases there are some specific needs that a general CMS does not always cover, or if it does, it does not provide the same facilities as a tool created for this function.

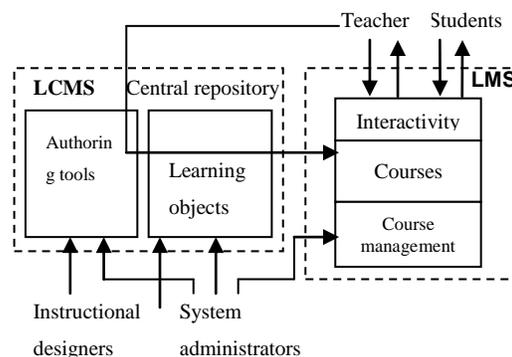
In recent years, CMSs have progressed in evolutionary stages.

- First Stage: CMS - Content Management Systems are in the most basic e-learning platforms and allow dynamic websites to be generated. The objective of these programmes is the creation and management of online information (texts, images, graphics, videos, sounds, etc.)
- Second Stage: LMS - Learning Management Systems come from CMSs and provide an environment which makes Web updating, maintenance and amplification possible with the collaboration of multiple users.
- Third Stage: LCMS - Learning Content Management Systems are platforms that integrate CMS and LMS functionalities, incorporate content management in order to personalise resources for each student and where companies become their own editing entity, with self-sufficiency in content publication, in a simple, quick and efficient manner, solving problems from previous platforms.

LCMSs add knowledge management techniques to the LMS model in structured and designed environments so that organizations can implement their processes and practices better, with the support of online courses, materials and content.

Finally, a LCMS is normally linked to a LMS (platform) or can integrate its applications, as can be seen in the following schematic diagram (Figure 1):

Figure 1: Schematic diagram of LMS and LCMS integration (adapted from [8])



## 4. E-LEARNING PLATFORM TOOLS

A good way to conceptualize e-learning platforms is to pay attention to the functions they carry out. Current e-learning platforms offer many functionalities via their numerous tools,

which can be grouped in the following manner: (1) Learning-oriented tools, (2) Productivity-oriented tools, (3) Tools for involving students, (4) Support tools, (5) Tools dedicated to publishing courses and content, (6) Knowledge management systems in the area of education [9].

Within learning-oriented tools the following stand out (1) discussion forums, tools that allow messages to be exchanged during the time the course lasts (or that the course-trainer estimates). Forums can be organised chronologically, by categories or by conversation topics (threads) and can allow or not for files to be attached to the message. Besides forums the following must exist: (2) synchronous communication tools for exchanging instantaneous messages among participants (Chat, Videoconference), and (3) asynchronous communication tools for exchanging mails and messages both on the course as well as using external mail addresses (Electronic mails, messaging) [10]. Another similar modality would be (4) blogs, a tool that enables students and teachers to make notes in a diary (edublogs), whether they be subject blogs or personal blogs of the students to whom they provide support and follow-up. We would also include in this section (5) Wikis, tools which facilitate the development of online documents in a collaborative manner. Within learning-oriented tools we would also find (6) E-portfolios or digital portfolios, tools which allow students to receive educational follow-up, having access to work done in their formative activities.

In the category of productivity-oriented tools, all those which provide the student with follow-up of activities and work would be included. The following stand out (1) bookmarks, which enable the student to return easily to the website previously visited whether it be that of the course or not and (2) calendar and review of progress which makes time planning easier, as regards course activities. The following would also be included (3) helping to use the platform for orientation and helping participants to use the learning system (tutorials, manuals, online help...), (4) course search-engine to facilitate course selection and location indicating a search pattern, and (5) synchronization and off-line work mechanisms, where students have the possibility of working disconnected from the platform, once the course or part of it has been uploaded onto their computer. The following could also be included (6) control of publication, expired websites and broken links to allow web-sites to be published up to a specific date and not to leave them accessible once the publication period has expired; they also carry out checks to locate and correct the existence of links to inexistent pages.

As concerns tools for involving students, typical utilities would be; (1) Work groups to organise a class in groups, so as to provide an area for each and everyone of them, where the teacher assigns tasks or corresponding projects, (2) self-assessment tools to practise or review online tests and be aware of their evaluations. These are normally not counted by the teacher, (3) lastly the following can be introduced into the system: a student corner or areas reserved for students (to display jobs and advertisements, see photographs, personal information, etc.)

In support tools, tools for user authentication would be included: each with a user name and password to access the platform. This allows privileges to be allocated according to role. Each user type has different privileges (students, teacher, assistant, tutor, administrator, etc.), which facilitates audits described at the desired level of detail, allowing for all actions carried out by platform participants to be consulted as well as obtaining

statistics about use. Student register tools are usually also incorporated in such a way that student course enrolment can be carried out in different ways (self-enrolment, manual, database...).

Within tools dedicated at publishing courses and content, the following would be included: (1) Test and automated results, which allow teachers to create, administer and evaluate finished tests. If the teachers wishes so, they can be self-corrected, and solutions, comments and explanations can be shown. (2) Course administration: these tools allow teachers to control class progression via course material. (3) Support to course creator to help and support them in administration of the following (forums, online help, electronic mail...). And finally, (4) student follow-up tools to provide additional analysis about the use made of course materials.

Tools for designing study plans would be made up of: (1) Conformity with accessibility, to meet standards of access to online information for the disabled, (2) re-use and content sharing, given that content created on a course can be shared by other teachers, on another course and in another centre, (3) standard course templates to create the structure of an online course, (4) vitae administration, allowing students to be provided with a personalised CV, based on the prerequisites of the educational programme or activities, in previous work or test results, (5) environment personalization: this allows graphic appearance and the way courses are seen to be changed. In this way the institution's image of these can be given. (6) Tools for education design help course creators to be able to do sequences of learning, templates or set-up assistance, and finally, (7) they ensure conformity with education design or standards (IMS, AICC, ADL) for sharing learning materials with other eLearning platforms.

Finally, in knowledge management systems in the education area [11], the following would be included: (1) integral knowledge systems to provide both mechanisms of collaborative work, as well as internal organization of common knowledge memory. Normally they represent the structure of knowledge in a hierarchical form, the tree of knowledge. It can also be represented by way of hierarchical networks of nodes interconnected by relationships. The function of facilitating interaction and collaboration among users is normally carried out via forums or discussion groups, (2) Mediating information systems, whose main objective is to provide users with an interface in order to search for information (generally using the web) about a specific domain, where knowledge sources are heterogeneous and distributed, although it seems as if a centralised and homogeneous system were being used, (3) digital bookshops or repositories, assembly of communication technologies and digital storage of information to reproduce, emulate and extend the service provided by conventional bookshops. They can collect information from distributed sources and the user is allowed to build their own digital library, (4) systems based on ontology or systems where the domain structure is known a priori. For this reason they support automatic knowledge search systems and facilitate decision-making, applying an inference engine to ontologically structured databases.

### **Advantages and drawbacks of e-learning platforms**

It is evident that e-learning platforms, understood as technology (digital platforms included), are only a means and not an end in the education process (be it virtual or otherwise). The added

value is fundamentally teacher-student interaction. In spite of all the advances in technology, a personal relationship continues to be the basis and the most efficient method. A teacher is not a “talking head” that can be substituted by a computer. Many universities have the necessary investments, sufficient materials and institutional support but the most important factor for these virtual models to work is to change relationships between teachers and students so that they do not only meet in classrooms but also in the area reserved in these Virtual Training Environment systems.

Let us go over the advantages and limitations found that e-learning platforms offer the two most important variables that intervene in the virtual education process: teachers and students.

The **advantages** would be [12][13][14]:

- They offer flexible and open learning, without limitations of space or time. Therefore rigid timetables and travel problems are avoided as well as improving access to information.
- They enable the teacher to incorporate new learning resources, to elaborate study materials, exams, to automate marking systems, to elaborate class assistance resources....., with greater ease and quality of presentation.
- Information can be digitalised and managed with relative ease and at reduced costs (cd-rom, dvd, external USB Memory, etc.).
- Teacher-student and student-student communication can improve with incorporated communication tools.
- Incorporating tools for student follow-up and evaluation.
- Constant updating of information is possible from teachers as well as students.
- Interaction with experts on any area is possible, without it mattering where they are (for example by means of videoconference).
- They promote collaborative work, in teaching as well as research. Work groups can be created with teams of people from different Universities, institutions or companies, who can be in different cities or countries.
- They allow research results to be published worldwide. Quicker knowledge circulation.

In accordance with the aforementioned authors, we can summarise the **limitations** found in virtual platforms for teachers and students, as follows [14][15][16] [17]:

- Lack of preparation on behalf of teachers when using these platforms. Mention is made to teacher “technophobia” when faced with new technology. They need to make an effort to learn, given that the same structures from classroom courses cannot be transferred. They can also generate fears about new expectations and possibilities of integration in teaching.
- They generate exclusion in those that have no access or possibilities to access them (the digital divide increases).
- Teachers need to carry out the learning process with greater previous detail than with traditional methods.
- A greater time investment is needed (especially at the beginning) when preparing a virtual course.
- Problems that a student might come across must be anticipated. Elements that help him or her to overcome them must be available.

- “Face to face” contact is lost. Technological relationships are “colder” meaning that personal social relationships are lost to a small extent.
- It can be difficult for teachers to motivate students to participate and to avoid students dropping out.
- Students must be mature, constant and responsible. They may prefer traditional methods in which they can play a more passive role.

From what we have analysed, e-learning platforms constitute the technical base on which a Training Programme’s design and implementation is maintained and developed. Therefore it constitutes the main issue of VTE systems.

These solutions provide students and Companies with greater accessibility as regards carrying out courses over the Internet, and in the same way, they facilitate programme development by the teachers. These tools are found in an increasingly wider market. Most have a commercial licence but free software ones that can be downloaded via the Internet also exist.

## 5. REQUIREMENTS WHEN SELECTING A E-LEARNING PLATFORM

We need to bear in mind that the conditions determining the adoption of a specific Platform depend on the characteristics of the educational environment in which it will be used. Web-based formative actions need to be developed in accordance with pursued objectives and need to work according to the expectations for which they are chosen. Choosing a e-Learning Platform will depend on many aspects [18]: number of users, characteristics of said users and their needs, types of courses to be given and their objectives, content to be taught and methodology to be used (more or less interactive) and type of evaluation to be used. The importance of each variable leads to choosing the most suitable Platform and at the same time concentrating on or emphasizing those aspects which most concern the Organization or University.

Currently Virtual Platforms are being used to develop and teach four online course modalities, which are differentiated in their extent of Web use [19]. We can find totally developed courses, where the Web is used as the only means of presenting content, interaction and student evaluation. The second modality is the dependent course, where fundamental parts of the course are only available on the Web. The third modality is the supplementary course that uses the web as a means of circulating materials. Lastly, informative courses only use the web to present information about the course.

Next we present some requirements that need to be taken into account when selecting a specific e-Learning Platform (table 1).

## 6. CONCLUSION

Distance-learning systems have taken over from previous distance-training systems, achieving substantial improvements thanks to the advantage provided by Internet from powerful communication and informative channels: a great deal of information which is constantly growing, can be very easily stored and is accessible with unlimited communication possibilities.

Table 1: Requirements when selecting a eLearning Platform [20]

<p><b>1 USERS</b></p> <ul style="list-style-type: none"> <li>▶ Number</li> <li>▶ Age</li> <li>▶ Socio-cultural characteristics</li> <li>▶ Skills or Internet use</li> <li>▶ Educational needs</li> </ul>
<p><b>2 COURSE</b></p> <ul style="list-style-type: none"> <li>▶ Type</li> <li>▶ Number of students</li> <li>▶ Course objectives</li> <li>▶ Course content</li> <li>▶ Course methodology</li> <li>▶ Course evaluation</li> </ul>
<p><b>3 TECHNICAL CHARACTERISTICS</b></p> <ul style="list-style-type: none"> <li>▶ Server Hardware or Software needs</li> <li>▶ Hardware or Software needs for user</li> <li>▶ Number of online users possible</li> <li>▶ Security and access control</li> <li>▶ Technological infrastructure needed by Platform</li> <li>▶ Necessary technical knowledge</li> <li>▶ Surfing via the platform</li> <li>▶ Design strategies</li> <li>▶ User licence contracting methods</li> </ul>
<p><b>4 MANAGEMENT OF FORMATIVE OFFER</b></p> <ul style="list-style-type: none"> <li>▶ Management of courses on offer</li> <li>▶ Management of student enrolment and cancellations</li> </ul>
<p><b>5 TEACHING/LEARNING PROCESS</b></p> <ul style="list-style-type: none"> <li>▶ Teacher learning method</li> <li>▶ Student learning method</li> <li>▶ Type of content to be taught:                             <ul style="list-style-type: none"> <li>• Index of contents</li> <li>• Glossary of terms</li> <li>• Content search tool</li> <li>• Content available in various formats</li> <li>• Multimedia possibilities</li> <li>• Possibility of consulting content <i>off line</i></li> </ul> </li> <li>▶ Evaluation of Teaching/Learning process                             <ul style="list-style-type: none"> <li>• Initial evaluation</li> <li>• Formative or continuous evaluation: tools for monitoring student activities, observation and follow-up tools, student self-evaluation tools.</li> <li>• Final or summative evaluation: tests- multiple choice, connection, fill the gaps, short and open answer, etc.</li> </ul> </li> <li>▶ Student help system</li> <li>▶ Course design tools for teachers</li> <li>▶ Course management tools for teachers</li> </ul>
<p><b>Communication tools</b></p> <ul style="list-style-type: none"> <li>▶ Must facilitate communication and collaborative work between teachers and students.</li> <li>▶ Asynchronous: Discussion forums, Distribution lists, Electronic mail, Tutorials, Notice boards, Calendar</li> <li>▶ Synchronous: Shared electronic board, Chat, Audio Conference, Videoconference</li> </ul>

Although we always need to bear in mind that e-learning goes far beyond its technological component and that we need to analyse and evaluate those learning, understand their demands, preferences and actions to understand electronic learning in its

entire dimension [16]. The choice of an appropriate e-learning platform is an essential previous condition.

A e-learning Platform needs to resolve questions of heterogeneity in the wide sense of the word (Alkouz and El-Seoud 2007), ie: heterogeneity of the channel's broadband in Internet/Intranet environments.

When the design of an EVE/A (Teaching/Learning Virtual Environment) or an e-Learning Platform is approached, at least two different levels must be taken into account [21]: (1) User interface, considered independently at both a hardware and software level, taking into account the fact that users will basically be of three types; Teachers - Students - System administrators. (2) Teaching/Learning Module: All services required for optimum development of teaching-learning processes need to be able to be implemented in this environment.

The possibilities and advantages that e-learning platforms offer e-learning are numerous. However, new difficulties and limitations need to be overcome in the learning process [22]. And merely incorporating ICTs into training does not guarantee effectiveness in results [23]. Given the selection of interactive means and resources and their incorporation in a global design of distance-learning environment, they must be supported on the basis of a learning theory that both justifies and defines them. Already in 1997, reflecting on its experience of progressively incorporating Internet functionalities into a course at the University of Twente (Holland), Collis [24], concludes by saying that "technology is not important per se, but as a tool to specify an educational philosophy". In this sense, the innovations that have been produced in the field of distance learning, do not consist so much of the presence of a new technology, rather than that of the review of the formative processes based on the latter.

## 7. REFERENCES

- [1] M. Damian, "Management and software solutions for distance learning", **Management of technological changes**, Vol. 1, 2009, pp. 477-480.
- [2] D. Dagger., A. O'Connor, S. Lawless, E. Walsh., V.P. Wade, "Service-Oriented E-Learning Platforms: From Monolithic Systems to Flexible Services", **Internet Computing, IEEE**, Vol. 1, No 3, 2007, pp. 28-35.
- [3] T. Okamoto, A. Cristea, M. Kayama, "Future integrated learning environments with multimedia", **Journal of Computer Assisted Learning**, Vol. 17, No 1, 2001, pp. 4-12.
- [4] D. Bri, M. Garcia, H. Coll, J. Lloret, "A study of virtual learning environments. WSEAS Transactions on Advances in Vol. 6, No 1, 2009, pp. 33-43.
- [5] A.W. Bates, **Technology, E-Learning And Distance Education**, London, UK, Routledge Studies in Distance Education, 2005.
- [6] S. Graf, B. List, "An evaluation of open source e-learning platforms stressing adaptation issues", **Fifth IEEE**

**International Conference on Advanced Learning Technologies**, pp. 163-165, 2005.

[http://www.cios.org/getfile\Chadwick\\_v9n199](http://www.cios.org/getfile\Chadwick_v9n199). Accessed 15 March 2008

- [7] A. Wahlstedt, S. Pekkola, M. Niemel, "From e-learning space to e-learning place", **British Journal of Educational Technology**, Vol. 39, No 6, 2008, pp. 1020-1030.
- [8] J. Laviña, L. Mengual **Libro Blanco Universidad Digital 2010**. Fundación Telefónica. Ariel, 2008.
- [9] B. Allan, **Blended learning tools for teaching and training**, London, UK, Facet Publishing, , 2007.
- [10] M.D. Roblyer, J. Freeman, M.B. Donaldson, M. Maddox, "A comparison of outcomes of virtual school courses offered in synchronous and asynchronous formats", **The Internet and Higher Education**, Vol. 10, No 4, 2007, pp. 261-268.
- [11] M Raman, T Ryan, L Olfman, "Designing Knowledge Management Systems for Teaching and Learning with Wiki Technology", **Journal of Information Systems Education** Vol. 16, No 3, 2005, pp. 311-20.
- [12] College Board, **Trends in College Pricing**, Washington, D.C., 2001.
- [13] J.F. Young, "Hybrid teaching seeks to end the divide between traditional and online instruction" **The Chronicle of Higher Education**, 2002.  
<http://chronicle.com/free/v48/i28/28a03301.htm>. Accessed 12 February 2009.
- [14] K.J. Kim, S. Liu, C.J. Bonk, "Online MBA students' perceptions of online learning: Benefits, challenges, and suggestions", **The Internet and Higher Education**, Vol. 8, No 4, pp. 335-344, 2005.
- [15] F. Colace, De M. Santo, M. Vento, "Evaluating on-line learning platforms: a case study System Sciences", **Proceedings of the 36th Annual Hawaii International Conference on System Sciences**, 2003.
- [16] S. Naidu (Ed.) **Preface. Learning and teaching with technology**. London, UK Principles and practices. Kogan, 2003.
- [17] P.C. Sun, R.J. Tsai, G. Finger, Y.Y. Chen, D. Yeh, "What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction", **Computers & Education**, Vol. 50(4): pp. 1183-1202, 2008.
- [18] C.W. Holsapple, A. Lee-Post, "Defining, Assessing, and Promoting E-Learning Success: An Information Systems Perspective", **Decision Sciences Journal of Innovative Education**, Vol. 4, No 1,: pp. 67-85, 2006.
- [19] S.A. Chadwick, "Teaching virtually via the web: Comparing student performance and attitudes about communication in lecture, virtual Web-based, and Web-supplemented courses", **The electronic Journal of Communication**, Vol. 9, No 1, 1999.
- [20] B. Collis, **Tele-learning in a digital world. The future of distance learning**. London, International Thomson Computer Press., 1997.  
<http://www.ub.es/forum/Conferencias/betty.htm>. Accessed 15 March 2005
- [21] S.J. Cho, S. Lee, "A module-based software framework for e-learning over internet environment", **Lecture Notes in Computer Science**, 3220, pp. 803-805, 2004.
- [22] A. Romiszowski, "How's the e-learning baby? Factors leading to success or failure of an educational technology innovation", **Educational Technology**, Vol. 44, No 1, 2004, pp. 5-27.
- [23] M. Bell, S. Farrier, "Measuring success in e-Learning - a multi-dimensional approach", **6th European Conference on e-Learning**, ECEL 2007.
- [24] B. Collis, **Tele-learning in a digital world. The future of distance learning**. London, International Thomson Computer Press, 1997.  
<http://www.ub.es/forum/Conferencias/betty.htm>. Accessed 15 March 2005

# Introduction to The Border Gateway Protocol – Case Study using GNS3

Sreenivasan Narasimhan<sup>1</sup>, Haniph Latchman<sup>2</sup>

Department of Electrical and Computer Engineering

University of Florida, Gainesville, USA

n.sreeni86@ufl.edu<sup>1</sup>, latchman@ufl.edu<sup>2</sup>

**Abstract** – As the internet evolves to become a vital resource for many organizations, configuring The Border Gateway protocol (BGP) as an exterior gateway protocol in order to connect to the Internet Service Providers (ISP) is crucial. The BGP system exchanges network reachability information with other BGP peers from which Autonomous System-level policy decisions can be made. Hence, BGP can also be described as Inter-Domain Routing (Inter-Autonomous System) Protocol. It guarantees loop-free exchange of information between BGP peers. Enterprises need to connect to two or more ISPs in order to provide redundancy as well as to improve efficiency. This is called Multihoming and is an important feature provided by BGP. In this way, organizations do not have to be constrained by the routing policy decisions of a particular ISP. BGP, unlike many of the other routing protocols is not used to learn about routes but to provide greater flow control between competitive Autonomous Systems. In this paper, we present a study on BGP, use a network simulator to configure BGP and implement its route-manipulation techniques.

**Index Terms** – Border Gateway Protocol (BGP), Internet Service Provider (ISP), Autonomous System, Multihoming, GNS3.

## 1. INTRODUCTION

Routing protocols are broadly classified into two types – Link State routing (LSR) protocol and Distance Vector (DV) routing protocol.. In Distance vector routing protocol, each node shares its routing table with the neighbors periodically in contrast to Link State where updates are event-triggered. Examples of Link State would be Open Shortest Path First (OSPF) and Intermediate System-to-Intermediate System (ISIS) while Routing Information Protocol (RIP v1 and 2) are examples of Distance Vector. BGP is called a Policy Based routing protocol because the route-selection is done based on routing policies of an Autonomous System(AS).

When BGP is running between routers belonging to different Autonomous Systems, it is called EBGp while BGP running between routers belonging to the same Autonomous system is called IBGP. Administrative Distance (AD value) is the first criterion that a router uses to determine which routing protocol to use if two protocols provide route information for the same destination [5]. BGP has an AD (Administrative Distance) value of 200 for IBGP (Inter-domain routes learnt by BGP) routes and 20 for EBGp (Exterior routes learnt by BGP) routes. BGP allows path- manipulations to be done by the AS. However, It is expected that the System Administrator has a clear understanding of its working.

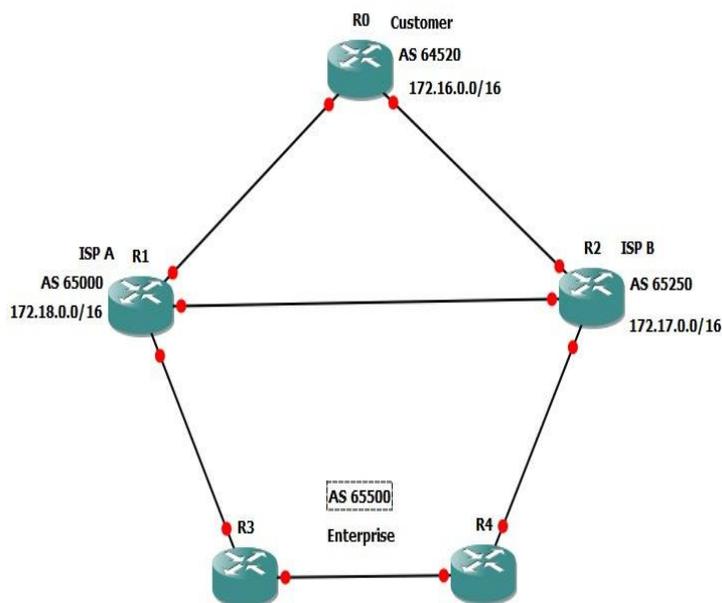


Figure 1. Internet using BGP [2].

In the figure, AS 65500 learns about the route 172.18.0.0/16 through ISP A. Suppose that route is announced to ISP B. ISP B may decide that the best path to 172.18.0.0/16 in ISP A is through AS 65500. Thus AS 65500 becomes a transit AS between 172.18.0.0/16 in ISP A and ISP B. This may not be acceptable for AS 65500 which is connected to both ISP A and ISP B in order to provide redundant connection to the Internet rather than to be a transit AS. BGP's policy based routing helps mitigate such problems.

Version 4 of BGP has been deployed in the Internet since 1994. BGPv4 supports Classless Inter-Domain Routing (CIDR) and Variable Length Subnet Masking (VLSM). CIDR incorporates VLSM techniques and aggregation wherever necessary so that the number of routes in the global routing table does not increase exponentially. With the growing use of the Internet, the routing table of a core router of a major ISP, without CIDR, would typically contain more than 2000000 entries. However, by using CIDR the BGP routes in the routing table would be reduced to just about 170000 routes, thereby reducing the memory and CPU power wasted on lookup.

## 2. BGP AND OTHER IGP PROTOCOLS

Firstly, IGPs are routing protocols which are used to route packets within an autonomous system. On the other hand, BGP is an example of Exterior Gateway Protocol which is used to route packets between autonomous systems. IGPs decide the best path based on a certain predefined metric. For e.g. RIP uses hop count (number of layer 3 devices to be passed). EIGRP uses composite metric. BGP is a policy based routing protocol used for traffic flow control between autonomous systems. Unlike IGPs, it has multiple metrics - which are called as "attributes" using which it decides the route to the destination.

BGP is a complicated routing protocol which should not be used unless one has a complete understanding of route filtering and BGP's path Selection process. It is not advisable to configure BGP on routers having low memory or when there is low bandwidth link between autonomous systems. An alternative is to use static (routes entered manually by the network administrator) or default routes.

BGP provides multi-homing options, that is, BGP can be configured when the flow of traffic entering and leaving the AS has to be manipulated. Multi-rom is another scenario in which BGP is used.

## 3. GRAPHIC NETWORK SIMULATOR 3 AND THE CISCO IOS

For implementing BGP, we use a software called as GNS3 which is a graphical network simulator that allows simulation of complex networks. It is an open source, free program [6].

Cisco IOS (originally Internetwork Operating System) is the software used on the vast majority of CISCO routers and switches. CISCO IOS uses a command-line interface. According to the privilege level of the user, CISCO IOS allows only a set of commands to be used in each mode to ensure security and efficient operation. That is, the mode you are in determines the command you can use. Upon first connecting to the router, you are by default in unprivileged mode (characterized by ">" sign). You could then log on to enter the privileged mode (characterized by "#" sign) if a password has been set. The privileged mode is parent to many sub modes like Global configuration mode (characterized by "#(config)" sign) etc which is used to configure all features [4].

It is assumed that the reader has a basic understanding of networking concepts like routers, Autonomous Systems, VLSM etc.

## 4. NEIGHBOR-SHIP IN BGP

There are thousands of routers all over the internet that run BGP representing over 21000 ASs. Any two routers that have formed a TCP connection to exchange BGP routing information are called BGP neighbors or BGP peers [2].

A BGP router has direct relationship with only a few number of BGP routers. As per rules, EBGP neighbors are to be directly connected. On the other hand, IBGP neighbors can be indirectly connected. The route to a particular neighbor can be learnt dynamically through any routing protocol or can be statically assigned.

Information exchanged between BGP neighbors is what enables them to learn routes to any advertised network. Every router which has BGP configured on it is called a BGP speaker. BGP peer on the other hand is a "BGP speaker that is configured to form a neighbor relationship with another BGP speaker for the purpose of directly exchanging BGP routing information with each other" [2]. They can be internal or external.

Two routers connected to each other and having BGP configured on them have to successfully pass the TCP three way handshakes before the neighbor-ship session can be established between them.

## 5. ATTRIBUTES

BGP has a number of metrics each of which is called an "attribute". When routers exchange routing information, attributes are also exchanged so that the path-selection process is based on a particular attribute known to both the routers.

Attributes are basically of four types:

1. Well-known- most popularly used attributes.
2. Mandatory- as the name suggests, are mandatory
3. Transitive or Non-Transitive.
4. Partial.

Combinations of these path attributes are also possible as in well-known - mandatory, well-known - discretionary etc.

Well-known attributes are usually the most manipulated attributes on all BGP implementations. They are usually propagated through the BGP routing information that is exchanged between BGP routers. They can be mandatory or discretionary.

Attributes that are not well-known are termed as optional attributes. Optional attributes are not required to be supported by BGP configured routers. They can be transitive or Non-transitive. Optional-transitive attributes are also called as partial attributes. Upon receiving an optional-transitive attribute, which it does not support, a router still has to pass it to its peers. On the other hand, if it receives a non-transitive attribute, it can be dropped.

BGP has the following attributes:

- a. **AS path** – Autonomous System-Path or AS-path is a list of AS numbers the packet traverses to reach the destination
- b. **Next-hop** –(if from a network belonging to a different AS) is an IP address of the entry point of the AS along the path to the destination
- c. **Origin** – is used to inform all AS in the Internet how the prefixes (in BGP routing table, see Fig.2) originated. Legal values are IGP (i) – by the use of network command, EGP (e) – redistributed from EGP and Incomplete (?) – redistributed from IGP or declared as a static route.
- d. **Local preference** – is local to the AS. Default value is 100. Higher the value, higher is the preference
- e. **MED** – Multi-Exit Discriminator or MED is used to advertise to EBGP neighbors an exit path to the destination network. Lowest MED is most desirable
- f. **Weight** (CISCO proprietary) – is not propagated to other routers. Local to the router only.
- g. **Others**

Items a, b, c are well-known mandatory attributes while d is an example of well-known-discretionary attribute. MED is an optional non-transitive attribute. Weight has the greatest priority among all the attributes listed above.

## 6. BGP PATH SELECTION PROCESS.

BGP supports Multi-homing and Multi-roaming. Hence, the BGP forwarding table has multiple options to choose from to reach a particular network. Unlike IGP, paths are chosen based on policy rather than hop-count or Bandwidth (or any other metric for that matter). The BGP path-selection process is based on the process of elimination until a single best path is found. If it has the lowest AD value among all the routes submitted, it is registered in the routing table.

The Route Selection process is validated first by verifying whether it has a valid hop or not with no AS loops. Then paths are considered according to their priorities in the following order:

1. Highest Weight.
2. Highest Local Preference.
3. Originated by local router
4. Shortest AS-path
5. Lowest origin code (IGP < EGP < Incomplete)
6. Lowest MED
7. EBGp path > IBGP path
8. Higher preference to path through closest IGP neighbor (when in "no synch" mode. see VII)
9. Oldest route for EBGp path
10. Higher preference to lowest neighbor BGP router ID
11. Higher preference to lowest neighbor IP address

A BGP routing table showing some of the metrics is as shown in Figure 2.

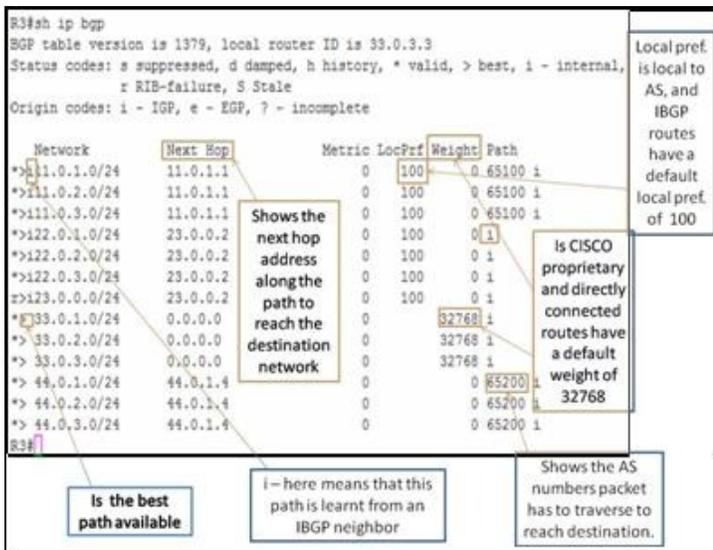


Figure 2. Interpreting BGP routing table.

### A. BGP Synchronization Rule

The BGP synchronization rule is "Never use or forward a route learnt from an IBGP neighbor to an EBGp neighbor unless the same is learnt from any IGP protocol".

If an autonomous system will be acting as a transit AS to pass traffic from one AS to another, all the routers in the transit AS have to learn routes to reach both the communicating AS through an IGP routing protocol. However, if the routes are advertised to the ASs before the routers in the transit AS have learnt the routes, then the packets would be dropped as they come in to the transit AS. To prevent this from happening BGP has to wait till IGP has propagated the routes to all the routers in the Transit AS. Thus BGP has to be synchronized with IGP. This mode is enabled by default.

However, When an AS is not configured to act as a transit AS, Synchronization can be disabled. You can also disable synchronization when all the routers in the AS have been configured with BGP.

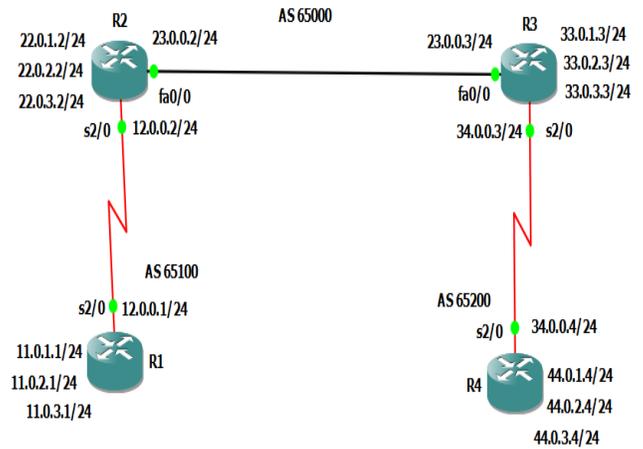


Figure 3. Network as configured on GNS3.

As shown in the above figure, We have configured 4 routers namely R1, R2, R3 and R4. Each of the routers is a BGP speaker. R1 is configured in AS 65100, R2 and R3 in AS 65000 while R4 is configured in AS 65200. Three loopback addresses are configured on each router as shown. 11.0.1.1, 11.0.2.1, 11.0.3.1 are in R1, 22.0.1.2, 22.0.2.2, 22.0.3.2 are in R2, 33.0.1.3, 33.0.2.3, 33.0.3.3 in R3 and finally 44.0.1.4, 44.0.2.4, 44.0.3.4 are configured in R4. Serial link exists between routers R1-R2 and R3-R4. A fast Ethernet link is used between R2-R3.

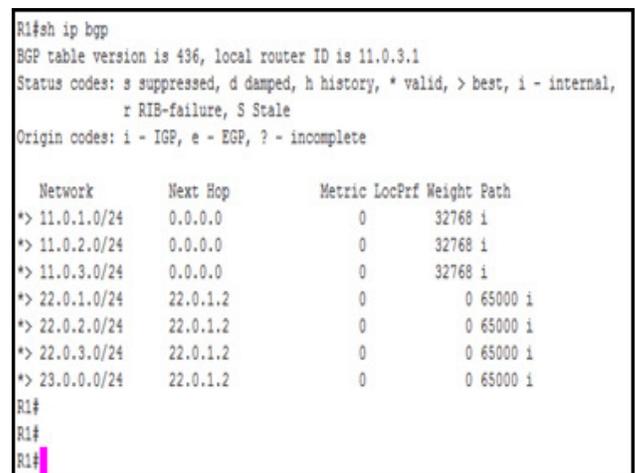


Figure 4. BGP forwarding table.

Synchronization is enabled by default. In Figure. 3, since synchronization is enabled at R2, it does not forward the paths learnt from its IBGP neighbor R3 and hence R1 does not know the routes to loopback addresses in R3 and R4.

However, when synchronization is disabled at R2 then the forwarding table looks like –

```
R1#sh ip bgp
BGP table version is 3077, local router ID is 11.0.3.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop           Metric LocPrf Weight Path
*> 11.0.1.0/24    0.0.0.0             0         32768 i
*> 11.0.2.0/24    0.0.0.0             0         32768 i
*> 11.0.3.0/24    0.0.0.0             0         32768 i
r> 12.0.0.0/24    22.0.1.2           0          0 65000 i
*> 22.0.1.0/24    22.0.1.2           0          0 65000 i
*> 22.0.2.0/24    22.0.1.2           0          0 65000 i
*> 22.0.3.0/24    22.0.1.2           0          0 65000 i
*> 23.0.0.0/24    22.0.1.2           0          0 65000 i
*> 33.0.1.0/24    22.0.1.2           0          0 65000 i
*> 33.0.2.0/24    22.0.1.2           0          0 65000 i
*> 33.0.3.0/24    22.0.1.2           0          0 65000 i
*> 44.0.1.0/24    22.0.1.2           0 65000 65200 i
*> 44.0.2.0/24    22.0.1.2           0 65000 65200 i
*> 44.0.3.0/24    22.0.1.2           0 65000 65200 i

R1#
R1#
R1#
```

Figure 5. BGP table with synchronization disabled.

With Synchronization disabled, R2 advertises the loopback network configured on R3 and R4 to R1. Synchronization has been disabled by default in Cisco IOS Software Release 12.2(8) and later.

### 7. ROUTE-MANIPULATION

Consider the network as shown in Figure 3. All of the routers are configured with BGP. Route-manipulation techniques are described as follows.

#### A. Using Local-preference

Local-Preference attribute determines the preference of an exit point from the AS. Thus, this attribute is used to select a particular exit point from the AS.

Suppose we have incoming as well as outgoing traffic from multiple neighbors and would want that the traffic to the destination be routed through a particular neighbor. Then, such route-manipulations can be done using the following commands in global configuration mode.

```
R1(config)#router bgp 65100
R1(config-router)# neighbor 22.0.1.2 route-map local in
R1(config-router)#route-map local permit 10
R1(config-route-map)#set local-preference 2000
```

Figure 6. Commands for route-manipulation using local-preference

Now, all the routes incoming from neighbor 22.0.1.2/24 have local preference of 2000. Local-preference is local to AS, i.e. Routers within the same AS exchange this attribute.

```
R1#sh ip bgp
BGP table version is 2584, local router ID is 11.0.3.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop           Metric LocPrf Weight Path
*> 11.0.1.0/24    0.0.0.0             0         32768 i
*> 11.0.2.0/24    0.0.0.0             0         32768 i
*> 11.0.3.0/24    0.0.0.0             0         32768 i
s> 12.0.0.0/24    22.0.1.2           0 2000   0 65000 i
*> 22.0.1.0/24    22.0.1.2           0 2000   0 65000 i
*> 22.0.2.0/24    22.0.1.2           0 2000   0 65000 i
*> 22.0.3.0/24    22.0.1.2           0 2000   0 65000 i
*> 23.0.0.0/24    22.0.1.2           0 2000   0 65000 i
*> 33.0.1.0/24    22.0.1.2           0 2000   0 65000 i
*> 33.0.2.0/24    22.0.1.2           2000    0 65000 i
*> 33.0.3.0/24    22.0.1.2           2000    0 65000 i

R1#
```

Figure 7. Route-Manipulation using Local-preference.

#### B. Using AS-path

When a route passes through an AS system, the AS number is added to an ordered list of AS the route has traversed.

Route Manipulation using AS-path is similar to manipulation using local preference.

Commands for route-manipulation using as-path are as follows

```
R1(config)#router bgp 65100
R1(config-router)#neighbor 22.0.1.2 route-map as in
R1(config-router)#route-map as permit 10
R1(config-route-map)#set as-path prepend 111 11
```

Figure 8. Commands for route-manipulation using as-path.

Here all the traffic incoming from neighbor 22.0.1.2 would have two AS numbers prepended in their path field as shown.

```
R1#sh ip bgp
BGP table version is 3605, local router ID is 11.0.3.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop           Metric LocPrf Weight Path
*> 11.0.1.0/24    0.0.0.0             0         32768 i
*> 11.0.2.0/24    0.0.0.0             0         32768 i
*> 11.0.3.0/24    0.0.0.0             0         32768 i
s> 12.0.0.0/24    22.0.1.2           0          0 111 11 65000 i
*> 22.0.1.0/24    22.0.1.2           0          0 111 11 65000 i
*> 22.0.2.0/24    22.0.1.2           0          0 111 11 65000 i
*> 22.0.3.0/24    22.0.1.2           0          0 111 11 65000 i
*> 23.0.0.0/24    22.0.1.2           0          0 111 11 65000 i
*> 33.0.1.0/24    22.0.1.2           0          0 111 11 65000 i
*> 33.0.2.0/24    22.0.1.2           0          0 111 11 65000 i
*> 33.0.3.0/24    22.0.1.2           0          0 111 11 65000 i
*> 44.0.1.0/24    22.0.1.2           0 111 11 65000 65200 i
*> 44.0.2.0/24    22.0.1.2           0 111 11 65000 65200 i
*> 44.0.3.0/24    22.0.1.2           0 111 11 65000 65200 i

R1#
```

Figure 9. Route-manipulation using as-path.

#### C. Selective Route-manipulation using MED.

The Multi-Exit Discriminator (MED) attribute is a suggestion to the external receiving AS about the preferred path into an AS which has multiple entry points.

Attributes pertaining to a particular route can also be edited. You can edit weight, local preference, as-path and also MED of the route as shown in the example below.

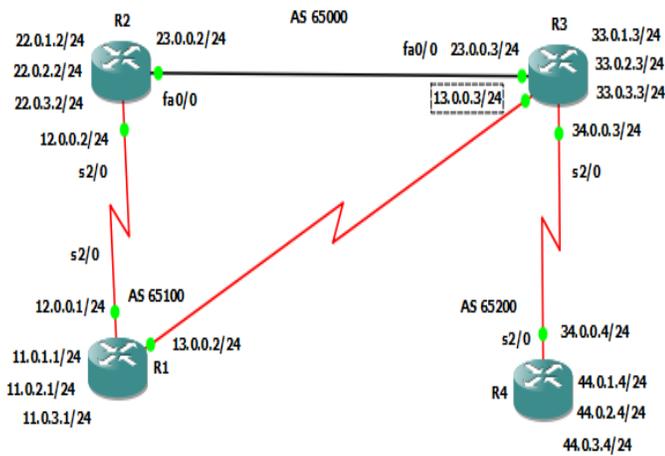


Figure 10. Example network for route-manipulation using MED

```
R1(config)#router bgp 65100
R1(config-router)#neighbor 13.0.0.3 route-map med in
R1(config-router)#access-list 1 permit 22.0.1.0 0.0.0.255
R1(config)#route-map med permit 10
R1(config-route-map)#match ip address 1
R1(config-route-map)#set metric 25000
R1(config-route-map)#route-map med permit 20
R1(config-route-map)#
R1(config-route-map)#end
R1#
```

Figure 11. Commands for route-manipulation using MED

Here in this example we needed that the traffic from ip address 22.0.1.0 be routed through 22.0.1.2 and not from 13.0.0.3. As you can see by increasing the metric of the route from 13.0.0.3 we have successfully made the route through 22.0.1.2 as the best path.

```
R1#sh ip bgp
BGP table version is 413, local router ID is 11.0.3.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop        Metric LocPrf Weight Path
  ** 11.0.1.0/24   0.0.0.0          0         32768 i
  ** 11.0.2.0/24   0.0.0.0          0         32768 i
  ** 11.0.3.0/24   0.0.0.0          0         32768 i
  ** 22.0.1.0/24   22.0.1.2          0         45000 1
  * 22.0.2.0/24   13.0.0.3        25000        0 45000 1
  ** 22.0.2.0/24   22.0.1.2          0         45000 1
  ** 22.0.3.0/24   13.0.0.3          0         45000 1
  ** 22.0.3.0/24   22.0.1.2          0         45000 1
  ** 23.0.0.0/24   22.0.1.2          0         45000 1
  ** 23.0.0.0/24   13.0.0.3          0         45000 1
  ** 33.0.1.0/24   22.0.1.2          0         45000 1
  ** 33.0.1.0/24   13.0.0.3          0         45000 1
  ** 33.0.2.0/24   22.0.1.2          0         45000 1
  ** 33.0.2.0/24   13.0.0.3          0         45000 1
  ** 33.0.3.0/24   22.0.1.2          0         45000 1
  ** 33.0.3.0/24   13.0.0.3          0         45000 1
  ** 44.0.1.0/24   13.0.0.3          0 45000 45200 1
  ** 44.0.1.0/24   22.0.1.2          0 45000 45200 1
  ** 44.0.2.0/24   13.0.0.3          0 45000 45200 1
  ** 44.0.2.0/24   22.0.1.2          0 45000 45200 1
  ** 44.0.3.0/24   13.0.0.3          0 45000 45200 1
  ** 44.0.3.0/24   22.0.1.2          0 45000 45200 1
  **
R1#
R1#
R1#
```

Figure.12 – route-manipulation using MED

**D. Route-manipulation using weight**

Weight is a CISCO proprietary attribute which has greater preference than any other attribute. It is assigned locally to a router and is not propagated to other routers. Route manipulation using weight as an attribute is similar to that of using MED. Simply substitute the command “set metric 25000” to “set weight (to any number between 0 and 32768)”. However unlike metric, Higher the weight higher is the preference.

**E. Aggregation and route-filtering using route-filtering using prefix list.**

Border Gateway Protocol (BGP) allows the aggregation of specific routes into one route in such a way that advertisement of that single route is possible. It is used to reduce the number of routes registered in the routing table. [5].

On R3 router, after aggregation, the routing table now contains 4 routes of 44 network – 3 routes having /24 mask and 1 aggregated route with /22 mask as follows.

```
R3#sh ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       I - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR

Gateway of last resort is not set

 34.0.0.0/24 is subnetted, 1 subnets
C    34.0.0.0 is directly connected, Serial1/0
 33.0.0.0/24 is subnetted, 3 subnets
C    33.0.1.0 is directly connected, Loopback1
C    33.0.3.0 is directly connected, Loopback3
C    33.0.2.0 is directly connected, Loopback2
 23.0.0.0/24 is subnetted, 1 subnets
C    23.0.0.0 is directly connected, FastEthernet0/0
 22.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
B    22.0.2.0/24 [200/0] via 23.0.0.2, 01:46:51
B    22.0.3.0/24 [200/0] via 23.0.0.2, 01:46:51
R    22.0.0.0/8 [120/1] via 23.0.0.2, 00:00:23, FastEthernet0/0
B    22.0.1.0/24 [200/0] via 23.0.0.2, 01:46:51
 11.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
B    11.0.3.0/24 [20/0] via 13.0.0.2, 01:43:19
B    11.0.2.0/24 [20/0] via 13.0.0.2, 01:43:19
B    11.0.1.0/24 [20/0] via 13.0.0.2, 01:43:19
R    11.0.0.0/8 [120/2] via 23.0.0.2, 00:00:24, FastEthernet0/0
R    12.0.0.0/8 [120/1] via 23.0.0.2, 00:00:24, FastEthernet0/0
 13.0.0.0/24 is subnetted, 1 subnets
C    13.0.0.0 is directly connected, Serial1/1
 44.0.0.0/8 is variably subnetted, 5 subnets, 3 masks
B    44.0.0.0/22 [20/0] via 44.0.1.4, 00:00:01
R    44.0.0.0/8 [120/1] via 34.0.0.4, 00:00:14, Serial1/0
B    44.0.1.0/24 [20/0] via 44.0.1.4, 00:00:01
B    44.0.2.0/24 [20/0] via 44.0.1.4, 00:00:01
B    44.0.3.0/24 [20/0] via 44.0.1.4, 00:00:01
R3#
```

Figure. 13 – Routing table

It is also possible to send only the summarized routes to the neighbor.

An alternate way to send summarized routes is through prefix list. Prefix list will filter routes on the basis of prefix mask on R4 router, the commands are as follows.

```
router bgp 65200
neighbor 33.0.1.3 prefix-list abc out
ip prefix-list abc seq 5 deny 44.0.0.0/8 ge 23
ip prefix-list abc seq 10 permit 0.0.0.0/0 le 32
```

This will filter all the routes of class A network 44.0.0.0/8 having mask greater than or equal to 23 from being sent to R3 router.

**F. Route-manipulation using Community list**

Community list is a tagging mechanism in which the routes in the list are tagged before they are sent. The receiving router matches the tag and performs route manipulation. Here in our e.g. on R4 router we will set a community string 8888 for all routes being sent to R3 router.

The commands would be

```
R4(config)#router bgp 65200
R4(config-router)#neighbor 33.0.1.3 route-map XYZ out
R4(config-router)#route-map XYZ permit 10
R4(config-route-map)#set community 8888
```

Figure14. Commands for setting Community list.

This will set community string 8888 to all routes being sent to R3 router but as community string is non transitive, tag will not be forwarded to other AS. However, In order to forcefully send it we can use the command

**neighbor 33.0.1.3 send-community**

```
R3#sh ip bgp community 8888
BGP table version is 2238, local router ID is 33.0.3.3
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop        Metric LocPrf Weight Path
  * 44.0.1.0/24    44.0.1.4         0         0 65200 i
  * 44.0.2.0/24    44.0.1.4         0         0 65200 i
  * 44.0.3.0/24    44.0.1.4         0         0 65200 i
R3#
```

Figure 15. Community list table

The community String thus created can be used for Route manipulation. For e.g.

```
R3(config)#router bgp 65000
R3(config-router)#neighbo 44.0.1.4 route-map XYZ in
R3(config-router)#ip community-list 1 permit 8888
R3(config)#route-map XYZ permit 10
R3(config-route-map)#match community 1
R3(config-route-map)#set weight 33333
R3(config-route-map)#
```

Figure 16. Commands for Route-manipulation using Community list

This will set weight 33333 for all routes having string 8888.

```
R3#sh ip bgp
BGP table version is 2295, local router ID is 33.0.3.3
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop        Metric LocPrf Weight Path
  *> 11.0.1.0/24    13.0.0.2         0         0 65100 i
  *> 11.0.2.0/24    13.0.0.2         0         0 65100 i
  *> 11.0.3.0/24    13.0.0.2         0         0 65100 i
  *>122.0.1.0/24    23.0.0.2         0 100     0 i
  *>122.0.2.0/24    23.0.0.2         0 100     0 i
  *>122.0.3.0/24    23.0.0.2         0 100     0 i
  *>123.0.0.0/24    23.0.0.2         0 100     0 i
  *> 33.0.1.0/24    0.0.0.0         0         32768 i
  *> 33.0.2.0/24    0.0.0.0         0         32768 i
  *> 33.0.3.0/24    0.0.0.0         0         32768 i
  *> 44.0.1.0/24    44.0.1.4         0         33333 65200 i
  *> 44.0.2.0/24    44.0.1.4         0         33333 65200 i
  *> 44.0.3.0/24    44.0.1.4         0         33333 65200 i
R3#
R3#
R3#
```

Figure 17. BGP table showing the edited weight.

**8. CONCLUSION AND FUTURE WORK**

BGP’s role in the Internet Routing Infrastructure is paramount. Policy-Based routing techniques enable Enterprises and ISPs to interact efficiently. The BGP protocol is being used by both Service providers and Enterprise networks. The architectural design goals of these two groups are very different which results in deployment of BGP in different environments. The idea is to break out the goals, and provide corresponding solutions for each group so as to assist effective operation. In this paper, we discussed the policy based operation of BGP and analyzed the various parameters involved in its operation to provide such optimized solutions.

Once a neighbor relationship has been established, the routers exchange routing information with each other. These routers are vulnerable to the Man in the Middle (MIN) attack as they have no way of authenticating the BGP update they receive from their neighbors. As discussed in section 6. , a BGP update contains information about all the Autonomous Systems the packet has traversed to reach a particular AS. However, the attacker can inject fake AS numbers thereby compromising the authenticity of the update. Validation of the source and path of the BGP update message without much change in the existing architecture is critical.

**9. REFERENCES**

1. CISCO IOS IP and IP routing configuration guide.
2. Building Scalable CISCO Internetworks, Volume 2, Student Guide.
3. <http://www.gns3.net/>
4. Router IOS used include c7200 and c 3600 series.
5. <http://www.cisco.com/en/US/docs/internetworking/technology/handbook/bgp.html>
6. <http://www.gns3.net/>
7. <http://tools.ietf.org/html/rfc1771> - RFC for BGP-4.
8. “Communication Networks” by Alberto Leon Garcia, Indra Widjaja.
9. “Secure Border Gateway Protocol (S-BGP)”, Stephen Kent, Charles Lynn, and Karen Seo, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 18, NO. 4, APRIL 2000
10. “Optimal Configuration for BGP Route Selection” Thomas C. Bressoud, Rajeev Rastogi and Mark A. Smith.
11. “Study on the AS relationship based inter-domain routing”, Ruijun Wang; Hongjun Wang; Cuirong Wang; Yuan Gao, Parallel and Distributed Computing, Applications and Technologies, 2003. PDCAT’2003. Proceedings of the Fourth International Conference, 2003, 42-46.
12. “A Case Study in Understanding OSPF and BGP Interactions Using Efficient Experiment Design”, D. Bauer M. Yuksel C. Carothers S. Kalyanaraman, Principles of Advanced and Distributed Simulation, 2006. PADS 2006. 20th Workshop, June 2006.

# Work Process Oriented Learning via Mobile Devices – Theoretical Basics and Examples for a (New) Didactical Approach

Georg Spöttl

Institute Technology and Education (ITB), University of Bremen  
28359 Bremen, Germany

and

Schulte, Sven

Institute Technology and Education (ITB), University of Bremen  
28359 Bremen, Germany

*Abstract:* The current technological change in the world of work is clearly visible. At the same time two problems can be identified which counteract the need for further training: On the one hand the clientele of skilled workers is not necessarily keen on further training. On the other hand the time and cost pressure within the sector does not offer any room for time-consuming further training measures far away from the workplace. This is why the project “Virtual Learning on the building site – (Vila-b)” is testing a concept which facilitates learning adapted to the occupational reality and supported by the advantages of digital media. The central didactical elements for the development of this further training course are the contextual and methodological orientation to real work processes as well as the use of digital mobile media which facilitate learning directly at the workplace. The present article starts with a description of the theoretical basics for learning within the work process and discusses the didactical elements which are necessary for work process oriented learning with digital and mobile media. This approach has been tested in research projects. The procedure for the development of a “virtual” further training concept is outlined. It encompasses learning on the building site, in in-firm schooling facilities and at home including the respective networking.

## Introduction

In European scientific circles, discussion of curricula is currently playing an increasingly important role [1] [2]. This discussion is taking place in an interdisciplinary way and is gradually increasing links to the world of work and to the employment system. The merging economic systems of the European Union provide the background and impetus for this discussion.

Education systems are thus challenged to safeguard the acquisition of competencies and to work according to standards known and acknowledged in all European countries. The further development of these systems is at the same time the political position of the European Commission, which has formulated five objectives:

1. The improvement of educational standards in Europe.
2. An easier access to learning in all phases of life.
3. The actualization of the definition of basic skills for the knowledge society.
4. The opening of general and vocational training to the local environment, to Europe and to the world.
5. The best possible use of resources [1].

In order to implement these objectives - above all, no. 4 - the quality of vocational education systems has to be improved so that they train students for widely accepted competency levels in the labour market. While the idea of a European educational policy has long been established, education and training in individual countries are carried out according to highly varied curricula structures with transfers only taking place with difficulty. There is a close interrelationship between employees' competencies and curricula. Curricula determine

the framework conditions and the prerequisites for competency development. They determine the extent to which the competencies to be developed will be broad, flexible and adequate for the labour market [3].

## 1. Learning during the work process – Basics and technological developments

The paradigm change in the world of work during the 1990ies entailed a new orientation in terms of the then strongly traditional and Tayloristic shaping of the conditions of production and work [4]. This was replaced by an enrichment of the occupational tasks particularly known as the concepts of “job enlargement” and “job enrichment” [5]. This new orientation and enrichment of work was supported by further accompanying measures resulting in the implementation of self-organized team or group work, the reduction of corporate hierarchy levels and in the continuous further development of the necessary specialist and individual competences of the staff. At the same time the companies were prepared to establish a new learning culture calling for a life-long and self-organized learning of the employees in order to successfully face the swiftly changing and ever more complex world of work.

These changes in the world of work are reflected by the structural change towards a more knowledge-intensive and complex society which is marked by a continuous and even increasing need for highly qualified skilled workers. With regard to the problems of a (at least in terms of sectors) shortage of skilled workers [6] and the long-term changes in the demographical structure, consequences become visible for both the companies and their employees. In order to safeguard a company location in a highly industrialized country and on the conditions of the swiftly progressing globalization, strategies are necessary which help to ensure and improve the ability for innovations and the quality of the produced goods and offered services. Therefore production, manufacturing and business processes of the companies must be continuously and dynamically adapted to the work-scientific or economical developments. These developments and also further changes, such as the dominance of data processing of the skilled work or the demographic change, have an impact on the qualifications and competences of the employees and thus also on initial and further training.

So far further training models mostly work with a sequential process: There are phases without any further training for the employees due to a lack of time, the current status of orders or because further training is not deemed necessary. And there are phases when further training courses are carried through as urgent “fire fighting measures”. This contradicts the basic rules of how further training measures should be organized within the framework of “lean management” which assigns a central importance to the philosophy of human resource development [7]. It calls for the establishment of a corporate culture with

competence development and knowledge transfer as central and continuous elements.

With regard to this changed importance of continuing further training of the employees, the systematic linkage of learning to the work and business processes represent one option. The orientation to work and to concrete work tasks promotes the development of increasingly relevant process oriented competences. At the same time the ability of self-relied learning is in high demand. Active learning and self-reflection in the context of concrete work processes enable the workers to acquire implicit and explicit learning experiences during and/or with the aid of an action. The core idea of the approach of learning during the work process is that learning processes start where they have the highest relevance and urgency for skilled workers and companies. Current mobile information and communication technologies support the shifting of learning into the work process. This development happened in the above mentioned research project and was optimized for concrete work processes (i.e. the ecological restoration of old buildings).

## 2. Work process oriented learning with digital and mobile media – the approach

### Principles of curriculum development

Learning in and during vocational work processes and/or work-oriented change in vocational training is being discussed everywhere triggering the question of sources for the structuring and shaping of the respective curricula. The task is to find out which contents should find their way into vocational education and training and how they should be structured. Contents with a special relevance for curricula are groundbreaking work interrelationships that have to be identified with the help of qualifications research. Such contents are the basis for conceiving complex learning and work arrangements for the qualification of employees for occupational fields. Numerous curriculum approaches - above all, when developed by work scientists - reveal a work orientation. This is, however, not the only performance index for a good curriculum.

“A curriculum must not only perform better than simply answering to the needs of the labour market and to ensure that the imparted qualifications can be made use of. It must also react to the changes in the labour market and take into consideration the multi-dimensional requirements of employees/apprentices and the learning process. These are all normative determinations for the quality of a curriculum which are subject to different criteria depending on their purpose.” [8]

In order to ensure a work orientation for curricula, vocational education scientific work-process analyses concentrate on the identification of work interrelationships and the dimensions of skilled workers. Scientifically oriented work-process analyses for vocational education pursue the following three aims:

1. To identify the *competencies* for coping with and shaping occupational work tasks;
2. To access the most important coherences for *competency development*;
3. To determine the *work-process knowledge* for the shaping of business and work processes.

With their three categories of objectives, i.e. *competency*, *competency development* and *work process knowledge*, these objectives hint at competing principles for the determination of the contents of curricula. Reetz and Seyd presume three different curriculum structures and approaches (science principle, personality principle and situation principle) and predicted well in advance of the introduction of learning fields that there would be an increasing importance of the principle of personality in vocational education practice and a continuous

dominance of the principle of science at the macro-level with its specialized scientific structures [9]. Work-process analyses take into consideration all three principles as conceived by vocational education science [10].

The following sections will outline the example of application of work process analyses for curriculum construction. The embedding of work-process analyses into a scientifically oriented set of research instruments for vocational educational and the basic methodological concept of work-process analyses are outlined.

### Didactical principles and framework conditions for learning during the work process

Since the 1990-ies the work process as a new didactical reference point has increasingly moved to the center of interest in research [11]. Within the framework of a work process oriented turning point for the areas of vocational and corporate education and training it is increasingly “demanded to make better use of the potential of the workplace as a place of self-relied and guided learning for learning processes and competence development respectively.” [12] [cf. 13] [cf. 14]. The development can be traced back to different framework conditions and targets as well as to some of the latest research questions and conceptual approaches.

**Technological progress and organisation development:** The parallelly proceeding technological and organizational developments in many companies led to the emergence of decentrally organized initial and further training concepts. Related measures such as knowledge organisation, continuing improvement processes, quality assurance etc. contain learning potentials immanent in work which are a direct part of skilled work [15]. The novelty of these concepts is the use of new learning opportunities and potentials which are present due to the development of digital media but also due to an interlinking of learning and working. The technical progress in the field of information and communication technologies also supports this development.

**Knowledge as a competitive factor:** The development towards more decentralized learning also has economical reasons as the resource of knowledge is of increasing importance in a knowledge and service society: For a lot of experts of different disciplines, learning during the work process and the thus acquired knowledge has currently become an important productive force in an increasingly customer oriented and globalised economy [cf. *ibid.*]. By shifting learning into the work process the companies can – contrary to classical learning situations – increase the effectiveness and efficiency of corporate further training. The employees as well can take advantages from this situation: learning processes are linked to work actions and thus have a positive motivational impact. Therefore learning during the work process is likely to gain further importance also in the future.

**Interlinking learning and work processes:** The term “learning during the work process” has rarely been used in this context. It is assumed that adequate (work)processes with learning contents can be identified and that learning can be encouraged by coping with tasks. The authors pursue the approach that learning can be consequently shifted into work processes. The work process as such represents a complete, multi-dimensional work process in order to cope with a corporate work order, corporate problem-solving or unknown tasks. The latter must always be seen with all their implications and imponderabilities and must always aim at a work result [16]. A complete and multi-dimensional approach also reveals previous and successive processes, objects and tools, as well as work methods as elements and takes into account that this complex process is important for the individual.

**From qualification to competence:** The rediscovery of learning during the work process includes a changed

understanding of learning objectives. Thus the focus of the learning process is no longer just on analytical qualifications and behavior imparted in educational institutions, qualifications and behavior necessary for the coping with concrete occupational work situations with the help of abilities, skills and knowledge. Moreover competence development is the primary target. The term “competence” also encompasses a person’s knowledge, abilities and skills. However, these properties represent “dispositions of self-organised acting” [17]. Another definition describes competence as “availability and adequate use of (motor, cognitive and emotional) behavior for an effective performance in concrete (occupational) situations” [18].

In this sense the understanding of competence includes subjective components in the form of self-organisation. On the other hand it is clearly shown that emphasis is placed on behavior required in the occupational context. Against this background the occupational scientific research aims at identifying and analyzing the typical tasks including the respective requirements with regard to their didactical importance for competence development [19]. With regard to the present case, this context implies that the complex work tasks to be found in real work processes should serve as the basis for learning situations. Learning takes place by confronting the learner with these tasks. The development of competences can then take place during the coping with core work tasks.

**Blended Learning – the integration of digital media in the learning process**

The pedagogical legitimation for the use of new media in the learning process reveals a number of advantages: Technology permits more flexibility in terms of time and place and supports an active and individual learning process. These benefits on the level of the individual learner also result in cost saving effects for the companies, as well as the option to better keep learning contents up-to-date and to make knowledge available where it is needed.

In spite of these advantages which altogether underpin the objective of an increase of efficiency and quality of learning processes, a critical reflection also reveals possible “dangers” or problems: given the various requirements, the use of new media is linked to a certain technical quality. The requirements for the individuals as well as the necessary acceptance within a company in terms of medially supported and/or self-organized learning process may hamper the desired learning success.

All in all the value-added of digital media does not result from their implementation but from the didactical concept, i.e. their targeted media-pedagogical use [20]. One possible approach is the integration of different learning environments and methods into a learning concept by making use of the respective benefits. Such a mixed setting is known as the concept of “Blended Learning”. It combines virtual learning phases with attendance phases in an effort to unite the potentials of e-learning with the potentials of traditional courses [21]. This is supported by a didactically adapted combination of media, medial learning offers and methods [cf. ibid]. As for the learning opportunities at the workplace, digitalized media play an important role and are being used as a support for learning.

The use of technology brings along a (qualitative and quantitative) extension of the interaction during a learning process: The teachers’ task is the pedagogical shaping of knowledge transfer to be reached by adequate didactical-methodological framework conditions. The learners’ learning strategy depends on the technical prerequisites, but also on the individual abilities of coping with technology.

Nevertheless the support of interactions during the learning process via the use of attendance and technological methods

represents the basic idea of „Blended Learning“. The interlinking of individual learning strategies for the different learning environments results in new teaching-learning-settings whose interactive structure helps the learners to mutually benefit from their experience knowledge.

The learning concept for further training developed within the framework of the research project Vila-b basically consists of such a “Blended Learning” approach. The following learning environments are interlinked:

- Seminar (attendance phase),
- Workplace: Learning during work processes on the building site, supported by a mobile device and
- PC-workplace, to be used in the company and also at home (cf. Figure 1).

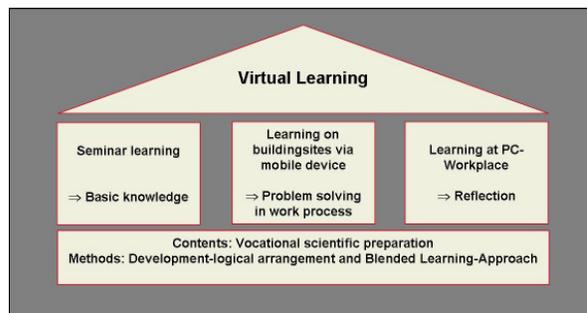


Figure 1: The learning approach

The target oriented and joint application of the three learning environments within a further training measures makes use of the respective advantages: Apart from specialized and methodological basics, in-class or attendance learning offers a direct interaction among the students and/or between students and teachers. Thus informal contacts between the students of the further training course will be established. Above all for learning processes shaped by the use of digital media it has been shown that in-class learning phases are of great importance for the later exchange of information and for the communication via the media [22]. The use of technology for learning on the building site offers new cooperative forms of learning and at the same time allows for learning processes oriented to problem solving [cf. ibid]. Learning at the PC-workplace calls for an active reflection of the just experienced learning processes and offers the opportunity to act in the sense of self-organized learning. Nevertheless a minimum of learning contents and learning times have to be achieved in order to meet the requirements for a successful completion of the further training measure.

The central didactical element is problem-solving oriented learning within the work process with the help of mobile terminals/ devices. Learning then takes place as soon as the skilled worker encounters a problem situation during the work process and during the execution of work tasks. At that moment, the skilled workers can access context-related information with the help of the mobile device, e.g. product data sheets, information of processing, expert knowledge of other skilled staff and other hints relevant for problem-solving. The existing detailed knowledge on the level of engineers and experts is then transferred to the skilled-worker level and is available at the required time and within the correct context. At the same time the relevance of these learning contents is increased.

**3. Implementation: Identification and processing of learning contents for mobile learning during the work process**

The central question for the development of contents for a further training measure is, how competence development

processes are designed and what they are based on in order to generate a domain-specific (expert)knowledge. Answers to this question can be accessed by the use of vocational scientific research methods. Four research levels and the use of adequate instruments are envisioned:

### **Sector analysis**

A sector analysis is planned as the first step of a measure which serves as a contentual preparation for the occupational context [19]. With the aid of a document analysis, specialist literature, occupational statistics, sector reports, market analyses, or descriptions of technological developments are taken into consideration. This measure is normally a long-term and comprehensive task and will be tackled in cooperation with experts.

### **Case studies - Organizational structures of occupational work processes**

Case studies are suitable to determine the basic interrelationship between work organisation and the relevant work tasks at company level. The result of the case studies reveals a multi-dimensional image of the company and shows work task structures, overall structure and personnel management, acting routines and change processes as well as organizational and in-firm processes.

### **Work process analyses**

The next step is marked by work process analyses. Based on the results of the sector description, adequate companies will be selected. These companies and their employees will be accompanied during their work on building sites. The monitoring will take place during the work process and in the form of a participating observation and interviews. Special observation and interview forms have been developed. The vocational scientists are especially interested in the corporate processes, the work tasks and their organizational structures as well as how the skilled workers are coping with core work tasks. This requires knowledge of the work structures in the companies and the involvement of the skilled workers. The objective is to document the corporate work processes, the applied and relevant tools and methods as well as the competences to be developed on the level of subjects in order to enable them to holistically tackle these work processes.

### **Expert-Skilled worker-workshops**

The documentation of the performed work process analyses are evaluated and summarized. At first core work processes will be identified which have emerged from a number of work process analyses and represent the central task for an occupational group. The observations will be evaluated with the aid of clustering.

These methods help to establish an approach for the identification of central core work tasks and the necessary appropriate core competences. However, also the next step, i.e. the didactical processing of the learning contents, must be performed.

The didactical basics for the design of the identified learning contents represent a development-logical didactical approach which concentrates on competence development by working on concrete work tasks and on problem-solving in challenging real work situations.

In terms of learning theory, development logical didactics constitutes a further development which - in contrast to behaviorism - applies some elements of the constructivist approach [23]. The orientation to concrete work tasks thus clearly differs from the stimulus-reaction scheme of learning on the one hand but also from learning processes based on the mainly experience-based construction of knowledge. The

orientation on real (work) tasks makes greater demands on the ability of self-learning in the context of work situations and contains active learning and self-reflection processes in the context of real work situations. In addition the perspective of holistic learning processes is in the focus of development logical didactics: Knowledge and acting form a unit and allow the gaining of both implicit and explicit learning experiences. This means that competence development is not only perceived as a behavioral change oriented to consequences or as a mere context-specific generation of knowledge and skills. With regard to an innovative use of the mobile terminals, the application of new media cannot only be regarded as a programmed instruction or as an acting oriented simulation. Moreover it promotes real problem solving during the work processes.

Based on these reflections the question must be raised how the learning contents may be didactically arranged in a development logical order which reveals a learning development. Considerations of a respective competence model lead to a multi-dimensional implementation which differs in terms of depth and/or levels of learning contents, with regard to the specter of the learning contents and in terms of the phases of a holistic action. Therefore the context orientation of the learning contents as a result of the different work tasks is relevant.

The competence acquisition model created by Dreyfus and Dreyfus illustrates the five-level development path from a "novice" to an "expert" [24]. A slightly modified version of this approach has been discussed for some time in the German speaking countries. The contents of learning processes can be formulated as competence requirements on different levels. Nevertheless the holistic approach is being pursued and the learning contents are always focused on concrete work processes.

Comprehensive studies carried through in the health sector [25] and in the automotive sector (car mechatronic) [26] confirm the model in its basic form, above all in terms of its horizontal arrangement of competence levels. In addition these studies show the relevance of the incorporation of competences into the domain specific work interrelationships: Occupational competences are typified by a close link to the respective occupational sector and to vocational education and training.

## **4. Previous findings of the research project and outlook**

The implementation of the approach of work process oriented learning with the aid of digital media and a further training measure based on this approach has been tested for craftspersons.

In practice a learning process was e.g. triggered by a problem resulting from a work situation of a skilled worker on the building site. The worker scanned the barcode assigned to the problematic component with the aid of a mobile terminal. The user surface then indicated all proposals or information available for the respective component. Due to the shaping of the surface, the explorative and development-logical structure of the learning contents and the intuitive operation, deeper and more detailed solutions could be found for the problems. The learning experience acquired during the execution of the work processes could later be retrieved, deepened and reflected at the PC workplace.

The mobile learning system to be developed offers additional social components to the user. The option for the generation and incorporation of User Generated Content (UGC) allows for the addition of comments on the indicated contents element. Apart from this, the skilled worker can highlight the information in order to facilitate a quick access to experience knowledge according to the current situation.

By linking of learning and working within the work processes, knowledge is generated where it originates. At the same time, active and acting oriented learning takes place on an individual level. This contains phases of holistic acting (planning, realization and evaluation) and thus facilitates the required holistic competence development process.

However, the organisational framework conditions have to be adequately adapted. Initial and further training concepts are needed which help to develop and to support the required competences. Above all in today's digitalized world, digital media are deeply integrated into the work processes due to an increased application of personal computers and the Internet. Thus it is crucial to extend the initial and further training measures in the companies by the use of E-Learning concepts.

The central question is how successful the learners are by adhering to this approach with respect to learning progress and increase in competences. But also the applicability (the present case concentrates on the learning environment "building site"), the satisfaction of the participants and companies as well as the contribution of the different learning environments and teaching-learning methods for the learning success are subjects to scientific evaluation questions. These are necessary to verify the options of a target-group adequate implementation of the learning concepts as well as the shaping of the further training contents according to a work process oriented didactical approach of learning with mobile end devices. The following surveys are in the center of further research activities:

#### The influence of work process oriented learning with digital media on competence development

Surveys have to be conducted in the context of work process oriented learning in order to identify the impact of Web2.0 technology on the one hand and of the individual media competences of the trainees/skilled workers in terms of competence development. Surveys on the competence development by the self-reliant and active use of a learning community - by keeping in mind the described change of skilled work - are an important research target from the vocational scientific and information technological perspective.

#### Options for the use and shaping of learning with digital media

The use of digital media brings along options for a qualitative and quantitative extension of the interaction during the learning process: Teachers (e.g. in the form of a tutor) must shape the pedagogical learning processes in an innovative and individually coordinated way. In addition the shaping depends on the technical prerequisites, but also on the individual prerequisites of the learners. The Web2.0 technology emphasizes the aspect of self-organized learning [cf. 27] and enables the opportunity to establish "communities of practice" (and the use their advantages) inside of companies [28]. At the same time the potential difficulties resulting from the requirements and the acceptances by the individuals must also be taken into consideration.

#### REFERENCES

- [1] Commission of the European Communities, "Report from the Commission: the concrete future objectives of education systems", Brussels, European Commission (COM 59, 2001), 2001.
- [2] European Training Foundation, "Development of standards in vocational education and training", Vol. 1-4, Luxembourg, Office for Official Publications of the European Communities, 1999.
- [3] G. Spöttl, "Curriculum Approaches and Participative Curriculum Development", in: R. Maclean, D. Wilson (Eds.): International Handbook of Education for the Changing World of Work, Bridging Academic and Vocational Learning, Netherlands, Springer, CCLVI, pp. 1627-1637, 2009.
- [4] H. Kern and M. Schumann, "Das Ende der Arbeitsteilung? Rationalisierung in der industriellen Produktion", Beck Verlag, München, 3. Auflage, 1986.
- [5] H. Schuler, "Lehrbuch der Personalpsychologie", Hogrefe Verlag, Göttingen, 2006.
- [6] M. Schnitger, L. Windelband, "Fachkräftemangel auf Facharbeiterebene im produzierenden Sektor in Deutschland: Ergebnisse der Sektoranalyse aus dem Projekt „Shortage of Skilled Workers“ (SOS)", Bremen, Institute of Technology and Education, 2007.
- [7] J. Liker, D. Meier, "Toyota Talent: Developing your people the Toyota way", McGraw-Hill Professional, 2008.
- [8] M. Becker, G. Spöttl, "Berufswissenschaftliche Forschung und deren empirische Relevanz für die Curriculumentwicklung.", in: bwp@, Berufs- und Wirtschaftspädagogik – online, Ausgabe 11, 2006.
- [9] L. Reetz, W. Seyd, "Curriculare Strukturen beruflicher Bildung.", in: R. Arnold, A. Lipsmeier, (eds.), Handbuch der Berufsbildung, pp. 203-219, Opladen, Germany, Leske & Budrich, 1995.
- [10] G. Spöttl, "Learning through the Work Process - Challenges and the Shaping of Skill Requirements.", in: G. Loose, G. Spöttl, Md. Sahir Yusoff, (eds.), "'Re-Engineering' Dual Training - The Malaysian Experience.", Frankfurt/M., Berlin u.a., pp. 31-44, 2008.
- [11] W. Georg., "Lernen im Prozeß der Arbeit", in: H. Dederich (Hrsg.), "Handbuch zur arbeitsorientierten Bildung", München, pp637-659, 1996.
- [12] W. Bauer, C. Koring, P. Röben, M. Schnitger, "Weiterbildungsbedarfsanalysen. Ergebnisse aus dem Projekt 'Weiterbildung im Prozess der Arbeit (WAP)", Institute of Technology and Education, Bremen, 2007.
- [13] P. Dehnbostel, H. Nowak, "Arbeits- und erfahrungsorientierte Lernkonzepte", Bielefeld, 2000.
- [14] K. Illerins, "Workplace learning and learning theory", in: Journal of Workplace Learning. Volume 15, Issue 4, pp. 161-178, 2003.
- [15] P. Dehnbostel, "Lernumgebungen gestalten", in: F. Rauner, (Hrsg.), "Handbuch Berufsbildungsforschung", 2. aktual. Aufl. Bielefeld, pp. 378-383, 2006.
- [16] J. Pangalos, S. Knutsen, "Möglichkeiten und Grenzen der Orientierung an Arbeitsprozess für die berufliche Bildung", In: J.-P. Pahl, F. Rauner, G. Spöttl, "Berufliches Arbeitsprozesswissen. Ein Forschungsgegenstand der Berufsfeldwissenschaften", Baden-Baden, pp. 105-116, 2000.
- [17] J. Erpenbeck, L. v. Rosentiel, "Einführung", in: J. Erpenbeck, L. v. Rosentiel, (Hrsg.), "Handbuch Kompetenzmessung", Stuttgart, pp. IX-XL, 2003.
- [18] J. Zabeck, "Schlüsselqualifikationen - Zur Kritik einer didaktischen Zielformel", in: Wirtschaft und Erziehung, 3, pp. 77-86, 1989.
- [19] M. Becker and G. Spöttl, "Berufswissenschaftliche Forschung – Ein Arbeitsbuch für Studium und Praxis", Peter Lang GmbH, Frankfurt am Main, 2008.
- [20] R. Reimer, "Blended Learning - veränderte Formen der Interaktion in der Erwachsenenbildung", in: Report: Zeitschrift für Weiterbildungsforschung, Heft 27, pp. 265-271, 2004.
- [21] H. Mandl, B. Kopp, S. Dvorak, "Blended-Learning als neues E-Learning-Konzept", in: Deutsches Institut für Erwachsenenbildung (DIE), "Aktuelle theoretische Ansätze und empirische Befunde im Bereich der Lehr-Lern-Forschung - Schwerpunkt Erwachsenenbildung", pp 55-59, 2004.
- [22] S. Kraft, "Blended Learning - ein Weg zur Integration von E-Learning und Präsenzlernen", in: Report: Zeitschrift für Weiterbildungsforschung, Heft 26, pp. 43-52, 2003.
- [23] T. Grantz, S. Schulte, G. Spöttl, "Virtuelles Lernen auf der Baustelle", in: bwp@. Berufs- und Wirtschaftspädagogik - Online, Ausgabe 15, 2008.
- [24] H. L. Dreyfus and S. E. Dreyfus, "Mind over machine. The power of human intuition and expertise in the era of the computer", Oxford, 1986.
- [25] Benner, "From Novice to Expert – Excellence and Power in Clinical Nursing Practice", Menlo Park, Addison-Wesley Publishing Company, 1984.
- [26] G. Spöttl, "Car Mechatronic – An Open Approach for Designing a New European Core Occupational Profile." in:

- Schlick, C. M. (ed.): Industrial Engineering and Ergonomics. Visions, Concepts, Methods and Tools. Dordrecht, Heidelberg: Springer, pp. 67-87, 2009.
- [27] H. D. Höpfner, H. Koch, “**Self-reliant learning in technical education and vocational training (TVET)**”, Berlin, Kuala Lumpur, 2003.
- [28] E. Wegner, R. McDemott, W. Snyder, “**Cultivating communities of practice: A guide to managing knowledge.**”, Harvard Business School Press, Cambridge, 2002.

#### AUTHORS

**Prof. Georg Spöttl** is Director of the Institute Technology and Education (ITB) at the University of Bremen, Am Fallturm 1, 28359 Bremen, Germany. e-mail: [spoettl@uni-bremen.de](mailto:spoettl@uni-bremen.de).

**Sven Schulte** is scientific assistant of the Institute Technology and Education (ITB) at the University of Bremen, Am Fallturm 1, 28359 Bremen, Germany. (e-mail: [sven.schulte@uni-bremen.de](mailto:sven.schulte@uni-bremen.de)).

# Virtual 3D Manipulation Using Cutting Plane Lines

Hidetoshi Nonaka

Graduate School of Information Science and Technology, Hokkaido University  
Sapporo, 060 0814, Japan

## ABSTRACT

Various techniques have been proposed for virtual 3D user interfaces. They are roughly classified into two categories, virtual 3D manipulation, and virtual 3D navigation. In this paper, we present a virtual 3D/6DOF manipulation system using conventional 2D pointing device, such as a mouse with single button.

**Keywords:** 3D user interface, Pointing device, 6DOF operation, Cutting plane line

## 1. INTRODUCTION

There have been various methods to realize virtual 3D manipulation using 2D pointing devices such as mouse.

In order to achieve the rotation of 3DOF, Hanson [1] proposed a method by assigning the movement of x-axis and y-axis to the positive (counter-clockwise) rotation around y-axis and the negative (clockwise) rotation around x-axis (Rolling ball). In the result, the positive rotation of a virtual 3D object around the z-axis is achieved by a negative rotation of the mouse. Chen et al. [2] proposed a 3DOF rotation method using a virtual track ball in addition to 3DOF rotation using the slider, the menu, and the button, and evaluate the operational performance. Shoemake [3] formulated the correspondence of the turn of the arc on a virtual sphere and virtual 3D object in geometrical consideration, and has achieved the mouse operation not to accompany hysteresis (ARCBALL). Henriksen et al. [4] reformulated the rotation operation method with the virtual track ball, and improve the error margin. Khan et al. [5] proposed a 3D widget as an orientation controller (View Cube)

In order to achieve the movement of 3DOF, many systems adopted “ray casting” method or “camera control” method, for example, Balakrishnan et al. [6] proposed a method using two-handed operation. Stuerzlinger et al. [7-8] introduced a translation method by restricting the movement to sliding on a surface for the mouse cursor (SESAME), and it was applied to a virtual 3D puzzle system [9]. Moreover, a lot of input devices specialized to 3D/6DOF manipulations have been also proposed [10-11].

We have presented a demonstration of a 3D/6DOF direct manipulation system, using mouse with single button [12]. It means 3D rotation and 3D movement of

virtual objects are achieved with only left button and none of right button, mouse wheel, or other devices are required. It also means this method is applicable to most 2D pointing devices such as pen-tablet, pointing-stick, trackball, touchpad, and so on.

Unlike the past techniques such as virtual trackball, virtual sphere, or 3D widgets, the user can rotate a virtual 3D object with the feeling like gripping it directly. By using cutting plane line, the user can move a virtual 3D object along a trajectory on a plane, which can be adjusted within a 3D space. The cutting plane line is also helpful for the perception of depth and positional relationship of the object and other objects.

In this paper, we present a virtual 3D/6DOF manipulation system using revised method of cutting plane line.

## 2. ROTATION IN 3DOF

In this section, we summarize our method described in [12]. By assuming that a virtual 3D object is adequately modeled and displayed to the user, and the user perceives it properly with the pictorial cue, user’s pointing on a perspective image of the object can be regarded as the pointing on the object in a virtual 3D space.

In the Figure 1, a virtual icosahedron is shown. The origin of the screen, the origin of the virtual 3D space, and the center of rotation are assumed to be identical for only the simplicity. The coordinate system of the virtual 3D space is right-handed. Suppose the user presses the left button when the mouse cursor is on the point P, the corresponding 3D position vector  $\tilde{p}$  is given by (1).

$$\mathbf{p} = \overline{\mathbf{OP}} = (p_x, p_y, 0)^t, \quad \tilde{\mathbf{p}} = \overline{\mathbf{O\tilde{P}}} = (\tilde{p}_x, \tilde{p}_y, \tilde{p}_z)^t$$

$$\tilde{\mathbf{p}} = \frac{\mathbf{n}^t(\mathbf{c} - \mathbf{h})\mathbf{p} + \mathbf{n}^t(\mathbf{c} - \mathbf{p})\mathbf{h}}{\mathbf{n}^t(\mathbf{p} - \mathbf{h})}, \quad (1)$$

where  $\mathbf{n}$ ,  $\mathbf{c}$ , and  $\mathbf{h}$  are the three-dimensional coordinate of the normal vector of triangle ABC, the coordinate of arbitrary point on the same triangle, and the coordinate of the viewpoint of the perspective transformation. The distance between the center of rotation O and the point P is given by (2).

$$\rho = \frac{1}{\mathbf{n}^t(\mathbf{p} - \mathbf{h})} \sqrt{(\mathbf{n}^t(\mathbf{c} - \mathbf{h}))^2 \mathbf{p}^t \mathbf{p} + (\mathbf{n}^t(\mathbf{c} - \mathbf{p}))^2 \mathbf{h}^t \mathbf{h}} \quad (2)$$

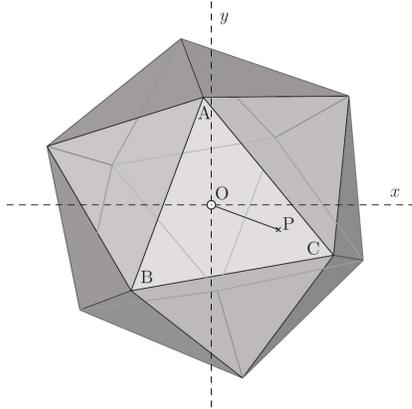


Figure 1. Pointing on the surface of a virtual 3D object.

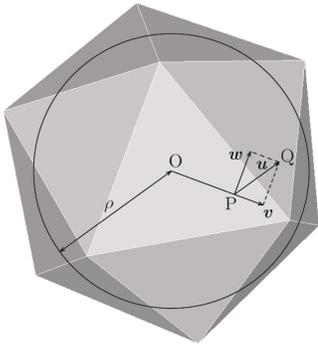


Figure 2. Dragging on the surface of a virtual 3D object.

In the Figure 2, suppose the user performs a dragging from P to Q. The axis of rotation  $\mathbf{r}$  and the angle of rotation  $\theta$  are calculated by (3) and (4).

$$\mathbf{q} = (q_x, q_y, 0)^t, \quad \mathbf{u} = \overline{PQ} = (u_x, u_y, 0)^t$$

$$\mathbf{r} = (r_x, r_y, r_z)^t, \quad \mathbf{v} = \mathbf{p}^t \mathbf{p} \mathbf{u} / \mathbf{p}^t \mathbf{p}, \quad \mathbf{w} = \mathbf{u} - \mathbf{v},$$

If  $0 \leq \mathbf{p}^t \mathbf{p} < \rho^2$

$$r_x = -v_y - \sqrt{1 - \mathbf{p}^t \mathbf{p} / \rho} w_y,$$

$$r_y = v_x + \sqrt{1 - \mathbf{p}^t \mathbf{p} / \rho} w_x, \quad r_z = \frac{1}{\rho} \mathbf{p} \times \mathbf{w},$$

If  $\rho^2 \leq \mathbf{p}^t \mathbf{p}$

$$r_x = -v_y, \quad r_y = v_x, \quad r_z = \frac{1}{\sqrt{\mathbf{p}^t \mathbf{p}}} \mathbf{p} \times \mathbf{w}, \quad (3)$$

$$\theta = \arctan(\sqrt{\mathbf{u}^t \mathbf{u}} / \rho). \quad (4)$$

Finally, the matrix of rotation is given by (5).

$$\bar{\mathbf{r}} \equiv \mathbf{r} / \sqrt{\mathbf{r}^t \mathbf{r}}, \quad [\bar{\mathbf{r}} \times] \equiv \begin{pmatrix} 0 & -\bar{r}_z & \bar{r}_y \\ \bar{r}_z & 0 & -\bar{r}_x \\ -\bar{r}_y & \bar{r}_x & 0 \end{pmatrix},$$

$$R = \frac{1}{\sqrt{\mathbf{u}^t \mathbf{u} + \rho^2}} \left( \frac{\mathbf{u}^t \mathbf{u} \bar{\mathbf{r}} \bar{\mathbf{r}}^t}{\sqrt{\mathbf{u}^t \mathbf{u} + \rho^2} + \rho} + \sqrt{\mathbf{u}^t \mathbf{u}} [\bar{\mathbf{r}} \times] + \rho I \right) \quad (5)$$

In the case that the center of rotation and the origin are different point, only the parallel translations of coordinate axes before and after the rotation are required. Consequently, the rotation of 3D object in 3DOF can be achieved by dragging using conventional 2D pointing device. In the rotation by (5), there are following characters:

1. During dragging, if the mouse cursor is on the object, the positional relation between the cursor and the surface of the object is roughly constant.
2. During dragging, even if the mouse cursor is off the object, the speed of rotation is roughly proportional to the speed of mouse.
3. During dragging, if the distance between mouse cursor and the origin is longer than  $\rho$ ,  $\mathbf{w}$ -component of the vector  $\mathbf{u}$  corresponds to the rotation around the  $z$ -axis.
4. The speed and the orientation of rotation depend on where on the surface of the object is dragged. Figure 3 and Figure 4 show the examples of the relation between dragging point and rotation.

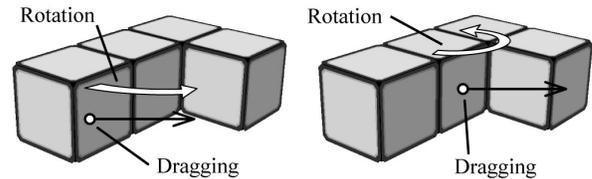


Figure 3. An example of the relation between dragging point and rotation. The speed of rotation depends on the distance in the 3D space between the dragging point and the center of rotation.

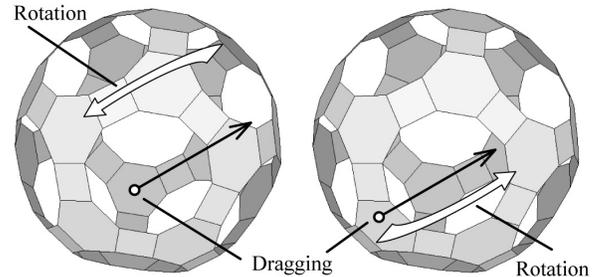


Figure 4. An example of the relation between dragging point and rotation. The orientations of rotations are opposite between lateral side and medial side of polyhedral faces.

### 3. MOVEMENT IN 3DOF USING CUTTING PLANE LINE

The mouse is one of *indirect-control* pointing devices. Cognitive processing for hand/eye coordination and sensori-motor operation are required to the mouse user. Especially, appropriate mapping between the surface under the mouse and the plane of the display are necessary. However, once the mapping is established, the user can operate the mouse even in the cases that the display is horizontal, the display is upright, the projector-screen is obliquely backward, and so on. It suggests that the user can operate the mouse if he/she can perceive a plane in a virtual 3D space.

We assume to approximate a curve or trajectory in three-dimensional space by the concatenations of curves or trajectories within two-dimensional subspaces (Figure 5). It can be considered to be the approximation of 3D curve by a continuous sequence of piecewise torsion-free curves, and the natural dimensional extension of piecewise linear approximation of a curve.

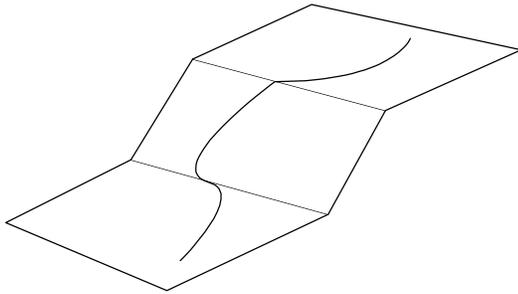


Figure 5. Approximation of 3D curve by 2D curves.

In our method, the operation of object manipulation is divided into three phases: the phase to rotate (Rotating), the phase to adjust the plane (Tilting), and the phase to move along the plane (Translating). The movement along arbitrary trajectory can be achieved by alternately executing the last two phases. The phase switches by releasing and pressing the left button within about 200 msec. Figure 6 shows simplified state transitions among three phases and the normal phase (Neutral). In the figure, “Press” stands for the pressing mouse left button, and “Release” stands for the releasing mouse left button.

In the second phase, the axes are tilted in 2DOF (roll and pitch) according to the mouse movement. The center of rotation is the virtual 3D position that was pointed by mouse at the moment of state transition. During the second phase, the plane is indicated by cutting-plane lines on all of the 3D objects in the screen. The three axes of the plane are calculated by (6). Again,  $u$  denotes the two-dimensional vector of dragging.

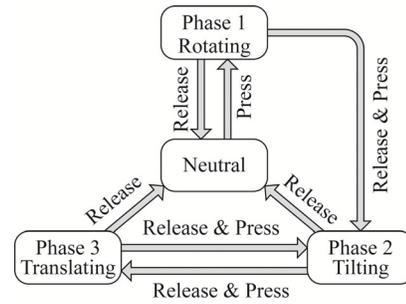


Figure 6. State transitions among four phases.

$$\theta = \mu\sqrt{u_x^2 + u_y^2}, \quad c = \cos \theta, \quad s = \sin \theta,$$

$$e_x = \begin{pmatrix} cu_x^2 + u_y^2 & (c-1)u_x u_y & -su_x \\ u_x^2 + u_y^2 & u_x^2 + u_y^2 & \sqrt{u_x^2 + u_y^2} \end{pmatrix}^t,$$

$$e_y = \begin{pmatrix} (c-1)u_x u_y & u_x^2 + cu_y^2 & -su_y \\ u_x^2 + u_y^2 & u_x^2 + u_y^2 & \sqrt{u_x^2 + u_y^2} \end{pmatrix}^t,$$

$$e_z = e_x \times e_y = \begin{pmatrix} su_x & su_y & c \\ \sqrt{u_x^2 + u_y^2} & \sqrt{u_x^2 + u_y^2} & \end{pmatrix}^t, \quad (6)$$

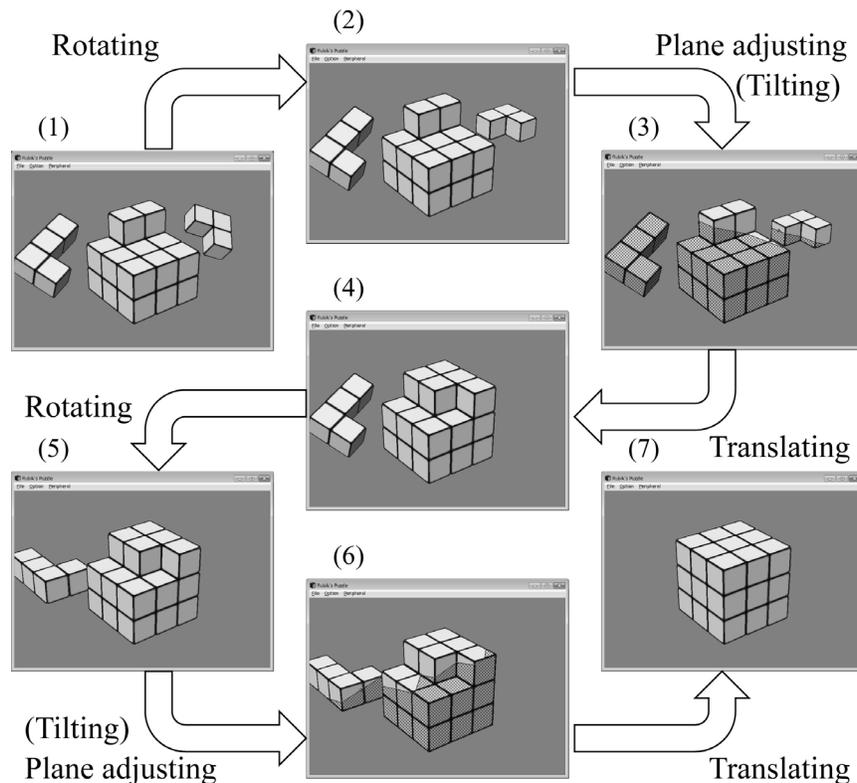
where  $\mu$  is a constant depending the cursor/device ratio, individual difference, and so on.

In the third phase, x-y movement of mouse corresponds to the cursor movement along the plane aligned in the second phase. The cutting-plane lines are hidden during the third phase.

Figure 7 shows an example of 3D/6DOF applications. Recently, 3D puzzle of this type was adopted as the problem of a UI contest [13]. In this example of puzzle, last two pieces remain uncompleted (1). By rotating the rightmost piece, it becomes parallel to the position of the target (2). By adjusting the tilt of the plane with the cutting plane line, the piece and the target are located on the same plane (3). The rightmost piece is translated to the target along the plane (4). In the same way, the last piece is also rotated and moved to the position of the target (5-7). Time required to complete such a virtual puzzle was less than 1.5 times of that to assemble the same real puzzle by hand, as the result of preliminary experiments for evaluation.

### 4. CONCLUSION

We proposed a method for virtual 3D/6DOF manipulation. The operation of object manipulation was divided into three phases: rotating, tilting, and translating. Especially, the translating of 3D object was realized using cutting plane line. Experimental evaluation is near future work.



**Figure 7.** An example of applications. A 3D puzzle can be operated with 3 phases: rotating, tilting plane, and translating.

## REFERENCES

- [1] Hanson, A.J.: The Rolling Ball, Graphics Gems III, pp.51-60, Academic Press (1992)
- [2] Chen, M., Mountford, S.J., and Sellen, A.: A Study in Interactive 3-D Rotation Using 2-D Control Devices, Computer Graphics, Vol. 22, No. 4, pp.121-129 (1988)
- [3] Shoemake, K.: ARCBALL: A User Interface for Specifying Three-Dimensional Orientation Using a Mouse, Proceedings of Graphics Interface 92, pp.151-156 (1992)
- [4] Henriksen, K., Sporning, J., and Hornback, K.: Virtual Trackballs Revised, IEEE Transactions on Visualization and Computer Graphics, Vol. 10, No. 2, pp.206-216 (2004)
- [5] Khan, A., Mordatch I., Fitsmaurice, G., Mtejka, J., and Kurtenbach, G.: ViewCube: A 3D Orientation Indicator and Controller, I3D 2008 Conference Proceedings: ACM Symposium on Interactive 3D Graphics, pp.17-25 (2008)
- [6] Balakrishnan, R. and Kurtenbach, G.: Exploring Bimanual Camera Control and Object Manipulation in 3D Graphics Interfaces, Proceedings of CHI99, pp.56-63 (1999)
- [7] Oh, J.-Y., Stuerzlinger, W. and Danahy, J.: SESAME: Towards Better 3D Conceptual Design Systems, Proceedings of ACM DIS, pp.80-89 (2006)
- [8] Stuerzlinger, W., and Wingrave, C.A.: The Value of Constraints for 3D User Interfaces, Coquillart, S et al. (eds.), Virtual Realities, Springer-Verlag, pp.203-223 (2011)
- [9] Shuralyov, D. and Stuerzlinger, W.: A 3D Desktop Puzzle Assembly System, IEEE 3D Contest, Proceedings of IEEE Symposium on 3DUI 2011, pp.141-142 (2011)
- [10] 3D connexion: SpaceNavigator, etc. <http://www.3dconnexion.com/>
- [11] Sudin, M., and Fjeld, M.: Elastic 6 DOF Input, International Journal of Human-Computer Interaction, Vol. 25, No. 7, pp.647-691 (2009)
- [12] Nonaka, H.: 3D/6DOF Manipulation Using Mouse with Single Button, Proceedings of Interaction 2011 (in Japanese)
- [13] 3DUI contest, IEEE Symposium on 3D User Interfaces 2011, Singapore, March 19-20, 2011 <http://conferences.computer.org/3dui/3dui2011/>

# Arctic Challenges – Developing the ICT Strategy for Greenland – a Case Study

**Soren Duus Ostergaard**

senior eGovernment Advisor,

External Lecturer, eGovt

IT University of Copenhagen

Rued Langgaardsvej 7, 2300 Copenhagen,

Denmark

soren@duus.com

## ABSTRACT

This paper describes the background, process to develop and the outcome of a new ICT strategy for Greenland following the referendum of autonomy from Denmark in 2008 and the formation of a new, progressive Government in 2009. Due to climate, the natural conditions and not least the scattered and small population inhabiting the largest island on the Globe, the need for deployment and take-up of internet based technology is a must and seen as a pre-condition for sustainable development of the society, and the paper outlines the phases, methodology and results of the steps to develop a consolidated strategy, that was accepted by the Greenland Government, the Nalaakkersuisut, this summer.

**Keywords:** Greenland, ICT Strategy, Governance, eHealth, eGovernment, Competencies, Innovation

## 1. INTRODUCTION

By 2011 we have seen a number of case studies of eGovernment from various countries – most of these come from the Western Industrialized countries, some are focusing on the challenges of managing public ICT projects, others are descriptive studies focusing on 'best practices' for ministries, cities. This case study differs as it is not commonly known or otherwise documented, as it is based on an on-going project initiated by the new Greenland Government, that took over in June 2009 as the first Government to manage a high level of independence from Denmark. The idea of promoting a new and efficient ICT strategy for the entire Greenland society was a part of the agreement between the coalition partners that formed the new Greenland government.

The agreement stated:

*'The citizen will have equal access to information, communications and service including optimal opportunities to participate in the democratic process in Greenland'*

*'We will create an effective, holistic and coherent administration based on a joint ICT-organization across the public sector'*

*'The ICT-strategy will have the citizens' needs as the starting point and also support the educational system, industry as well as the municipalities of Greenland'*

The scope that makes this case study different from most other 'e-Government' cases is that it deliberately aims at developing a total

ICT strategy for the Greenland society by establishing an umbrella-like, joint strategy for:

- Developing Competencies and Skills on all levels
- Innovation
- Health
- Municipalities and Social Services
- Public IT-Governance and Citizen Services

To ensure consensus and cooperation the strategy was developed in close cooperation with private sector organizations and communities as well across the Public Sector organisations.

## 2. GREENLAND – BACKGROUND

Greenland covers more than 836.000 square miles, of which 81% are covered with ice. Originally a part of the Norse kingdom since 836, the Viking population vanished after 4-500 years; contact was re-established and Greenland became a colony under the Danish Kingdom in mid 1700's. After 1979 Greenland was granted Home Rule, and in 2008 a referendum was held leading to autonomy within the Kingdom of Denmark for all matters except defense, protection of sea, justice and police. Foreign policy is a joint area. During the last part of the so-called home rule period, the social-democratic party Siumut was leading the country including key industries like Royal Greenland (Fishing), KNI (Retail), Royal Arctic (Boat traffic), Air Greenland and TeleGreenland. This led to a lot of accusations of nepotism and corruption, and the situation worsened as the party tried to promote the Greenland Inuit language, which resulted in a severe reduction of the language skills of the younger Greenland population, as only a very limited number of intermediate and higher education possibilities are offered in Greenland itself. The leading administrative, technical and skilled staff of the public sector and more advanced private sector has traditionally been imported as more or less temporary resources from Denmark.

The Greenland society has developed in just 3 generations from a country dominated by hunters and small fishers to a semi-industrialized country with a dominating industry in fishing and fish plants. At this point Greenland is expecting a boom in mining and oil drilling, and negotiations with Alcoa has led to a DOU for construction of a large scale aluminum plant powered by hydro power stations.

Greenland has a population of 56.000 of which 6.000 are temporary workers, specialists, mostly Danish, but as the Government has sold exploration rights for oil and minerals, an increasingly higher number of English speaking oil and mining specialists are coming to Greenland. 80% of the population lives in cities of which the capital, Nuuk, has about 20.000 inhabitants. Less than 9000 inhabitants are living in 69 small settlements and villages.

The Siumut party that was the dominating party since 1970 was replaced by a new, progressive coalition under IA (Inuit Ataqatigiit) with the small Democratic Party in June 2009. The new Government declared that they wanted a fresh start for Greenland, focus on creation of new jobs (mining, oil, aluminum), improved social conditions, growth and education for all. An important area was the modernization of the administration, reorganizing the public sector and improving services to the citizens and the industry plus opening up government to become transparent and responsive.

Late 2008 a reorganization of local Government/Central Government took place; 14 municipalities were replaced by 4 and tax offices moved from local to central Government.

## 2. ICT Strategy - Objectives

### Project Initialization & Organization

The ICT Strategy work was launched by the joint ICT Steering Committee for Central and Local Government in Greenland in November 2009. At the kick-off meeting the overall time schedule for the ICT plan was decided:

*Phase 1* – Clarify base-line status and Define Visions and Barriers for each Focus Area – November 2009 – February 2010

*Phase 2* - Define Overall Strategies for each Focus Area – March – April 2010

*Phase 3* – Define Key Strategic Action Plans, High priority Tactical Plans and initial High Priorities for Fiscal Year 2011 - May – July 2010

*Phase 4* – First draft of a Consolidated National ICT Strategy – Spring 2011 – Steering Committee/Hearing of Stakeholders

*Phase 5* – Final Proposal 2011-2015 to Government – May/June 2011

The original time schedule aimed at delivering a final proposal in December 2010, but due to lack of a skilled project manager – a recurrent problem in Greenland – the project was delayed. For each of the strategic areas a team leader was appointed, and the Joint Public Sector ICT Steering Committee appointed the ICT Secretariat as the responsible coordinating unit. The author of this paper was hired as an external facilitator based on his experience as a former member of the Greenland ICT Council from 2006 – 2009.

### Developing the Vision – Phase 1

The initial vision for the strategy was defined as:

*In Greenland Information and Communications Technology is a crucial factor in securing a coherent society and is the foundation for a globally oriented development of the society that will result in growth and welfare for everybody.*

This is obviously far from being an operational vision, so in order to develop, qualify and test the specific visions and ambitions for all of the key focus areas, the methodology developed by the IBM Institute for Business Value was used.

This methodology is particularly well suited for national strategy development and is derived from the fact that a number of global drivers are creating impacts on all nations, but the way these drivers will occur, differs widely, leading to specific effects for individual countries; also national, cultural social, and economic barriers vary between the countries and are all leading to practical limitations for the visions and ambitions that are actually possible and obtainable/feasible.

The global drivers are: Demographic Changes, Globalization, Climate Changes, Impact of Network society, Social Changes, Technology. The manifestations of these drivers of course vary enormously, leading to opportunities or challenges that are unique for the individual countries; For example, Global Warming at least in an intermediate future will offer Greenland more opportunities for mining, while the impact on traditional fishing and hunting is negative as the sea ice disappears.

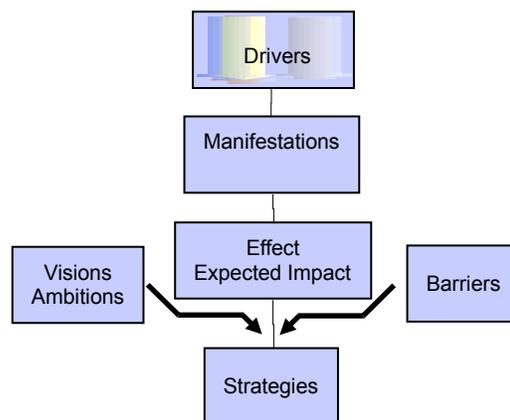


Figure 1. Developing realistic Strategies (Source IBV<sup>1</sup>)

As a result of the initial brainstorming sessions with each of the 5 focus teams, almost 90 vision statements were identified together with more than 100 perceived and real barriers. The barriers were then sorted according to origin and nature, and the structural barriers that could be removed by appropriate actions were boiled down to a few key areas of which the most important are:

1. Telecommunications infrastructure and Internet pricing structure

TeleGreenland operates a radio based infrastructure and only in 2 areas – Nuuk, Qaqortok – high speed WANS connected to Canada and Scotland via high capacity submarine cables. Everywhere pricing of Internet services are based on volume pricing at a very high level compared to other countries, thus reducing traffic and take-up of the use of internet - particularly in the education sector that has limited the access to internet for pupils due to budgetary constraints.

<sup>1</sup> <http://www-935.ibm.com/services/us/gbs/bus/html/government2020-pcm.html?cntxt=a1000055>

2. Lack of ICT competencies, particularly in villages
3. Lack of ICT coverage – 55% of the population in villages has never used a computer compared to 28% of city inhabitants.

### **Main Strategies – Phase 2**

For each of the focus area teams the next phase called for a much more precise definition of strategy and sub-strategies to realize the overall vision and to remove the most important barriers. So in parallel a team discussing the key Tele-infrastructure challenges worked to specify possible actions to overcome the specific Tele-sector barriers – capacity, pricing, competition. For each of the focus areas additional 3 – 5 specific sub-strategies were then defined. All in all 23 sub strategies have been described and prioritized-

#### *Focus Area: Competencies*

The ICT Strategy aims at proving the population with sufficient ICT competencies and to assist in providing a major, general increase in competencies through education and through in-job training offerings.

Sub strategies: Improving ICT competencies in villages, improving ICT competencies for employees as well as language skills, Integrate ICT and improve all levels of youth education, Increase skill levels of teachers

#### *Focus Area: Innovation and Industry*

ICT is a fundamental part of the foundation for development of innovation and culture and will support the promotion of talents to help obtain Greenland's new role as an actor in a global economy.

Sub strategies: Use the public Sector as a partner/customer for new services and products, exploit ICT as a generator of innovation through industry CSR-programs, Partnership, Transparency and ease of establishing new business initiatives and companies in GL. Also emphasis on identifying 'Talent of the Year' and 'Innovative Company of the Year' as well as establishing a promotional, international website to market and sell Greenland products and service.

#### *Focus Area: Health*

The overall strategy is to ensure that GL's effective and efficient health ICT infrastructure secures a high level of information, prevention of diseases as well as individualized treatment.

Sub strategies: Establish a new, country-wide general Electronic Patient Record/clinical workplace system, Improve and expand use of Tele-medicine to offer efficient health services in remote areas, Improve Administrative and Management Health Information Systems to increase quality, and focus on increased quality of living by promoting healthy life styles and understanding of health issues by every citizen.

#### *Focus Area: Citizen Service/Administration*

Establish a cross-agency IT service platform focusing on the need and possibilities for the citizens based on effective handling of knowledge and data to provide a fair and equal administration and establish a Financial management system/management information system to assist political prioritization and manageability of all aspects of the Greenland society.

Sub strategies: A common IT Architecture for the Public Sector including common ICT principles, standards for interfaces, data, and governance. An important sub strategy is to ensure availability of all public information and services in the Greenland language to ensure inclusion and promoting understanding of the citizens role in the Greenland democracy. develop and expand a citizen portal solution to increase level of competencies for civil servants and promoting citizen self services. Implement a suitable ERP financial management system common to all agencies and public institutions in Greenland.

#### *Focus Area: Tele-Infrastructure*

The overall strategy for the tele-sector is to ensure an infrastructure and a pricing model that will support the up-take and optimal use of ICT in the GL society at large and that offers the citizens a fair and equal access to relevant communication services at an affordable price.

Sub strategies: Increase competition where it makes sense and expand service offerings, define future need for internet capacity and related investments, and – most urgently - obtain a common agreement on pricing and price differentiation for sub-sectors (Education, Health) to ensure fast take-up. Evaluate ways to support the educational sector and the health sectors' budget for ICT.

### **3. Action Plans/Key Projects – Phase 3**

A number of possible Key Projects were suggested to meet the strategies defined in phase 2, and the selection criteria were identified as well as the complexity and logical order of projects. Complexity is a very critical factor in a country where skills and competencies are limited and external assistance has to be brought in from outside while ensuring that local participants are trained in the necessary key disciplines. To ensure manageability of the portfolio within each strategy area, each project was split into a logical sequence of smaller project steps, where each step would provide value to the defined set of stakeholders.

#### *Evaluation Methodology – Portfolio Selection*

For each of the defined sub strategies the first step was to identify the value created for the stakeholders by the suggested set of projects be implemented. At this stage the description is qualitative, not quantitatively measured. For each sub strategy a more detailed plan-of-action was developed, typically covering 4-5 projects either in a logical progression or as a result of dependency on other sub strategies/action plans/projects.

As the challenge at this stage is to select a feasible portfolio of projects and de-select projects of low value and high cost and complexity, a scoring methodology was used ranking each project according to the following criteria:

Potential Value: Resources saved + Value to Users

Complexity: Time to deliver + Level of Change

Expense: Investment Level + Operational cost Level

For each criteria - for instance 'Resources Saved' - a score from 1 to 5 was defined, and the combined average yield was then

calculated as an average between Resources Saved and perceived Value to Users. (citizens, companies, students, in some cases employees - teachers, health care personnel).

The financial scoring at this stage is very preliminary, but nevertheless applied to give the participants a change to reflect on the potential costs.

To illustrate how the scoring was used we will look at 2 different sets of strategies/project proposals, one for the citizen services/public administration sub-strategy and one for the Health Care strategy.

*Example: Projects for Citizen Services*

Project	Potent. savings	User value	Duration	Change level
Employee support system and knowledge based hot line support	3	3	1	1
Mobile Banking solution plus 'easy bank account' for all	4	4	1	2,5
Citizen Portal Solution – Citizen self service capabilities	3	4	3	3
Digital Identity & Single Sign on	3	2	3	2,5
Video communication with Citizens and remote Employees	2	2	3	2 .

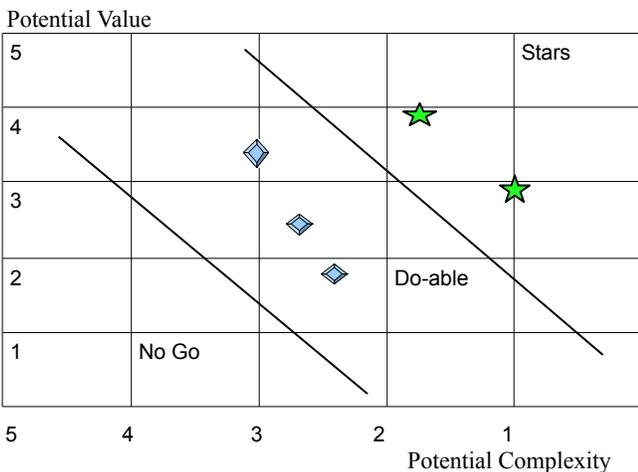


Figure 2: Citizen Service Project Prioritization

The graph illustrates that the 'low hanging fruits' are the first 2 projects, while the other because of complexity and the time to develop solutions are lower rated, yet still feasible and acceptable.

In a subsequent phase where more details were described and an initial business case developed for each of the projects, the costing and the availability of critical resources was added to the evaluation, leading to a detailed timing of the projects to ensure a logical sequencing and moving high prioritized projects forward.

*Example: Health Care Projects*

Project	Potent. savings	User value	Duration	Change level
Asynchronous Tele-medicine	3	3	1	1
Synchronous Video Telemedicine	4	4	1	3
One Electronic Patient Record	5	4	3	3
M.I.S. Based on financial, clinical and administrative data	2	5	4	2,5
Health Portal for Citizens and for professionals – incl. access to EHR	2	4	3	2

For the Health projects the corresponding graph looks like this:

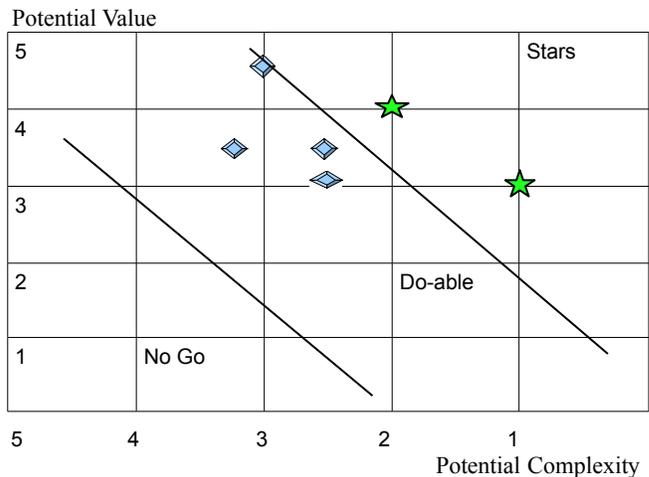


Figure 3: Health Care Project Prioritization

The 2 Telemedicine projects come out on top of the more complicated projects that require a long time to specify, select possible standard solutions and to implement (EHR). This is of no surprise in a country where the costs of sending out doctors and evacuating patients by helicopter are considerable, and because of to-days maturity of solutions and experience from pilot projects, these initiatives will pay off soon.

*Selection of projects - Summary*

The benefit of the methodology applied lies in the interactive discussions within the project teams; If a project turned out to be located in the no go area – too much complexity for value – the idea was then either dropped or refined, and the final result at this stage was the ability to come up with an efficient and feasible portfolio of projects. During the process the dependency between the projects were identified, and the initial timing of the projects could be derived.

An important discussion was also to define the priority of the projects regardless of value and complexity. Some of the projects should be regarded as prerequisites for others, and some projects

turn out in reality to be a process leading to a needed political decision. In both cases the projects were marked with a star.

At this stage the financial indicators were only used as an information, and the purpose is to elaborate on this before each budget year to ensure proper allocation of funds.

This may lead to an annual cycle of decisions, where the outcome of the projects implemented during the past fiscal year could be compared with the requested funding for the on going projects and/or for the funding of new projects in subsequent years.

#### 4. Consolidated ICT Plan – Phase 4

The result of the prioritization step was the first comprehensive total plan for the ICT strategies consisting of high priority decisions that had to be taken, the projects identified as pre-requisites or foundations for all other projects and the logical sequencing of projects over the strategic horizon from 2011 to 2015.

##### *High priority Decisions*

In countries with adverse geography and climate conditions such as Greenland, Telecommunication must play a very important role in keeping the country together. This calls for a massive investment in tele-infrastructure as well as an ordered set of policies for pricing, capacity allocation and also for a certain level of competition, controlled by agreements on level of public service. In the Greenland Strategy paper these factors and decisions is a fundamental prerequisite for all other ICT strategies and projects, so the list of High Priority Decisions contain the following main items:

- Political agreement to guarantee a 'Basic ICT service level' to everybody including settlements and villages
- Political acceptance to offer increased capacity and a higher service level for growth centers
- Ensuring competitive pricing of ICT services by regulation and (some) competition where feasible
- Investment plan for further development of Tele-infrastructure including extension of submarine cable and extension of radio chain for Internet traffic
- Pricing and budget decision to ensure that adequate budget and acceptable set of prices are offered to schools and health care sector

##### *Foundation Projects*

There are two sets of projects that are simply prerequisites for a successful implementation and management of the entire portfolio of projects and for the general acceptance of the sector-oriented sub-strategies:

- Establishing a joint ICT Governance model and develop an overall Enterprise Architecture for Greenland
- Focus on the Greenland language in every single project, set up support structure and systems to assist the project teams in meeting this challenge.

These 'foundation' initiatives have to be accepted by both central and local government, by major educational and cultural institutions and by the major companies in Greenland.

Enterprise Architecture is a journey, not a precise target, but without a set of jointly accepted principles for ICT systems, management principles, standards for acquisitions, there is a real risk that individual organizations and institutions will make uncoordinated decisions and hence erode the basic principles of the strategy: To develop an infrastructure that ties the country together.

##### *'Stars' – Key Projects*

For each of the Focus Areas the 'star'-marked projects are considered as flagships, where priority and a high level of political attention should be allocated. These projects are for instance:

- Support for public servants in remote areas (hot line, knowledge based guidelines etc)
- Development of a Citizen portal with access to all public sector info and self-service capabilities
- Implement and adapt a Public Sector Financial Management/ERP system to ensure manageability
- Develop educational systems for distance learning, teacher training and access to ICT based teachware
- Develop community centers in villages to support remote areas with ICT competencies
- Implement Mobile Cash to reduce need for physical money (social benefits, salaries, etc.)
- Expand the use of NUNA-GIS – information on geology, minerals, natural resources
- Telemedicine including video communication
- Implementing one common Electronic Patient Record and a health portal for citizens
- Demand partnership with local companies for all public tenders
- Establish a company portal to ease visibility for innovators, potential new companies and companies that want want to invest in Greenland.

This list is not complete but illustrates the breadth of the strategy.

##### **The Strategy Acceptance Process**

The final ICT strategy was presented in June 2011 to the Naalakkersuisut – the Greenland Government after initial discussions with the committee for finance. This process led to some changes in the timing of the various projects to ensure that the budget was within the targets for the coming years.

The name of the strategy that was presented to the Government was decided to be:

##### **ATTAVIGISSAARNEQ**

This translates to 'Good Connections' and as such covers the objective to ensure connectivity between all members of the Greenland Society, as well as across the public and private sectors and between agencies and institutions in the Public Sector.

## 5. Perspective – Lessons Learned

Greenland faces major challenges on its road to self determination and a sustainable economy; in a country almost the size of a continent with a small population and adverse geography and climate, the role of telecommunications cannot be underestimated. Where railroads and roads connect the population in other countries, the internet will be an increasingly more important factor to tie the population together, to support e-Democracy and e-Inclusion and not least to provide a platform for collaboration between networks of companies, innovators, employees. Most important is undoubtedly the focus on competencies from K-12 to higher levels of education and for on-the-job training.

The entire process of formulating the strategy in groups representing stakeholders from the public sector, industry and cultural institutions was indeed a first attempt to formulate a strategy that was also based on a strong consensus.

It was tempting to introduce the entire arsenal of strategy selection tools to help the process, but to ensure that all participants were offered opportunities to express and discuss the view points, we had to be very selective in our choice of methodologies.

### Tools – Summary

For the initial phase – the brainstorming phase leading to a first list of visions and barriers – we based the discussions on the methodology originally developed by the IBM Institute for Business Value as shortly illustrated by figure 1.

This proved to be an effective way of discussing firstly how the global drivers are affecting Greenland, secondly a fruitful discussion on which kind of barriers exists and not least how these barriers could be defined either as 'way of nature' and 'temporary' that could be removed or eased by proper tactical strategies.

From the vision phase to strategies we found inspiration in the tools developed by some of the recent EA researchers, particularly Chris Potts, that has developed a very useful set of guidelines for strategy definition including tests to determine if the selected set of sub strategies has the desired balance between organization, operations, culture and technology.

For the definition of possible projects we got the inspiration from the [US OMB exhibit 300 set of methodologies](#) although a number of the more detailed descriptions are still awaiting the political decision and subsequent project initialization process.

The portfolio evaluation phase and the scoring of the individual project to produce an efficient portfolio of projects were supported by the so-called Focus Methodology that was developed by the Danish Ministry for Science and Education. This methodology supports not only the initial phases where projects are defined and scoped, but also – like the OMB tools – the subsequent phases including business case development and ex post reviews.

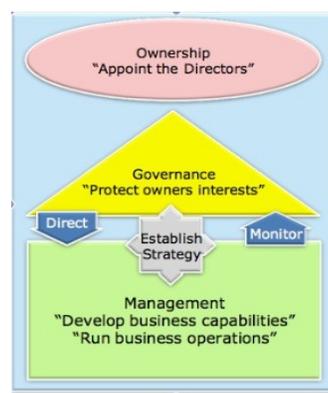


Figure 4 – Governance and Strategy Alignment

The final preparatory work includes suggestions for the organization of an effective Governance Model. For this we used inspiration from Chris Potts (2).

### About the author

Soren Duus Ostergaard was hired by the Greenland Government as a moderator to support the work towards the final strategy. He was a member of the Greenland IT Council from 2006-2009. He has retired from IBM after 40 years, the last 15 a senior eGovernment Advisor for IBM in Europe, Middle East and Africa. He is an external lecturer at the IT University of Copenhagen teaching eGovernment at the Master's level.

### REFERENCES

- [1] IBM Institute for Business Value 2010: Government 2020 and the perpetual collaboration mandate: [http://www-01.ibm.com/industries/government/ieg/pdf/govt\\_2020\\_report.pdf](http://www-01.ibm.com/industries/government/ieg/pdf/govt_2020_report.pdf)
- [2] Chris Potts: Enterprise Architecture is more than IT: <http://coherencyarchitect.com/2010/10/25/enterprise-architecture-is-more-than-it/>
- [3] Executive Office of the President office of Management and Budget 2010: Exhibit 300 planning and budgeting guidelines for acquisition and Management of Public Capital Assets: [http://www.modernisering.dk/fileadmin/user\\_upload/documents/Projekter/Redskaber\\_og\\_vejledninger/Projektmodel/Opdateret\\_projektmodel\\_ver\\_1.0.pdf](http://www.modernisering.dk/fileadmin/user_upload/documents/Projekter/Redskaber_og_vejledninger/Projektmodel/Opdateret_projektmodel_ver_1.0.pdf)
- [4] Danish Ministry of IT & Science 2010: General Model for Public IT projects: [http://www.modernisering.dk/fileadmin/user\\_upload/documents/Projekter/Redskaber\\_og\\_vejledninger/Projektmodel/Opdateret\\_projektmodel\\_ver\\_1.0.pdf](http://www.modernisering.dk/fileadmin/user_upload/documents/Projekter/Redskaber_og_vejledninger/Projektmodel/Opdateret_projektmodel_ver_1.0.pdf)

## **Professional Development of Teacher Educators through Developing their Computer Skills**

**Esther ZARETSKY**

**Giv'at Washington Academic College of Education**

**D.N. Evtah, 79239, Israel**

### **ABSTRACT**

The aim of this research was to study the professional development of college teacher educators by teaching them computer skills. The study covered 25 instructors and lecturers teaching at a teacher training college who participated in a course of computer applications in teaching. Each one of the teachers applied the skills acquired in the course in his classroom in his area of specialization. The data was gathered by means of an open-ended questionnaire and a survey of the teachers' opinions about the connection between the use of technology in teaching and professional development. These teachers reported progress in the courses they taught thanks to the content, the methodology learned in the course, and the personal attention they received from the course instructor. The findings also showed the teachers expressed positive opinions about using computers in teaching. Their responses to the questionnaire stressed the positive contribution the computer had made to communication, knowledge and study communities. One of the more striking effects of teaching these teachers computer skills for their work at the college was the great increase in cooperative group learning in accordance with the suggested strategies of this research.

Keywords: College teachers, computer, integration, online forum, professional development, teaching, technology.

### **1. INTRODUCTION**

Much recent research has been carried out on the connection between integrating technology in teaching and professional development. While most of this research has focused on teacher training, very little of it has examined the professional development of computer teaching skills of college teachers entrusted with teacher training, or the computer or aspects of teleprocessing in areas such as combining knowledge with learning, knowledge communication, and knowledge management, personal professional development, and pedagogical and didactical applications.

### **2. THEORETICAL REVIEW**

In the last few years, instructors and lecturers have been required to use technology in teaching and have also been required to teach online courses in a variety of disciplines both in this country (1) and abroad (2). Computer-assisted teaching makes use of new pedagogical paradigms refocusing a teacher's function to that of guiding and advising his students, while the students are refocused to carrying out research-type activities such as gathering data, selecting relevant data, processing data, etc. (3). The research literature attributes to guides/advisors the management of a multifunctional computerized forum. Mason (4) stresses the importance of a guide/advisor managing the forum in building a virtual learning community and in creating a fruitful discussion in the forum. It is imperative that he react to the participants' questions and that he enter the forum very frequently. This is necessary to stimulate the participants' enthusiasm and increase their motivation to participate in the forum. Consequently, the functions of the online instructor

include the functions of a teacher, conference head, mediator and organizer of a learning community. The community is the apparatus that must be "activated" in order to enable an environment supporting sharing of knowledge and creating new knowledge (5).

The lecturers of Boston University's School of Education and School for the Arts were called upon to participate in a project sponsored by the U.S. secretary of education entitled PT3 (Preparing and Training Tomorrow's Teacher (6)) to develop courses integrating technology in teaching. In the framework of this project, the lecturers participated in a one-hour-a-week course for professional development where they learned strategies to use a tool for building a site for the course, prepared material for the courses they taught, received individual technical support and held round table meetings where each one of them presented the materials he had developed during the professional development course. The lecturers reported on the importance of the project for them from the professional point of view. The findings of research done on this project show that the moment the lecturers became aware that technology helped them in their professional development, not only did their use of technology increase, but they also established a community whose function was to develop computerized and teleprocessed teaching tools (7). In research conducted by Shea, Pickett & Sau Lee (8), the lecturers attributed a lot of importance to the influence of technology on their professional development. Sahin and Thompson (9) stress the necessity of providing individual guidance to lecturers in subjects relevant to them and given to them to apply in their teaching process as much as the necessity of encouraging communication, cooperation and team work through learning communities providing technological and pedagogical support for colleagues in the community. Other projects dealing with training teaching staffs found a connection between the lecturers' progress in teaching as a result of the use of technology and its increased use, as well as their willingness to assimilate it (10).

Goldstein's, Kozaminsky's and Simca's research (11) was intended to examine the components that helped a group of education college teachers create a learning community ("siachnet") and study the characteristics of the online discourse of the group and their contribution to learning. The "siachnet" framework created a common learning culture stressing mutual relationships in constructing relevant knowledge for teacher training. One of the researchers' conclusions was the need to provide individual support for all the participants, both active and passive, in order to train them in guiding learner groups in subjects of their specialization (12).

The virtual meetings can provide a participant with a massive amount of information since the amount of knowledge in the world continues to double itself with increasing speed (13). Gray (14) claims that creating virtual learning communities increases the participants' social involvement, cooperation with the others in the group, motivation and expectations of themselves.

Innovative research checking to what extent teacher-training colleges in Israel have assimilated communication and

information technologies (15) has examined the processes by which teacher educators and student teachers have assimilated these technologies at four of the leading teacher-training colleges in the country. Lasting three years, the research combined qualitative and quantitative methodology. As far as teacher educators are concerned, among the advantages of using teleprocessing in teaching that were revealed by the research, the model that stands out is the interactive teaching model applied through a site accompanying and supplementing the teacher's course. The site offers teleprocessed activities and tasks, the use of varied and up-to-date information sources, development of information control skills, the use of electronic tools and experiencing a new teaching environment as the teacher develops professionally. The advantages mentioned here point to the foci of training that teacher educators need as they introduce technology in teaching.

The contribution of the Internet learning site called "The Virtual Campus" to teacher educators and to teachers in general (16) consists of the following: development of learning strategies including peer teaching, focusing teaching, access to new knowledge products of the in-service trainees, references to relevant information and professional information, access to new knowledge created as a result of discussions in the forums, mindfulness in using knowledge including making knowledge available to all participants, user-friendly access and availability of information, a change in the perception of the functions of learner and teacher, and presentation of all learning proceedings that took place during the in-service training at any given place and point in time.

From the point of view of computerization, in the framework of the evaluation survey of the multimedia presentations integrated in teaching-learning processes in the teacher-training colleges (17), the following items were examined: suggestions made by teacher-counselors (5), by trainees (11) by elementary and middle-school teachers (18) and 10<sup>th</sup> grade pupils (18) in order to get a picture of the evaluation and the concepts connected to the evaluation of the production process and its products. An analysis of the more noteworthy items by the frequency with which they occurred in the teacher educators' and the trainees' answers yielded the following: use of multimedia (teacher educators 90%, trainees 89%), cooperation in learning and production (teacher educators 90%, trainees 89%) and use of special features for programs and management (teacher educators 92%, trainees 84%). The lesser items that were noteworthy for the frequency with which they occurred in the teacher educators' and trainees' answers were as follows: presentations necessitate verbal and non-verbal data processing as well as representational processing in various media, connections and branching out, interactivity, up-to-date, reliable and valid variety (teacher educators: 100%, trainees: 93%). Among the unique characteristics of the multimedia presentations integrated in learning processes we wish to stress the production of the presentation versus its use only, alternatives to presentation of information by means of different media, production through social interaction, receiving immediate results, focus and integration of information, developing awareness of variety among one's target audience and cooperation in the learning discourse by means of the computer.

Generally, the findings show that all the interviewees attributed great importance to producing presentations as part of an on-going and dynamic process requiring integrating the computer, a teacher/counselor and peers. On the other hand, there were differences between pedagogical counselors and teachers, and some of the learners concerning the points relating to the unique characteristics of presentations. The learners paid a lot more attention to the visual aspects of presentations, while the counselors and teachers focused more carefully on the process of handling information. Evaluating presentations, the

counselors looked at things from a different perspective: they put emphasis on different aspects of the subjects learned and stated that the presentations require consideration by the producers and the viewers. They stated that the presentations should represent a variety of levels of thinking, should be up-to-date, should be clear and should be well-organized. The researchers also found that the trainees and the 10<sup>th</sup> graders had a tendency to use terminology taken from the program management and used terms denoting actions focusing on the products, while the counselors and teachers used terms taken from education theories and concepts, focusing on the process. The cooperative setting led the counselors and teachers to discussions about the content and the process involved in their work. Seemingly, the functions of forum coordinators are also needed in guiding teachers preparing multimedia presentations integrated in the teaching-learning process in the framework of professional development.

### 3. METHOD

#### Purpose

The main purpose of this research is to arrive at a universal model intended for teacher educators for the purpose of teaching particular learning contents in various disciplines. The universality of the model consists of the possibility of adapting the teaching program to all teacher educators in all disciplines according to their special characteristics and needs.

#### Goals:

- Presentation of innovations not otherwise available in integrating the computer in the teacher educators' field of expertise.
- Familiarizing educators with new technological tools applicable to teacher training.
- Brainstorming in data processing and turning it into information relevant to teacher educators.

#### Vision

The vision for the future in the domain of professional development is to come up with a universal model that would help actualize in practice technological teaching tools for teacher educators.

The study covered 25 instructors and lecturers teaching at teacher training college who participated in a course of computer applications in teaching. The study took place in MOFET Institute, a school for research and development of teacher training programs located in the center of the country. The areas of specialization of the instructors and lecturers include a wide range of disciplines such as science, mathematics, language, special education, early childhood, music, and so on.

We conducted interviews with and surveyed the opinions of 10 teacher educators to assess their success in teaching. We interviewed teacher educators from our college as well as some who had been taught by the author of this research at the MOFET Institute. In the chapter dealing with the results, we will present selected responses that seemed more relevant and interesting.

### 4. FINDINGS

According to Guskey's model (19), the processes involved in professional development are lengthy, systematic and directed. The effects of professional development on learning in the classroom are felt in the immediate influence they have on the teacher educators' and teachers' knowledge and practical experience, in the interaction taking place during counseling and in rising self-esteem.

The assessment of the results concerning management processes and reaching of goals was carried out through interviews with a

representative sample of participants, who were given an open questionnaire including identical questions; a survey of their positions was conducted, followed by evaluations and a presentation of planning versus practical execution of the planning.

A sample of the teacher educators' answers is quoted here. The sample answers were selected for their potential contribution to professional development research or just because they were interesting:

- **What appeals to you most in this course?**

First and foremost, the respondents stressed the appeal of the lecturer's personality, her approach to the discipline she taught and the way she taught it. As a result, their interest in innovative technologies was sustained, and continued even after the course was over. They adopted some of these innovations in their teaching as well. Concerning the lecturer's personality, here is what two of them had to say:

1a. *The appeal of the course was "the fact that the course coordinator was a 'teacher' and not just a computer person. In the context of her didactical and pedagogical skills, it was clear to her that her aim was to impart me the skills and not 'show me how.' It was also clear to her that I came to the course to become 'computer literate,' and not a technician; that is to say to be a consumer of the advanced technologies that could help me as a lecturer."* (Answer of a respondent who participated in the course 'computer applications in teaching')

1b. *"When I joined the course, I did not know how to use a computer, and I also felt great anxiety. I did not wish to use a computer because I was afraid. So, as far as I am concerned, the most important thing was the coordinator's professionalism with her wise pedagogical approach and great devotion. The fact that she was able to break through the wall of fear I had of the computer stemmed from the winning combination of the right dosage of help to the learner at the right moment side by side with allowing the learner freedom to experiment and make a great number of mistakes, of her ability to allow the learner to work in the domain of his specialization and give him some intelligent advice pertaining to his domain while simultaneously ensuring his progress in computer applications, let alone the close relationship with the learner, which she expressed in remaining with me, and those like me, for an hour, at least, after each lesson. In addition, the group of participants in the course was heterogeneous, which could have caused a lot of friction among the participants and could have worn out the teacher; however, the teacher's willingness to work with each one individually and without time constraints neutralized potentially negative occurrences such as friction and burn-out. Each one of us showed the others his products (at the beginning I did not have very much to show...), which created a thirst for more."*

- In your opinion, what were the course's more successful aspects?

1. Forming new perspectives on pedagogy and on the domain of disciplines taught.

### **Pedagogy**

1a. One respondent answered: *"The fact is that thanks to the course I became (finally...) computer literate."* She added: *"Today, I cannot even imagine myself working without using computer applications."*

1b. *"The course helped me overcome my 'compophobia' and developed the visual aspects of the subject I teach."*

*"Fine-tuning of the material taught in the course to meet the needs of the participants is a very important point in counseling teacher educators and teachers. Learning new content and new strategies arouse enthusiasm and increase motivation to take part in the courses and workshops, especially when the*

*instructors are very busy. In contrast, general subjects are tiring and boring. More emphasis should be put on the subject of teaching the students cognitive skills. The tools acquired in the course help in planning the contents of the syllabi and to plan specific lessons according to the curriculum. The presentation of the material taught in the course in tables and charts helped me organize it in an esthetic, clear and interesting way. In addition, you could tell the students in the course looked satisfied."*

2. The long-term influence of the course on the teacher educators' work

2a. The response of an English teacher:

*"The main benefit I derived from the course is the fact that since studying in the course, I have continued learning computer applications that I need as a teacher. I did not take additional courses except for Freehand and PhotoShop. All the rest of the computer applications I needed I have learned by myself, thanks to the computer application course for teacher educators that 'paved the way,' meaning using the Internet, e-mail and computerized information sites. Today, too, thanks to the computer, I discover additional applications every day."*

2b. *"No other lecturer before this one was able to do it. All those who tried failed. I have developed in the direction of using PowerPoint presentations and e-mail."*

The literature stresses the importance of uninterrupted life-long learning by teachers and lecturers as professionals (20).

3. Reason for surprise

Response of a physical education teacher:

*"The frontal alignment of the presentation page and the fonts used for teaching body movement [were surprisingly effective] although [at first] they seemed contradictory since using computers and body movement are two totally different domains."*

- Which materials from the course did you use in your teaching?

*"One of the most important tools was the teleprocessed forum site attached to the course;"*

The rest of the answer adds further information linked to this teacher's reorganization of ideas, which occurred simultaneously with her change of perspective of the materials that developed during her learning in the course and her subsequent teaching.

*"Not only did connecting my subject to computer applications create new connections between the two domains, but it also changed the way I looked at the learning materials I have used up to now. Most of the year I racked my brain together with the lecturer on how to reorganize ideas and trends in learning that up to now had not been expressed visually."*

*"In the course of the last few years I have also used the materials I produced in the course".*

Another answer stressed the accessibility and the intelligibility of the material: *"The material became organized and became accessible and intelligible by being transferred to the screen."*

- What were, in your opinion, the successful points of your teaching?

1. Accessibility seems to have been the most successful point.

*"The students' access to the material learned, which was taught visually through a screen presentation, was a great success. The students were satisfied with my teaching and with me as their lecturer."*

2. Changing learning into a dynamic and interactive experience  
*"The change from teaching by using mainly the blackboard to computer-aided teaching done dynamically and interactively."*

3. Cooperative activity in the forums to prepare for tests

*"I was successful in using the forum in order to help the students go over the material before the tests. They answered sample questions and shared their answers. In addition, I was able to go over their answers and point out certain misunderstandings of the questions or the material learned in the course. In this joint activity the students helped each other improve and deepen their understanding of the material. The students enjoyed this joint activity, and the lesson was very dynamic."*

The use of the forum for studying for a test "was a novel and unexpected way to study English literature" because "the computer was usually used in English academic writing lessons."

4. Feedback cycle

*"The ability of the students to formulate [possible] questions before test lowered the pressure during the test since they knew how to deal with similar questions. They discussed the test among themselves and realized that the test was not so difficult after all. They cooperated directly and indirectly through the teleprocessed forum."*

5. Development and planning of teaching material

*"Integrating computer applications in my teaching enables me to vary my teaching methods, gives me access to materials available on the Internet, and enables me to prepare efficient and esthetic work sheets. In addition, I can easily find materials I prepared in the past and modify them very easily according to my present needs. Moreover, the possibility to communicate with my students by e-mail enables me to help them prepare assignments, to give them quick feedback and to give them a chance to continuously improve their assigned work."*

6. Varying evaluation methods in the courses

Some lecturers allow students to prepare summary presentations instead of regular summary papers:

- Which one of the difficulties you experienced eventually turned into a 'springboard' for succeeding in the course?

The teacher educators' answers that follow below stress the positive change that occurred in them from an emotional point of view following their participation in the course:

*"I came to the course with no knowledge and a lot of frustration with everything that had to do with computers. I had participated in quite a few computer training courses, but I was unable to apply anything that I had learned. After this course, I was highly motivated, and typed a 70-page paper I had to submit to the university in the framework of my MA studies. The paper included not only the text, but tables, diagrams, flow-charts and pictures."*

An additional respondent wrote: *"Since participating in the course, I have begun to use computerized presentations in my teaching."*

- Describe at least one change that has occurred in your professional development since participation in the course.

### 1. Change in teaching methods

*"First, introducing PowerPoint presentations in my lessons. Second, varying methods of evaluation in my courses."* In varying evaluation methods, too, we see the feedback cycle, which is essentially a repetition of the process of developing materials for evaluating students' work, application and evaluation, and then starting all over again.

### 2. Overcoming mental blocks

*"I am more confident in my capability and my students' capability to use technology in teaching for learning purposes." "I had 'compophobia,' but I have moved on to the 21<sup>st</sup> century confidently. I have prepared presentations for the courses I teach too."*

3. Use of new information banks as a basis for motivating students

*"Before taking the course, I could only dream about computerized information banks. The course imparted me the skills that enabled me to teach myself how to use the various information banks. The capabilities I acquired in the course enabled me to prepare work and exercise sheets that include sources derived from computerized information banks, assess them properly from a didactical point of view, using tables and flow charts."*

The last part of this answer stresses the change in the organization and layout of work sheets.

- Describe a computer-assisted learning situation that was significant to you.

1. *"Preparing for a test using the e-forum and sending class e-mail as an addition to the material learned in class in order to give students advice and clarify the subject learned."*

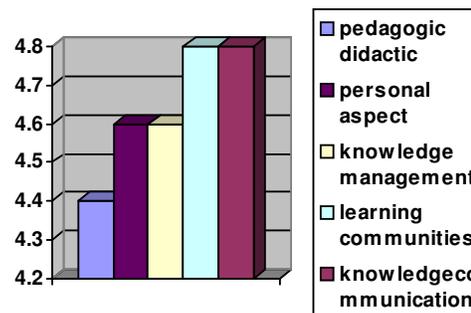
2. The computer as a planning tool and a teaching tool  
*"I will use the computer in almost every lesson both in the lesson planning stage and the actual teaching stage."*

3. Feedback cycle: application by students  
*"The presentations lead the students to ask questions that I have never anticipated."*

The lecturers apply what they have learned in the course in their own teacher training courses, and the students, in turn, apply what they have learned in their practice teaching (student teaching); they then go back to their classes at college and learn new material and new methods, and so the cycle begins all over again.

The teacher educators' direct cooperation in the virtual forum enabled them to compare their respective educational points of view, discuss ways of applying them, and enabled them to organize their materials. They could also comment on their colleagues' materials, add notes and connect them to their own documents or to other sources. (21). Sharing information also engendered social interaction among these teacher educators, which could turn out to be vitally important in dealing with problems in real time in their day to day work.

Diagram No. 1: The grading of the college teachers' positions toward / upon the contribution of professional development



The data displayed by the diagram no. 1 show the contribution of the professional development to knowledge communication and learning communities, then to knowledge management and personal aspect, and later to the pedagogic didactic aspect and its implication to teaching.

## 5. DISCUSSION

The utterance relating to integrating computer technologies in professional development, as far as sharing information is concerned and participating in learning communities, received the highest score (4.8). The second highest score (4.6) went to the utterance concerning the contribution of computerized online activity for professional development as far as information management is concerned, and to the utterance relating to the contribution of the use of computer technology to personal development. The utterances regarding the contribution of the use of the computer being significant in professional development from the didactical and pedagogical point of view, and teaching applications also received a high score (4.4). Seemingly, teacher educators attribute more importance to the domain of their disciplines and data processing processes.

*"The beginning of the management cycle must include a definition of the result to be attained. To manage information without defining the goal is like shooting an arrow without setting a target"* (Trigger, 2009, from a presentation by Avidav, 2009 (22)). In our case, the goal was defined and all our efforts were directed at its attainment: the professional development of teacher educators through the use of the computer.

### The involvement of lecturers

The teacher trainers also cooperated in writing the syllabi for developing teaching staffs. This was done in order to increase their involvement in practical application since a greater involvement of teachers as planners is likely to increase their motivation and commitment to teaching and learning, engendering a higher sense of self-efficacy.

Continuity and support for the professional development process should be anchored in 3-year development programs (23). The first step is to collect a large amount of information and use it efficiently. Professional development is based on allocation of information and transcripts of meetings concerning courses. The material is collected systematically and is continually analyzed and examined.

### Understanding theory

The face to face and virtual meetings created an opportunity for understanding the theory behind the knowledge and skills learned. The findings of the research were accessible for discussion to the teaching staffs.

### Professional development as a part of comprehensive view of the processes

The teachers' commitment to organization led to their willingness to apply the contents in the courses as needed while they evaluated the difficulties and the successes at the level of the (educational) system (24).

A sense of urgency should be created among the course participants. Among other things, subjects like the following should also be discussed: engendering a vision for change, communicating the vision to others, strengthening the vision and clearing obstacles, producing short-term results, continuing with the changes and anchoring the changes in the culture of the organization. The importance of the courses should be made clear to the staff expected to participate in the courses.

### **Emphases:**

- Peer learning through sharing information in face to face meetings and virtual forums, joint planning of interventions and promoting joint discussions.
- Processing information into knowledge.
- Tools for thinking and leading in the field.
- Creative thinking, finding creative solutions (Levy (25), or what Stamatis (26) calls Structure Inventive Thinking (SIT)). Such thinking increases teachers' motivation to

dare find solutions and examine their uniqueness, to define and reformulate the solution, to find ways to save time and resources by uniting, multiplying, subtracting, dividing, reversing, and adding dimensions.

- Dynamic and interactive learning.
- Familiarization with opportunities, flexibility and intentness on the needs of the group, commitment and involvement throughout the whole learning process.

### **The contribution of teacher educators using computer-aided teaching to the field of education and the domain of professional development**

Among the innovations of this research, we wish to emphasize the opening of new perspectives for the lecturers participating in the course. The opening of new perspectives contributed to knowledge management and joining it to the possibility of exposing the teacher educators to knowledge from many other domains and a variety of ideas. In addition, they learned new teaching strategies, course planning and practice in applying them. The importance of looking at material from an innovative point of view is a centerpiece of the constructivist approach to teaching and learning. This approach focuses on the tools in the computerized learning environment in order to help the learners to clarify the various perspectives on the world by forming perspectives of their own (27). In the framework of this approach, an understanding of the learning process becomes more likely when tools and environments for computerized learning are used. Participating in the development of a computerized environment enables educators to increase the use of this technology and may even affect the whole education system. Learning in such an environment should focus on authentic assignments that are relevant to the non-virtual, real world. Actually, the course used assignments taken from the curricula, which had a reasonable degree of difficulty (28).

Teacher training should be done through familiarizing the trainees with innovative computerized and teleprocessed tools, which should be continually updated in order to enable the innovations to penetrate the whole system of teacher training, and from teacher training to teaching in the education system. Communication should also be continually going on, meaning participating in the virtual forums independent of time and place. This is crucial because knowledge and innovations accumulate quickly. New knowledge should be used unhesitatingly, but in accordance with the needs of the environment and in accordance with prevailing circumstances, stressing the involvement of data managers, sharing the work, the products of their work and their willingness to contribute to the process of decision-making in the organization.

In order to heighten the lecturers' awareness of using computers in teaching, the connection between their knowledge of subject matter and their pedagogical and didactical knowledge should be strengthened.

The success of the course is also contingent on adapting its content and methodology to the needs of the various learners, including teacher educators. This adaptation, according to the respondents and researchers that have investigated the subject (8) (15) (29), speeds up and encourages changes in methodology and in adaptation of technology to teaching. Such adaptation promotes the professional development of teacher educators in the pedagogical and didactical domain, and in the implementation of changes in their own classes. Their critical reflection on their work will spur them on to continue to find more and better ways to use technology in the classroom for the benefit of their trainees, who in turn use it in their student-teaching, and so on until we come full circle back to the teacher educators. The more these teachers are able to overcome the

obstacles of 'compuphobia,' the more their motivation to widen their knowledge of technology in teaching increases.

## 6. IMPLICATIONS

### The Implications of Professional Development for College Teaching

1. Increase in academic cooperation according to the strategies suggested in this research.
2. Keeping in touch with colleagues through the peer networks.

Learning was based on past empirical experience (30), translated by Diamant, (16). For past results, see for example Zaretsky (31).

Training lecturers in computer technology should be adapted individually to each member of the staff, starting with courses, individual guidance, material for self-learning, and learning communities that draw into them additional members of the staff. Seemingly, the most important factor is the benefit of using the computer for advancing professional goals (15).

## 7. REFERENCES

- [1] The Government Ministry of Education, **Learning in an Online Environment**. The Department of Elementary Education, The Department of Novel Learning Environments in the Elementary Education, 2004.
- [2] R. M. Epper & A. Bates, **Teaching Faculty How To Use Technology: Best practices from leading institutions** (pp. 115-140). USA: American Council on Education, Oryx Press, 2001.
- [3] N. Law, W.J. Pilgrim & T. Plomp (Eds.), **Pedagogy and ICT use in schools around the world: Findings from the IEA SITES 2006 study**. Hong Kong: CERC-Springer, 2008..
- [4] R. Mason, **Moderating Educational Computer Conferencing**. 1997. URL: <http://star.ucc.nau.edu/~mauri/papers/mason.html>
- [5] Y. Hamish, "Campaign, essay and burden – observation on organizations, management and knowledge". **The Department of Foundations Support Knowledge**, 8.14.2004. . <http://www.notes.co.il/yigal/7170.asp>
- [6] Preparing and training tomorrow's teacher :PT3.
- [7] D. Whittier & S. Lara, "From faculty development to curriculum development: Assessing the Boston University PT3 program for preparing tomorrow's teachers to use technology". In C. Crawford et al. (Eds.), **Proceedings of Society for Information Technology and Teacher Education International Conference, 2004** (pp. 3783-3790). Chesapeake, VA: *AACE*, 2004.
- [8] P. Shea, A. Pickett & Li,C. Saul, "Increasing access to Higher Education: A study of the diffusion of online teaching among 913 college faculty". **IRRODL**, Vol. 6, No. 2, 2005, pp. 3831-1492.
- [9] I. Sahin & A. Thompson, "Analysis of Predictive Factors That Influence Faculty Members' Technology Adoption Level". **Journal of Technology and Teacher Education**, Vol. 15, No. 2, 2007, pp.167-190.
- [10] S. G. Gibson, M. L. Harris & S. M. Colaric, "Technology Acceptance in an Academic Context: Faculty Acceptance of Online Education". **Journal of Education for Business**, Vol. 83, No. 6, 2008, pp.355-359.
- [11] L. Kozminsky, O. Goldstein & M. Simka, "Computerized learning environments: The contribution of online conversation to developing learning community". **literacy: Research, Study and Action**, Vol. 7-8, 2004, pp. 103-126.
- [12] P. Senge, **The Fifth Discipline: The Art and Practice of the Learning organization**. New York: Doubleday, 1990.
- [13] C. Mann, "New Technologies and Gifted Education". **Roeper Review**, Vol.16, (1994, pp. 172-176.
- [14] T. Gray, "Teacher professional development romancing the virtual stone. On- Line". **Journal of the Massachusetts Computer Using Educators**, Vol. 10, No. 2, Spring, 2000.
- [15] O. Goldstein, M. Sheinfeld & A. Forkush-Baruch, **Information and communication Technologies in Training Teachers in Israel**. Haifa: Meita'l., 2008.
- [16] I. Diamant, **so what we actually evaluate here? Presentation meeting**, Mofe't, Tel-Aviv, 2009.
- [17] I. Barkaey, G. Alexander & E. Kozminsky, "Evaluating multimedia presentations produced by learners: evaluation conceptions of pupils, teachers, moderators and in-service teachers". **Script-literacy: Research, Study and Action**, Vol. 7-8, 2004, pp. 127-150.
- [18] O. Goldstein, L. Kozminsky & M.Simka, "'Yahadnet' - forum of online dialog among college teachers". In: R., Zozovsky, T. Ariav, & A. Keinan, (Eds.), **Preparation and professional development of . teachers: ideas exchanges**. Mofe't Institute, Tel-Aviv, 2001.
- [19] T. Guskey, **Evaluating Professional Development**. London: Sage, 2000.
- [20] D. Kfir, S. Avador, R. Reingold & M. Hishrik, "Primary preparation and professional developing of teachers – a sequential and continuing process". **Paths of Research**, Vol. 14, 2006, pp. 46-49.
- [21] T. Hampel & R. Keil-Slawik, "Experience With Teaching and Learning in Cooperative Knowledge Areas", **Proceedings of The Twelfth International World Wide Web Conference** , 20-24 May 2003, Budapest, Hungary.
- [22] O. Avidav, **Designing the Role of the Professional Development' Coordinator in Colleges of Education**. Presentation of final meeting .Mofe't Institute, Tel Aviv, 2009.
- [23] H.L.B. Hodges, "Using research to inform practice in urban school: 10 key strategies for success". **Educational Policy**, Vol. 10, No.2, 1996, pp. 223-252.
- [24] National Research Council, **How People Learn: Brain, Mind and Experience and School**. Washington, DC: National Academy Press, 1999.
- [25] G. Levy, **Creative Thinking**, Presentation Meeting, Mofe't, Tel Aviv, 2009
- [26] D.H. Stamatis, **Advanced Quality Planning: A Commonsense Guide to AQP and APQP**. Structure Investive Thinking (SIT). New York : Quality Resources, 1998.
- [27] D.H. Jonassen, **Objectivism vs. Constructivism: Do We Need a New Philosophical Paradigm?** University of Colorado in Boulder (unpublished), 1990.
- [28] D.H. Jonassen, Evaluating constructivistic learning. **Educational Technology**, Vol. 31, 1991, pp. 28-33.
- [29] N. Hativa, "The university teaching, alternative evaluation and models for its improvement". In: A. Yogev, R, Shapira & D. Chen, (Eds.), **The Education toward the 21<sup>st</sup> century**. (pp. 447-459). Ramot, Tel-Aviv, 2001.
- [30] J. Mayne, "Challenges and lessons in implementing results based management". **Evaluation**, Vol. 13, p. 87, The Tavistock Institute, 2007.
- [31] E. Zaretsky, " Determining norms (standards) of teachers licensing and grading, and proposing an alternative set of norms". **Giv'a Research**, 2006, pp. 123-164.

## **On Specialized Third-party Shipbuilding Logistics Companies In Shanghai**

Youfang HUANG  
Shanghai Maritime University  
Shanghai, China

and

Siqin YU  
Shanghai Maritime University  
Shanghai, China

and

Shunquan HUANG  
Shanghai Maritime University  
Shanghai, China

**Abstract:** Shipbuilding logistics service providers are mainly shipbuilding companies (self-operating logistics) and third-party logistics (logistics outsourcing). Dongchuan Logistics Co., Ltd, a subsidiary to Hudong-Zhonghua Shipbuilding (Group) Co., Ltd, is the only specialized shipbuilding logistics company in Shanghai. With shipbuilding company as the operator, Dongchuan falls into the category of self-operating logistics. When it comes to logistics management in other shipbuilding companies, they are mainly run by the company's logistics department in cooperation with other outside non-specialized logistics companies. On the one hand, shipbuilders have realized that logistics costs can be reduced by cooperating with third-party logistics companies, which in turn boosts logistics outsourcing. Nevertheless, ordinary third-party logistics companies are not capable of offering highly efficient service. In consequence, this kind of specialized logistics service can't keep up with the market demand, hindering the development of shipbuilding logistics and shipbuilding industry in Shanghai as well. This paper is mainly concerned what kind of logistics shall shipbuilding companies, as the core of supply chain for shipbuilding, choose to meet its demand, improving production efficiency and competitiveness. And then the paper puts forward some suggestions and countermeasures on the development of

specialized third-party shipbuilding logistics enterprises in Shanghai.

**Key words:** Shipbuilding Logistics; Third-party Logistics; Logistics Organizational Forms; logistics outsourcing; Shanghai

### **1. INTRODUCTION TO SHIPBUILDING LOGISTICS**

#### **1.1 Definition of Shipbuilding Logistics**

Shipbuilding logistics refers to all the logistics taking place during the process of shipbuilding, including the circulation, planning, implementation and management of materials, including raw materials, equipment, etc, warehousing, and information concerned. It is aimed at meeting demands from shipbuilding companies.

#### **1.2 Characteristics of Shipbuilding Logistics**

**1.2.1 Complex demands on materials.** Shipbuilding logistics involves complex demands on materials, featuring great variety but small quantity, For example, the 8,530-TEU super large container ship independently developed by Hudong-Zhonghua Shipbuilding (Group) Co., Ltd involves 2,000 kinds of sheets, 1-2 sheets for each type weighing 10-20 tons

apiece. However, building a 5,000-TEU container ship only requires 300-500 kinds of sheets.

**1.2.2 Different standards for supporting products.** Lots of supporting parts have to be imported, though reasons are different, including technological superiority overseas or ship-owners' requirements, resulting in the co-existence of different standards, i.e. American standards, European standards, international standards and Chinese standards. Since there are different standards for the same parts, those parts become different materials, falling into different categories and requiring different management.

**1.2.3 Duality in logistics demands.** Vessels, usually built as the owners require, are either built as one of a series of ships or just a unique one. Likewise, shipbuilding companies can fall into two categories: one mainly building series ships and one mainly building individual ships which differ from each other, causing different requirements on logistics, that is, duality in logistics demands.

When it comes to series ship manufacturers, we can take Shanghai Waigaoqiao Shipbuilding Co., Ltd as an example. The company only builds three types of vessels, featuring small variety and large quantity, which only requires simple logistics. However, so far as individual ship manufacturers are concerned, it is the other way round. All the ships have to be equipped with different facilities as a result of different demands from different owners, making the quantity and variety of materials much more complex. Thus flow process is impossible, making logistics management a challenging job.

**1.2.4 Difficulty in cooperation on logistics.** As shipbuilding companies and suppliers differ in production techniques, inconsistency occurs between supplies and demands, making it difficult for companies on the supply chain to cooperate on logistics. As far as steel sheets purchasing is concerned, steel companies prefer to accept orders that feature few category but huge quantity, for producing different steel sheets requires different moulds and changing moulds frequently will undoubtedly lowers production efficiency. In contrast, shipbuilding companies' orders feature small amount but great variety as a result of different configuration for different ships. To produce on a large scale, steel companies may sometimes choose not to produce until other similar orders are placed or even refuse to take the order to produce just a piece of steel sheet. As a result, shipbuilding companies' supply of raw materials is met with great uncertainty.

**1.2.5 Large investment in fixed assets.** Shipbuilding

logistics often involves transportation of large-size parts, such as steel sheets, ship-oriented pumps and generator sets, which demands special transportation facilities. Due to great variety in the types of materials, late shipment by suppliers or work suspension to wait for materials, it is highly necessary that shipbuilding logistics companies own large-floor area warehouses to ensure that works can go on continuously. Besides, a certain plants are required to do pre-treatment works, such as painting and removing oxidation scale from steel sheets. As a result, more investments in fixed assets are needed to guarantee above-mentioned facilities, making it harder for third party shipbuilding logistics companies to enter this sector.

## **2. ANALYSIS OF SHIPBUILDING LOGISTICS ORGANIZATIONAL FORMS**

### **2.1 Development Trend**

Logistics outsourcing is the result of division of labor and specialization, which also increases trading costs as division intensifies, hindering division consequently. Though division of labor and specialization increase efficiency for third party logistics companies and shipbuilding companies, they also cause higher cost for trading. The two combined hinders the development for logistics outsourcing in shipbuilding companies.

The upper limit for the rise in trading costs allowed by shipbuilding companies is benefits the company can obtain from logistics division and specialized third party logistics. In other words, shipbuilding companies will choose self-operating logistics instead of logistics outsourcing if increases in trading costs exceed the profits the company makes from logistics outsourcing.

In terms of shipbuilding logistics, third party logistics does not have advantages over self-operating logistics in technology and other factors (exogenous comparative advantage), which makes it impossible for third-party logistics companies outside the industry to appear as specialized shipbuilding logistics enterprises in the short terms. That is to say, specialized third-party shipbuilding logistics companies are the result of the development of non-specialized third-party logistics companies. The development lies in the endogenous comparative advantage that third-party logistics companies possess. Endogenous comparative advantage, which develops along with the division of labor, is the driving force of accumulation of knowledge and improvement of efficiency. Shipbuilding companies' demands on logistics outsourcing will boost third party logistics, resulting in

knowledge accumulation and endogenous development of the efficiency of logistics operation.

Presently, ships are either built as one of a series or individually as a unique one. As for the former, ships can be built on a large scale, creating relatively stable demand on materials in terms of both quantity and variety. Since scale economy is possible, it is fitting for third party logistics companies to take part in. Then the process of the forming of specialized third-party shipbuilding logistics companies can be described as follows: taking advantage of economy of specialization and economy of scale, third party logistics company shall get involved in logistics business in series ship manufacturers first and become specialized third-party logistics company for series ship builders after knowledge accumulation and increase of efficiency. After that, they can involve themselves in logistics services for individual ship manufacturers, becoming specialized third-party shipbuilding logistics company.

In the final analysis, it is quite probable for specialized third-party logistics company to come into being and develop further. Their existence results from demands on logistics outsourcing from shipbuilding enterprises. So far as shipbuilding industry is concerned, whether logistics outsourcing has advantages over self-operating logistics becomes a key issue in the fostering of third-party shipbuilding logistics companies.

**2.2 Inevitability of the Development of Specialized Third-party Shipbuilding Logistics Companies**

Analysis shows that third-party logistics companies will take part in logistics services in series ship manufacturers during the course of the emerging of specialized third-party shipbuilding logistics enterprises. The following section will divide the course of third-party logistics companies evolving into specialized third-party logistics providers for series ship manufacturers into two stages: the early stage and the late stage. Through illustrating different advantages third-party logistics has in terms of cost-benefit during different stages, we can be convinced that the trend is inevitable.

**2.2.1 Early stage.** At the early stage when third party logistics evolves into specialized third party logistics for series ship manufacturers, it features the fact that cost of running self-operating logistics is higher than that on logistics outsourcing ( $C_{AC} > C_{TC}$ ). Thanks to its distinct advantages, third party logistics grow apace at the stage.

Provided that shipbuilding company’s marginal cost is MC, we can use  $MC_1$  to refer to the company’s marginal cost under

the condition of self-operating logistics and  $MC_2$  under the condition of logistics outsourcing.

As third-party logistics has advantages in economy of specialization and economy of scale, cost can be reduced to some extent for shipbuilding companies. Meanwhile, since the supply chain for shipbuilding involves different nations, many parts and fittings have to be imported, though the reasons are diverse. Undoubtedly, costs for customs declaration can be greatly reduced if shipbuilding companies outsource all businesses that involve customs declaration to third-party logistics service providers. On the other hand, when outsourcing logistics business, shipbuilding companies will be able to pay more attention to their core business, making it possible to improve efficiency and lower production cost. With the two factors functioning together, shipbuilding company’s total marginal cost under self-operating logistics will be higher than that under logistics outsourcing, namely,  $MC_1 > MC_2$ .

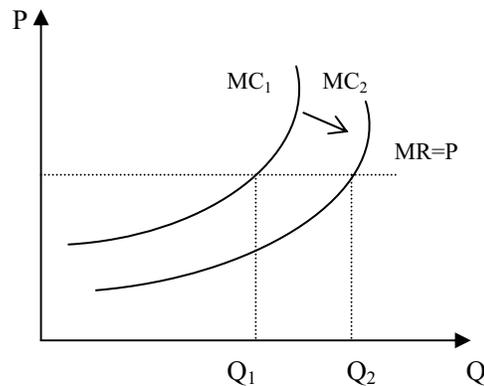


Figure 1. Equilibrium production under self-operating logistics and third-party logistics

As is shown in figure 1, shipbuilding company’s equilibrium production is  $Q_1$  when it choose self-operating logistics but it will increase to  $Q_2$  when the company choose third-party logistics since under this condition curve  $MC_1$  will shift to curve  $MC_2$  while marginal revenue MR (equal to price P) remain unchanged in the short term. Then there are two production strategies for shipbuilding company to choose: the strategy of profit maximization, namely keeping production at  $Q_1$ ; and the strategy of market share maximization, namely increasing production. As for the former strategy, the company can obtain super-normal profit since  $MR > MC$ . When it comes to the second strategy, the company can increase its market share for they can produce at the level  $Q_2$ .

In the final analysis, we can see that series ship

manufacturer can reduce cost, achieve different goals with different strategies when choosing third-party logistics.

**2.2.2 Late stage.** During the late stage when third-party logistics evolve into specialized third-party logistics for series ship manufacturer, the cost will increase as logistics outsourcing increases. As a result, costs for logistics outsourcing exceed those for self-operating outsourcing, i.e.  $C_{TC} > C_{AC}$ , losing almost all the advantage on costs enjoyed at the early stage. It is this stage that is essential to the transformation from third party logistics to specialized third-party logistics for series ship manufacturers.

Since production and logistics volume can form a linear function, we can see production as logistics volume in analysis. Provided that shipbuilding company has a benefit of  $B_{TC}$  (excluding logistics outsourcing cost  $C_{TC}$ ) when choosing third party logistics and a benefit of  $B_{AC}$  when choosing self-operating logistics. At this stage, logistics outsourcing may influence production in two different ways:

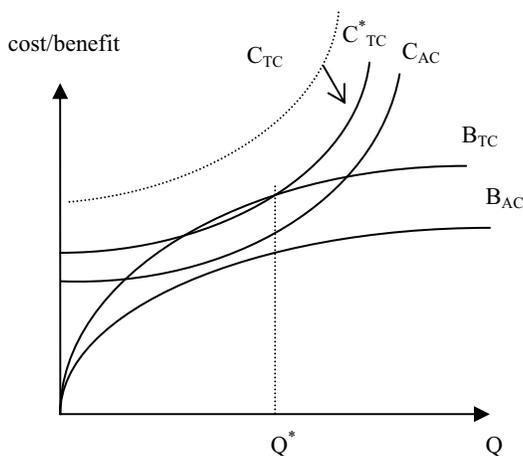


Figure 2. Influence I on production from logistics production

As is shown in figure 2, shipbuilding company may choose to purchase logistics outsourcing, i.e. purchasing highly specialized logistics service from outside, to increase benefits ( $B_{TC} > B_{AC}$ ) since company's production may be impossible under self-operating logistics ( $C_{AC} > B_{AC}$ ). However, at this stage logistics outsourcing is also impossible because the cost is extremely high. Production can be possible only when the costs are reduced, achieving a production of  $Q^*$  by shifting  $C_{TC}$  downward to  $C^*_{TC}$ , intersecting with  $B_{TC}$ . Nevertheless, to reduce cost, shipbuilding company has to cooperate with third party logistics companies by building up long term cooperative relationship and having trust towards each other or increase efficiency by adopting more efficient management methods and tools.

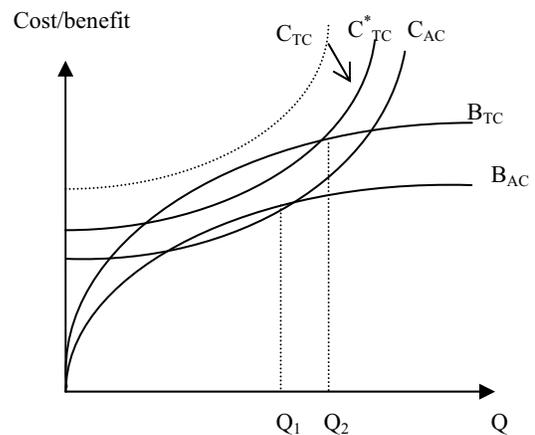


Figure 3. Influence II on production from logistics outsourcing

As is shown in figure 3, curve  $B_{AC}$  intersects with  $C_{AC}$  when the company choose self-operating logistics, achieving a production volume of  $Q_1$ . When it comes to the late stage, production may not be possible or production volume is less than  $Q_1$  because of too high cost ( $C_{TC} > B_{TC}$ ). Then it is more profitable for the company to choose self-operating logistics rather than third-party logistics. However, if the shipbuilding company has a long term or predicted long term cooperative relationship and trust with third-party logistics companies and come up with more efficient management methods and tools to increase trading efficiency, reducing cost, curve  $C_{TC}$  will shift downward to  $C^*_{TC}$  and intersect with curve  $B_{TC}$ , obtaining production volume of  $Q_2$  ( $Q_2 > Q_1$ ). In other words, third party logistics is not only possible but also can improve production.

As the analysis says, third-party logistics companies have to lower trading costs to further growth in series shipbuilding logistics sector at the late stage and the lower the cost, the higher production the shipbuilding companies can enjoy.

There are two methods than third-party logistics firms can use to reduce trading costs, building up long term relationship with shipbuilding companies and having trust with each other due to (predicted) partnership; increasing trading efficiency by taking more efficient management methods and tools. It goes without saying, keeping long term relationship with shipbuilding companies will help third party logistics firms gain endogenous comparative advantage, promoting knowledge accumulation and enhancing operation efficiency, which makes it possible to take more efficient management methods and tools to increase efficiency. However, the two methods are basically the same in essence, that is, helping third party logistics firms become specialized third-party logistics enterprises for series ship

manufacturers.

As we can see from analysis made above, the emergence of specialized third-party series ship building enterprises benefits not only shipbuilding companies but also third-party logistics firms. And it is inevitable so long as third party logistics enterprises keep getting involved in series ship manufacturers' logistics business.

### **3. SUGGESTIONS AND COUNTERMEASURES ON THE DEVELOPMENT OF SPECIALIZED THIRD-PARTY SHIPBUILDING LOGISTICS ENTERPRISES IN SHANGHAI**

Presently, there is no specialized third-party shipbuilding logistics company in Shanghai, resulting from high entry barrier caused by the fact that shipbuilding logistics are highly specialized and requires huge sums of investments on fixed assets. At this stage, preferential policies shall be made to encourage and guide the development of shipbuilding logistics, improving logistics service, network construction, and influence and abolishing external barriers. Besides, relevant laws and regulations shall be made to regulate third-party logistics market, creating a good, healthy development room for third party logistics.

Analysis shows that shipbuilding logistics service providers will develop in three stages and finally appear as specialized third-party shipbuilding logistics enterprise. Though it is the trend, it will take a long period to complete the evolution if only market force is resorted, making it difficult to meet market demand on specialized logistics from shipbuilding industry in Shanghai. To help develop Changxing Island Shipbuilding Center and seize present opportunity, we should make great efforts to foster and catalyze the birth of specialized third party shipbuilding logistics enterprise.

According to the evolution process, the author thinks shipbuilding logistics service provider can be fostered in three phases as follows:

#### **3.1 Encouraging Third Party Logistics Service Provider to Get Actively Involved in Shipbuilding Company's Logistics Services**

China State Shipbuilding Corporation can establish a logistics company based on the Supplies Department of the group at headquarters, integrating all the resources in logistics department of the group and all other supplies and warehousing

departments, transportation departments and logistics firms from shipbuilding companies. The company shall be responsible for the management of the purchasing of shipbuilding equipment, materials and logistics business for the group. Meanwhile, third party logistics companies can be invited to take part in the business, responsible for specific transportation. Then social resources can be utilized to reduce logistics costs and help manage conveyance fleet, maintaining balance between supply and demand and increasing transportation efficiency.

#### **3.2 Building up New Style Cooperative Relationship with Third-party Logistics Companies**

After stable and efficient cooperative relationship between series ship manufacturers and third party logistics companies was established at the first stage, series ship manufacturers can further their cooperation with third party logistics firms by holding shares or leasing their current warehouses, transportation equipment and other logistics facilities. Enjoying convenience brought by external logistics services, shipbuilding companies can also help third party logistics firms alleviate pressure from huge investments on fixed assets. Third party logistics can improve their logistics services rapidly within short period and enlarge their service scale and scope by making full use of their own logistics components.

#### **3.3 Establishing Specialized Third-party Shipbuilding Logistics Enterprise**

After the first two stages during which third party logistics firms finished knowledge accumulation and improvement of logistics service efficiency, these companies can become specialized third party logistics companies for series ship manufacturers, gradually replacing the logistics companies established by China State Shipbuilding Corporation during the first phase and finally emerging as specialized third party shipbuilding logistics enterprises.

Specialized third party shipbuilding logistics enterprises have the following three characteristics: firstly, they fully involve in shipbuilding companies' production process and become an organic component of the process; secondly, they have advantages on purchasing and possess a purchasing network that covers almost all the materials shipbuilding company requires; thirdly, they are able to offer fast and powerful warehousing and transportation services, meeting shipbuilding company's demands of timely distribution and other value-added services.

#### **4. ACKNOWLEDGMENT**

This research was supported by science & technology program of Shanghai Maritime University No.20110078.

#### **REFERENCES**

- [1]Dietrich, M. (1999) Economics of Transaction Cost, Economic Science Press, Beijing.
- [2]Richard W. Bolton(2001), Enabling shipbuilding supply chain virtual enterprises, Journal of Ship Production, 17(2).
- [3]Yemisi A . Bolumole(2001), The Supply Chain Role of Third-Party Logistics Providers, The International Journal of logistics Management, 12(2).
- [4]L.Yanpin(2005) ,Theory Analysis of Logistics Outsourcing Transaction Cost, Business Economics and Management,9.
- [5]S.Bin(2006), Logistics Industry Economics, Science Press, Beijing.
- [6]Y.Qiankuan(2007), On the Development Trend of Modern Ship Logistics, Ship Supplies and Market,3.

## Design of customized service using data mining - Electronic coupons

Chung-Wei Yeh

Department of Management Information System

Kao Yuan University

Kaohsiung, Taiwan 821, ROC

T90026@cc.kyu.edu.tw

and

Kee-Rong Wu

Department of Marine Engineering

National Kaohsiung Marine University

Kaohsiung, Taiwan 811, ROC

krwu@mail.nkmu.edu.tw

*Abstract* – This study presents a customized service approach by using data mining technology to provide the decision makers for profit-marketing policy for a local chained mid-size supermarket. The proposed approach mines all customer transaction data to extract the top 20% of the profit-making customers and then categorize the rest customers into  $n$  consumption clusters. Programs then generate bi-weekly customized electronic coupons (e-coupon) based on their consumption preferences for each cluster customer. Clustering technology and association rule are used herein in connection with relational database. Moreover, algorithms are developed to perform extraction of the valuable customers and implementation of the results of rules to generate e-coupons. Hence, the implementation of our approach will enable the local supermarket decision makers to determine promotion strategy to gain profit by analyzing transaction data regularly. This work provides the solutions via four steps: (1) database design and data transfer, (2) customer clustering, (3) rule mining and (4) customized e-coupon generation. A comparison is conducted to test the effectiveness and shows that there is an improvement obtained from the proposed approach against the conventional one.

**Keywords** - data mining, marketing promotion plan, customerized e-coupon, association rules, cluster technology

### 1. INTRODUCTION

Incorporating information technology and marketing strategy to solidify the relationship between the enterprises and their customers has become one of the top priority tasks for the decision makers in recent years. Due to the growth of information technology, enterprises are easily capable of collecting instant huge data. To convert meaningless data into useful information and in-depth knowledge is a critical challenge to the business of information technology. Data mining technology has been widely utilized to predict customer behaviors for the past few years. The technology

contains all activities to use the collected data to discover implicit information and analyze the historical records to gain valuable knowledge. Fayyad *et al.* [1] have defined data mining as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Meanwhile, John [2] has defined it as the process of discovering advantageous patterns in data. Data mining technology routinely searches for patterns presented in a particular representation set, including classification rules, decision trees, clustering, regression and sequence modeling. In the application of business application, new pattern discovery can be implemented to make prediction of customer's individual purchase preference. The presentation of promotional product information to customers via web has become increasingly important for many companies. Thus, research on the mining data to generate customer-oriented topics has been developed [3-4]. Marketing strategy aims to gain the most profits for a business. One of the key strategies includes maintaining a long-term relationship with customers and building their loyalty to an enterprise. It could be achieved by providing customer services and further discovering customers' purchase patterns and predicting their change. All these information is obtained by the result of marketing survey and analyzed by statistic software/experts in the past. Since now, data mining technology has been used; such as Giudici *et al.* [5] use statistic association and graphic models to study customer behaviors and discuss model selection procedures.

Though, data mining technology is mostly applied in large-scale enterprises in Taiwan. Many local mid-size organizations still run their businesses by traditional promotion strategy. This study aims to improve the promotion strategy design for a local chained mid-size supermarket by consulting their manager and utilizing the mining technology. We analyzed customer consumption patterns, developed programs and used data mining techniques to help manager discover consumption preference patterns and then generate e-coupons to improve marketing

sales. Cluster and association rule algorithms in Microsoft SQL Server 2005 were used in this research. Customers and transaction data were retrieved between Year 2008 and 2010. This paper is organized as follows. The background of the empirical case is discussed in Section 2. Methodology is given in Section 3. Section 4 illustrates the analyzing result. Conclusions are discussed in Section 5.

## 2. EMPIRICAL CASE – TUNCHI SUPERMARKET

### 2.1 Profile of TUNCHI Supermarket

TUNCHI Supermarket was established in Kaohsiung County in 1990. It has three branches located in Kon-Sin, Lu-Chi and Chi-Toe Shiung, suburbs of Kaohsiung City. TUNCHI mainly sells thousand items of groceries, fresh foods, fruits, drinks and fresh breads. Its enterprise size is very similar to Wegmans in the Northeastern USA. Prior to the grand opening of the worldwide Carrefour supermarket in TUNCHI’s territory three years ago (2008), TUNCHI was the leading provider of groceries and foods in its sale territory. In years 2008 and 2009, TUNCHI decreased 25% and 10% of total sales, respectively, because of the newly strong competitor, Carrefour. In order to increase TUNCHI’s total sales, a promotion strategy “e-coupon” was proposed (discussed in Section 2.2). Based on their membership database, the distributions of monthly household incomes (NT dollars) and education levels associated with their ages are organized as listed in Tables 1 and 2. Besides, the distributions of customer gender and daily e-mail user are as follows: 55% female, 45% male; 75% yes, 25% no, separately.

Table 1: Monthly household incomes

income (NT dollars)	percentage (%)
< 20,000	15
20,001~30,000	35
30,001~40,000	30
> 40,000	20

Table 2: Distributions of education levels and ages

Education level age range	grad – school (%)	college (%)	high school (%)	element. school (%)	Total (%)
< 20	0	2	15	10	27
21~35	1	10	5	0	16
36~50	2	15	17	1	35
> 50	0	4	13	5	22
Total (%)	3	31	50	16	100

### 2.2 Strategy

At the request from the manager of TUNCHI, the short-term marketing strategy is to focus on stopping the decrease in the total sales. In the long term, on the other hand, TUNCHI may consider to construct its website for cost consideration. Based on the requirement, we recommend

TUNCHI to adopt the Pareto principle (also known as the 80-20 rule, the law of the vital few, and the principle of factor sparsity) [6]. It is a common rule of thumb; for many events, roughly 80% of the effects come from 20% of the causes. In our case, 80% of the business sales come from 20% of the clients. Thus, the proposed approach provides two functions: (1) search of the 20% most profit-making customers and cluster the rest 80% customers to several groups of similar patterns; (2) design of the promotion plan for TUNCHI. The current advertisement approach of TUNCHI is mailing 16-page flyers to their members and non-members monthly. The proposed approach is designed to find out the 20% top-sale customers (we call them VIP) as a group and then cluster VIP to sub-clusters by mining their data to generate customized e-coupons for each sub-cluster members. The other 80% customers (Non-VIP) also receive their customized e-coupons by mining their purchasing behaviors.

## 3. METHODOLOGY

This study analyzed customer purchase experience data to extract knowledge for proposing promotion strategy. The methodological framework for this research could be divided into four essential phases: (1) database design and data transfer; (2) customer clustering; (3) rule mining; (4) e-coupon generating. Figure 1 depicts the framework of our approach. We redesigned the database structure, generate program to select VIP and mine rules for TUNCHI. Procedures to extract VIP customers were designed according to the cluster and association rules analysis. The executing steps are listed followed.

Step 1: Database design and data transfer. Customer basic data, transaction data and product data are imported into SQL server database.

Step 2: Customer clustering. We designed procedures to extract the members of VIP and then clustered them as sub-clusters (named VIP-A, VIP-B,...) to mine the possible customer’s purchasing intention. The other Non-VIP customers were then grouped into different clusters and named class 1, class 2, ..., and class n.

Step 3: Rule mining. Association rule in SQL Server 2005 analysis service was used to mine rules of specific purchase interest.

Step 4: Generation of the customized e-coupons. According to the content of rules, consulted with business consideration, programs were designed to generate the customized e-coupons for each cluster.

The four steps are discussed in the following sections.

### 3.1 Database design and data transfer

The collected data shows 267,204 transaction records which belongs to 6,223 customers within the transaction period between 1/1/2008 and 12/31/2010. The total sales of membership customers count for 75%, which indicates that

the collected data can highly represent the customers of TUNCHI. Data collection was based on Relational database structure applied in the system. The original data of TUNCHI were stored in Access 2007. Five main tables (Customer, Trans\_detail, Trans\_date, Product, Product\_price) were redesigned to mine the purchase behaviors of customers. This logical design was physically mapped into tables with relationship shown in Fig. 2. We established relational database on Access 2007 and transferred them into SQL 2005 Server within an ODBC environment to implement analysis service. The functions and attributes of tables are explained as follows.

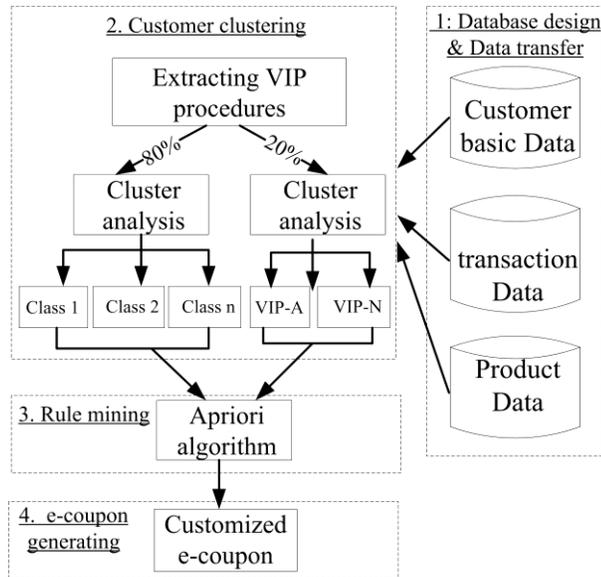


Fig. 1 System frame

**Customer:** It maintains customer’s basic data and includes Id, name, gender, postal and e-mail addresses, birthday, academic degree, professional career, monthly income (NT dollars), VIP member and household number for each member customer.

**Product:** It maintains product’s data and includes Id, name, brand, the belonging category (such as food, drink,...etc.) and update\_period of the products. The field update\_period is used to determine the loyalty of the customer to the specific product.

**Trans\_detail:** It records the transaction details and includes customer Id, product Id, transaction date and time and each sold quantity for any specific product.

**Product\_price:** It shows the sale price of a product. Since the sale price for every product varied depends on the wholesale price at the transaction time period, this table contains product Id and transaction date, sale price and whole sale price. The profit of one product comes from the difference between sale price and wholesale price.

**Category:** It contains category Id and name.

### 3.2 Customer Clustering

In this phase, a series of algorithms are designed to extract VIP cluster customers. These algorithms consider attributes in database tables as key attributes, such as household number, income, purchasing product category, professional career and education of each member. An attribute named “profit” is calculated from Product\_price table (discussed in Section 3.1) to calculate monthly profit. The following algorithm is the main program to output the VIP and Non-VIP members to class 1 and class 2.

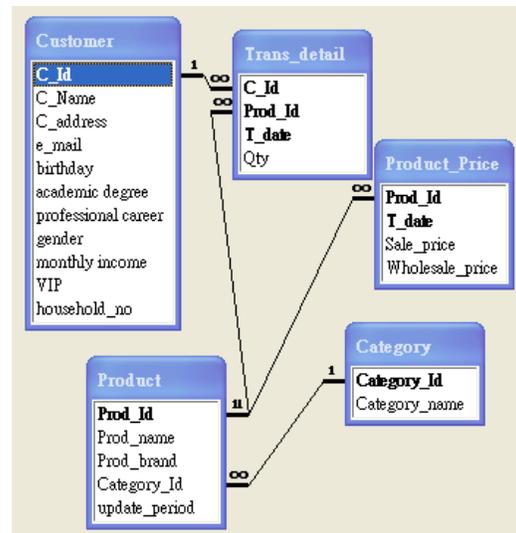


Fig. 2 Database design

Algorithm VIP (Training data D)

1. Create a customer profit node
2. Do while not EOF
3. call profit\_calculate (i)
4. End while
5. call sort(D)
6. call analyze\_category(class 1)
7. if call review\_term(class 1)
8. call SQL\_table\_product → output Table\_R
9. if T\_bale\_R(i,j) > 1.5 \* product.update\_period
10. then call review\_income
11. if customer.income < standard(income)
12. then {if convert(customer.career) < standard(career)
13. then class1.customer → class2.customer}
14. endif
15. endif

We then apply cluster algorithm to classify these two group customers in SQL Server 2005 separately and output sub-clusters. Cluster detection is derived from a map of self-organizing neural networks. In the processing of cluster, data are partitioned into subsets based on the degree of similarity. Cluster algorithm automatically detects difference between the groups in the database. The cluster algorithm can be used to cluster data to groups before running further analysis while processing mass data and variables. The K-means algorithm is one of the best implementation of cluster

technology. It basically includes three steps. Step 1: It randomly selects  $K$  numbers of observations as centers of initial clusters from  $N$  observations. Step 2: It finds the nearest cluster in terms of the Euclidean distance with respect to  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  for each of the remaining  $(N-K)$  observations. Step 3: It calculates the Euclidean distance between each observation and calculate the cluster's center point to verify that whether they have been allocated to the nearest cluster when all observations have been allocated.

### 3.3 Rule mining

Association rule in SQL Server 2005 is applied to generate purchasing rule for each sub-cluster in this phase. The association rule algorithm is basically used to search related items or features in database. The association rule is implication of the form  $X \rightarrow Y$ , where  $X \subset I, Y \subset I$  and  $X \cap Y = \phi$ . If  $D$  represents a set of transactions, then the rule  $X \rightarrow Y$  holds in the transaction set  $D$  according to two measurements, i.e. support and confidence. Support represents the rate of transactions in  $D$ , that contains item set  $X$ , and is denoted as  $Sup(X, D)$ . Support is used to evaluate the statistical importance of  $D$ . Thus, the rule  $X \rightarrow Y$  which has support  $Sup(X \cup Y, D)$  represents the rate of transactions in  $D$  containing  $X \cup Y$ . Confidence represents the rate of transactions in  $D$ , that contains  $X$  and  $Y$ , and is denoted as  $Conf(X \rightarrow Y)$ . The relationship between support and confidence is:  $Conf(X \rightarrow Y) = Sup(X \cup Y, D) / Sup(X, D)$ . It means the probability of that the transaction includes  $X$ , the chance that the transaction also contains  $Y$ . Confidence is used to evaluate the level of confidence of the association rule  $X \rightarrow Y$ . Given a set of transaction  $D$ , association rule generates all transaction rules that have certain Minimum support ( $Min sup$ ) and Minimum confidence ( $Min conf$ ) [7, 8]. Based on [9], two steps are used to process association rule. Step 1 detects a large item set whose support is greater than  $Min sup$ . Step 2 uses the large item set to generate association rules. It can be expressed as:  $Sup(X \cup Y, D) \geq Min sup$  and  $Conf(X \rightarrow Y) \geq Min conf$ . Many improved association rule algorithms, inspired from Apriori's algorithm, have been proposed such as using hash table technology and frequent pattern growth [10].

### 3.4 Customized e-coupon generation

The contents of e-coupons are generated based on the rule content. The IF-THEN rule in a rule-based system is typically formed as  $X \rightarrow Y$  (Fig.3a), in which  $X$  is an antecedent fact clause (node) and  $Y$  is a conclusion node. Two nodes, atomic and compound, exist in rule bases. One study [11] found that compound antecedents in rules are only presented in the conjunction format, such as [If ( $X_1$  and  $X_2 \dots$  and  $X_n$ ) then  $Y$ ], and that only the atomic node is allowed to be a conclusion (Fig. 3b). The association rule in SQL Server 2005 only applies basic rule.

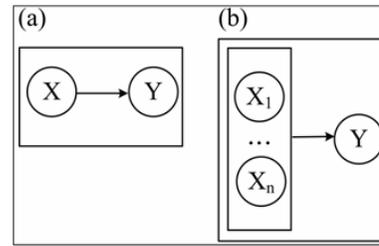


Fig. 3 (a) Basic rule; (b) Compound antecedent rule.

Rules of similar calculation features are considered of the same category. For each category rule, we generate a procedure to determine the discount and items shown on coupon. The following algorithm, for instance, generate coupons based on the calculable features to generate product coupons for cluster customers. All antecedent facts and conclusions of rules act as parameters to execute program, such as discount, sale price and sold quantity (QTY).

Procedure Discount\_rule (rule\_pre, rule\_con)

1. For i = 1 to n (cluster)
2.     get each customer\_Id
3.     if customer.attribute.value  $\in$  rule\_pre.value
4.     Then {if output today(>) > Trans\_detail.next\_T\_date
5.         new\_P = Trans\_detail.sale\_price \*(1-rule\_pre)
6.         if new\_P > wholesale\_price
7.         Then email.list  $\leftarrow$  Trans\_detail.Prod\_id, new\_p, rule\_con
8.         else next rule}
9. Next i

## 4. RESEARCH RESULTS

By following the proposed four phases, we constructed five relational tables on phase one and allocated the VIP to three sub-clusters (Fig. 4) and Non-VIP to six sub-clusters on phase two. Ninety-six rules were generated from the association rule mining on phase three. The followings are rules related to price discount, product item and sold quantity.

R1: If (price discount is between 0.12 and 0.27) then (product = food) (c=0.92)

R2: If (price discount > 0.27) then (product = milk) (c=0.54)

R3: If (price discount is between 0.12 and 0.27) then (QTY = 2~6) (c=0.57)

R4: If no discount then (QTY <2) (c=0.78)



Fig. 4 VIP cluster distribution

Part of the rules is shown in Fig.5.



Fig. 5 Partial Rule output

Under bi-weekly basis consideration, purchasing amount, sold items and associated category are attributes for procedures designed to extract the VIP customers in whom we considered the 80% profit-making. The antecedent facts and conclusions of rules are parameters applied in series of programs, such as Discount\_rule, to generate C\_Id, discount percentage and Prod\_Id. Currently, the content criteria of e-coupons are determined by the user (i.e. manager) based on the conclusion of rules. Basically, we apply association rule algorithm to find out the valuable customers' purchase patterns and predict their next purchasing and then promote associated products which might attract their attention and encourage their purchasing intention. Customers receive the coupon either via e-mail or receive the printout while customers make purchasing. The result of experiment shows the performance of the proposed approach. In comparison with the average month profit data over year 2008 and 2009, the proposed approach achieves about 10% better performance in the monthly profit on 2010 (Fig. 6). For example, the profit of Nov, 2009 and Nov, 2010 are NT \$288,900 and NT \$325,000, respectively;

a gain of 12%.

### 5. CONCLUSIONS

Analysis of customer purchase patterns and generation of next promotion strategy could potentially benefit the mid-sized enterprises to run business better in the competition of worldwide organization. The main contribution of the experiment result shows that the proposed approach achieves 10% improvement in the monthly profit, as compared with the conventional mean for the past two years. In this study, data mining tool was used to cluster customer groups based on data over years between Jan, 2008 and Dec, 2009. The VIP algorithms were programmed to segment the top 20% and the rest of the profit-making customers. To analyze customer consumption preference, we used the cluster and association rule of SQL Server 2005 analysis service to cluster groups and mine the rules. The programs then generated e-coupons. The approach is currently implemented as a prototype in TUNCHI. Further research will focus on two directions. (1). Either combine the programs output to SQL Server 2005 analysis service, or design customized mining system for the retail business; (2). Provide more customized promotion plan for the retail business by adopting the output of rules and by designing data mining algorithms.

### ACKNOWLEDGEMENTS

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 99-2221-E-244-009.

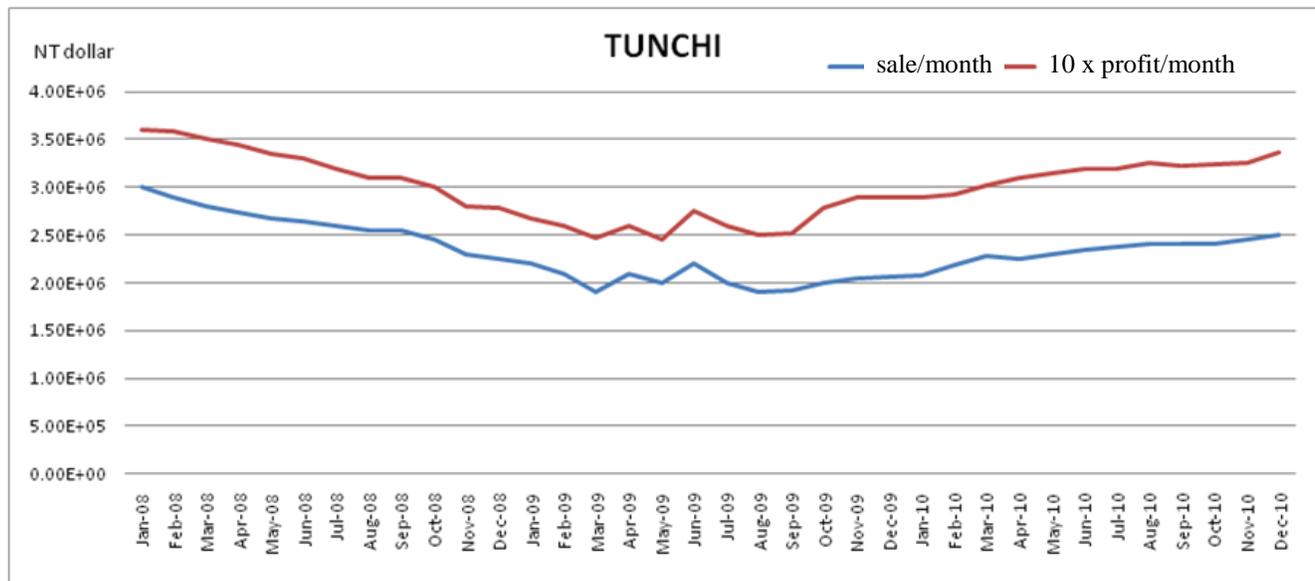


Fig. 6 Sales and profit distribution curves

## REFERENCES

- [1] U. Fayyad, G. Piatetsky, P. Smyth, "From data mining to Knowledge discovery: an overview", *Advances in Knowledge Discovery and Data mining*, MIT Press, 1998.
- [2] G. John, "Cross-validated C4.5: Using error estimation for automatic parameter selection", *Technical Note*, CS Dept., Standard Univ., 1996.
- [3] I. Mahdavi, M. Movahednejad and F. Adbesh, "Designing customer-oriented catalogs in e-CRM using an effective self-adaptive genetic algorithm", *Expert Systems with Applications*, vol. 38, 2011, pp. 631-639.
- [4] S. Liao, Y. Chen, Y. Lin, "Mining customer knowledge to implement online shopping and home delivery for hypermarkets", *Expert Systems with Applications*, vol. 38, 2011, pp. 3982-3991.
- [5] P. Giudici, G. Passerone, "Data mining of association structures to model customer behavior", *Computational Statistics and Data Analysis*, vol. 38, 2002, pp. 533-541.
- [6] V. Pareto, What is 80/20 Rule, Pareto's Law, Pareto Principle, <http://www.80-20presentationrule.com/whatisrule.html>
- [7] Y.F. Wang, Y.L. Chuang, "A personalized recommender system for the cosmetic business", *Expert systems with Application*, vol. 26, 2004, pp. 427-434.
- [8] I.N. Kouris, C.H. Makris, A.K.Tsakalidis, "Using Information retrieval techniques for supporting data mining", *Data and Knowledge Engineering*, vol. 52, 2005, pp. 353-383.
- [9] R. Agrawal, J. Shafer, "Parallel mining of association rules", *IEEE transactions on knowledge and data engineering*, vol 8(6), 1996, pp. 962-969.
- [10] R. Agrawal, T. Imilienski, A. Swami, "Mining association rules between sets of items in large databases. In proceedings of the 1993 ACM SIGMOD international conference on management of data, 1993, pp 207-216.
- [11] M. Ramaswamy, S. Sarkar, and Y.S. Chen, "Using Directed Hypergraphs to Verify Rule-Based Expert Systems", *IEEE Trans. Knowledge and Data Eng.*, vol.9, no. 2, 1997, pp. 221-236

# Study on Scheme for University Information System Integration Based on SOA

Lixia Zhou

Economics & Management College, Nanchang Hangkong University

Nanchang, Jiangxi, ZIP:330063, China

## 2. SOA AND WEB SERVICES

### ABSTRACT

With the characteristics of independent from bottom application software and platform and loose coupling, flexibility and reusability, Service Oriented Architecture (SOA) provides a solution for the information isolated island problem among information systems. This paper firstly explores the application of SOA and Web services. Then, proposes a Web service-based SOA model across different platforms, Then a scheme on university information system integration based on SOA is elaborated on, including the design for systems integration architecture and data exchanging platform based on SOA, and then the detail implementation of the scheme is proposed, which demonstrates an improvement on the scalability, reliability and efficiency of university information system.

**Keywords:** SOA, Web service, scheme, university, information system integration.

### 1. INTRODUCTION

Internationalization is the current trend of higher education, which has become a central strategy for sustainable development of China's university. In recent years, many universities have established multiple information systems such as teaching affairs management system, personnel management system, financial management system, and library management system, but they are applied in different departments without data sharing. Meanwhile, as developed in different time, running on different platforms, using different technology and no any standard, these are many problems for data sharing and exchange among them. The administrators always feel difficult to obtain a whole data view for comprehensive data analysis and use. Therefore, How to integrate the application software from different manufacturers with different technology standards is the key for further informatization of China's universities. In fact, the approach of Service Oriented Architecture (SOA) provides a good solution for the above problems, especially, the mature of Web service enhances the advantages of the SOA scheme.

#### A. SOA

SOA is software system architecture for organizing and utilizing distributed capabilities that may be under the control of different ownership domains in the Internet/Intranet environment. It provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations. SOA is a component model, it connects the different functional units (services) of application software by well-defined interfaces and contracts among these services. Services can therefore be run on various distributed platforms and be accessed across networks. This can also maximize reuse of services. So SOA features a loose coupling with any specific operating system, programming language, and other technologies, because the resources is consumed in an independent, standardized way.

SOA promotes the goal of separating users (consumers) from the service implementations, that's why some enterprise architects believe that SOA can help businesses respond more quickly and cost-effectively to changing market-conditions.

#### B. Web Service

Web service is web based applications that use open, XML-based standards and transport protocols to exchange data with clients. It's a standardized way of integrating Web-based application using the XML, SOAP, WSDL (Web Services Description Language) and UDDI open standards over an Internet protocol backbone. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.

#### C. SOA Model Based on Web Services

Thanks to the mature standards and popular application of Web service, it is a relatively suitable technology for SOA realization, namely it can realize SOA in many aspects, therefore it becomes the excellent choice for the development of SOA application software. In a Web service-based SOA model, software developers are grouped into three parties in terms of their responsibilities: the service consumers (requesters), the service brokers, and the service providers, as shown in Figure 1.

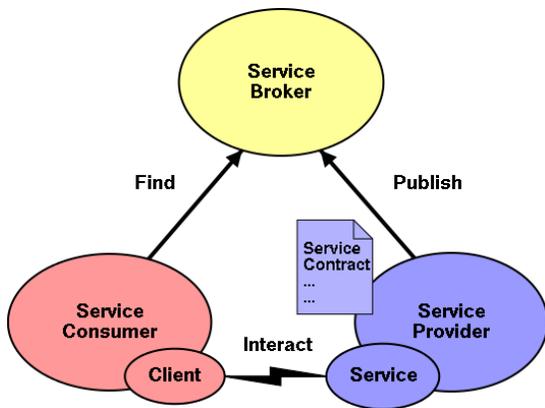


Figure 1. Web Service architecture

SOA Web services are supported by most major software vendors and industry analysts. Unlike RPC Web services, loose coupling is more likely, because the focus is on the "contract" that WSDL provides, rather than the underlying implementation details.

### 3. A SCHEME OF UNIVERSITY INFORMATION SYSTEM INTEGRATION BASED ON SOA

As mentioned above in introduction part of this paper, the university information system integration based on SOA is urgently needed for further informatization of China's universities.

#### A. Data sharing demand of university information systems

**1. Data connections among systems:** There are a lot of data exchanges among different information systems of different departments of universities. Herein we demonstrate the specific connections among information systems by teaching affairs management information system. It is one of the most important information system in university, which includes course arrangement system, course selection system, performance management system, teacher evaluation system and teaching material management system, etc. Its information is widely needed by many other information systems, for example the personnel management system gets teacher's workload information from its course arrangement system and teaching quality evaluation data from its teacher evaluation system; Students management information system needs the scores information from its performance management system when makes awards decision and graduate evaluation; The bursary tuition management system calculates the students charges according to the course information from its course selection system.

**2. Business connections cross systems:** Except of the data exchanges, there are business connections among

different systems. The main cross systems business processes are illustrated as follows: when a boy student is registering, registration system will audit his payment situation by the tuition management system; when he is selecting courses, course selection system will also audit his payment situation by the tuition management system; when he is graduating, graduation audit system will audit his exams result through the performance management system and inquire if his arrears is existing through the tuition management system; When he is graduating and leaving university, leaving management system will delete his corresponding account by E-mail system and campus BBS system, and the library system and dormitory management system will audit whether he has dealt with the university leaving formalities; And so on.

#### B. Integration Architecture

Concerning the University information systems, their integration architecture based on SOA is shown in figure 2, in which Enterprise Service Bus (ESB) is in the core status, it is the intermediary of intelligent integration and management between services.

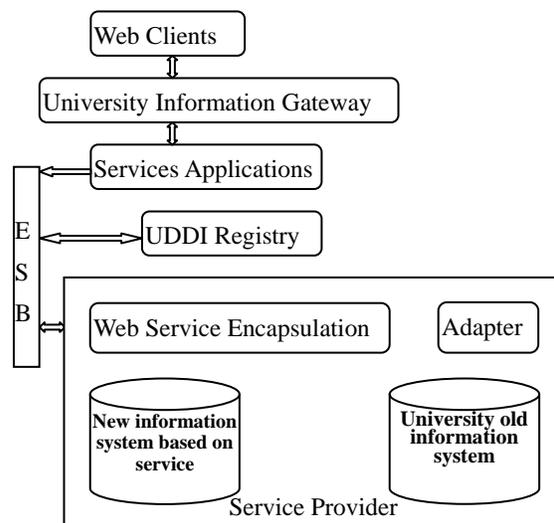


Figure 2. University information system integration architecture based on SOA

Event-driven is the important characteristic of ESB. Message passing between services usually have two kinds of forms, one is invoke, namely request/ respond, this is a common synchronous mode. Another is called as one-way news, whose purpose is often triggering asynchronous events, in which the sender doesn't need immediate reply. Considering some application services are long time running, the messages interaction of this asynchronous service must also be supported by ESB.

#### C. Data Exchange Platform Architecture Based on SOA

Because the university information systems involves different hardware and software platforms, database systems and developing languages, it is hard to achieve data unity and sharing by adopting the traditional software development methods. We will establish a service oriented data exchange platform, each business system collects data by the interface module, and saves them into raw data pool by XML file, the XML files in the raw data pool can be written directly into middle database by Web service interface, or they can be analyzed by the core processing modules, then written into the middle databases or output directly into peripheral systems. The architecture of data exchange platform is shown as Figure 3. (System A,B,C,...X refer to different university information systems such as teaching affairs management system, financial management system, personnel management system, library management system, etc.)

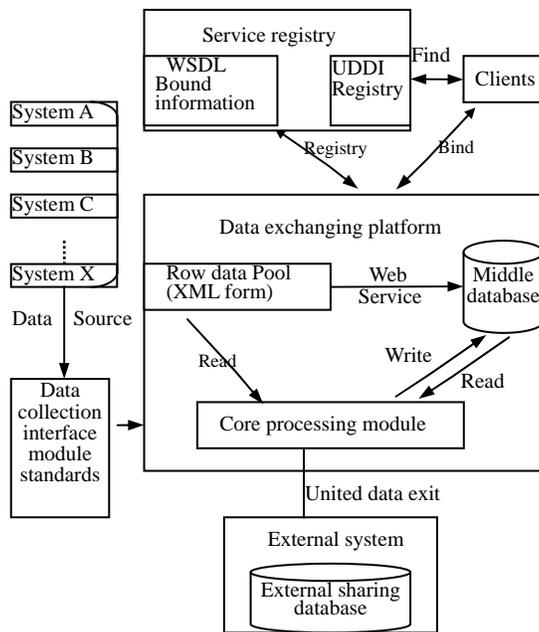


Figure 3. Data exchange platform architecture based on SOA

#### 4. IMPLEMENTATION OF THE SCHEME

##### A. Service Partition

We divide public services according to the following principles: core status of the system, core status of the data, function and stability of the service. Herein we take some important systems as an example to illustrate each divided specific service, which is shown in table 1.

Table 1. Service Partition

Public Services	Services Function	Service Provider	Service Requester
Login and authentication	United authentication service supports university teacher and students login and authentication	Teaching affairs system	Each system needing login and authentication
Curricular service	Support courses inquiry and arrangement, etc.	Teaching affairs system	Each system needing curricular service
Teacher info. service	Support teacher info. inquiry, etc.	Personnel management system	Each system needing teacher information
Experimental result	Support experimental result report and inquiry, etc.	Lab exam system	Teaching affairs system, etc

##### B. Service Encapsulation

For the new developed systems, such as laboratory management system, Mentor and Graduation paper topic selection system, they can be encapsulated into standard Web service components for function invoking. For the existing application systems based on J2EE or .NET, they will be encapsulated into single service by Web service technology, service interface is defined by WSDL, and the contents of WSDL files are published to university UDDI database, namely register various Web services in UDDI registry. Service encapsulation principle is shown in figure 4.

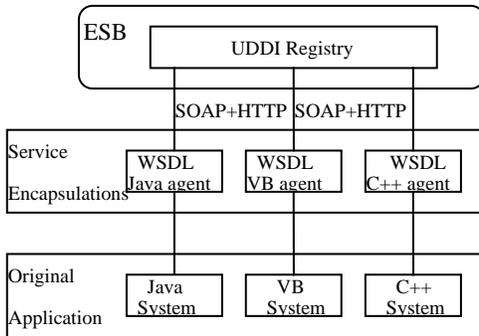


Figure 4. Service Encapsulation

When users access resources system through university web portals, the system should find the WSDL files corresponding with the user interface offered by system in the UDDI registry, then according to the WSDL files, invoke system user management module for user authentication. Requesting data sharing and business invoking, first should access the Web services registration information requesting service in UDDI registry, through the registration information access corresponding Web service interface description WSDL files, and generate local proxy objects. Then through this proxy objects, each needed business can use interactive information resource and invoke business function from Web services. If internal structure and realization of Web services are changed, we just modify WSDL descriptions, maintain original interface, and no need other changes.

### C. Development of the Core Processing Module in Data Exchange Platform

As a data exchange medium, XML is widely used in Java Message Service and Web Services, in order to accelerate the developing speed and reduce developing difficulty, we adopt XML Bean. Before using it, we should generate XML Schema, then generate Java Classes by scomp, finally we can program to use XML Bean read, write, modify and inquire XML files, convert the files in shared database to XML files and transfer to platform middle database or raw data pool. Conversely, the XLM files in both of latter can also be converted into database files and saved in shared database.

## 5. CONCLUSIONS

This paper focuses on a university information system integration scheme based on SOA. Except for the inherent advantage of SOA: coding flexibility, clear developers roles, supporting multiple clients type, easier to maintain, Higher scalability and higher usability, as using ESB, this scheme gets the characteristics of good encapsulation, loose coupling, expansibility, reliability, flexibility, reusability, autonomy, optimization, discoverability, etc, and can be integrated

step-by-step effectively. After the implementation of the scheme, university application systems can get better interoperability and openness, realize the data sharing in university campus, which promoting the development of university information. Actual application shows that SOA architecture can markedly improve university information system integration and software reusability.

## 6. REFERENCES

- [1] Lixin Han, Sanglu Lu and Li Xie, "A Model of Distributed Information Retrieval Systems for Internet Applications", **ACTA ELECTRONICA SINICA**. China, vol. 30, No.8, Aug. 2002.
- [2] **Web Services Architecture**, W3C.org. Note 11. Feb 2004. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/>.
- [3] Russ Basiura, Mike Batongbacal, **ASP.NET Advanced Programming of Web Services**, Beijing: Publisher of Tsinghua University, 2002.
- [4] Tu Huijun, **Campus Integrated Design**, Beijing: China Building Industry Press, 2007.
- [5] Parto: **Primer, W3C Web Service Description Language (WSDL) Version 2.0**. <http://www.w3e.org/IRIwsdl20-primer/>, 2004
- [6] Eric Newcomer, Greg Lomow, translated by Xu Han, **Understanding SOA with web Services (Chinese version)**, Beijing: Publishing House of Electronic Industry, 2006.
- [7] Xu KeFeng, "Enterprise application integration technology and realization based on Web Services", **computer application**, vol.24, No.3, Mar. 2004.
- [8] Thomas Erl, **Service-Oriented Architecture (SOA): Concepts, Technology and Design**, NJ: Prentice Hall, August 12, 2005.
- [9] **Regulations of the People's Republic of China on Chinese-Foreign Cooperation in Running Schools**, Higher Technical & Vocational Education in China, from <http://www.tech.net.cn/en/pl/10754.shtml>, September 2003, Retrieved 19 February, 2008.
- [10] Chris Kelly, "Does Distributed Innovation fit with current innovation theory and policy?", **PUBP6803 – Technology, Regions, and Policy**, 2006, P1-13.
- [11] Coombs, R. and Metcalfe, S., **Organizing for Innovation: Coordinating Distributed Innovation Capabilities**. In Foss, N. J. and Mahnke, V. **Competence, Governance, and Entrepreneurship**. Oxford: Oxford University Press, 2001.
- [12] **Simple Object Access Protocol (SOAP) 1.1**. W3C. org. Note 08. May, 2000. <http://www.w3.org/TR/2000/NOTE-SOAP-20000508/>.
- [13] **UDDI 101. UDD & XML**. org. Aug, 2006. <http://uddi.xml.org/uddi-101>.
- [14] M. L. Liu, **Distributed Computing: Principles and Application**, Beijing: Publisher of Tsinghua University, 2004.
- [15] Thomas Erl, **SOA Principles of Service Design**, NJ: Prentice Hall, 1 edition, July 28, 2007.

# Exchange Flow Rate Measurement Technique in Density Different Gases

Motoo Fumizawa and Syuhei Ohkawa

ME Dept., Shonan Institute of Technology, Fujisawa, 251-8511 Japan

## ABSTRACT:

*Buoyancy-driven exchange flows of helium-air through horizontal and inclined small openings were investigated. Exchange flows may occur following the opening of a window for ventilation, when fire breaks out in a room, as well as when a pipe ruptures in a high temperature gas-cooled nuclear reactor. The numerical analysis and experiment in this paper was carried out in a test chamber filled with helium and the flow was visualized using the smoke wire method. The flow behavior was recorded by a high-speed camera combined with a computer system. The image of the flow was transferred to digital data, and the flow velocity was measured by PTV software. The detected data was arranged by the densimetric Froude number of the exchange flow rate derived from the dimensional analysis. As the results, densimetric Froude number  $Fr$  by smoke wire (SW)-method is higher than that by mass inclement (MI)- method.(4) Numerical calculations predicts the circulation flow occurs in the narrow flow path.*

**keywords:** Buoyancy, Exchange flow, Helium, Smoke wire, Interferometer PTV Densimetric Froude number, HSMAC-method, Numerical analysis

## 1. INTRODUCTION

Buoyancy-driven exchange flows of helium-air were investigated through horizontal and inclined small openings. Exchange flows may occur following the opening of a window for ventilation, the outbreak of fire in a room or over an escalator in an underground shopping center, as well as when a pipe ruptures in a modular high temperature gas-cooled nuclear reactor. The fuel loading pipe is located in an inclined position in a pebble bed reactor such as the Modular reactor [1,2] and AVR [3,4].

In safety studies of High Temperature Gas-Cooled Reactor (HTGR), the failure of a standpipe at the top of the reactor vessel or a fuel loading pipe may be one of the most critical design-based accidents. Once the pipe ruptures, helium immediately rushes up through the breach. Once the pressure between the inside and outside

of the pressure vessel has balanced, helium flows upward and air flows downward through the breach into the pressure vessel. This means that buoyancy-driven exchange flow occurs through the breach, caused by the density difference of the gases in the unstably stratified field. Since an air stream corrodes graphite structures in the reactor, it is important to evaluate and reduce the air ingress flow rate when a standpipe rupture occurs. To date, studies have been performed on the exchange flow of two fluids with different densities through vertical and inclined short tubes. Epstein [5] studied the exchange flow of water and brine through various vertical tubes, experimentally and theoretically. Mercer et al.[6] experimentally studied an exchange flow through inclined tubes with water and brine. The latter experiments were carried out in the range of  $3.5 < L/D < 18$  and  $0 \text{ deg} < \theta < 90 \text{ deg}$ , and indicated that the length-to-diameter ratio  $L/D$ ,

and the inclination angle  $\theta$  of the tube are the important parameters for the exchange flow rate. Most of these studies were performed on the exchange flow using a relatively small difference in the densities of the two fluids (up to 10 per cent). However, in the case of a HTGR standpipe rupture, the density of the outside gas is at least three times larger than that of the gas inside the pressure vessel. Few studies have been performed so far using such a large density difference. Kang et al.[7] studied experimentally the exchange flow through a round tube with a partition plate. Although one may assume that the partitioned plate, a kind of obstacle in the tube, would reduce the exchange flow rate, Kang found that the exchange flow rate was increased by the partition plate because of separation of the upward and downward flows.

The objectives of the present study are to investigate the behavior of the exchange flow, then to evaluate the exchange flow rate using the PTV and PIV methods, as well as mass increment with the helium-air system. The following methods are applied the present study.

- (1) numerical analysis
- (2) smoke wire method (SW)
- (3) mass increment method (MI)
- (4) optical system of the Mach-Zehnder interferometer

## 2. NUMERICAL ANALYSIS

Two dimensional unsteady code of HSMAC method is adopted to the buoyancy-driven exchange flow system [1]. In the code, thermal hydraulic program is described basic equation of mass, momentum and energy. The energy equation is transferred to mass diffusion equation, as shown in Figure 1. Analysis coordinate is shown in Figure 2. The left part is test chamber filled with helium gas and the right part is outside region filled with air. Typical calculation result is shown in Figure 3, where is the narrow channel, between the left and right. The exchange flow occurs with vortex in the narrow channel. Therefore, the center flow rate  $Q_1$  is

larger than the right edge flow rate  $Q_2$ , as shown in Figure 4.

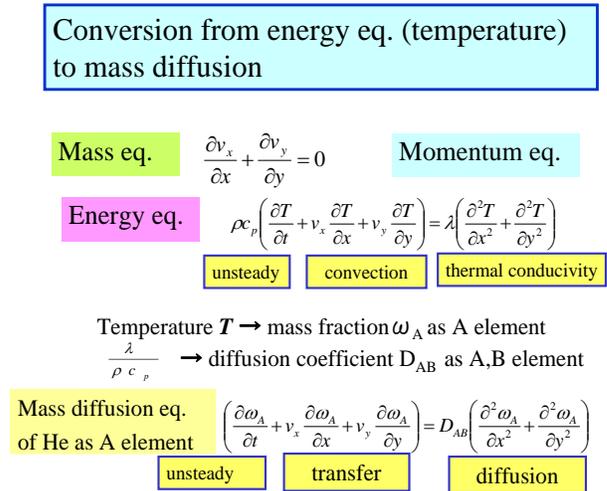


Fig.1 Basic equations of energy and mass diffusion

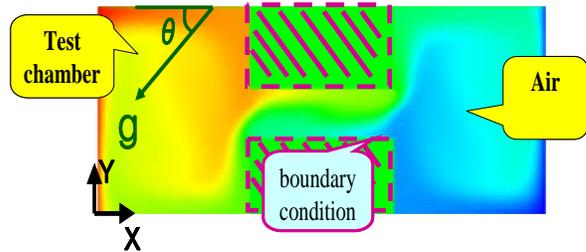


Fig.2 Analytical coordinate and conditions

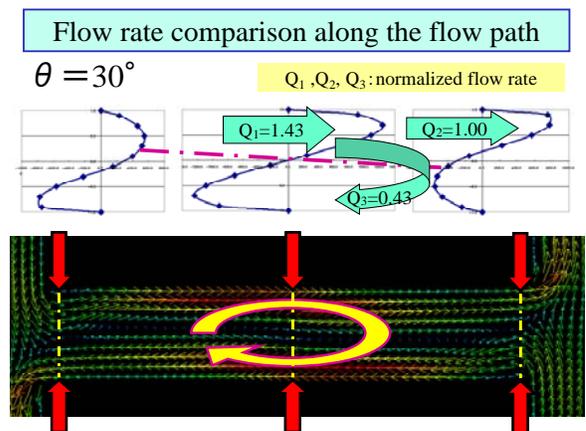


Fig.3 Analytical result of narrow flow path of Exchange flow

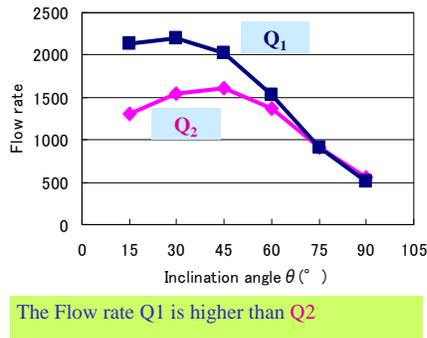


Fig.4 Relation of Exchange flow rate and inclination angle of analytical results

### 3. SMOKE WIRE MWTETHOD

#### 3.1 Experimental apparatus and procedure

The smoke wire method (SW) was used for the present investigation. Figure 5 shows a typical sketch of the apparatus. It consists of a smoke pulse generator, a thin Nichrome wire coated with oil and a test chamber. This figure also shows the high-speed camera system, which transfers the visual digital data to the personal computer for data acquisition. The experimental procedure is as follows. The test chamber is filled with pure helium. By removing the cover plate placed on the top of the tube, the exchange flow (i.e. exchange flow of helium and air) is initiated. Under such conditions, the smoke pulse generator ignites the high voltage. Immediately smoke appears and the helium up flow and air down flow are visualized in the flow path of the long tube. The test chamber diameter and height are 350 mm, the diameters of the long tubes ( $=D$ ) are 20mm and 17.4mm, the lengths of the tubes ( $=L$ ) are 200mm and 319mm, respectively. These conditions indicate  $L/D=10$  and  $L/D=18.3$ , respectively. The inclination angle  $\theta$  is 30 deg. The smoke wire conditions are as follows. The voltage is approximately 250 (V), the current duration is 30 m sec, and the oil on the thin wire is CRC-556. The high-speed camera system, which uses a D-file, records the visual data up to

1600 frames per second. The upward flows peak velocity is measured by PTV method.

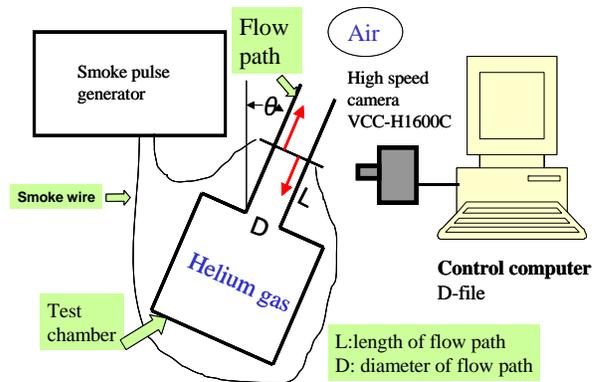


Fig.5 Exchange flow apparatus of smoke wire method high speed camera system

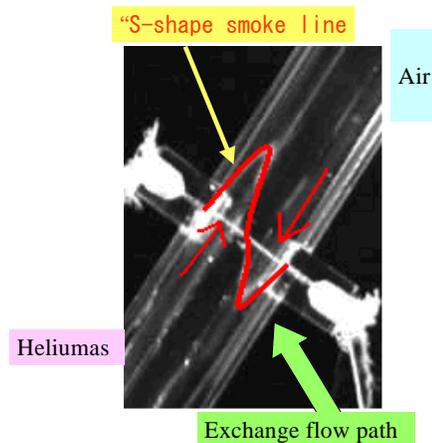


Fig.6 Mechanism of exchange flow

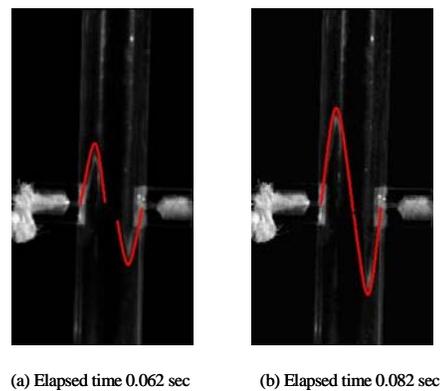


Fig. 7 Visualized data listed along the elapsed time ( $L/D=10$ , frame rate=500 f/s,  $\theta=30$  deg)

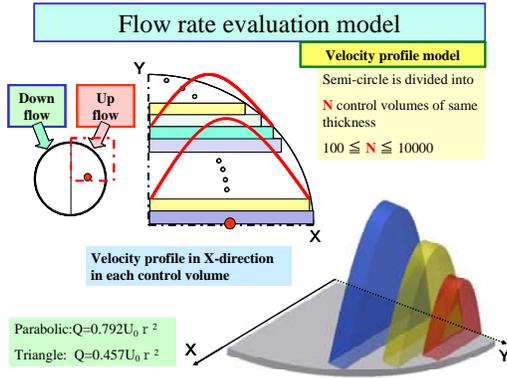


Fig.8 Flow rate evaluation model in smoke wire method

### 3.2. Results

The typical exchange flow in the tube was visualized in Figure 6. The visualized exchange flow resembles an S-shape. The frames are set at 200 and 500 frames per second in the high-speed camera. The visualized data is listed along the elapsed time in Figure 7, the upward flow peak velocity measured by PTV. In the case of  $L/D=18.3$ , the average velocity value  $U_0$  is evaluated as 0.315 m/s. In the case of  $L/d=10$ , the average velocity value  $U_0$  is evaluated as 0.662 m/s. This signifies that high exchange velocity is detected in a low  $L/D$  ratio. Assuming flow model of a parabolic flow profile, as shown in Figure 8, the exchange flow rate  $Q$  is derived as follows, where  $r$  is the radius of the flow path in a horizontal direction. Result is shown in Figure 9. The flow rate  $Q$  is the highest in the 30 degree.

$$Q = 0.792 U_0^2 \quad (1)$$

The densimetric Froude number is defined by the following equation derived from the dimensional analysis suggested by Keulegan [8]:

$$Fr = \frac{Q}{A} \sqrt{\frac{\rho}{gD \Delta \rho}} \quad (2)$$

In the PTV method, the exchange flow rate  $Q$  is calculated as  $1.47 \times 10^{-5} \text{ m}^3/\text{s}$  under the condition of  $L/D=18.3$ . Therefore, the densimetric Froude number  $Fr$  is evaluated as 0.202, and in this condition Reynolds number  $Re$  is 79.2. When  $L/D=10$ , the densimetric Froude number  $Fr$  is evaluated as 0.287.

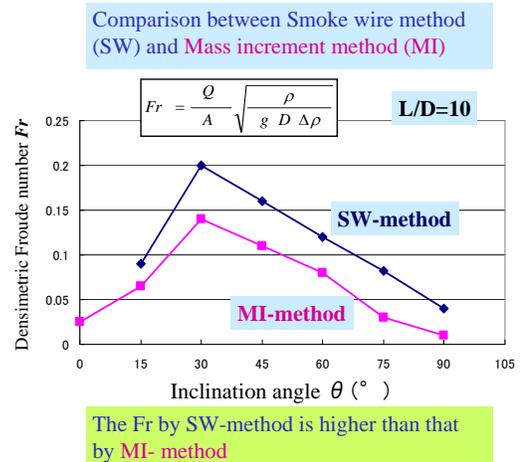


Fig.9 Relation of Densimetric Froude number and inclination angle in both experiments

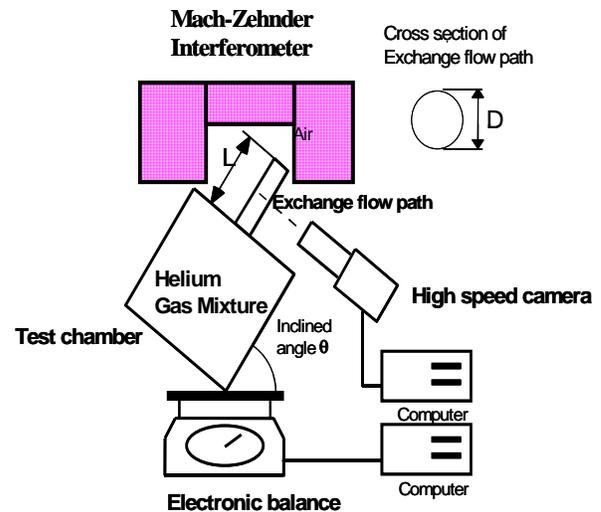


Fig.10 Experimental apparatus of optical system and mass increment

## 4. METHOD OF MASS INCREMENT

### 4.1 Experimental apparatus and procedure

The mass increment method (MI) was used for the investigations. Figure 10 shows a rough sketch of the apparatus. It consists of a test chamber, an electronic balance and a personal computer for data acquisition. The experimental procedure is mentioned in Sec. 2.1. Air enters the test chamber and the mass of the gas mixture in the test chamber increases. The mass increment  $\Delta m$  is automatically measured by the high accurate electronic balance. The density increment of the gas mixture  $\Delta \rho_L = \Delta m/V$  is calculated from mass increment data. The density increment signifies the difference between densities of the gas mixture from the density of pure helium in the test chamber. Subsequently, the volumetric exchange flow rate is evaluated using the following equation:

$$Q = \frac{V}{\rho_H - \rho_L} \cdot \frac{d(\Delta \rho_L)}{dt} \quad (3)$$

In the above equations,  $V$  is the volume of the test chamber,  $\rho_H$  the density of air,  $\rho_L$  the density of the gas mixture in the test chamber,  $\Delta \rho_L (= \rho_H - \rho_{He})$  = the density increment of the gas mixture,  $t$  the elapsed time,  $U(=Q/A)$  the exchange-velocity,  $\rho (= (\rho_H + \rho_{He})/2)$ ,  $D$  the diameter and  $g$  the acceleration of gravity. The experiments are performed under atmospheric pressure and room temperature using the vertical and inclined round tubes, and using a vertical annular tube. The density of the gas mixture is close to that of helium in the present experiment. The sizes of the tubes are as follows. The diameter of the round tube  $D$  is 20 mm, which is much smaller than that of the test chamber. The inclination angle  $\theta$  ranges from 15 to 90 deg and the height  $L$  ranges from 0.5 to 200 mm.

### 4.2. Results and discussion

It is already known that the densimetric Froude number is regarded as constant within a time

duration when the gas in the upward flow is assumed to be helium [10]. Figure 8 shows the relationship between  $Fr$  and inclination angle  $\theta$  with  $L/D$  as a parameter. For inclined tubes,  $Fr$  is larger than that for vertical tubes. The black circles show the experimental data for the orifice (i.e.  $L/D = 0.05$ ) and the black rhombuses for the long tube (i.e.  $L/D = 5$ ). The densimetric Froude number reaches the maximum at 60 deg for the orifice and 30 deg for the long tube. It is found that the angle for the maximum  $Fr$  decreases with the increase of  $L/D$  in the helium-air system.

## 5. OPTICAL SYSTEM OF MACH-ZEHNDER INTERFEROMETER

### 5.1 Experimental apparatus and procedure

The optical system of the Mach-Zehnder interferometer, MZC-60S to visualize the exchange flow is shown in Figure 10. After being rejoined behind the splitter, the test and reference laser beams interfere, and the pattern of interference fringe appears on the screen. If the density of the test section is homogeneous, the interference fringes are parallel and equidistant [9]. If the density is not homogeneous, the interference fringes are curved. Inhomogeneity in the test section produces a certain disturbance of the non-flow fringe pattern. The digital camera and high-speed camera using a D-file can be attached to the interferometer.

### 5.2. Results

Figure 11 shows the typical interference fringes for an inclined long tube ( $L/D=5$ ). The curved interference fringes indicate that the lighter helium flows in the upper passage of the tube. The straight fringes indicate that the heavier air flows in the bottom of the tube. It is clearly visualized that the exchange flows take place smoothly and in a stable manner in the separated passages of the tube. This leads to less resistance for the exchange flow in the inclined tubes compared to the vertical ones. In the case of a 30

deg. angle, the curvature of the interference fringes is larger than that at other angles, indicating that the exchange flow rate and the densimetric Froude number are the largest at 30 deg.

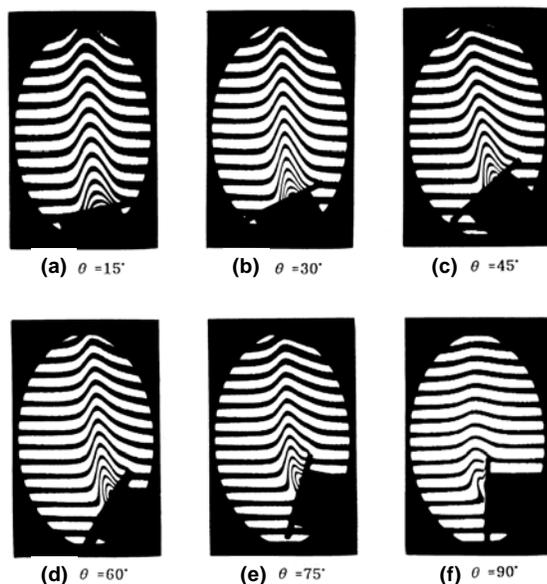


Fig.11 Typical interference fringes for an inclined long round tube ( $L/D=5$ )

## 6. CONCLUSION

- (1) Flow visualization results indicate that the exchange flows through the inclined round tube take place smoothly and in a stable manner in the separated passages of the tube.
- (2) Assuming velocity profile, exchange flow rate ,i.e., densimetric Froude number are evaluated.
- (3) Densimetric Froude number  $Fr$  by SW-method is higher than that by MI- method.
- (4) Numerical calculation predicts the circulation flow occurs in the narrow flow path.

## ACKNOWLEDGEMENTS

The authors are deeply indebted to Dr. Makoto Hishida, professor of Chiba University in Japan,

for his unfailing interest and helpful input to this study.

## REFERENCES

- [1] Fumizawa,M. ; Proc. HT2005 ASME Summer Heat Transfer Conference, HT2005-72131, Track 1-7-1 (2005) , pp.1-7.
- [2] Kiso et.al.;JSME Annual MTG, (1999) pp.339-340.
- [3] M.M.El-Wakil, Nuclear Energy Conversion, Thomas Y. Crowell Company Inc., USA (1982)
- [4] Arbeitsgemeinschaft Versuchs-Reactor AVR, "Sicherheitsbericht fuer das Atom -Versuchskraftwerk „Juelich" (1965)
- [5] Epstein,M., Trans. ASME J. Heat Transfer, 110, (1988) ,pp.885 -893
- [6] Mercer, A. and Thompson.H., J. Br. Nucl. Energy Soc., 14, (1975) , pp.327-340
- [7] Kang,T. et al., NURETH-5, (1992), pp.541-546
- [8] Keulegan,G.H., U.S.N.B.S.Report 5831 (1958)
- [9] Merzkirch.W., "Flow Visualization", Academic Press (1974)
- [10] Fumizawa,M. et. al., J. At. Energy Soc. Japan, Vol.31, (1989), pp.1127-1128

## Hardware System for Robot and Proprioceptive Algorithms

**Authors: Elmo B. de Faria, Mauricio F. L. Pereira, Jesús F. Bueno, Carla M. Silva, Sidney R. D. Carvalho, Sayuri A. Joazeiro, Wellington S. E. Santos**

**Universidade Federal de Mato Grosso**

**IARA: Intelligence Artificial, Robotics and Automation groups**

<http://www.ic.ufmt.br/>

[elmo@ufmt.br](mailto:elmo@ufmt.br)

### Abstract

The proposal of this work is to develop hardware for algorithms completely integrated with the sensorial systems and of performance of robots in the intention to adjust its movements in sensorial environments. In this context, the algorithms are called Proprioceptive Systems, where these algorithms must act as adjustment of the controllers who provide the movement of robots.

This work initially introduces some concepts of control and intelligent controllers, approach an introduction on sensors and its more varied types and finally he considers a diagram to integrate the systems of movements that are used in robots.

A considered experiment is the simulation of imperfections in joints of robots and to use the Proprioceptive algorithms, to identify the imperfection and to provide a modification in the action of control to compensate this effect and guarantee that the system continues functioning normally.

### 1-Introduction

*“Proprioception from Latin proprius, meaning “one's own” and perception, is the sense of the relative position of neighbouring parts of the body. Unlike the exteroceptive senses, by which we perceive the outside world, and interoceptive senses, by which we perceive the pain and movement of internal organs, proprioception is a third distinct sensory modality that provides feedback solely on the status of the body internally. It is the sense that indicates whether the body is moving with the required effort, as well as where the various parts of the body are located in relation to each other.”[5]*

The position-movement sensation was originally described in 1557 by Julius Caesar Scaliger as a 'sense of locomotion'. Much later, in 1826, Charles Bell expounded the idea of a 'muscle sense' and this is credited with being one of the first physiologic feedback mechanisms. Bell's idea was that commands were being carried from the brain to the muscles, and that reports on the muscle's condition would be sent in the reverse direction. Later, in 1880, Henry Charlton Bastian suggested 'kinaesthesia' instead of 'muscle sense' on the basis that some of the afferent information (back to the brain) was coming from other structures including tendons, joints, and skin. In 1889, Alfred Goldscheider suggested a classification of kinaesthesia into 3 types: muscle, tendon, and articular sensitivity.

In 1906, Charles Scott Sherrington published a landmark work that introduced the terms 'proprioception', 'interoception', and 'exteroception'. The 'exteroceptors' are the organs responsible for information from outside the body such as the eyes, ears, mouth, and skin. The interoceptors then give information about the internal organs, while 'proprioception' is awareness of movement derived from muscular, tendon, and articular sources. Such a system of classification has kept physiologists and anatomists searching for specialised nerve endings that transmit data on joint capsule and muscle tension (such as muscle spindles and Pacini corpuscles)[5].

In many robots is common the use of controllers, sensors, complex mechanics and intelligent programs. All this for create machines capable of moving like humans. These robots can develop tasks important and to substitute the action human in activities of risk and repetitive effort. These complex robotic systems moving with the use of links and innumerable other subsystems how communication, control movement and intelligent programs.

One difficulty consists on development of systems capable to integrate all these elements and still to promote the interaction between them. These systems have information of the environment and the system robotic to take decisions and to provide information for interaction with real world[1][2].

These robots systems initially can be composites of three great groups of systems involving the hardware and software, is they: Systems of security; Systems of movement and controlled action; Proprioceptive and sensorial systems. All the systems need the perception with the use sensors of innumerable classrooms and categories. Since simplest sensor of position and speed until the most sophisticated and complex as the sensors of torque and gyroscopes. The first group of systems acts in the measurement of destined sensors to keep the system in functioning. It among others consists of temperature sensors, positioning, that determine the stoppage so that yours hardware is not damaged seriously. These systems also guarantee the limits for the movements and correct functioning of engines and mechanical components of the robots[3][4].

The second group of systems is composed of programs to simulate common human tasks. It is a module programmed for the user for interaction and to execute the predetermined tasks. They consist in algorithms of movement, perception and execution of tasks.

The third basic group of algorithms is composed for the Proprioceptive algorithms. They are the algorithms that monitor the signals produced in the two previous processes, beyond using specific sensors, to identify to anomalies and situations where the algorithms of tasks must be adjusted.

This process of adjustment consists mainly of modifying the action of control on the movement engines. Producing modifications in the tasks that are executed but guaranteeing that they occur normally.

For this, the controllers must be intelligent systems and that provides the possibility of adaptation on the actions of control. In this work used Fuzzy controllers, therefore he allows to the adjustment of the actions based on rules of the fuzzy logic. The Proprioceptive algorithms also are created using base of rules fuzzy.

The Proprioceptive algorithms have as input the signals of the controllers and the values measured for sensors. The all signals of components are also evaluated, as the engines, sensors, actuators and values of its physical and mechanical measures.

The development of these algorithms can allow that the robots have one better perception of its proper metallic body, knowing the dimensions of the arms, height, width among others information that before were processed for the system of execution of tasks. The tasks are the programs executed normally for the robots in controlled environment, and that they are many times successfully executed. But if some unexpected situations happen, an inherent characteristic of dynamic environments, the Proprioceptive system can act and through adjustments to provide that task is executed.

The robots have difficulties to decide simple situations, as the simple fact to suffer a push, where the same ones tumble and fall. When this happens need to identify this new situation and to use another algorithm to be arisen. The proposal of the Proprioceptive algorithms is that when perceiving that they are being pushed, the Proprioceptive system detects this uncommon situation and acts on controllers to prevent the fall. Preventing this unexpected situation.

**2-Controllers**

The control systems are in practically all the modern equipment since simple bikes until sophisticated equipment like automobiles. In this brief summary will be presented enters some used techniques of control, initiating with classic control and techniques of intelligent control.

A control system can be defined as [21] and [22]:

*“A Control System is a device, or a collection of devices that manage the behavior of other devices. Some devices are not controllable. A control system is an interconnection of components connected or related in such a manner as to command, direct, or regulate itself or another system.”[10]*

**2.1- Control System Design**

A basic control system is shown in Figure 1. The process P (or “plant”) is the object to be controlled. Its inputs are  $u(t)$ , its outputs are  $y(t)$ , and the reference input is  $r(t)$ . One of the fundamental reasons for adding feedback control to a system is that steady-state errors are reduced by the action of the control system.

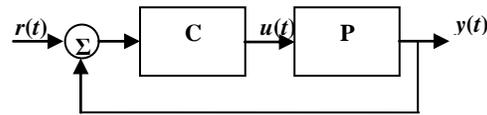


Figure 1- Control System [10]

**2.2-Mathematical Modeling**

For SISO systems we have the following partial list of typical classical performance specifications. Linear models such as the one in Equation (1.1) have been used extensively in the past and the control theory for linear systems is quite mature [10],[13] and [16].

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{1.1}$$

In this case  $u$  is the  $m$ -dimensional input;  $x$  is the  $n$ -dimensional state ( $\dot{x} = dx(t)/dt$ );  $y$  is the  $p$  dimensional output; and  $A$ ,  $B$ ,  $C$ , and  $D$  are matrices of appropriate dimension. Such models, or transfer functions ( $G(s) = C(sI - A)^{-1}B + D$  where  $s$  is the Laplace variable), are appropriate for use with frequency domain design techniques (e.g., Bode plots and Nyquist plots), the root-locus method, state-space methods, and so on. Sometimes it is assumed that the parameters of the linear model are constant but unknown, or can be perturbed from their nominal values.

**2.3-Fuzzy Control System Design**

Basically, the difficult task of modeling and simulating complex real-world systems for control systems development, especially when implementation issues are considered, is well documented. Even if a relatively accurate model of a dynamic system can be developed, it is often too complex to use in controller development, especially for many conventional control design procedures that require restrictive assumptions for the plant (e.g., linearity).

Fuzzy control provides a formal methodology for representing, manipulating, and implementing a human’s heuristic knowledge about how to control a system[11].

The fuzzy controller block diagram is given in Figure 2, where we show a fuzzy controller embedded in a closed-loop control system. The plant outputs are denoted by  $y(t)$ , its inputs are denoted by  $u(t)$ , and the reference input to the fuzzy controller is denoted by  $r(t)$ , where P is the process or “plant”.

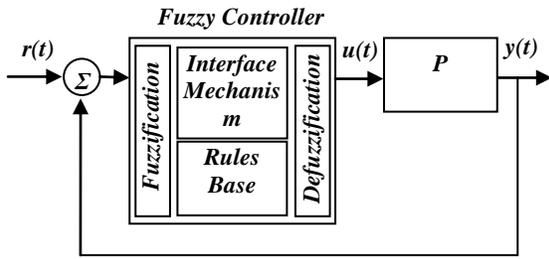


Figure 2 - Fuzzy Control System

The fuzzy controller has four main components: (1) the “rule-base” holds the knowledge, in the form of a set of rules, of how best to control the system. (2) The inference mechanism evaluates which control rules are relevant at the current time and then decides what the input to the plant should be. (3) The fuzzification interface simply modifies the inputs so that they can be interpreted and compared to the rules in the rule-base. And (4) the defuzzification interface converts the conclusions reached by the inference mechanism into the inputs to the plant. Basically, you should view the fuzzy controller as an artificial decision maker that operates in a closed-loop system in real time. It gathers plant output data  $y(t)$ , compares it to the reference input  $r(t)$ , and then decides what the plant input  $u(t)$  should be to ensure that the performance objectives will be met, figure 2.

**2.4-Fuzzy Controller for Robot Joint**

Consider the simple robot manipulator illustrated in Figure 3 , the equations of motion for this system are quite simple to derive, and take the form of the standard “manipulator equations” (1.2):

$$H(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = B(q)u \quad (1.2)$$

According to Newton, the dynamics of mechanical systems are second order ( $F=ma$ ). Their state is given by a vector of positions,  $q$ , and a vector of velocities  $\dot{q}$  and (possibly) time. The general form for a second-order controllable dynamical system is:

$$\ddot{q} = f(q, \dot{q}, u, t); \quad (1.3)$$

where  $u$  is the control vector. As we will see,

the forward dynamics for many of the robots that we care about turn out to be affine in commanded torque, so let’s consider a slightly constrained form:

$$\ddot{q} = f_1(q, \dot{q}, t) + f_2(q, \dot{q}, t)u \quad (1.4)$$

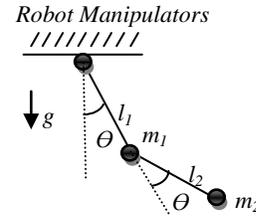


Figure 3 - Simple double pendulum

It is well known that the inertial matrix,  $H(q)$  is (always) uniformly symmetric and positive definite, and is therefore invertible. Putting the system into the form of equation 1.4 yields:

$$\ddot{q} = H^{-1}(q) [C(q, \dot{q})\dot{q} + G(q)] + H^{-1}(q)B(q)u \quad (1.5)$$

Because  $H^{-1}(q)$  is always full rank, we find that a system described by the manipulator equations is fully-actuated if and only if  $B(q)$  is full row rank. For this particular example,  $q = [\theta_1, \theta_2]^T$  and  $u = [\tau_1, \tau_2]^T$ , and  $B(q) = I_{2 \times 2}$ . The system is fully actuated. Now imagine the somewhat bizarre case that we have a motor to provide torque at the elbow, but no motor at the shoulder. In this case, we have  $u = \tau_2$ , and  $B(q) = [0, 1]^T$ . This system is clearly underactuated, for both cases we can consider the following fuzzy controller in figure 4.

Where  $r(t)$  is the reference input,  $e_1$  the reference error,  $g$  is the gain of velocity, position and  $V(t)$ , last one is the control variable.

The use of fuzzy controllers to this procedure can be done in the following steps.

- The first step consists on execute the normal program of position the manipulator.
- In the second phase, simulate a fail in join 1 (interrupt the energy in the servo-motor 1). In this time the Proprioceptive System Fuzzy Control implement the action on the local controllers. This action was be detected by current sensor of motor and send de Proprioceptive System.

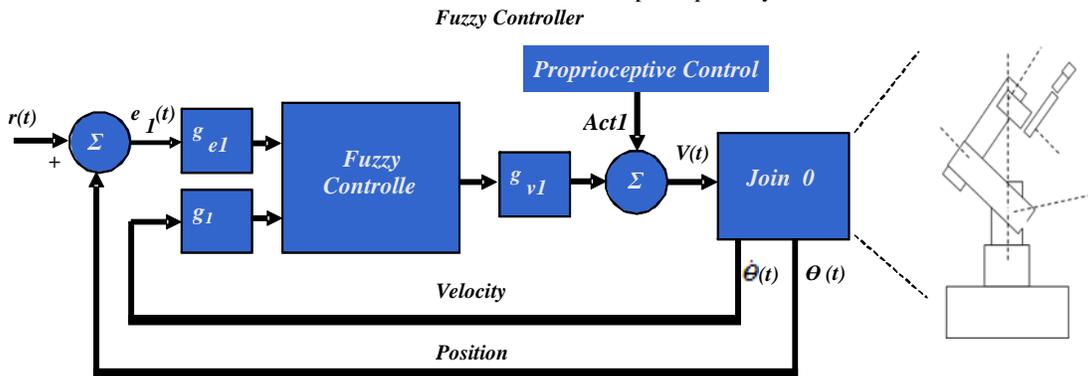


Figure 4 - Single Controller for one joint

### 3-Lab implementation

The proprioceptive algorithms are implemented with the use of Fuzzy controllers. The structure presented in the figure 5.

A typical arrangement used for control the Robots Manipulator has been used to illustrate the approach proposed in this work. It consists of an robot

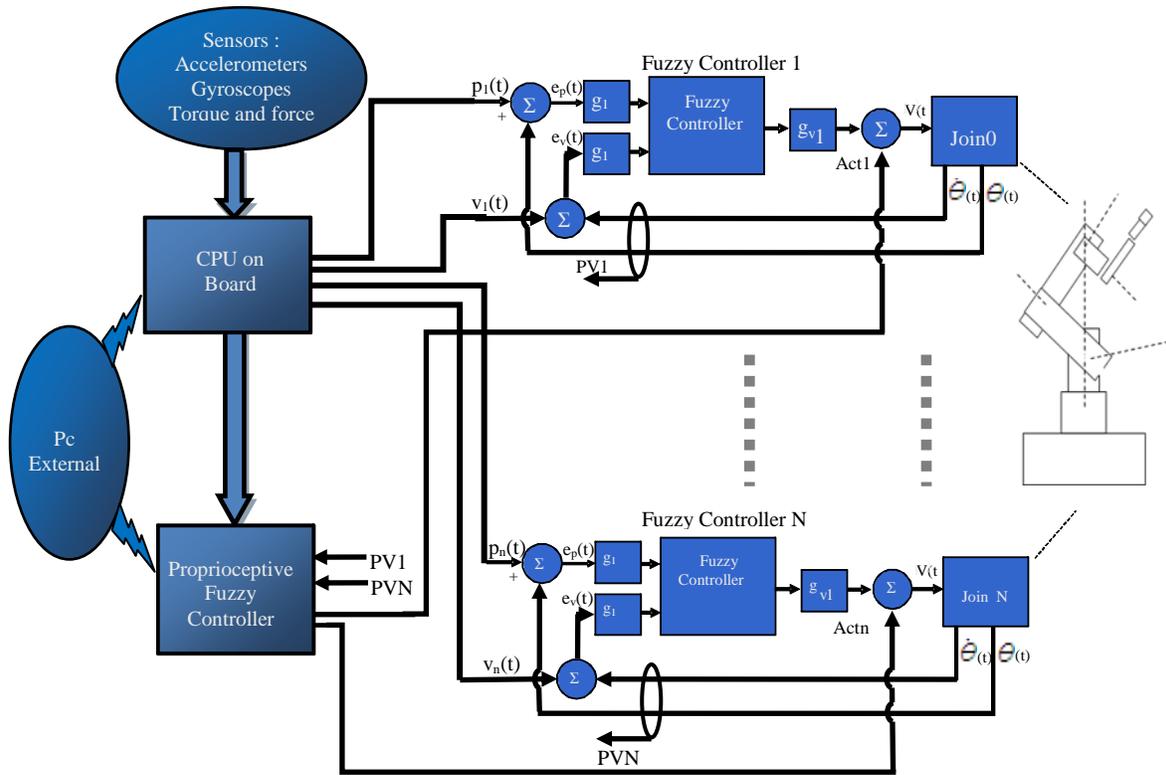


Figure 5- Fuzzy Control for position and velocity

Development:

General diagram of the robotic system and controllers is presented in figure 5.

with 4 DOF driven by D.C. motors. A potentiometer is used to measure the position the each joints. The system is excited by values and the input-output data is used by Proprioceptive System of Control. A figure-6 illustrates the real system of proposal.

#### 3.1 Real robotic System

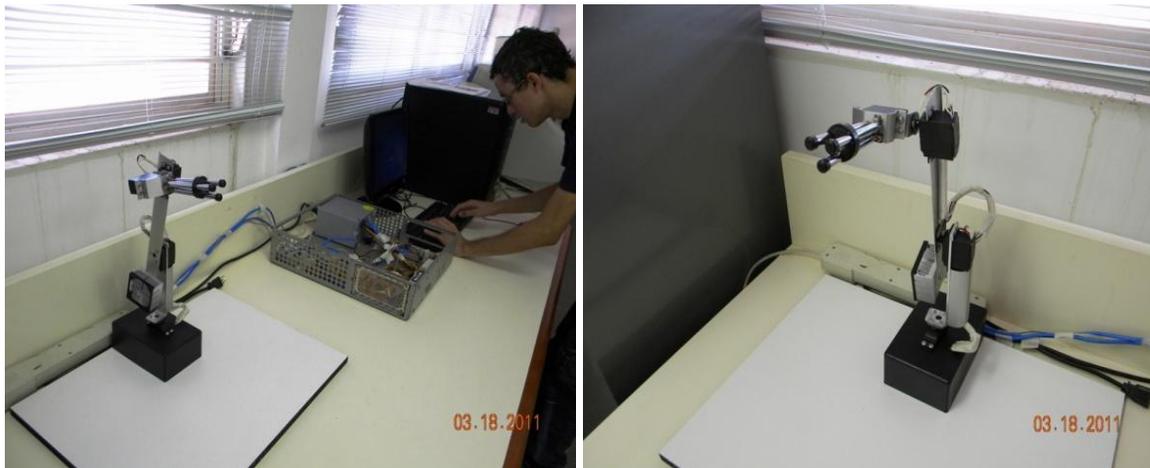


Figure 6 – Real system for fuzzy controller

### 3.2 Servo Motors

The internal circuits of the servo-motors had been removed and inserted a precision potentiometer, keeping the mechanism of reduction and engines. The used motors had been the Futaba 3003, figure 7. The Linear Potentiometer had been modified.



Figure 7- Servo Motors Modified

### 3.3 Structure

All structure was mounted in aluminum with a resistant hard plastic base, afixado on a MDF plate. The material of aluminum pipes was molded specifically for this project. The System possesss 4 DOF with an actuator with 3 clamps. The system is capable to get rubber balls, figure 8.

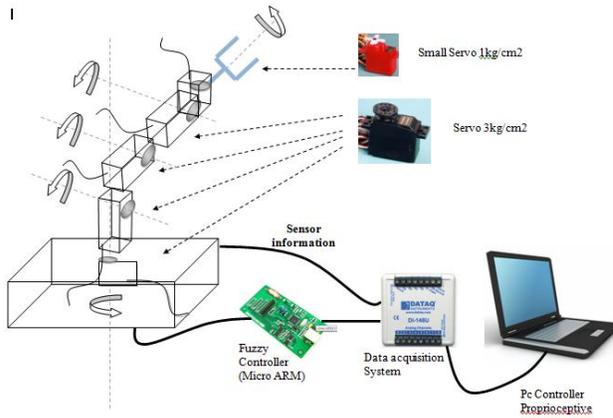


Figure 8 – General Control system

### 3.4 Fuzzy Control Board

The control Board use ARM 7 Microcontroller, equipped with protocol TCP/IP and the digital systems, PWM, converters AD and DA. Have protocol for memory card and pendrive. Have real time clock and innumerable digital communication port, figure 9.



Figure 9 – Microcontroller Arm Board

### 3.5 Computer Interface

External Board of acquisition Dataq Instruments DI-148U, with 5 ports of Digital I/O and 8 A/D ports.



Figure 10-Dataq Board

USB connection with the PC and compatible software with systems that use the technology ActiveX, and C++ programming.

### 3.6 Circuit driver

Actuators for the engines using a circuit with bridge H modified for use as PWM. The circuit uses transistor FET for engines of until 10A, figure 11.

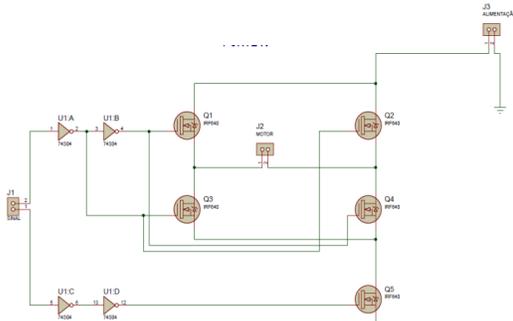


Figure 11 – Diagram driver

### 3.7 Software for position reading and control.

The software for test the algorithms are showed in figure 12. Where ADG is the signal read from actuator, AD2 is de results of join 2, AD1 is result of

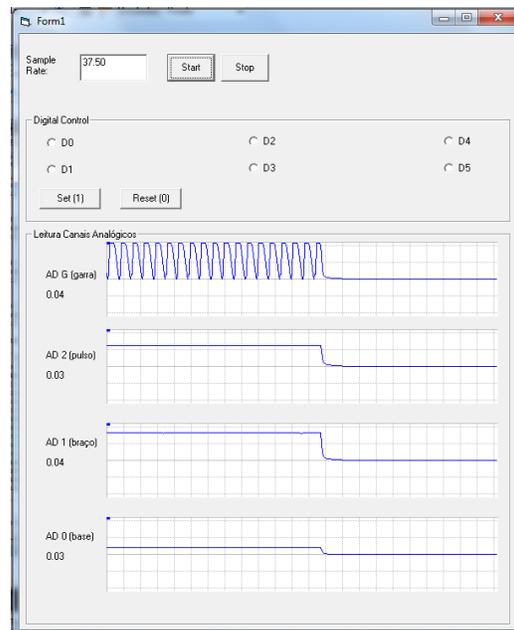


Figure 12 – Software for Proprioceptive Control

join 1 and AD0 is the result from join 0 or join from base. In results all join moved on same instant.

#### 4. CONCLUSIONS

The results obtained from the lab experiments show that robot system using de fuzzy control is a arrangement possible. The fuzzy control sends to Proprioceptive software and the robot system dealing the best trajectory for the robot.

All circuits and hardware developed for this project have been showed efficient. The result obtained for enhancement using fuzzy logic techniques, show us effective, very speed and of the simple implementation.

In figure 12 the robot start the process of movement, in this moment the algorithm compute all nodes of links and search for a best trajectory.

A new stage in the robotic system must avoid obstacles perceived by sensors attached to the arm, in order to avoid those obstacles and accomplish their tasks.

#### 4-Bibliography and References

- [1] Okko H. Bosgra, Huibert Kwakernaak and Gjerrit Meinsma, “**Design Methods for Control Systems**”, 2009.
- [2] Hunt, R., Su, R. Meyer, G. - “**Global Transfomation of Nonlinear Systems**”, IEEE Trans. Aut. Control, AC-28, N°.1, pp 24-31, 1983.
- [3] Markus Eich, Felix Grimminger, Frank Kirchner DFKI Bremen “**Proprioceptive Control of a Hybrid Legged-Wheeled Robot (German Research Center For Artificial Intelligence)**“, 28359 Bremen, Germany
- [4] Dmitry Kaynov, “**TESIS DOCTORAL Open motion control architecture for humanoid robots**” Leganés, octubre 2008
- [5] <http://en.wikipedia.org/wiki/Proprioception>
- [6] Jakubczyk, B. “**Feedback Linearization of Discrete Time Systems**”, Systems And Control Letters, N°. 9, pp 411-416, 1987a.
- [7] G. Benet\*, F. Blanes, J.E. Simó, P. Pérez, “**Using infrared sensors for distance measurement in mobile robots**” Robotics and Autonomous Systems 40 (2002) 255–266.
- [8] Meyer, G., Cicolani, L. - “**Application for Nonlinear System Inverses to Automatic Flight Control Design-System Concepts and Flight Evaluation**”, Theory and Aplications of optiomal Control in Aerospace systems, P Kent. Ed. Nato, 1980.
- [9] Matej Andrejasi, “**MEMS ACCELEROMETERS**”, University of Ljubljana - Faculty for mathematics and physics - Department of physics (seminar)
- [10] Andrew Whitworth ,“**Control Systems And Control Systems Engineering With Classical and Modern Techniques And Advanced Concepts**”, [http://upload.wikimedia.org/wikipedia/commons/e/e4/Control\\_Systems.pdf](http://upload.wikimedia.org/wikipedia/commons/e/e4/Control_Systems.pdf)
- [11] Elmo B. Faria, Ghilson R. Corrêa & Edilberto P. Teixeira, “ **Implementation of Neural Networks controller form Robot Manipulates** ”, Proceedings of the IV Brazilian Conference on Neural Networks - IV Congresso Brasileiro de Redes Neurais pp. 350-355, July 20-22, 1999 - ITA, São José dos Campos - SP – Brazil.
- [12] Martin T. Hagan, Howard B. Demuth, “**Neural Networks for Control**” School of Electrical & Computer Engineering Oklahoma State University and Electrical Engineering Department University of Idaho.
- [13] K.S. Narendra, and K. Parthasarathy, “**Identification and Control of Dynamical Systems Using Neural Networks,**” *IEEE Transactions on Neural Networks*, vol. 1, pp. 4-27, 1990.
- [14] K.S. Narendra and B. Balakrishnan, “**Improving Transient Response of Adaptive Control Systems Using Multiple Models and Switching,**” *IEEE Transactions on Automatic Control*, vol. 39, no. 9, pp. 1861-1866, 1994.
- [15] R.S. Sutton, and A.G. Barto, “**Introduction to Reinforcement Learning**”, Cambridge, Mass.: MIT Press, 1998.
- [16] E.P. Nahas, M.A. Henso and D.E. Seborg, “**Nonlinear Internal Model Control Strategy for Neural Models**”, *Computers and Chemical Engineering*, vol. 16, pp. 1039-1057, 1992.
- [17] M. Kawato, “**Computational Schemes and Neural Network Models for Formation and Control of Multijoint Arm Trajectory,**” *Neural Networks for Control*, W.T. Miller, R.S. Sutton, and P.J. Werbos, Eds., Boston: MIT Press, pp. 197- 228, 1990.
- [18] L.G. Kraft and D.P. Campagna, “**A Comparison between CMAC Neural Network Control and Two Traditional Control Systems,**” *IEEE Control Systems Magazine*, vol. 10, no. 2, pp. 36-43, 1990.
- [19] W.T. Miller, “**Sensor-Based Control of Robotic Manipulators Using a General Learning Algorithm,**” *IEEE Journal of Robotics and Automation*, vol. 3, no. 2, pp. 157-165, 1987.
- [20] E. B. Faria, “**How Linear Is A System Feedback Linearized By A Neural Networks. International Conference On Neural Networks**”, ICNN -IEEE\1997, 1997.
- [21] Michael B. Histan & david G. Alciatore, “**Introduction to Mechatronics and Measurement Systems**”, McGraw-Hil, 1998
- [22] Dan Neacsulescu, “**Mechatronics**” Prentice Hall, 2002.
- [23] Sastry, S. & Isidori, A.- “**Adaptative Control of Linearizable Systems**”, IEEE Trans. On Autom. Control, Vol.34, N°. 11, pp.1123-1131, November 1989.

## Tracking Performance Improvement for a Parallel Robot based on Synchronized Control

Hyun Min DO    Chan Hun PARK    Dong Il PARK    Byung In KIM    Jin Ho KYUNG  
Robotics and Mechatronics Research Center, Nano Convergence Systems Research Division,  
Korea Institute of Machinery and Materials,  
104 Sinseongno, Yuseong-gu, Daejeon, 305-343, Korea  
hmdo@kimm.re.kr

### ABSTRACT

This paper presents a design method of synchronization error based control scheme for a parallel robot. Since the parallel robot has the closed-loop kinematic structure, the tracking error of one arm was tightly affected by the other arms. Thus, the coordination of multiple arms should be emphasized. To improve the overall performance of a parallel robot, the control approach using the synchronization error is proposed. In the proposed method, the tracking error and the synchronization error are considered at the same time. Some simulation results are included to verify the performance of the proposed method.

Keywords: Parallel Robot, Synchronization Error, Tracking Error, Coordination and PID Control.

### 1. INTRODUCTION

Parallel robots have advantages of high stiffness, high speed, high accuracy and high payload compared to their counterparts, the serial robots, thanks to their closed-chain mechanism [1, 2]. These merits make parallel robots be proper for high performance motion platforms and thus parallel robots are mainly used in the industrial area.

To realize the high speed and high accuracy motion, the important things are the design of a mechanism and the design of control systems. Optimal design problem was studied in [3]. In this paper, we focus on the design of the control algorithm. Amongst most of conventional control method, an independent control scheme is widely used, i.e., each actuator is controlled by each controller independently to track the reference trajectory. Thus the controller of each actuator does not receive the information of the tracking error from the others. However, since the mechanism of a parallel robot is closed-loop kinematic chain, each actuator

takes effect on each other. Thus the tracking error of other manipulators should be considered to control each actuator precisely.

Many researches have been conducted in the field of control systems for parallel robots. Design of independent joint controllers was suggested in [4]. Some studies address the design for control to overcome difficulties of control problem [5]. Several studies on synchronized control of parallel robot have been done [6-9]. However, the target model of those works was planar type parallel manipulator. A synchronization error was defined in several forms according to the purpose of a control in each approach. In this paper, we propose a new control approach to improve the accuracy of the tracking performance of 3-DOF parallel robot. A synchronization error was defined to measure the difference of errors between one joint and the others. A synchronization error and a tracking error have merged to form a coupled error which will be used in the feedback signal to PID controllers. Due to the PID controllers, both errors can be reduced and the position accuracy of each manipulator can be improved.

This paper is organized as follows. Section 2 describes the parallel robot which is considered as the target model of this paper. Section 3 addresses the newly proposed control algorithm based on synchronization error. In Section 4, some numerical results are suggested to demonstrate the validity of the proposed approach. Conclusions and future works are drawn in Section 5.

### 2. PARALLEL ROBOT MODEL

In this paper, the considered model of a parallel robot is 3-DOF manipulator consisted of 3 arms (Delta robot). Overall structure is like Fig. 1. The base of a robot requires high stiffness, and thus it is arranged in a circle with 3 motors. The base of a robot and the

moving platform is connected with 3 arms, where each arm is consisted of 2 links. Inner arm is a first link which is connected to a motor. Outer arm has a pair of links and they make a parallelogram. Components of outer arm are ball joints and aluminum pipes. The moving platform is a part in which 3 links are closed and will be equipped with a tool for an operation.

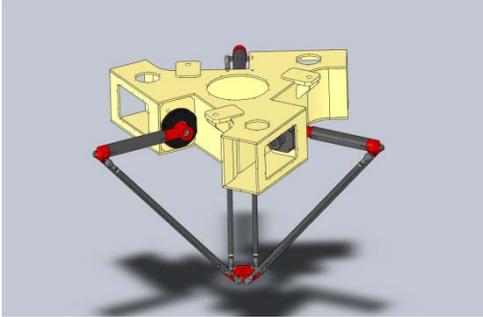


Fig. 1. The structure of the considered parallel robot

### 3. SYNCHRONIZED CONTROL

#### A. Synchronization error

As for parallel manipulators with  $n$  actuators, let  $q_i$  ( $1 \leq i \leq n$ ) be the position of  $i$ -th active joint and  $q_i^d$  ( $1 \leq i \leq n$ ) be the desired trajectory of  $q_i$ . Then the tracking error can be defined as

$$e_i = q_i^d - q_i, \quad 1 \leq i \leq n. \quad (1)$$

From the concept of the synchronization among multiple manipulators, the synchronization error  $\varepsilon_i$  of the  $i$ -th active joint between two neighboring joints ( $(i-1)$ -th joint and  $(i+1)$ -th joint) can be defined as follows.

$$\begin{aligned} \varepsilon_i &= (e_i - e_{i-1}) + (e_i - e_{i+1}), \quad 2 \leq i \leq n-1 \\ \varepsilon_1 &= (e_1 - e_n) + (e_1 - e_2), \quad i = 1 \\ \varepsilon_n &= (e_n - e_{n-1}) + (e_n - e_1), \quad i = n. \end{aligned} \quad (2)$$

Then, the tracking error and the synchronization error can be combined into the coupled error  $e_i^{coupled}$  as eqn. (3).

$$e_i^{coupled} = e_i + \gamma \cdot \varepsilon_i, \quad 1 \leq i \leq n, \quad (3)$$

where  $\gamma$  ( $0 \leq \gamma \leq 1$ ) is a constant parameter which means the portion of the synchronization error of overall error. Finally, the coupled error will be fed back to each active actuator, which contains the tracking error of each actuator and the degree of coordination of among multiple actuators.

#### B. Design of controller

For the trajectory tracking control of a parallel robot, model-based control and model-free control can be considered in large. In this paper, we adopt a PID control, which is a conventional model-free feedback control and a feedforward control to achieve a faster settling time for high speed motion in addition. Without consideration of the synchronization error, only the tracking error  $e_i$  of each active joint is fed back to controller. For the feedforward controller, also, the desired trajectory (including position, velocity and acceleration trajectory) of each joint is used for the control input respectively.

In the proposed schemes, the coupled error is used instead of the tracking error. Thus, the control input for  $i$ -th active joint can be written by eqn. (4),

$$\begin{aligned} u_i &= u_i^{feedback} + u_i^{feedforward} \\ &= \left[ K_p e_i^{coupled} + K_d \dot{e}_i^{coupled} + K_i \int e_i^{coupled} \right] \\ &\quad + \left[ K_{pff} q_i^d + K_{vff} \dot{q}_i^d + K_{aff} \ddot{q}_i^d \right], \end{aligned} \quad (4)$$

where  $K_p$ ,  $K_d$ ,  $K_i$  are feedback gains for PID control and  $K_{pff}$ ,  $K_{vff}$ ,  $K_{aff}$  mean gains for the position trajectory, velocity trajectory and acceleration trajectory in the feedforward control. Then, the overall scheme in the case of the synchronized control is depicted in Fig. 2.

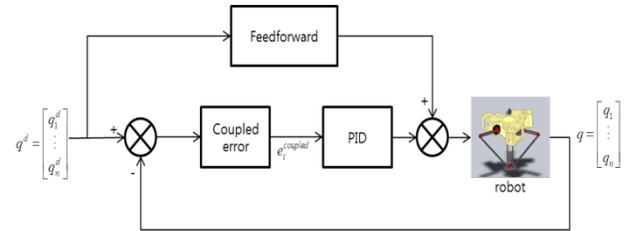


Fig. 2. Overall control scheme

Since the considered parallel robot is a 3-DOF manipulator as described in section 2, the synchronization error can be defined as eqn. (5),

$$\begin{aligned} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} &= \begin{bmatrix} (e_1 - e_3) + (e_1 - e_2) \\ (e_2 - e_1) + (e_2 - e_3) \\ (e_3 - e_2) + (e_3 - e_1) \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = S \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}. \end{aligned} \quad (5)$$

Then the coupled error can be written as

$$\begin{bmatrix} e_1^{coupled} \\ e_2^{coupled} \\ e_3^{coupled} \end{bmatrix} = (I + \gamma \cdot S) \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}, \quad (6)$$

where

$$S = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

#### 4. PERFORMANCE EVALUATION

The performance of the proposed approach is now shown through some numerical simulations. The model of a parallel robot was implemented using RoboticsLab developed by SimLab Co., Ltd. For the test motion, the well-known ‘‘Adept cycle’’ was implemented. It represents a typical pick-and-place task, and first requires a vertical motion at the picking location (up 25mm), then a linear horizontal motion (305mm), a vertical motion at the placing location (down 25mm), and the same trajectory back [3]. The trajectory is shown in Fig. 3.



Fig. 1. Adept cycle motion

Then, tracking results were suggested with synchronization control and without synchronization control. Also, simulations were made for two kinds of payload, i.e., no payload and 3kg. Fig. 4 shows the simulation model of a parallel robot. The payload is shown in the bottom of Fig. 4.

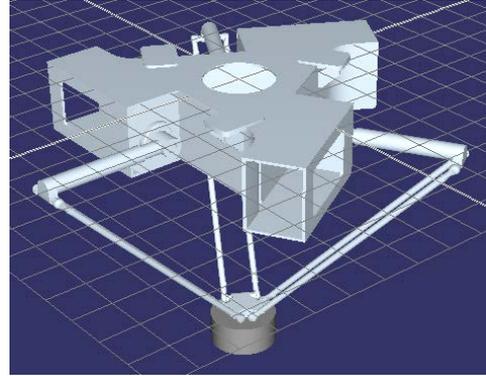


Fig. 4. The simulation model

#### 4. SIMULATION RESULTS

##### 1) In the case of no payload

Since the reference trajectory is a linear vertical and horizontal motion and the x-axis of a coordinate system of a moving platform was matched with joint 0, the motion of joint 1 and joint 2 are same. Thus we suggest a result of joint 0 and joint 1. Fig. 5 shows the motion trajectory of each joint when the synchronized control was adopted.

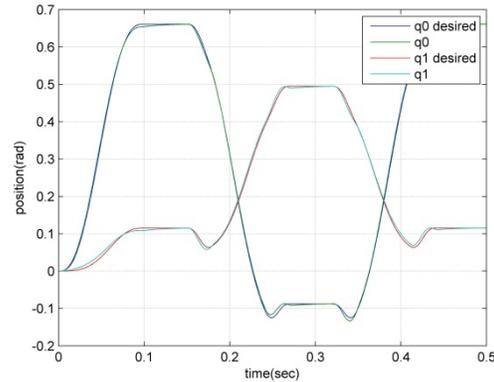
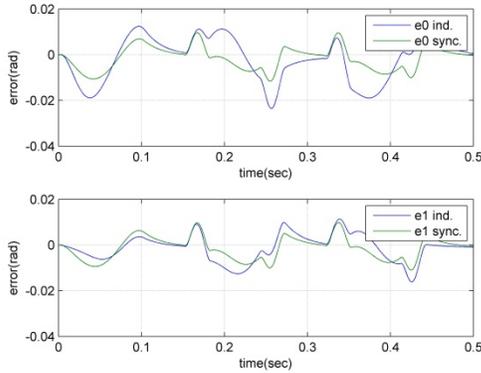
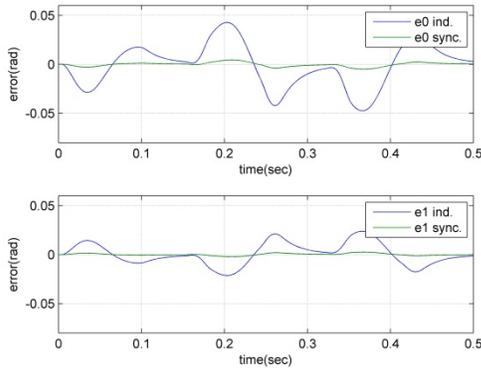


Fig. 2. The motion trajectory of synchronized control (no payload)

In this experiment, the error trajectory and synchronization error trajectory of each joint were shown in Fig. 6 and Fig. 7 respectively, where upper one is for joint 0 and lower one is for joint 1. From the error trajectory, we can see that the tracking performance is improved in the case of synchronized control. Also, the synchronization error has greatly improved as shown in Fig. 7. The blue line means the case of independent control and the green line means the case of synchronized control.



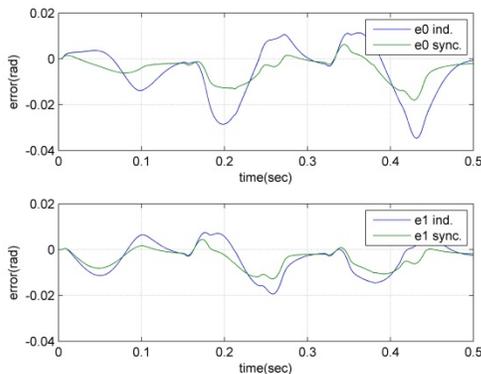
**Fig. 3. The error trajectory (no payload)**



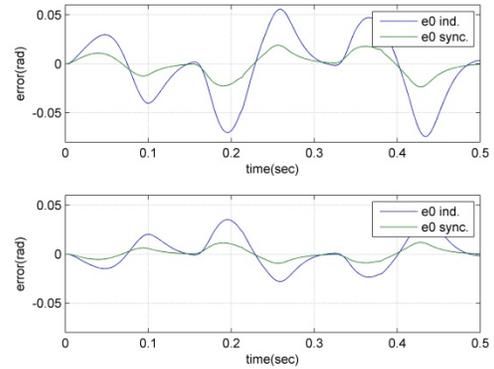
**Fig. 4. The synchronization error trajectory (no payload)**

2) *In the case of 3kg payload*

Now the payload is increased to 3kg and the same simulation was conducted. As the payload is increased, the error has grown worse. However, the performance of the proposed approach can still be verified. The error trajectory and synchronization error trajectory of each joint were shown in Fig. 8 and Fig. 9 respectively. We can see that the tracking performance is improved in the case of synchronized control.



**Fig. 5. The error trajectory (3kg payload)**



**Fig. 6. The synchronization error trajectory (3kg payload)**

In table I, the maximum and minimum errors were suggested in each case. Thus the simulation results can be compared clearly.

Table 1 COMPARISON OF MAXIMUM AND MINIMUM ERRORS (RADIAN)

PAY LOAD	ERROR TYPE	JOINT #	MAX /MIN	INDEPENDENT CONTROL	SYNCHRONIZED CONTROL
0KG	TRACKING ERROR	0	MAX	0.0133	0.0096
			MIN	-0.0236	-0.0116
		1	MAX	0.0114	0.0099
			MIN	-0.0162	-0.0110
	SYNC ERROR	0	MAX	0.0427	0.0044
			MIN	-0.0476	-0.0050
1		MAX	0.0238	0.0025	
		MIN	-0.0214	-0.0022	
3KG	TRACKING ERROR	0	MAX	0.0114	0.0062
			MIN	-0.0346	-0.0180
		1	MAX	0.0078	0.0045
			MIN	-0.0194	-0.0128
	SYNC ERROR	0	MAX	0.0556	0.0189
			MIN	-0.0744	-0.0236
		1	MAX	0.0372	0.0118
			MIN	-0.0280	-0.0095

**5. CONCLUSIONS**

This paper proposed the control approach based on the synchronization error to improve the tracking performance and to coordinate multiple arms of a parallel robot. The synchronization error was defined and merged into the coupled error with the tracking error and the coupled error was used as feedback signal to conventional PID controller. The performance of the proposed scheme was verified through some numerical simulations. As future works, we address the

application of the proposed method to the prototype of a parallel robot.

#### REFERENCES

- [1] C. R. Boër, L. Molinari-Tosatti and K. S. Smith(Eds), *Parallel Kinematic Machines*, Springer, 1999.
- [2] J.-P. Merlet, *Parallel Robots*, Kluwer Academic Publishers, 2000.
- [3] F. Peirrot, V. Nabat, O. Company, S. Krut and P. Poignet, "Optimal Design of a 4-DOF Parallel Manipulator: From Academia to Industry," *IEEE Trans. Robot.*, Vol. 25, No. 2, pp.213-224, 2009.
- [4] P. Chiacchio, F. Pierrot, L. Sciavicco and B. Siciliano, "Robust Design of Independent Joint Controllers with Experimentation on a High-Speed Parallel Robot," *IEEE Trans. Ind. Electron.*, Vol. 40, No. 4, pp.393-403, 1993
- [5] Q. Li and F. X. Wu, "Control performance improvement of a parallel robot via the design for control approach," *Mechatronics*, Vol. 14, pp.947-964, 2004.
- [6] Y. X. Su, D. Sun, L. Ren, X. Wang and J. K. Mills, "Nonlinear PD Synchronized Control for Parallel Manipulators," in *Proc. IEEE Int'l Conf. Robot. Automat.*, pp.1386-1391, 2005.
- [7] L. Ren, J. K. Mills and D. Sun, "Performance Improvement of Tracking Control for a Planar Parallel Robot Using Synchronized Control," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems*, pp.2539-2544, 2006.
- [8] L. Ren, J. K. Mills and D. Sun, "Experimental Comparison of Control Approaches on Trajectory Tracking Control of a 3-DOF Parallel Robot," *IEEE Trans. Contr. Syst. Technol.*, vol. 15, no. 5, pp.982-988, Sep. 2007.
- [9] W. Shang, S. Cong, Y. Zhang and U. Liang, "Active Joint Synchronization Control for a 2-DOF Redundantly Actuated Parallel Manipulator," *IEEE Trans. Contr. Syst. Technol.*, vol. 17, no. 2, pp.416-423, Mar. 2009.

# On the Design of a Bluetooth Data Acquisition Card for the Control of Manipulator Robots

A. Donado, J. Castañeda-Camacho F. Reyes, G. Mino and G. Muñoz-Hernández.  
 Facultad de Electrónica de la Benemérita Universidad Autónoma de Puebla.  
 Puebla, Puebla, 72000, México.

## ABSTRACT

A simple and successful design of a Bluetooth data acquisition card for the control of manipulator robots is presented. The integration of software through a program in Visual C# language and the use of a microcontroller PIC18F4550 embedded with a bluetooth module, give rise to a wireless data acquisition card with digital and analog inputs and outputs. This embedded system is applied to the control of a manipulator robot with three degrees of freedom.

**Keywords:** Data acquisition card, manipulator robots and bluetooth.

## 1. INTRODUCTION

Bluetooth wireless communication is a wireless LAN technology designed to operate in an environment of many users to connect devices with different functions such as telephones, computers, cameras, printers, etc [1]. A Bluetooth LAN is an ad hoc network that is formed spontaneously and provides support for three general application areas using short range wireless connectivity.

**Data and voice access points.** Bluetooth facilitates real-time voice and data transmissions by providing effortless wireless connection of portable and stationary communications devices.

**Cable replacement.** Bluetooth eliminates the need for numerous, often proprietary, cable attachments for connection of practically and kind of communication device. Connections are instant and are maintained even when devices are not within line of sight. The range of each radio is approximately 10 m, but can be extended to 100 m with an optional amplifier.

**Ad hoc networking.** A device equipped with a Bluetooth radio can establish instant connection to another Bluetooth radio as soon as it comes into range.

The Bluetooth technology regulated by the protocol IEEE 802.15 is defined as a layered protocol architecture consisting of core protocols, cable replacement and telephony control protocols, and adopted protocols [1, 2]. The Bluetooth core protocols form a five-layer stack consisting of the following elements as it is shown in Fig. 1.

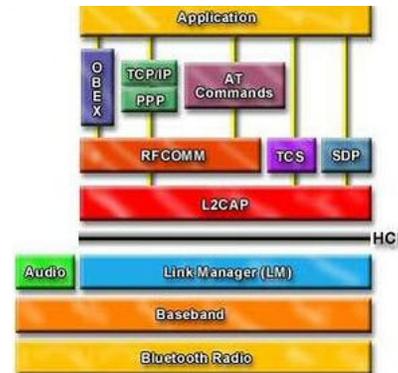


Fig. 1. Bluetooth Stack

**Radio.** Specifies details of the air interface, including frequency, the use of frequency hopping, modulation scheme, and transmit power.

**Baseband.** Concerned with connection establishment within a piconet, addressing, packet format, timing and power control.

**Link manager protocol (LMP).** Responsible for link setup between Bluetooth devices and ongoing link management. This includes security aspects such as authentication and encryption, plus the control and negotiation of baseband packet sizes.

**Logical link control and adaptation protocol (L2CAP).** Adapts upper-layer protocols to the baseband layer. L2CAP provides both connectionless and connection-oriented services.

**Service discovery protocol (SDP).** Device information, services and the characteristics of the services can be queried to enable the establishment of a connection between two or more Bluetooth devices.

RFCOMM is the cable replacement protocol included in the Bluetooth specification. RFCOMM provides binary data transport and emulates EIA-232 control signals over the Bluetooth baseband layer. EIA-232 (formerly known as RS-232) is a widely used serial port interface standard.

In this work we design of a Bluetooth data acquisition card for the control of manipulator robots, concentrating on the RFCOMM protocol which is a radio frequency emulator oriented to a computer COM port, and the AT commands which are used for configuring Bluetooth devices [1- 3].

The authors would like to thank the PROMEP project 103.5/08/3343 and anonymous reviewers for their valuable comments and suggestions that enhanced the quality of this work.

## 2. DESIGN OF DATA ACQUISITION CARD (DAQ)

The data acquisition card is designed through the module Parani ESD 1000 with 8 digital inputs, 8 digital outputs, 6 analog inputs and 2 analog outputs for pulse-width modulation (PWM). This includes USB interface, Bluetooth interface and dedicated lines for USB PIC boot loader, in circuit serial programming (ISCP).

As we can see in Fig. 2, the core of the card is the PIC18F4550 microcontroller [4-6].

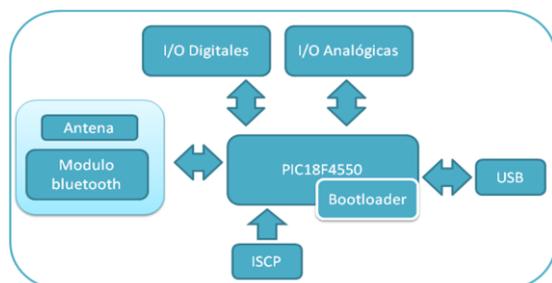


Fig. 2. Block Diagram of the DAQ

This family of devices offers the advantages of all PIC18F24550 microcontrollers namely, high computational performance at an economical price with the addition of high endurance, enhanced flash program memory [7]. In addition to these features, the PIC18F24550 family introduces design enhancements that make these microcontrollers a logical choice for many high-performance, power sensitive applications. All of the devices in the PIC18F4550 family incorporate a range of features that can significantly reduce power consumption during operation. Devices in the PIC18F4550 family incorporate a fully featured Universal Serial Bus communications module that is compliant with the USB Specification Revision 2.0. The module supports both low-speed and full-speed communication for all supported data transfer types. It also incorporates its own on-chip transceiver and 3.3V regulator and supports the use of external transceivers and voltage regulators. All of the devices in this family offer twelve different oscillator options, allowing users a wide range of choices in developing application hardware.

Asynchronous dual clock operation, allowing the USB module to run from a high-frequency oscillator while the rest of the microcontroller is clocked from an internal low-power oscillator.

Besides its availability as a clock source, the internal oscillator block provides a stable reference source that gives the family additional features for robust operation:

- Fail-Safe Clock Monitor: This option constantly monitors the main clock source against a reference signal provided by the internal oscillator. If a clock failure occurs, the controller is switched to the internal oscillator block, allowing for continued low-speed operation or a safe application shutdown.
- Two-Speed Start-up: This option allows the internal oscillator to serve as the clock source from Power-on Reset, or wake-up from Sleep mode, until the primary clock source is available.

### Other Special Features

- Memory Endurance: The Enhanced Flash cells for both program memory and data EEPROM are rated to last for many thousands of erase/write cycles – up to 100,000 for program memory and 1,000,000 for EEPROM. Data retention without refresh is conservatively estimated to be greater than 40 years.

- Self-Programmability: These devices can write to their own program memory spaces under internal software control. By using a bootloader routine, located in the protected Boot Block at the top of program memory, it becomes possible to create an application that can update itself in the field.

- Extended Instruction Set: The PIC18F24550 family introduces an optional extension to the PIC18 instruction set, which adds 8 new instructions and an Indexed Literal Offset Addressing mode. This extension, enabled as a device configuration option, has been specifically designed to optimize re-entrant application code originally developed in high-level languages such as C.

- Enhanced CCP Module: In PWM mode, this module provides 1, 2 or 4 modulated outputs for controlling half-bridge and full-bridge drivers. Other features include auto-shutdown for disabling PWM outputs on interrupt or other select conditions and auto-restart to reactivate outputs once the condition has cleared.

- Enhanced Addressable USART: This serial communication module is capable of standard RS-232 operation and provides support for the LIN bus protocol. Other enhancements include Automatic Baud Rate Detection and a 16-bit Baud Rate Generator for improved resolution. When the microcontroller is using the internal oscillator block, the EUSART provides stable operation for applications that talk to the outside world without using an external crystal (or its accompanying power requirement).

- 10-Bit A/D Converter: This module incorporates programmable acquisition time, allowing for a channel to be selected and a conversion to be initiated, without waiting for a sampling period and thus, reducing code overhead.

- Dedicated ICD/ICSP Port: These devices introduce the use of debugger and programming pins that are not multiplexed with other microcontroller features. Offered as an option in select packages, this feature allows users to develop I/O intensive applications while retaining the ability to program and debug in the circuit.

### Configuration and Data Transmission of the DAQ.

The DAQ can be configured to be connected with other bluetooth device knowing the Media Access Control (MAC) address. Fig. 3 shows a schematic of the card. This device can operate in two different modes DAQ Bluetooth and DAQ USB – Bluetooth.

- 1) DAQ Bluetooth: Receives and transmits information on the card. The Bluetooth module configuration is made directly by reading a dedicated input pin for this purpose.

- 2) DAQ USB – Bluetooth: The card is connected via USB to the computer as a data acquisition card USB, sending and receiving the information to another Bluetooth device.

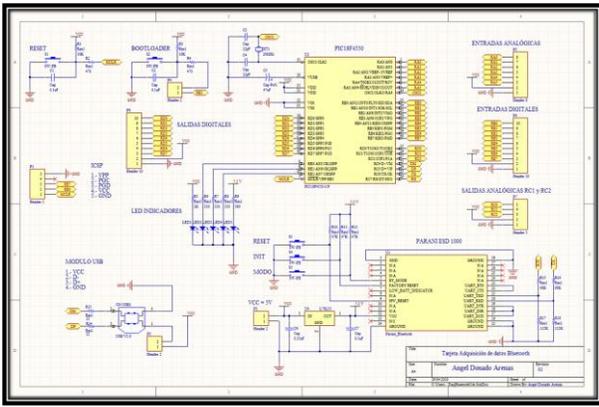


Fig. 3. Bluetooth DAQ Schematic

The programming of the firmware is developed in C, using the PIC C compiler for the recognition of the card in USB mode through drives of Microchip mchpusb.cat, mchpusb.sys, mchpusb64.sys, and mchpusb.inf. The configuration and ad hoc connection with another Bluetooth device is designed via AT commands, either stored in the microcontroller or transmitted through the user interface working in USB mode [8]. All the information needed in the Bluetooth DAQ mode is stored in the ROM of the PIC which is activated through the sensing PIN C0. The DAQ in any of two modes of operation works as the master who initiates the connection. Others Bluetooth devices in a computer can be used as slaves. The algorithm of the microcontroller has four important steps. It is shown in Fig. 4.

1. Configuration of the Bluetooth module, through the detection of the C0 PIN, which executes the code stored in the CIP for this purpose.
2. Connection or disconnection of the module, with another Bluetooth device.
3. Configuration and mapping in the microcontroller, for the analog and digital inputs.
4. Configuration and mapping in the microcontroller, for the analog outputs - PWM and digital outputs.

**3. SOFTWARE**

User interface for controlling the card is held under the programming language Visual C. This consists of three parts Initialization and recognition of the card, Data Acquisition and Robot Control [9-13].

**Initialization and recognition of the card**

As we can see in Fig. 5, on this step, the card recognizes possible failures and connects the configuration via USB or via Bluetooth COM port. In USB mode it works as a Communications Device Class. When the card is installed and recognized by Windows, the Bluetooth module can be configured to transmit and acquiring data [9].

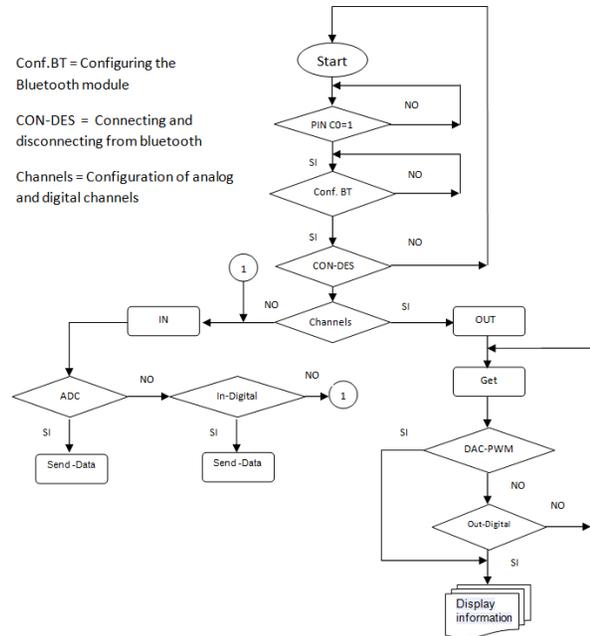


Fig. 4. Firmware algorithm

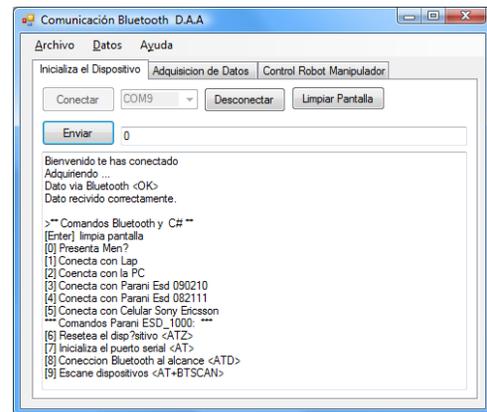


Fig. 5. Initialization and connection

**Data Acquisition**

The data acquisition window is depicted in Fig. 6. This stage is focused on acquiring and transmitting data in the Bluetooth data acquisition mode or as data acquisition USB-Bluetooth. The outputs are selected and the distance where the card is located is displayed using LEDs. Digital inputs are read from the card but only one analog channel can be read by sweeping 6 possible analog inputs. The analog outputs are transmitted through pulse width modulation.

**Robot Control**

In this step the card sends the paths, positions and commands to run the robot. For example, the position of the shoulder, elbow and wrist. Control and address data, are transmitted as a floating, detached by nibble, so that the robot can interpret them. Fig. 7 depicts a specific routine to control a robot.

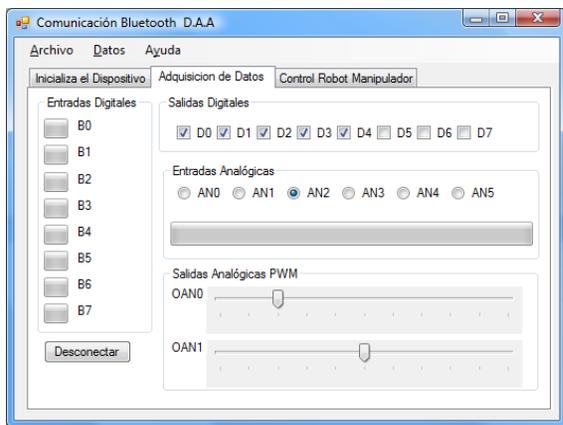


Fig. 6. Data Acquisition in C #

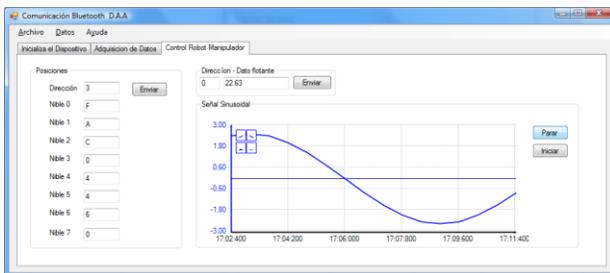


Fig. 7. Robot control

#### 4. APPLICATION

The control of an arm is a training platform for robot manipulators. This has theoretical and practical interest for experimental validation of new controller designs. An arm has been manufactured and built in the “Facultad de Ciencias de la Electrónica” of the “Benemérita Universidad Autónoma de Puebla”. The robot has a drive with three degrees of freedom [13, 14]. In the Fig. 8 we can observe the communication system.



Fig. 8. Communication System

The control algorithm, programming in Borland C, is loaded in the robot console. It receives external data through the parallel port. In Fig. 9 we can observe the experimental robot.



Fig. 9. Experimental Robot

In addition, the control of the robot needs a digital signal conditioning step. The circuit connected to the digital output port of the card is shown in Fig. 10.

Fig. 11 shows the complete Bluetooth DAQ.

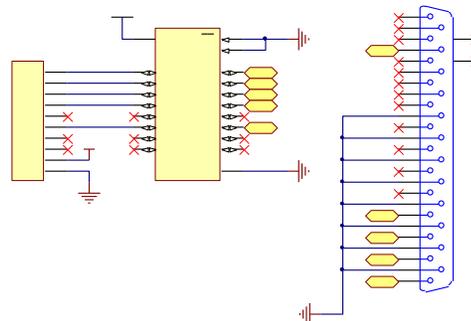


Fig. 10. Digital signal conditioning

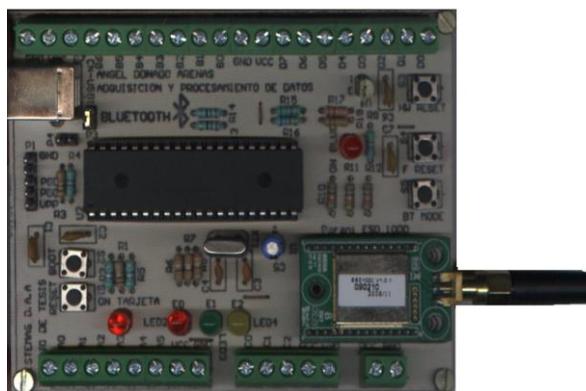


Figura 11. Bluetooth DAQ

#### 5. CONCLUSIONS

This paper has shown a simple but successful design of a Bluetooth data acquisition card for the control of manipulator robots. Industrial applications involving this system can benefit from the use of wireless communication technologies. The localization and tracking of components, the coordination of autonomous transport vehicles and mobile robots, as well as applications involving distributed control are all areas in which

wireless technologies could be used in an industrial environment.

### REFERENCES

- [1] W. Stallings, *Wireless Communications and Networking*. Prentice Hall.
- [2] IEEE Computer Society. Part 15.1: Wireless medium access control (MAC) and physical layer (PHY) specifications for wireless personal area networks (WPANs). IEEE Std 802.15.1 , New York, 14 June 2005.
- [3] Bluetooth SIG, Inc. BLUETOOTH SPECIFICATION Version 2.1 + EDR [vol 0]. The United State, 26 de July de 2007.
- [4] Bluetooth SIG, Inc. RFCOMM with TS 07.10 Serial Port Emulation, The United State, March 2003.
- [5] Amar Palacherla, Microchip Technology Inc. Using PWM to Generate Analog Output, U.S.A, 2002, pp. 1-4.
- [6] Ross M. Fosler and Rodger Richey, Microchip Technology Inc. A FLASH Bootloader for PIC16 and PIC18 Devices, U.S.A, 2002, pp. 1-38.
- [7] Microchip. PIC18F2455/2550/4455/4550 Data Sheet, 28/40/44-Pin, High Performance, Enhanced Flash, USB Microcontrollers with nano Watt Technology.
- [8] Sena Technologies, Inc. User Guide for the Parani-ESD1000, Korea, November 2009.
- [9] Custom Computer Services, Inc, PCD C Compiler Reference Manual, U.S.A, February 2010.
- [10] Rawin Rojvanit Microchip Technology Inc. Migrating Applications to USB from RS-232 UART with Minimal Impact on PC Software, U.S.A, 2004, pp. 1-16.
- [11] Axelson, J. *USB Complete: The Developer's Guide*. United States of America: Distributed by Independent Publishers Group, 2009.
- [12] Kelly R & Santibáñez V. *Control de Movimiento de Robots Manipuladores*. Madrid España: Person Educación, 2003.
- [13] Craig, J.J. *Introduccion to robotics: Mechanics and control*. Addison-Wesley, Reading, MA. 1989.
- [14] Arturo Corona, Jorge Moctezuma, Fernando Reyes, Roberto Campos, "Nueva plataforma de Entrenamiento para Robots Manipuladores", *El HOMBRE Y LA MAQUINA* numero 28, Colombia, enero-junio 2007, pp. 108-115.

# A Systems Approach for Disclosing the Dialectics of Emergent Technology-Based Organisational Performance along Axiological Dimensions

Stephen Corea  
University of Warwick,  
Coventry CV32 5NR, United Kingdom.  
steve.corea@wbs.ac.uk

## ABSTRACT

A major focus of information systems research is the emergent nature of organisational action around the use of information technologies: its contextual, evolutionary, and often unanticipated character. Many previous studies have approached this topic from the viewpoint of emergence as a process, led by social actors. However, the investigation of organisational emergence as a property has been neglected. The systems thinking approach is concerned with emergent properties, but has hitherto been inadequately developed for socio-technical analysis of organisational performance around the use of IT. This paper outlines a theoretical framework of interpretive systems inquiry for performing such analysis. It enables an understanding of technology-based organisational action as formed by the interaction of various elements/factors that are in relations of contrariety, contradiction or association. The emergent nature of an organisation's IT-based activities might thus be illuminated in terms of principle tensions or contradictions which may be easy to miss or difficult to pinpoint, and that shape the trajectory of organisational transformation, or constitute a persistent pattern of functioning. Such understanding thus provides a different but complementary perspective to the micro agency-centred model of emergence which has been dominant in recent information system studies.

Keywords: systems thinking, interpretive methodology, organisational performance, technology use, information

## 1. INTRODUCTION

In recent times, the emergent nature of organisational functioning around the use of information technologies (IT) has grown in importance as a research topic in the field of Information Systems (IS). Prior to this, the integration of IT into organisational processes had been widely promoted in from the standpoint of planned, deliberate changes, based on ideal visions/plans and rational projections [e.g. 14]. In practice however, it has been apparent that organisations are widely differentiated in their ability to effectively carry out such planned co-adaptations of work processes and IT (Scott-Morton, 1991). Numerous IS studies have evoked strong recognition that it is quite often the emergent nature

of such practices – their contextually-shaped, evolutionary, dynamic and often unanticipated nature – that significantly conditions the processes and outcomes of IT based work performance [3, 8]. This area of study is considered as “... particularly relevant today as unprecedented environmental, technological and organisational developments facilitate patterns of organising which cannot be explained or prescribed by appealing to a priori plans and intensions,” [25, p. 65].

Consequentially, a large amount of IS research studies, employing theoretical frameworks like structuration theory and actor-network theory [6, 12, 26], have investigated the way in which IT-based work practices in organisations are unexpectedly shaped by the unique socio-historical context that they are embedded within. These studies, collectively known as the ‘emergence paradigm’ [13], primarily stress an action-oriented, contextually-situated understanding of organisational functioning, that is based upon an enacted first-person perspective, and on studying emergence as a process. In other words, these emergence studies highlight the way in which the viewpoints, decisions and actions of key organisational actors can, as a ‘process’ over time, lead to emergent phenomena in IT and organisational practices. As a result, the IS field has gained considerable knowledge of organisational emergence in the use of IT tools from a processual, agency-based perspective.

Nevertheless there remain key areas of inadequacy in the understanding of this area. Firstly, as strongly argued by Shulman [4], past ‘IS emergence’ studies have produced rich insights on the structuring of human *action* around the use of IT, but they offer little guidance on understanding how effective organisational *performance* is enabled (or constrained) by the use of IT tools. In other words, the processual agency-based perspective developed in previous IS studies has failed to give an adequate account of the outcomes of emergent organisational functioning from the standpoint of the qualitative effectiveness (i.e. in terms of the value/significance) of IT usage and work performance.

Secondly, and more importantly, the use of systems thinking as a method for studying IS emergence has been conspicuously missing from recent theoretical approaches used to elaborate the emergent nature of IT based activity.

Systems thinking stresses the importance of understanding organisational functioning holistically as irreducibly-integrated ‘wholes’, and lays key emphasis on identifying *emergent* properties i.e. properties exclusive to such wholes and which are not deducible from any of their constituent elements. Salient reasons might be adduced for the relative absence of systems based inquiry in previous IS studies of the emergent nature of IT-based organisational functioning. The traditional systems theory approach came under strong criticism during the 1970/80s for being unable to deal with the social complexities and emergent meaning structures of organizations [7]. It was seen to be inadequate because of its essential premise of treating organisations as analogous to natural (i.e. biological) or mechanical systems, that have fixed, pre-determined goals: this key assumption foreclosed the consideration of any emergent, evolutionary modes of functioning [5, 7]. This premise significantly undermined the validity of traditional systems thinking as a theoretical platform for providing insightful description of the nature of dynamic organisational phenomena. Recent innovations in systems theory [16, 17] however have made significant strides towards repairing this deficiency. Nevertheless there remains much scope for developing the utility of modern systems thinking, specifically in clarifying the emergent nature of IT based organisational practices.

The research which I have undertaken [20, 21, 22] has sought to address these theoretical requirements: to develop and extend a systems-oriented understanding of emergent IT based organisational functioning, that permits qualitative appraisal of performance capabilities or effectiveness. The next section broadly summarises the approach I have taken.

## 2. THEORETICAL APPROACH

A fundamental premise of my approach is that in order to understand IT-enabled organisational *performance* (over and above organisational *action*), a theory of signification is required (i.e. to illuminate the structuring of significance or value of performance). I have accordingly turned to the use of theoretical concepts and analytical tools found in the field of socio-semiotics, which has developed advanced ways of conceptualising social/use value. In particular, I draw from the ideas of a major French semiotician who worked within a systems perspective, A. J. Greimas [1], whose work was concerned with the structuring of human behavior from a systemic, inter-relational viewpoint (i.e. in terms of contrary, contradictory and associative relations of significance) that helps distinguish axiological dimensions of value. I have adapted his work and ideas to frame the study of emergent IT based organisational action within the analytical lens of such ‘signification systems’ (patterned inter-relationships of value/significance).

In my studies [20, 21, 22], I have found the conceptual ability of Greimasian socio-semiotic analysis (i.e. to frame

positive and negative dimensions of significance/value in a co-constitutive way) to have been a very useful theoretical resource for generating insights, insofar as the emergent nature of organisations functioning and IT use appears to be due to the value or significance of performance being *phenomenologically reversible* – what is ‘good’ or effective performance to one group of stakeholders may be ‘bad’ performance to another – such reversibility is implicit in the important concept of the ‘interpretive flexibility’ of IT established by past IS emergence studies [e.g. 26].

The new systems framework adapted for sociotechnical analysis of organisational functioning is also (in addition to Greimasian socio-semiotic analysis) partly underpinned by Ricoeur’s [18] work in the hermeneutics (interpretation) of texts and social action. A significant achievement by Ricoeur had been to articulate the ontological basis for a fruitful integration between the hermeneutic method and structural-analytic techniques like systems thinking [24]. Therefore this adapted framework of ‘interpretive systems thinking’ [20, 23] combines the use of system-oriented theoretical concepts from Greimas with the methodological basis of Ricoeurian hermeneutics. This approach reflects an interpretivist ontology, in which an understanding of organisational reality is considered to be an inter-subjective construction of the researcher and the social actors who are involved in making sense of such reality. This furnishes the IS sociotechnical systems research [2] with an interpretive approach to data analysis (as an alternative to its reliance hitherto on the action research method). Butler [24] notes that while interest in hermeneutic theory has been created by prominent IS researchers [2, 15], there has hitherto been a lack of exposition on its use during empirical analysis of IT use in organisations.

## 3. UTILITY OF APPROACH

The primary utility of the method introduced above lies in advancing the application of an inter-relational, systems thinking approach for illuminating the social complexity of IT based organisational enterprise. It embodies the idea of system thinking promoted by key IS researchers [2, 14, 16, 19]. Systems thinking refers to “thinking in terms of facts and events in the context of wholes, forming integrated sets with their own properties and relationships” [9, p. 19]. This salient principle is the fundamental basis of the structured systems analysis method, traditionally used in design and construction of software, in which only *technical aspects* are considered i.e. the human behavior, work processes and information flows of an organisational domain are modeled as purely functional elements, and their interactions viewed in terms of cause-and-effect linkages or transformations. Although it has been successfully enlisted for developing IT systems, traditional systems analysis has been criticised as being conceptually ill-equipped for analyzing the use of IT in human milieus, because it only stresses the functional

or technical aspects, while neglecting the social aspects, of such activity [5, 10]. Traditional systems analysis has been seen as incapable of elucidating the social complexities of organizations, or addressing human processes of interaction and sense-making [5, 7].

By comparison, the interpretive systems analysis method I have sought to develop is, in a limited respect, analogous to the traditional systems analysis method: in which an area of activity is reduced to a combination of sub-processes analysed inter-relationally. However, this approach differs significantly from traditional systems analysis in that *social aspects* of human activity are also considered (e.g. power, legitimization, motivation), in addition to functional or technical aspects. Moreover, it is the analyst/researcher's understanding of a socially-situated domain of IT based organizational operations that is reduced to several subsets that are known as 'value programs' [23], and it is the inter-relationships of elements within and between the programs that are examined. The complex nature of an organisation can thus be analytically disengaged into 'whole' assemblies of actions, events, means and conditions, ordered by their correspondence to the prevailing social and/or technical concerns of stakeholders. This approximates the essence of organisations as being made up of nets of relations [11]. IT enabled organisational functioning can therefore be treated as a sociotechnical process of pluralistic constitution [3,], entailing multiple courses of action (either complementary and inconsistent) that elaborate it into an emergent form, having evolutionary, unexpected outcomes. The use of this approach to disentangle the disparate nature of social organisations, and to clarify the way in which various key features, functions or capacities of IT take on relevance for particular courses of action, can assist in identifying the trade-offs that management must juggle with [11].

Furthermore, the relationships between elements are not considered in terms of causative linkages or interactions, but as relations of significance/value. The core concept of 'signification systems' (i.e. patterned relations of value or significance) supports the representation and evaluation of organisational functioning in terms of a synoptic *whole*. A composite, inter-relational appraisal may thus be obtained, framed in terms of critical incompatibilities or tensions that shape IT-based organisational transformation/performance.

The early application of systems thinking in IS studies was hinged on the root idea of *natural/mechanical systems*. Recent useful innovations in this school of thought, such as the Soft Systems Methodology [16, 17], have significantly advanced its application towards the study of *designed systems*, in which systems concepts and ideas are used to clarify the learning process by which social actors decide on the feasibility of change in IT based work environments. The socio-semiotic approach that I have taken, may be seen as developing the focus of contemporary systems theory in

the direction of illuminating emergent organisational action or information systems usage as *signification systems*.

#### 4. CONCLUSION

The IS field has gained considerable knowledge of organisational emergence from a processual, agency-based perspective. However I have taken a different, relatively unexplored starting-point/emphasis, by studying emergence from 'the other side of the coin': a third-person analytical perspective, and a focus on emergence as a property.

Rather than accounting for the process of organisational emergence (in the use of IT) from the viewpoint of actors, I have developed an approach by which an analyst/researcher can usefully characterise the contextual shaping of IT based activities in an organisation from an 'outside-looking-in' viewpoint, and from an understanding of emergence as an outcome (i.e. a property) of those operations being seen as a complex whole. Systems thinking is the methodological basis for structuring this third-person, property-focused account of emergence in organisational functioning. This is fundamentally not incompatible with, or in strict opposition to, the accounts of past IS studies. However, this approach does provide significant new benefits and fresh insights. Its distinctive focus on dialectics and disequilibria contributes to the growing arsenal of organizational analysis techniques based on contemporary systems thinking.

#### 5. REFERENCES

- [1] A. Greimas, **On Meaning: Selected Writings in Semiotic Theory**, (trans. Perron, P. and Collins, F.), London: Frances Pinter, 1987.
- [2] A. Lee, "Researching MIS." In **Rethinking Management Information Systems**, Currie, W. and Galliers, B. (eds.), Oxford: Oxford University Press, 1999, pp. 7-27.
- [3] A. Lin and T. Cornford, "Socio-technical perspectives on emergence phenomena", In **The New Sociotech: Graffiti on the Long Wall**, Coakes, E., Lloyd-Jones, R. and Willis, D. (eds.), London: Springer-Verlag, 2000.
- [4] A. Shulman, "Putting Information Technology in its Place: Communication and Good Work Performance", In **Managing Organizations: Current Issues**, Clegg, S., Hardy, C. and Nord, W. (eds.), London: Sage, 1999, pp. 107-124.
- [5] C. Ciborra, "Management Information Systems: a Contractual View", In **Information Analysis:**

- Selected Readings.** Galliers, R. (ed.), Sydney: Addison-Wesley, 1987, pp. 125-137.
- [6] C. Ciborra, "A Theory of Information Systems based on Improvisation", In **Rethinking Management Information Systems**, Currie, W. and Galliers, B. (eds.), Oxford: Oxford University Press, 1999, pp. 136-155.
- [7] D. Silverman, **The Theory of Organizations: A Sociological Framework**, London: Hienemann, 1970.
- [8] D. Truex, R. Baskerville and H. Klein, "Growing Systems in Emergent Organizations", **Communications of the ACM**, Vol. 42, No. 8, 1999, pp. 117-123.
- [9] E. Lazlo, **The Systems View of the World**, New York: Braziller Press, 1972.
- [10] F. Emery and E. Trist, "Socio-Technical Systems", In **Management Science, Models and Techniques**, Churchman, C. and Verhulst, M. (eds.), Vol. 2, London: Pergamon, 1960, pp. 83-97.
- [11] G. Vickers, "Human Systems are Different," **Journal of Applied Systems Analysis**, Vol. 10, 1983, pp. 3-13.
- [12] G. Walsham, "Actor-Network Theory and IS Research: Current Status and Future Prospects," In Lee, A., Liebenau, J. and DeGross, J. (eds.), **Information Systems and Qualitative Research**, London: Chapman & Hall, 1997, pp. 466-480.
- [13] H. Klein, "Preface: the Potential Contribution of Semiotics and Systems Theory to the Continuing Evolution of Information Systems Research", In **Signs of Work: Semiosis and Information Processing in Organisations**, Holmqvist, B., Andersen, P., Klein, H. and Posner, R. (eds.), Berlin: Walter de Gruyter, 1996, pp. V-XVII.
- [14] M. Hammer, M. and J. Champy, **Reengineering the Corporation**, New York: Harper Business, 1993.
- [15] R. Boland, "Information system Use as a Hermeneutic Process," In **Information Systems research: contemporary approaches and emergent traditions**, Nissen, H., Klein, H. and Hirschheim, R. (eds.), Amsterdam: North-Holland, 1991, pp. 439-459.
- [16] P. Checkland, "Systems Thinking", In **Rethinking Management Information Systems**, Currie, W. and Galliers, B. (eds.), Oxford: Oxford University Press, 1999, pp. 45-56.
- [17] P. Checkland, **Systems Thinking, Systems Practice**, Chichester, UK: John Wiley, 1981.
- [18] P. Ricoeur, **Time and Narrative, Volume I**. (Trans. McKlaughlin, K. and Pellauer, D.), Chicago: University of Chicago Press, 1984.
- [19] R. Galliers, J. Mingers, and M. Jackson, "Organization Theory and Systems Thinking: the Benefits of Partnership", **Organization**, Vol. 4, No. 2, 1997, pp. 269-278.
- [20] S. Corea, "Refocusing Systems Analysis of Organizations through a Semiotic Lens: Interpretive Framework and Method", **Systemic Practice and Action Research**, Vol. 18, No. 4, 2005, p. 339-364.
- [21] S. Corea, "Mounting Effective IT based Customer Service Operations under Emergent Conditions: Deconstructing Myth as a basis of Understanding", **Information and Organization**, Vol. 16, No. 2, 2006, pp.109-142.
- [22] S. Corea, "Information Technology and the Modalisation of Organisational Behaviour: a Theoretical Framework", **Journal of Information Technology**, Vol. 21, No. 2, 2006, pp. 86-98.
- [23] S. Corea, "Value-Program Analysis: Capturing the Pluralistic Nature of IT Enabled Organizational Action", **Proceedings of IBIMA 2008**, the 10th International Business Information Management Conference, Kuala Lumpur, Malaysia, 2008.
- [24] T. Butler, "Towards a hermeneutic method for interpretive research in information systems", **Journal of Information Technology**, Vol. 13, No. 4, 1998, pp. 285-300.
- [25] W. Orlikowski, "Improvising organizational transformation over time: a situated change perspective", **Information Systems Research**, Vol. 7, No. 1, 1996, pp. 63-92.
- [26] W. Orlikowski, "Using Technology and Constituting Structures: a Practice Lens for Studying Technology in Organizations", **Organization Science**, Vol. 11, No. 4, 2000, pp. 404-428.

## Is Reality digital or analog?

Jeremy Horne

<http://home.earthlink.net/~jhorne18>

e-mail: [mindsurgeon@hotmail.com](mailto:mindsurgeon@hotmail.com)

### **Abstract**

Reality is presented to us both in a digital and analog manner, the first as evidenced by the findings about the nature of space and sub-atomic entities, and the latter by the uncertainties at the quantum level. It is not sufficient to regard reality, our universe, simply as both but dialectically, one in terms of the other. The most fundamental law of understanding is that we apprehend something in terms of what it is not. We need contradiction to discern anything. Our understanding is process based and has deep historical roots extending back more than 4500 years. In modern times, scientists have relied upon Cartesian reductionism to discern the nature of our world, but deduction also is inherently dialectic, relying upon induction for its integrity. Logic, itself, is the language of innate order in the universe, but its digital aspect is bi-valency that describes what digital physicists have found to be the case in our reality. Evidence is in the form of how the syntax of the binary system and relationships within it reflect what happens here. Such has major implications for us in the form of inherent computations, phenomena as illusions, and complexity arising from simplicity.

### **The question**

A brief answer to the 2011 WFXI challenge question, "Is Reality Digital or Analog?" is that reality seems to be both. The etymology of "digital" is "discrete," or distinct. "Digit" and its derivatives come from the Latin "digitus," meaning finger, the simplest form of expression being binary. "Analogue" refers to continuity - ἀνά, "up to" + λόγος "word, speech, reckoning," this etymology suggesting approximation or finding a likeness but not an exact match. An analogon is a comparison, hence, ratio. However, the distinction between digital and analog is expressed in more complicated ways, such as in particle versus wave and position versus momentum, the former of the pairs being the discrete and the latter continuous. Classical thinking focused on an "either-or" exclusionary description of something. Post Einsteinian thinking regarded the entity as both. A third way of thinking considers each but one in terms of the other.

### **Some problems inherent in answering the question and my caveats**

This non-trivially awesome question focuses upon determining the nature of reality, and it would be arrogant to think that one would have a definitive answer. Inextricably bound to an answer are issues such as epistemology (how we know), objectivity, and truth. The philosophical literature entangles one in endless debates about these, so while my reply to the question is not meant to be a contribution to that forum, I am aware of the basic issues not being finally resolved [1].

To claim what reality is flies in the face of Plato's famous assertions (cave allegory and divided line in the seventh book of *The Republic*, for example) that it only can be represented. Aristotle retorted throughout his *Metaphysics*, among other places, that reality is in front of us. Neither philosopher

provides anything more than speculation. If people know what reality is, they should be able unequivocally bring others to the same conclusion. Yet, this is not the case, or the question about the nature of reality and whether we are in it would not be persistent now. Not being certain shouldn't stop us from moving ahead with an answer, for if our world is real, then so our answer should be. If it is something else, then, it will not bring us any closer to being there. Physicists have their ways of attempting to ascertain its nature, these being riddled with complex mathematics that seems to have yielded only more questions. Perhaps my method using the philosophy of logic can offer a productive route. If I end up where the physicists have, at least it might elucidate some matters where digital physics may have been lacking. I add my philosophy of binary logic because I consider it different; no one to my knowledge has presented my approach to answering a question like this before. I cannot worry about whether my answer is "absolutely" right or wrong, for whatever comes after me probably will be the determinant of that. Yet, it may be a key to finding a meaningful answer. My suspicion is that the question will linger long after I am gone. We still are confined in the proverbial philosophical fishbowl, looking from the inside out without benefit of perspective or reference frame. The exercise inherently is solipsistic.

I will for convenience skirt the rat's nest of epistemological debates by bootstrapping, the tool that logicians use and one that is consistent with the idea of digitization, i.e., analysis. This approach will become clearer as we proceed. To me "reality" can mean "universe," as philosophers from the ancients to the moderns seem to think that something called "the universe" is, in fact, reality.

Let us descend like Orpheus down to Eurydice to retrieve our other. Once we come out into the sunlight of Plato, we should not look back to the shadows of all the prejudices and human constructs of reality, lest the journey be all in vain. We begin with how we are going go there.

## **The most fundamental law of understanding**

In front of us is a panorama of phenomena that we want to call reality. Is what makes it up discrete or continuous? Besides focusing upon a phenomenon as an object of understanding, I reflect on the process in apprehending it. We are often caught up in – no, I'll say obsessed with objects as objects. Naïve observers have no sense of context or history. After drawing something on a blank chalkboard, I used to ask my students how they knew what I drew was there. They would answer "because you drew it," "I see it," and so forth. I would turn out the lights and ask, "What do you see?" "Nothing," the students usually would say. I would ask, what would happen if they were placed in an environment in which everything is the same shade of a color. The same answer would come back. Only after much prodding would they see that without distinction, contrast, or contradiction, one apprehends nothing. Distinguishing something and its "other" is the core process in discerning what is in our environment. This is the fundamental law of dialectics, the realization of something in terms of what it is not, resulting in a peculiar form of mutual containment. It is just as much a law as is the law of gravity.

Such is a process, a way of knowing which ancient philosophers in South Asia knew 4500 years ago. The *Creation* hymn in the *Rig Veda* says, "Whence all creation had its origin, he, whether he fashioned it or whether he did not, he, who surveys it all from highest heaven, he knows--or maybe even he does not know. [2]" For Samkhya, the oldest form of Hinduism, the soul (purusha) is counterpoised against matter (prakriti), one in terms of the other; neither has it own identity in isolation. In the West, it is the dualism of mind and matter.

In the Samkhy philosophy, everything started with an eternal unconscious as the universe and from it unfolded everything we have today. From a whole emerges diversity, a law of cosmic order (*rta*), according to the Rig Veda. In the Vedic view [3], creation emanates from the self-consciousness of the primeval being (Purusha) that modern philosophers of consciousness could equate with the universe, the universe, itself, being conscious [4].

A world emerging from the inchoate also is told of in ancient Western philosophy. Hesiod (ca. 750 and 650 BCE) wrote of everything being born of chaos [5]. Others, such as Anaximander (c.610—546 BCE), stated, “...some other nature which is indefinite, out of which come to be all the heavens and the worlds in them.[6]”

Lucretius (ca. 99 BC – ca. 55 BC) stated that without differentiation there could be no harmony, i.e.:

In that long-ago  
The wheel of the sun could nowhere be discerned  
Flying far up with its abounding blaze,  
Nor constellations of the mighty world,  
Nor ocean, nor heaven, nor even earth nor air.  
Nor aught of things like unto things of ours  
Could then be seen--but only some strange storm  
And a prodigious hurly-burly mass  
Compounded of all kinds of primal germs,  
Whose battling discords in disorder kept  
Interstices, and paths, coherencies,  
And weights, and blows, encounterings, and motions,  
*Because, by reason of their forms unlike  
And varied shapes, they could not all thuswise  
Remain conjoined nor harmoniously  
Have interplay of movements.* But from there  
Portions began to fly asunder, and like  
With like to join, and to block out a world,  
And to divide its members and dispose  
Its mightier parts--that is, to set secure  
The lofty heavens from the lands, and cause  
The sea to spread with waters separate,  
And fires of ether separate and pure  
Likewise to congregate apart [7]. (emphasis added)

All these writings have one thing in common, a description of a whole and parts coming from it. Yet, looking at a whole is not possible without looking at its parts and *vice versa*. Indeed, this recursive property is represented by figures such as the Moibus Strip, Necker Cube, and Klein Bottle; they illustrate the nature of dialectical symmetry. It would not be surprising to see time in this way, where future can influence the present, as the present can influence the future [8].

Are these accounts of our universe emerging from the unformed so far from the modern concept of our universe having its origin in a singularity, where the four basic forces of nature – strong, weak, electromagnetic, and gravitational – were as one, where there was no distinguishing anything, a

continuum? These are forces, or processes were bound up with the singularity giving rise to our universe and thus are its essence. There appears to be a deep structuralism in thinking throughout history, starting with Hindu Brahma, Vishnu, and Shiva corresponding to the inchoate – potential for creation, the whole that emerges from chaos, and division – or destruction.

## The analytical method for answering the question

To produce a distinction in an undifferentiated whole, we simply could point to what our universe is not, such as another entity in an environment of multiverses, a highly speculative exercise. Another whole universe such as ours can be added, but this still would not take us any closer to comprehending what we have presently.

To arrive at building blocks of our world, we start with Descartes (1596 – 1650), who, continuing on the path of the ancient philosophers about cosmology, talked about the beginning of the world coming from “...no other form than that of chaos. [9]” His method of understanding was to “...to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution, [10]” by showing that we cannot conceive body unless as divisible;” [11]. To look at the totality of phenomenon around us, we analyze it, or cut it up. It is a bootstrapping method that alleviates the solipsistic, or fishbowl problem alluded to at the beginning of this essay. A standard method of proof is assuming and seeing what conclusions follow. This deductive Cartesian method is digitization – making something discrete from the whole. The process is just as much a part of the object as the object, itself. It is truly dialectic. Welcome to digital physics.

Subdivision towards an infinitesimal leads to quantum states and an equal amount of uncertainty of the same quality as met by attempting to effect the dialectic by adding a universe to ours. Yet, we can accomplish addition by the Cartesian method. That you add by division is an apparent paradox. However, dialectics is like that, where paradoxes often are removed.

There is a *caveat* here. In logic, uncertainty lies with both analysis, or deduction, and synthesis, bringing together entities to form something new. With the first we identify definitions, rules, and primitive expressions by which the analysis can be made, but Gödel demonstrated that that there will be statements about natural numbers not provable within that system (also expressible by logic), and the consistency of the system cannot be demonstrated from within that same system. With induction, one does not know the realm of the unknown universe from which new elements are introduced. Yet, when we think we are seeing continuity, it is the rods and cones in our eyes as discrete sampling devices that provide the basis of our visual induction. Both deduction and induction, then, are not closed systems. We should not be prevented from discovering how deduction can describe our universe, however.

Logic as a language of order is indifferent to whether there is a world of shadows or reality. As a hallmark of empiricism and scientific methods, the process is repeatable, accepted, and used. Whether our world as a shadow or “reality” is answerable not, it is, nonetheless our world. Our bootstrapping approach now will treat what we have around us phenomenologically as “real.” With hypothetico-deduction [12], we can start with a hypothesis based on theory, test it, and see what results. The manner of testing what it produces still is in the realm of theory in many respects, but there is sufficient information to think that the conclusions merit at least a further exploration, if not acceptance.

## **The result of the analysis**

Logic is a process, an interplay between digital and analog, the former being discrete (analogous to a particle) and the latter being continuous (analogous to a wave). The dialectical aspect of logic, deductive as digital and inductive as analog reflects this. To appreciate this interplay, it is helpful to review the relationship between continuity and particularity.

The problem of continuity is as old as Zeno's (ca. 490 BCE? – ca. 430 BCE?) paradox of the tortoise and the hare, where two objects coming together never can touch, as no one can stop subdividing the distance between them. Of course, this issue was overcome mathematically by calculus and the convergent series. Yet, a fundamental precept of calculus is the idea of limit, something set by a human. For example, approximating the area of a circle depends upon how precise we want pi, the ratio of the circumference to the diameter, suggesting that pi represents a process displaying a relationship, rather than a number. We decide how close to a boundary, or limit, the value of a function is to be and set how close to a value we want the function to approach. That is, the limit depends on us. Another example is determining the area under a curve, where set the size of the polygons we want to approximate the area. A boundary, itself, is problematical at best, but at the quantum level (as in the double-slit experiment, where an attempt at measurement destroys the interference pattern), a human seems to determine reality. In measuring anything; one has to know the degree of precision required, and this is set by the one doing the measuring.

It is appropriate to interject the boundary issue in determining the extent of our universe. A limit may not always be a physical edge but a process, such as Feynman's "Character of Physical Law" [13] framing an edge. Here, a process is the limit, the speed of light allowing us to see across a distance of about 46.6 billion light years or  $8.80 \times 10^{26}$  meters. It is conceivable that it may be larger, but such would involve getting outside the current dimension or going faster than the speed of light, and transcending time.

Now comes the simplest form of measurement in dialectic form, a measure of two - binary logic, the most basic form of being digital. It is two things. First, it is the language describing innate order in the universe. Second, within this language itself are processes that are found in the universe that have been discovered by other methods of inquiry [14]. The idea that the universe is binary is not new, as widely diverse scientists such as Wheeler [15], Wolfram [16], and Piaget [17], among many others, have argued, but how far does this idea carry in our ability to apprehend our environment that we take to be reality? More succinctly, is reality that of "it from bit," or is it continuous?

We now make a decision of how far subdivision is to be made before there is little utility in going further. Mathematical physics reduces us to the world of Planck volume in a world of vacuum space that allows virtual particles to come in and out of existence. Beyond this distinction seems to disappear, and we truly are at where we were in attempting to discern boundary for the universe. Again, the dialectic manifests itself, the Planck world being defined in terms of the largest of all, the universe. At both extremes is uncertainty as to size, and it seems that only that the character of physical law as a process provides us with any way of assessing a boundary, one that emerges concurrent with our ability to measure.

Then, what is the dialectic "other" of Planck volume ( $4.22419 \times 10^{-105}$  m<sup>3</sup>)? If Planck volume is one aspect of the dialectic underscoring our binary world, then we look at vacuum space as the other aspect.

Loop quantum gravity says that the space is, quantized, i.e., discrete. This idea isn't new, as Zeno of Elea said:

The arrow which is moving forward is at every present moment in a space equal to itself, accordingly it is in a space equal to itself in all time; but that which is in a space equal to itself in the present moment is not in motion. Accordingly it is in a state of rest, since it is not moved in the present moment, and that which is not moving is at rest, since everything is either in motion or at rest. So the arrow which is moving forward is at rest while it is moving forward, in every moment of its motion [18].

The key words are “every moment of its motion,” indicating a discreteness of space. Modern physicists, such as Zizzi, also argues for space being discrete [19]. Planck volume is in terms of vacuum energy, a fundamental binary and dialectical relationship [20].

At and below the Planck scale, we are in the realm where all four forces seem to be unified and even time, itself, lacks. At such point, we barely can even talk about a binary world, but we seem to reasonably sure that this is the scale at which everything in our universe started as the singularity. Time no longer appears relativistic, as everything is one, the singularity. Hence, to for binary expression, we stop at the Planck scale with a unit of Planck volume in terms of vacuum space.

## What the language of binary logic tells us about our universe.

Two of my favorite aspects of binary logic illustrate the power of that system to express the nature of the universe. First and foremost is the primary inference relationship in logic: “If something, then something.” Logic instructors often tear their hair out in explaining this “material implication” operator, the truth table of which is,

<b>p</b>	<b>q</b>	<b>p <math>\supset</math> q</b>
0	0	1
0	1	1
1	0	0
1	1	1

where 0 = false and 1 = true. Ordinary language translations of the operator have given logicians such a fit that modal logic was born [21]. However, there is a consistent interpretation, as indicated by the word “containment,” stated in several ways. It is clear that elements contain themselves, and if we consider “0” standing for potential information, an unknown, or an origin (perhaps vacuum space), then, it can be seen that it is true that 1(Planck volume), representing a known or information, emerges from that 0. However, “1” standing for potential information cannot go backward and itself become the unknown; once it is known, it is known. As 1, something coming from an origin 0 does not produce that origin, although it can be the origin for something else. The 0 can stand for the undefined, or chaos, and 1 as order. The 0 can remain itself as 0 or generate order as 1. Order does not generate chaos. The 1 can be itself or generate order. In cosmological terms, we may treat 0 as a singularity, the inchoate, as in the Vedas, Hesiod, and Lucretius, and 1 as the universe emerging from that chaos.

In returning to recursive representations of our universe, such as the Klein Bottle [22] we find that such recursion is represented by the sixteen functions comprising basic logical space. Two basic entities,

let's say represented by p and q, bear a permutation of four relationships between them, as indicated by the truth table above. From this extends the sixteen four-place functions in the following manner, known commonly as the Table of Functional Completeness, namely,

p	q	f <sub>0</sub>	f <sub>1</sub>	f <sub>2</sub>	f <sub>3</sub>	f <sub>4</sub>	f <sub>5</sub>	f <sub>6</sub>	f <sub>7</sub>	f <sub>8</sub>	f <sub>9</sub>	f <sub>10</sub>	f <sub>11</sub>	f <sub>12</sub>	f <sub>13</sub>	f <sub>14</sub>	f <sub>15</sub>
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

When the outputs of each of these functions are forward-fed as inputs, the function will repeat itself. Perforce, because our universe is binary digital in character, and spaces are comprised on one or more of these sixteen functions, space is recursive [23].

There are many more processes within the binary system that describe our universe, but space now only allows me to move to what some of the implications are.

### ***Implications of the answer that the universe is digital***

Our path of descent has reached a world containing entities that would take more energy than is contained in the universe to split them. It is where matter is mind and vice versa. It is Aristotle's substratum out of which everything comes. It is the world of quantum processes and no thing is reality with any lasting quality. Only process seems to be reality, where the Hindus say that all is maya, or illusion. All phenomena are comprised from this substratum, and it is in constant flux. It appears that knowing, knowledge, and, everything else – even ideas – at this level are uncertain. Uncertainty describes reality. Boundaries clearly have no identity. At the Planck scale, only Feynman's "Character of Physical Law" exists, where it is process that that determines what is. It is not that the universe is such and thus but only what that physical law allows it to be, regardless of whether we are able to apprehend it empirically.

Here, computations are implicit just by the nature of binary space, as in cellular automaton and the mere juxtaposition of fundamental binary functions so as to produce others. Unfolding from these arrangements are patterns, or regularities that allow one to ascertain what will happen if there is a continuation of those arrangements, as illustrated by "gliders" in cellular automata and fractals. Even a reputedly random juxtaposition of functions reveals patterns [24]. It is plausible that all complexity obtains from the simplicity of the binary world [25]. It is becoming clear that simplicity often translates itself into complexity, as suggested by Turing's reaction-diffusion model expressed by patterns in nature [26]. Indeed, one example of the markings on a typical conus textile snail shell is the same as discussed by Wolfram resulting from a simple cellular automaton [27].

What a digital universe doesn't seem to tell us at first is anything about time and our consciousness. As we look out over space and see distant objects, we know that these are in the past. Alpha Centauri is 4.37 light years away. We are seeing the sun as it was eight minutes ago. Carrying this process further, this "perceptual lag" is all around us, where even our neighbor is seen as she or he existed petaseconds or less ago. Going still further, we ask where does it stop; what is the focus of our consciousness? Perhaps it is not within us at all, but as Zizzi[28], Kafatos (cited above), and the ancients suggest, it is in the universe itself. It may be that ideas are passed in and out of the universe to and from another

dimension. Again, one should be reminded about the meaning of boundary. Perhaps there is something about that character of physical law that accounts for consciousness, as well. If consciousness is immanent in the universe and each of us is merely partaking of it, then, we may need to think differently about the ego.

### **End words**

In light of all that has been said, I conclude that the answer to whether reality is digital or analog is that dialectics frames the answer. An entity exists because of its other and vice versa. Digital is seen in terms of the analog, just as a particle, or discreteness has its existence in terms of a wave, or continuity. It is the same with deduction in terms of induction. I am reminded of the argument in Marcia Eliade's *Myth of the Eternal Return* about people regarding circular time as becoming linear with the advent of the written word. As I write this, and wonder, given dialectics and all I have said, whether, like the hoop snake or the closed timelike curve, I am just looking at my tail.

$$x \supset 0 \subset x$$

**References** (all websites accessed 5 February 2011- Almost everything cited is accessible on-line.)

- [1] *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/epistemology>.
- [2] *Rig Veda*, Hymn CXXIX – Creation, Book the Tenth.
- [3] *Rig Veda – Oxford Companion to Philosophy*, [http://www.answers.com/topic/ancient-philosophy#Indian\\_philosophy](http://www.answers.com/topic/ancient-philosophy#Indian_philosophy).
- [4] M. Kafatos and R. Nadeau, *The Conscious Universe*, New York: Springer-Verlag, 1999.
- [5] Hesiod, *Theogony*, Line 116.
- [6] Simplicius, *Commentary on Aristotle's Physics* (24.13-21), *Aristotle, Physics* VI:9, 239b5.
- [7] Lucretius, *The Nature of Things*, “Formation Of The World and Astronomical Questions”.
- [8] M. Zeeya Meral, “Back from the future,” *Discover*, [http://discovermagazine.com/2010/apr/01-back-from-the-future/article\\_view?searchterm=Tollaksen&b\\_start:int=2](http://discovermagazine.com/2010/apr/01-back-from-the-future/article_view?searchterm=Tollaksen&b_start:int=2).
- [9] R. Descartes, *Discourse On the Method of Rightly Conducting the Reason, And Seeking Truth in the Sciences*, Part V, <http://infomotions.com/etexts/philosophy/1600-1699/descartes-discourse-124.txt>.
- [10] *Ibid.* Part II, p. 7.
- [11] R. Descartes, *The Method, Meditations, and Philosophy*, “Discourse on Method,” Part II, p. 122, [http://oll.libertyfund.org/?option=com\\_staticxt&staticfile=show.php%3Ftitle=1698&Itemid=27](http://oll.libertyfund.org/?option=com_staticxt&staticfile=show.php%3Ftitle=1698&Itemid=27).
- [12] William Whewell, [http://en.wikipedia.org/wiki/Hypothetico-deductive\\_model](http://en.wikipedia.org/wiki/Hypothetico-deductive_model).
- [13] R. Feynman, *The Character of Physical Law*, Modern Library, 1994.
- [14] J. Horne, *Logic as the language of innate order in consciousness*, <http://home.earthlink.net/~jhorne18/>.
- [15] C.W Misner, et al., *Gravitation*, New York: W.H. Freeman and Company, 1973.
- [16] S. Wolfram, A New Kind of Science, <http://wolframscience.com/nksonline/toc.html>.
- [17] J. Piaget, *Logic and Psychology*. New York: Basic Books, Inc., 1958.
- [18] A. Fairbanks, ed. and trans., *The First Philosophers of Greece*, “Zeno Commentary,” 1898, – page 10, <http://history.hanover.edu/texts/presoc/zeno.htm>.
- [19] P. Zizzi, “Spacetime at the Planck scale,” 2004, [arXiv:gr-qc/0304032v2](http://arxiv.org/abs/gr-qc/0304032v2), pg. 345-358.
- [20] Vacuum Energy, [http://en.wikipedia.org/wiki/Vacuum\\_energy](http://en.wikipedia.org/wiki/Vacuum_energy).
- [21] H. MacColl, *Symbolic logic and its applications*, Longmans, Green & Co. Londres, 1906, Chapter IX, p.67 et seq., <http://www.archive.org/details/symboliclogicand00maccuoft>.
- [22] D. L. Rapoport , *Surmounting the Cartesian Cut*, Foundations of Physics, Vol. 39, 2009, <https://docs.google.com/viewer?url=http://vixra.org/pdf/1006.0007v1.pdf&embedded=true&chrome=true>.
- [23] J. Horne, “Recursion of Logical Operators,” <http://home.earthlink.net/~jhorne18/> .
- [24] A.Wuensche, "The Ghost in the Machine: Basins of Attraction of Random Boolean Networks . (University of Sussex at Brighton: *Cognitive Science Research Papers*, CSRP 281, 1993, [https://docs.google.com/viewer?url=http://www.informatics.sussex.ac.uk/users/andywu/downloads/papers/ghost\\_in\\_machine.pdf&embedded=true&chrome=true](https://docs.google.com/viewer?url=http://www.informatics.sussex.ac.uk/users/andywu/downloads/papers/ghost_in_machine.pdf&embedded=true&chrome=true) .
- [25] Wolfram, *passim*.
- [26] S. Kondo et al., “Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation,” *Science* 329, 1616 (2010); DOI: 10.1126/science.1179047.
- [27] Wolfram, p. 29.
- [28] P. Zizzi, “Emergent Consciousness: From the Early Universe to Our Mind”, 2000, [arXiv:gr-qc/0007006](http://arxiv.org/abs/gr-qc/0007006).





## AUTHORS INDEX Volume I

Ahmed, Shaimaa B.	1	Gaud, Nicolas	200
Amar Bensaber, Boucif	238	Gayoso, Miguel	216
Andersone, Ieva	82	Gil, Ignacio	309
Aramina, Daira	106	Ginevičius, Adomas	88
Artut, Selcuk	285	Ginevičius, Romualdas	88
Atac, Goktug	149	Gray, Tom	293
Balch, Robert	299	Güzererler, Ahmet	285
Biskri, Ismail	238	Ham, Chan	19; 24; 28; 44
Bistrova, Julia	59	Ham, Sang Yong	34; 40
Bueno, Jesús F.	365	Ham, Woonchul	254
Burton, Matthew	19	Hamilton, Mark S.	49
Camp, Joseph	154	Helkin, Steven	28
Carvalho, Sidney R. D.	365	Higashiyama, Yoichi	178
Castañeda-Camacho, J.	376	Hillali, Hassane	238
Cermakova, Hana	212	Horne, Jeremy	184; 385
Chandramouli, Ramaswamy	166	Huang, Shunquan	343
Chase, Phillip H.	243	Huang, Youfang	343
Chen, Hao	228	Joazeiro, Sayuri A.	365
Chen, Meng	289	Joo, Young Hoon	19; 28
Chen, Te-Yu	249	Jung, Kyung Kwon	196
Choi, Sang-Kyu	34	Kamejima, Kohji	260
Ciemleja, Guna	65	Kamenicky, Jan	212
Corea, Stephen	381	Kim, Byung In	371
Dag, Tamer	149	Kim, Hyeong-Seok	196
de Faria, Elmo B.	365	Ko, Won	24
Descoteaux, Steve	238	Kocmanova, Alena	94
Devece, Carlos	309	Koke, Tatjana	112
Do, Hyun Min	371	Koukam, Abderrafiâa	200
Docekalova, Marie	94	Kozlovskis, Konstantins	100; 143
Doma, S. B.	7	Kupwade Patil, Harsh	154
Donado, A.	376	Kwon, Hyeokjae	254
Elgarawany, H.	13	Kyung, Jin Ho	371
Enkhbaatar, T.	254	Lace, Natalja	59; 65
Eriņš, Ingars	71	Lacis, Jurijs	100
Fathy El Dessouky, Naglaa	303	Lapina, Inga	106
Fernández, Rocío	309	Lapinskaitė-Vvohlfahrt, Indrė	123
Fries, Terrence P.	172	Latchman, Haniph	315
Fumizawa, Motoo	359	Lee, Jong-Hyun	24
Gaile-Sarkane, Elīna	76	Lee, Min Goo	196
Gaile-Sarkane, Elina	82	Lee, Sung Whee	40

Lim, Hyung Chul	40	Santos, Wellington S. E.	365
Lin, Kuo-Chi	28; 44	Schillo, Kevin	44
Lin, Kurt	19	Schulte, Sven	321
Lohrenz, Maura	293	Segall, Richard S.	206
Lokaj, Zdenek	160	Silva, Carla M.	365
López, Joaquín	216	Simberova, Iveta	94
Lu, Shen	206	Son, Young Su	40
Luo, Dean	289	Soto Toro, Felix A.	49
Ma, Yulin	222	Soušek, Radovan	212
Michnevič, Edvard	123	Spöttl, Georg	321
Mino, G.	376	Srinivasan, Sambasivam	233
Mozayani, Naser	200	Srinivasan, Satish Mahadevan	233
Muñoz-Hernández, G.	376	Stankeviciene, Jelena	137
Němec, Vladimír	212	Stasytytė, Viktorija	131
Nachtgael, Mike	266	Stasytytė, Viktorija	123
Narasimhan, Sreenivasan	315	Straub, Jeremy	272
Nemecek, Petr	94	Szygenda, Stephen A.	154
Nonaka, Hidetoshi	327	Taraškevičs, Ronalds	71
Oganisjana, Karine	112	Titko, Jelena	143
Ohkawa, Syuhei	359	Trenchard, Michael	293
Oliva, Jan	277	Trifas, Monica	272
Ostergaard, Soren Duus	331	Tripathi, Abhishek	233
Ou, Hsia-Hung	249	Tseng, Jing-Hang	249
Palacios, Daniel	309	Uturytė-Vrubliauskienė, Laura	123
Park, Chan Hun	371	Valle, Christopher	44
Park, Cheol Hoon	34; 40	Vázquez Ramos, Gabriel	53
Park, Dong Il	371	Vlcek, Josef	277
Park, Yong Kuk	196	Volf, Jaromír	277
Parker, Andrew	299	Vu, Bruce T.	49
Paz, Enrique	216	Wang, Jin-You	228
Pereira, Mauricio F. L.	365	Wang, Lan	289
Pérez, Diego	216	Wu, Kee-Rong	349
Pfisterer, Michael	44	Wu, Qing	222
Pilvere, Irina	118	Yang, Yong-Jay	249
Podvezko, Valentinas	88	Yeh, Chung-Wei	349
Razavi, Seyed Naser	200	Yoo, June Jae	196
Reyes, F.	376	Yu, Siqin	343
Rivza, Baiba	118	Yuan, Jiann-Shiun	53
Rivza, Peteris	118	Zaretsky, Esther	337
Rivza, Sandija	118	Zelinka, Tomas	160
Rompré, Louis	238	Zhang, Shuqun	281
Rumchev, Ventsi	178	Zhang, Yong-De	228
Rutkauskas, Aleksandras Vytautas	123; 131	Zhao, Yan-Jiang	228
Ščeuļovs, Deniss	76	Zhou, Lixia	355
Saleh, M. M.	13	Zhu, Guopu	281